

TRANSLATIONAL ARTICLE

Using mobile phone data for epidemic response in low resource settings—A case study of COVID-19 in Malawi

Dylan Green^{1,2,*} , Michael Moszczynski¹, Samer Asbah¹, Cassie Morgan¹, Brandon Klyn¹, Guillaume Foutry¹, Simon Ndira¹, Noah Selman³, Maganizo Monawe⁴, Andrew Likaka⁴, Rachel Sibande⁵ and Tyler Smith¹

¹Cooper/Smith, Washington, District of Columbia, USA

²Department of Epidemiology, University of Washington, Seattle, Washington, USA

³Harris School of Public Policy, The University of Chicago, Chicago, Illinois, USA

⁴Quality Management Division, Ministry of Health—Malawi, Lilongwe, Malawi

⁵Country Outreach, Digital Impact Alliance, United Nations Foundation, Washington, District of Columbia, USA

*Corresponding author. E-mail: greend7@uw.edu

Received: 25 September 2020; **Revised:** 24 May 2021; **Accepted:** 17 June 2021

Key words: COVID-19; Malawi; mobile network data

Abstract

The COVID-19 global pandemic has had considerable health impact, including sub-Saharan Africa. In Malawi, a resource-limited setting in Africa, gaining access to data to inform the COVID-19 response is challenging. Information on adherence to physical distancing guidelines and reducing contacts are nonexistent, but critical to understanding and communicating risk, as well as allocating scarce resources. We present a case study which leverages aggregated call detail records into a daily data pipeline which summarize population density and mobility in an easy-to-use dashboard for public health officials and emergency operations. From March to April 2021, we have aggregated 6-billion calls and text messages and continue to process 12 million more daily. These data are summarized into reports which describe, quantify, and locate mass gatherings and travel between subdistricts. These reports are accessible via web dashboards for policymakers within the Ministry of Health and Emergency Operations Center to inform COVID-19 response efforts and resource allocation.

Policy Significance Statement

This study presents a case study of the use of aggregated mobile phone data to better understand population mobility and connectedness—a key indicator of COVID-19 response. In this analysis, we demonstrate how mobile phone data can be used to identify the occurrence and size of potential mass gatherings, which could be events in which significant COVID-19 transmission occurs, including spread to previously unaffected areas. Ultimately, this information can be leveraged to assist policymakers in making tough trade-offs in terms of mitigation strategies and allocating resources.

1. Introduction

The COVID-19 global pandemic has spread to 192 countries causing 155-million infections and 3.2-million deaths through April 2021 (Dong et al., 2020). High-income nations with older populations and

high prevalence of cardiovascular diseases - each large risk factors for COVID-19 severity - have been hardest hit by the pandemic (Booth et al., 2021; Li et al., 2021). Sub-Saharan Africa has been less severely impacted to date, and likely sustained community spread much later than other more connected nations in North America, Europe, and Asia (Nordling, 2020). While testing capacity in Africa was extremely limited in the early outbreak, the continent was quick to lockdown in March and April, which likely helped to mitigate and suppress early spread of COVID-19 (Cooper/Smith, 2020a; University of Oxford, 2020). However, policies in recent months have relaxed, and mobility patterns have returned to near normal levels in many places as questions continue to mount over the trade-offs between COVID-19 risks and the economic impact of mitigation policies (Carlitz and Makhura, 2020; Cooper/Smith, 2020b). Nonetheless, COVID-19 cases and deaths continue to grow and indeed accelerate across Africa (Africa Center for Strategic Studies, 2020). Indeed, the “South Africa Variant,” lineage B.1.351, has been determined to be a variant of concern and is believed to have in part fueled the second COVID-19 wave in Africa (WHO Africa, 2021).

The COVID-19 response in Malawi, a low-income country in southern Africa, has been complex. Malawi did not detect its first case of COVID-19 until April 2—among the latest countries globally to diagnose a case, 181 of 192 countries (Fiko, 2020; Wikipedia, 2021b). Government response was swift, COVID-19 was pronounced a national disaster more than a week before the first case was detected, and the government subsequently banned gatherings of 100 persons or more, closed schools, and restricted regional and international travel (Chilunga, 2020). Weeks later, as the president announced a three-weeklong lockdown, Malawian courts struck down the policy as it could bring harm to the economically vulnerable (Al Jazeera Staff, 2020). The Malawian COVID-19 response was further complicated as presidential elections were being held for a second time in June after courts ruled that elections in the prior year had widespread irregularities (Chitete, 2020). Whistle-stop tours were prevalent throughout the country, drawing huge crowds to rallies, with recommendations to limit large gatherings seen as a potential attempt to undermine and suppress the political opposition (Kondowe, 2020). Finally, in May and June, thousands of Malawian migrant workers throughout South Africa were repatriated to Malawi (Masina, 2020; Nyasulu et al., 2021). The management of the returnees was challenging, with many returning to their homes without testing, or without receiving their test results—and ultimately at risk of causing outbreaks (Agence France Presse Staff, 2020). While the epidemic was certainly slow to spread in Malawi, there is clearly extensive community spread, with a cumulative total of 34,151 confirmed cases and 1,152 confirmed deaths through April 2021 (Ministry of Health—Malawi, 2021).

As COVID-19 quickly spread, traditional means of contact tracing to achieve epidemic control quickly became untenable and had limited effectiveness as the number of tests per capita far lower than the World Health Organization’s recommended levels (Cooper/Smith, 2020a). Experts in public health, informatics, and data science began weighing the potential impact of leveraging mobile phone data to facilitate more effective contact tracing (Ferretti et al., 2020; Joseph, 2020). Technical approaches have varied, but as many as 50 countries have implemented some sort of COVID-19 smartphone app for voluntary download and use—although with mixed utilization and impact, and notably only useful where smartphone use is high (Ahmed et al., 2020; Munzert et al., 2021; Wikipedia, 2021a). Alternatively, aggregated call detail records (CDRs)—phone calls and text messages—are available for all mobile phone types and have been used for research and monitoring purposes (Blondel et al., 2015; Jones et al., 2018). These data pose less risk to individual privacy—if used responsibly, with utmost management security, adequate aggregation, and strict access limitations—and can be utilized passively, without significant numbers of individuals opting in en masse as was demonstrated in the West African Ebola Outbreak of 2014 (GSMA, 2014; Jones et al., 2019). These data are even more attractive as a public health tool in low-income settings, such as Malawi, where smartphone ownership (and therefore potential COVID-19 app penetration) is at only 15% (GSMA, 2018). This study aims to successfully develop a daily pipeline of CDR data in Malawi, including analytics and visualizations to improve targeting of limited resources and public health response in a resource-limited setting.

2. Methodology

2.1. Population

The target population of interest in this study is persons of all ages residing in Malawi any time between March 2020 and April 2021. We summarize population density and movement relative to administrative disaggregations including the 28 districts and 433 traditional authorities (TAs).

2.2. Call detail record data

2.2.1. Project background and ethics

The Digital Impact Alliance (DIAL) began engaging the Malawi government's Digital Health Division and local telecom firms to establish use cases for mobile network data for public good. In 2018, an initial project using anonymized and aggregated CDR data sought to understand population density and distance from health facilities to inform the placement of future health facilities to improve access and equity (Cooper/Smith; DIAL; Infosys, 2019). The project was extended into 2020 to develop additional use cases in both health and nonhealth sectors. As the COVID-19 pandemic unfolded, the opportunity arose to leverage prior work and plans to meet the needs of the Malawian public health response to the novel coronavirus. We obtained authorization from the Malawi Communications Regulatory Authority as well as approval from Malawi's Institutional Review Board (IRB), the National Commission for Science and Technology (Ref. No. NCST/RTT/2/6, Protocol No. P.02/18/253).

In April 2020, Malawi's second largest mobile network operator, Telekom Networks Malawi (TNM) agreed to share anonymized CDR data with Cooper/Smith for the explicitly agreed purpose of supporting Malawi's Digital Health Division and Public Health Institute within the Ministry of Health to combat the COVID-19 crisis. TNM has approximately 42% of the mobile phone market share in Malawi (GSMA, 2012). TNM shared limited and aggregated CDR data from subscribers stretching back to January 1, 2020, which allowed for us to establish a baseline estimate of call patterns, mobility, and distribution of subscribers before COVID-19 was geographically widespread and triggered individual and government policy changes to mitigate spread. After establishing the connection between TNM's data warehouse and the Malawian Ministry of Health for historical data, we developed an automated daily data pipeline which would capture a daily export of anonymized CDR data on the previous day's events.

2.2.2. Mobile phone events

Each observation in the anonymized CDR data represents a mobile phone event—a call or text message—as well as the accompanying cell phone tower latitude and longitude through which the connection was made. This cell tower location is usually (but not always) the nearest tower geographically to the mobile phone user, and therefore represents an accurate estimate of an individual's location. To preserve privacy, the timestamp attributed to each mobile phone event is a 4-hr time block, rather than the precise time of day. Finally, each subscriber's phone number was de-identified using salted-hashing, with procedures in alignment with industry standards for anonymization behind TNM's firewall before delivery to the research team. No names, phone numbers, or other personal identifiers were included in the data pipeline. Briefly, salted-hashing combines two tools to encrypt and protect data. First, hashing is a technique that involves a mathematical algorithm which encrypts the underlying data. While hash functions are common and effective cryptographic tools, they can be susceptible to decryption attacks, and therefore “salting,” or added random data to hashed encryption, is a further protection from brute force attacks or use of rainbow tables (Gauravaram, 2012; Stevens et al., 2017). Without access to the salting key, the encrypted data are extremely difficult to decrypt.

2.2.3. Cell tower to administrative area mapping

We generated a map of Voronoi cells which represent each cell tower's probable coverage area (Online Appendix). However, the Voronoi cells do not map effectively to Malawi's administrative districts and TAs which are ultimately more meaningful for understanding individual mobility and clustering of

individuals. Therefore, for each record, we attribute a fraction of an individual's location to the proportion of population, as determined by 1-km-squared density tiles, shared by the Voronoi cell and each TA it covers. This allows us to estimate the total number of subscribers in each district and TA throughout Malawi every 4 hours.

2.2.4. Validation of CDR-derived population estimates

To validate the estimates of population derived from CDR data, we leverage two other sources of population size in Malawi—the recently performed 2018 census and WorldPop 1-km population size estimates (National Statistics Office, 2018; WorldPop, 2020). The 2018 census produces population estimates at the district level, while WorldPop creates population estimates at a 1-km-squared level. For validation purposes, we intersect the WorldPop estimates with shapefile maps of TAs in Malawi to aggregate population estimates.

2.2.5. Identifying potential mass gatherings

Using estimates of subscribers by TA over time, we can detect when there may be a mass gathering in excess of normal variability. We define such mass gatherings as 4-hr time blocks when the number of subscribers is more than two standard deviations greater than the mean number of historical subscribers. Following a mass gathering, we can estimate which TAs potential attendees traveled through and to following the event. We classify mass gathering events as either recurring or nonrecurring. For example, a village market might be active regularly, a single day of the week, and represents a routine event. Alternatively, there are apparent “one-off” gatherings which are less cyclical, such as a political or sporting event. This is important in situations where known transmission has been linked to an event, as individuals and communities at risk identified through these data can scale public health messaging and testing services. We expect this method of identifying mass gatherings has a higher sensitivity than specificity and may falsely identify a mass gathering event when none occurred. However, considering the public health risks, this seems a reasonable metric for focusing limited preventative resources to potential areas of transmission and investigation. Reports of mass gatherings are generated and visualized for public health officials using Tableau Server daily and are deployed in the Emergency Operations Center. The Tableau interface allows for end users to explore and customize their analysis of the data, while preserving subscriber privacy and eliminating the need for policymakers to access individual-level information, as the reports draw on aggregated data completely separated from the raw CDR dataset.

2.2.6. Estimating mobility and geographic spread

We also sought to describe individual mobility to understand to what extent Malawians might be sheltering in place, or otherwise reducing their movements to curb risk of COVID-19. For each subscriber, we can determine, at a minimum, how many physical movements they make by counting the number of different cell towers utilized by the subscriber over a time span. From one perspective, this is likely an underestimate of mobility, as it only captures movement when individuals use their mobile phone in multiple locations. However, it could also be an overestimate of mobility, particularly in urban or dense settings, where mobile phone traffic might be redirected to other cell towers which are further away when local utilization exceeds tower capacity (i.e., load sharing). Using estimates of mobility, we can estimate the interchange between TAs, as a proxy for the connectedness between locations. These data are important to understand how a disease could geographically spread through Malawi.

2.2.7. Simulation of COVID-19 spread along the cell phone network

COVID-19 has been characterized with high overdispersion, meaning that only a small proportion of infected individuals cause onward transmission (Bi et al., 2020; Endo et al., 2020; Zeller et al., 2021). While the dispersion factor for COVID-19 is difficult to estimate, a common figure is 0.1, which means that approximately 10% of infections cause 80% of transmissions and superspreading events are key to sustaining community spread (Sneppen et al., 2021). The implication for high overdispersion is that the

geographic spread of COVID-19 can be uneven and lagged in time considerably. This has important implications for the mathematical modeling of COVID-19 during the early introduction of the disease to new countries, especially settings with limited testing and low rural mobility such as Malawi (Porter, 2002).

We sought to estimate the start date of sustained community spread in each TA through using the anonymized, aggregated CDR data. We constructed a graph of the mobile network in Malawi, with each node representing a cell tower and the distance between nodes representing the average percentage of people that move between them—these links are unidirectional, with the distance from A to B being unequal to the distance from B to A. Note that in some cases, particularly in densely populated areas, towers are clustered together in the same location to accommodate high levels of usage. In these cases, tower clusters with the same coordinates are considered the same node. From the node and link structure, we can simulate the spread of the disease based on routine disease surveillance.

We run a simulation based on this structure, with each node exerting infection pressure on other nodes, and a node exerts internal pressure on itself without division. This infection pressure is expressed as a probability which is a product of the magnitude of travel between locations and the prevalence of COVID-19 in those locations. For example, suppose we estimate 100 individuals travel from location A to B, and the estimated prevalence of COVID-19 in location A is 5%. We then estimate that a total of five travelers from A to B are infectious, and are added to the pool of infectious individuals already present in location B.

As a result, the simulation shows spread reflecting the movements of people, for example, moving along main roads to large cities before spreading to rural areas, rather than spreading based on geographic distance. We then estimate the date community spread is sustained in a TA when the number of infections reaches 100 cumulative infections.

2.2.8. Data processing, pipeline, architecture, and security

We leverage Amazon Web Services to manage the anonymized CDR data, including the daily exchange of data from TNM, processing, and ensuring data are secure (Figure 1). We launch Amazon Elastic MapReduce, Apache Ranger, Active Directory, and Amazon Simple Storage Service within a Virtual Private Cloud (VPC) to isolate access and potential intrusion from the Internet. Within the VPC is a private subnet which performs the cluster analytics on the primary dataset. These results are transferred securely (still within the VPC) by launching a bastion host to a public subnet accessed with a weblink using a Ranger server. Through this Internet gateway, only aggregated data are transferred to our Tableau Server using Tableau Bridge and the bastion server. Throughout the entire process, security groups are used to manage the incoming and outgoing crowd, thus creating a defensive border.

3. Results

From March 2020 to April 2021, we processed more than 6-billion mobile phone events, with an average of 12-million events per day. After removing probable marketing accounts, which is done via the MNO's records, we captured a total of 1.5-million unique subscribers per day on average. The geographic population distribution as measured by the anonymized CDR data was highly correlated with empirical measures of population at the district level (from the 2018 census) and at the TA level (as estimated from WorldPop 1-km data), with a 5% deviation between CDR-derived figures and census and WorldPop estimates.

We generate daily reports of mass gatherings as identified by the anonymized CDR data. For each date, and each 4-hr time block, we identify the number of active subscribers in that place and time, as well as the average number of active subscribers, and variation from that average (Figure 2). These reports are prioritized in rank order, with areas of the largest deviance from normal subscriber volume at the top of the report to quickly identify potentially problematic areas.

For each event, users can see the normal patterns of subscriber volume for a location and time, as well as the locations where attendees of the event traveled afterward (Figure 3). In this example in Chimbiya TA, in Dedza district, we can see a recurring market event on Tuesdays during daytime hours. From the

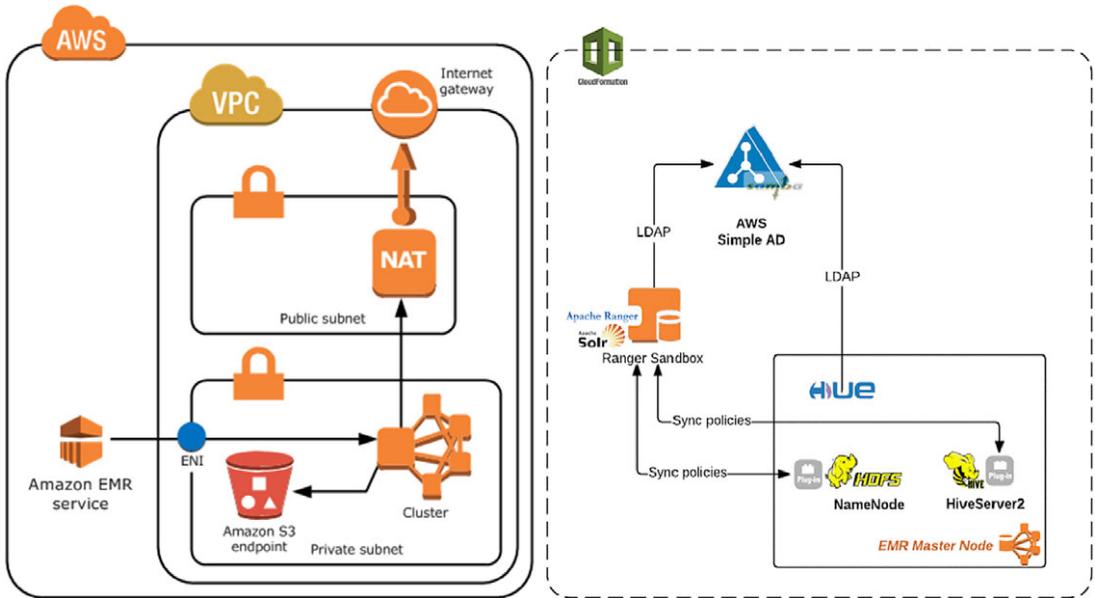


Figure 1. Schematic of (MNO) network topology (left) and system infrastructure (right).

Spread of Risk map, we can see that attendees of the event traveled from nearby in Dedza district, as well as Lilongwe to the north, and further south to Ntcheu and Balaka districts. With ongoing COVID-19 community transmission in Lilongwe, these data present a probable chain of spread via market commerce on a weekly basis in Dedza.

In another example from May 10, we can see a nonrecurring mass gathering in Mzuzu, likely due to an event at Mzuzu stadium (Figure 4). Historically, we can see very little variation in this area, with an extremely large spike in subscribers on May 10 from midday to evening, with nearly 8,000 subscribers identified in an area which usually has no more than 2,000 subscribers on average.

Finally, the aggregated Mobile Network Operator (MNO) data are used to estimate interchange, or physical connectedness between TAs. The Public Health Institute of Malawi and its partners have primarily used these data to inform infectious disease modeling efforts for COVID-19. An important epidemic feature to understand is when the epidemic begins in each jurisdiction or in this case TA, as not adjusting for variable epidemic start dates changes the shape of epidemic curves tremendously. Empirical testing data (as well as tests) are very limited in Malawi, and therefore modeling the epidemic start dates based on physical connectedness is a viable solution. Using the CDR data and attributing a location to the first 100 infections in Malawi, we could estimate the probable start date of epidemic spread in each TA (Figure 5). Although community spread was documented in urban areas of Malawi in early April 2020, we estimate that many peri-urban and suburban TAs did not have sustained community spread for another 6 weeks, with a few most isolated TAs not seeing sustained community spread of COVID-19 until 10 weeks after the epidemic began.

4. Discussion

This study demonstrates the process and potential utility of leveraging anonymized, aggregated CDR mobile phone data for a public health response during an emergent outbreak in a resource-limited setting. Mounting a successful response to an epidemic requires high-quality data on individual behaviors, policy compliance, and risk—each of which can be difficult to ascertain on a continuous basis in a low-income country. For COVID-19, this means understanding to what extent individuals are compliant with physical distancing or shelter-in-place recommendations as reducing physical contact with others remains the

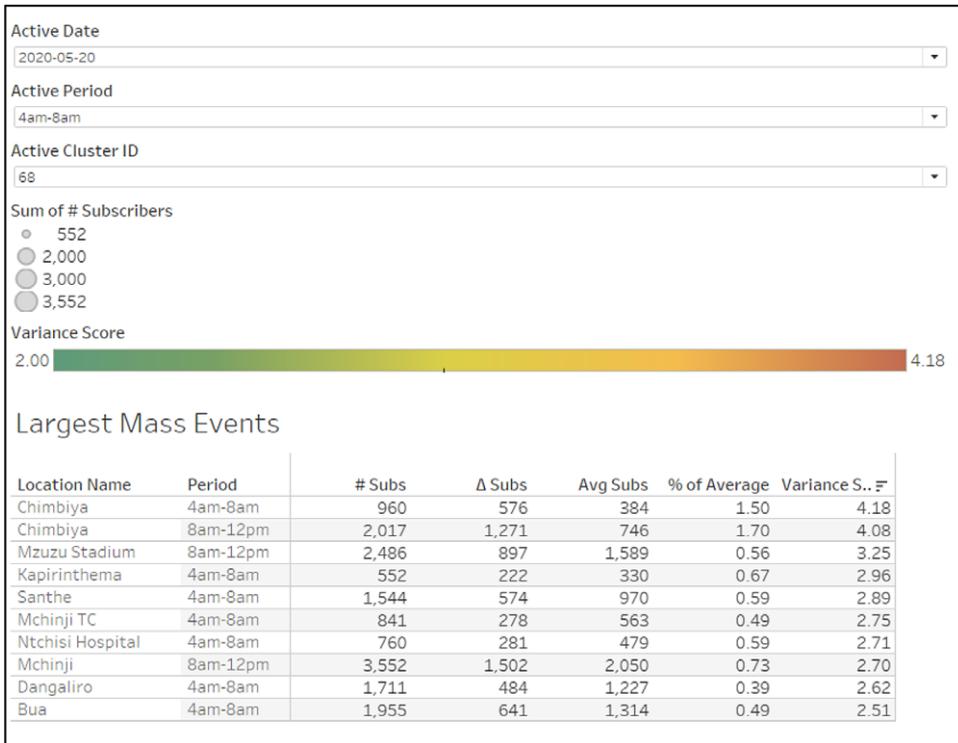


Figure 2. Example mass gathering daily report, May 20, 2020.

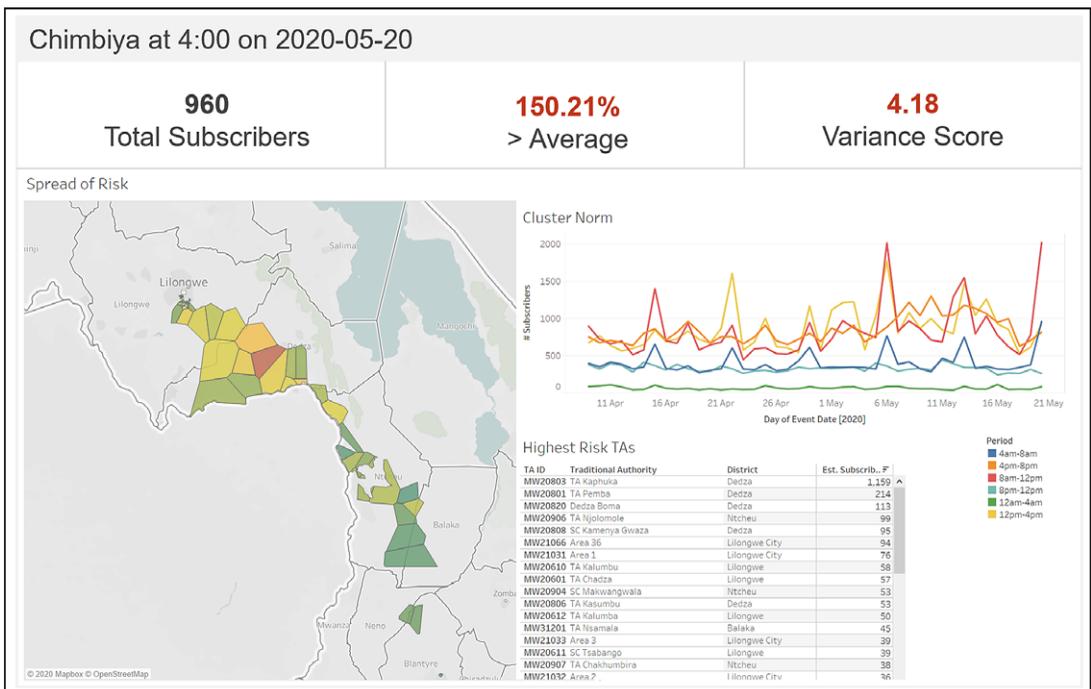


Figure 3. Example mass gathering event detail, Chimbiya, May 20, 2020.

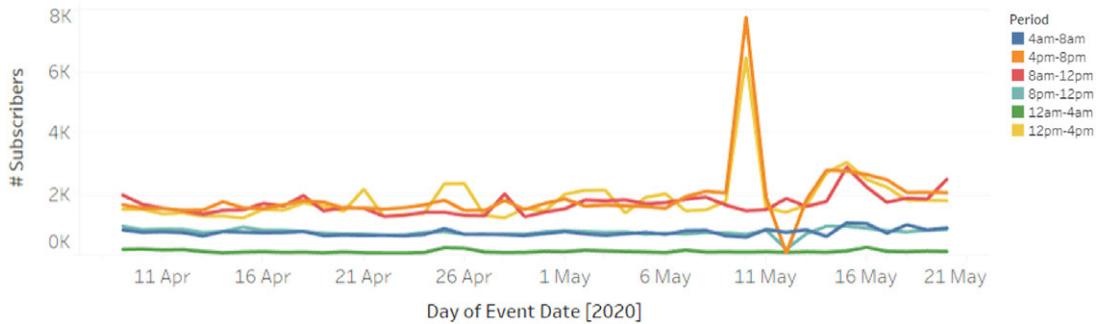


Figure 4. Example of a nonrecurring, potential mass gathering, Mzuzu, May 10, 2020.

primary prevention strategy across the globe. Furthermore, it is important to understand risk of transmission, particularly in the early phases of an outbreak when risk is not uniformly distributed geographically. Understanding geographic risk allows public health officials to give communities the correct level of guidance to mitigate possible transmission, while weighing the indirect negative effects of policies, such as shelter-in-place and banned gatherings, can have on mental and economic well-being—especially critical in a largely subsistence economy like Malawi.

Each of the analyses in this study has been automated and integrated into available dashboards for policymakers and surveillance staff to monitor trends and to assist in informed decision-making. While these tools were under development and testing during the early days of the epidemic, with limited availability or use, they were fully available to leverage during the intense second wave which occurred in late 2020 and early 2021 (Salyer et al., 2021; UNICEF, 2021).

To date, there are no data available on population-wide mobility or estimates of physical distancing compliance in Malawi. In high-income settings, researchers and public health officials have leveraged mobile phone mobility estimates from Apple and Google; however, no such estimates are available in Malawi due to limited data via low smartphone utilization levels (Apple, 2020; Google, 2020). In this analysis, we bridge this data gap through the development of an automated data pipeline which aggregates CDR data on a daily basis from a mobile network operator and produces succinct and clear reports for public health officials—all while preserving individual privacy.

The daily mass gathering reports allow health officials to understand where large groups of individuals have likely gathered which could have led to high levels of transmission or even become a superspreading event. The Public Health Institute of Malawi and their Emergency Operations Center have scarce testing and tracing resources, and these data are essential to understand which subdistricts might be of most concern, while reflecting also on recent test positivity data. Furthermore, the reports provide data on where potential attendees traveled to after the event, which creates a potential trail of infections, which are of particular interest especially for subdistricts which have had no known infections to date and could suddenly be at risk of community transmission.

These data also served to fill data gaps in mathematical modeling efforts, a key undertaking to ascertain the current and future burden of disease, and counterfactually what might happen under various policy scenarios. Namely, prior to the utilization of MNO data, mathematical models in Malawi assumed either universal concurrent spread and mixing geographically, or that spread radiated to nearest adjacent TAs first, without regard for transit corridors and connectivity. The aggregated CDR data in this study allowed modelers to ascertain to what extent TAs are physically connected by travelers, and quantify that into an estimate of risk, therefore improving the accuracy and relevance of infectious disease models. The official COVID-19 mathematical model used by the Malawian Ministry of Health leverages the CDR-derived epidemic start dates for TAs in its calibration.

Finally, an additional value of these data is their potential policy and research use outside of COVID-19. There are several applications including improving disaster management scenarios, understanding

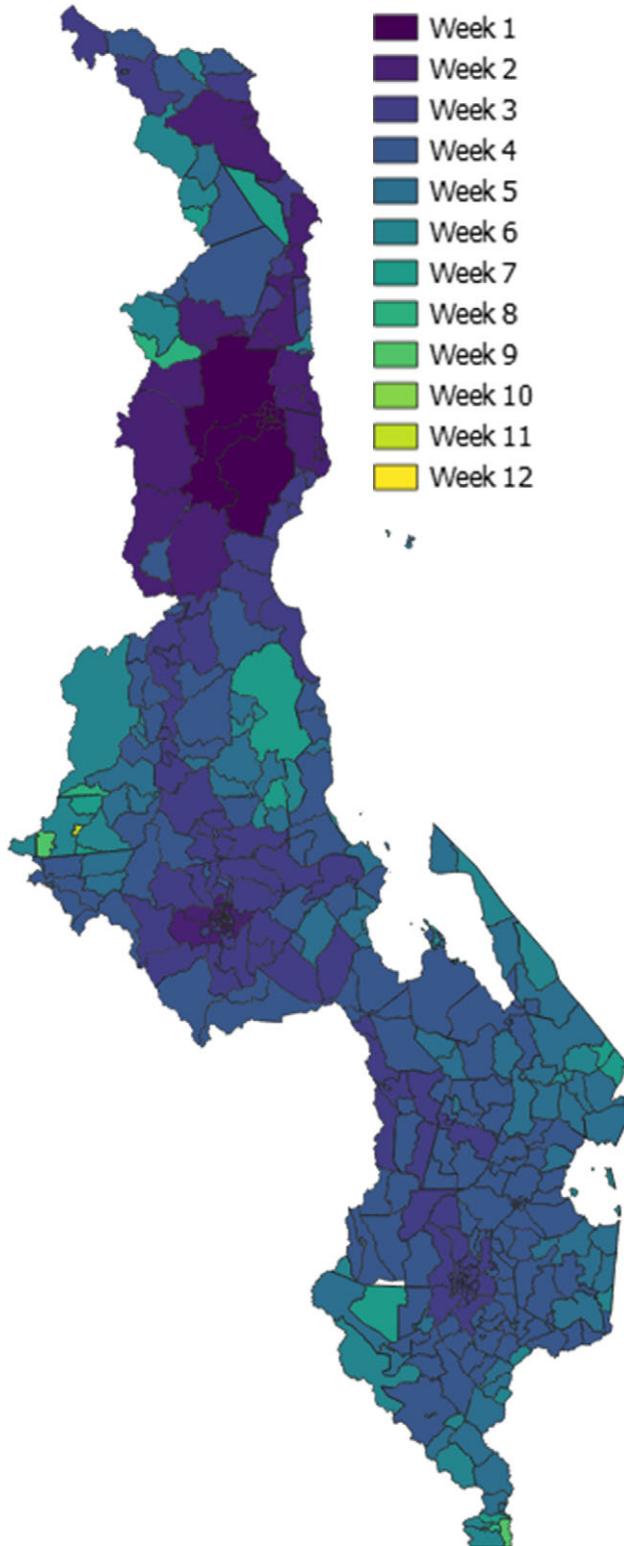


Figure 5. Map of modeled epidemic start weeks by TA. Darker color represents earlier COVID-19 spread and Epidemic Week 1 begins on April 26, 2020.

resiliency, improving equity and access to health services, better understanding mobile populations like seasonal laborers, and understanding social and sexual networks. As this data pipeline is now established, additional research and use can lead to improvements in policymaking, and ultimately welfare of Malawians who may benefit from better decision analysis.

Our study has several limitations. First, our analysis relies on a presumption that a mobile phone's data are representative of a single person. This is often not the case in Malawi, and in many low-income nations. Phone sharing is common with individuals sharing their mobile devices with friends, family, and their community, which represents a discontinuity in ownership and representation in truth—but not our dataset. One study estimates that as many as 40% of Malawians with mobile phones report sharing their device (Marron et al., 2020). Similarly, SIM card swapping is common, as individuals might change mobile-network providers, particularly when special pricing may be available. Another challenge is how movement with mobile phones is defined in our study. In urban areas, there are higher concentrations of cell towers to meet underlying demand, and therefore even small movements may be easily captured in the CDR data. However, in rural and sparsely populated areas, a single tower may cover large areas of land, meaning individuals in that range may travel large distances without detection in our dataset. Finally, there are substantial selection effects in our data. Mobile phone ownership is not universal in Malawi with nearly 50% reporting owning a phone, but this skews male, urban, educated, and young (GSMA, 2012; Steinfield et al., 2015). The possible spatiotemporal fidelity of CDR data is reliant on the frequency of phone use among subscribers, which can vary significantly relative to individual's preferences, technology uptake, socioeconomic status, and social network (Hoteit et al., 2017).

Privacy concerns and ethical challenges still exist with these data, despite aggregation, anonymization, and best security practices. We made some sacrifices to the detail of the data to further preserve privacy. For example, we chose to attribute each mobile event to a 4-hr time block, rather than the exact time, as an individual tracked under physical surveillance, coupled with precise time-stamp CDR data could be identified, even with the anonymization process. Furthermore, there are concerns that even the mass gathering data could be misused, for political reasons, which was of particular concern in early to mid 2020 when reelelections were ongoing. These reports could be used inappropriately to identify large political gatherings of incumbent or opposition political affiliates. These are valid concerns to remain cognizant of and keep top of mind, including maintaining close oversight of even the aggregated report data and its use.

Despite these challenges, we believe these data are invaluable to government officials who are seeking information to formulate an effective COVID-19 response which minimizes direct harms due to COVID-19, while also not causing excessive indirect harms through disruption of routine medical services or economic upheaval. We have used these data to understand the frequency, geographic location, and size of mass gatherings which could be substantial transmission events which require public health action including improved health communication and testing. Furthermore, we demonstrate the capability to estimate population-level mobility and use those data for infectious disease modeling which also inform policy response and resource allocation. Aggregated CDR data can be feasibly analyzed on a routine basis in a way that is accessible to policymakers, even in resource- and data-limited settings.

Supplementary Materials. To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/dap.2021.14>.

Data Availability Statement. The primary data used to generate these results, the CDRs, are not publicly available nor are they released as part of this analysis for a variety of reasons, including the preservation of privacy of mobile phone users and their sensitive information as well as ensuring the security of any mobile network market insights which could be gleaned from the data and used as a commercial advantage by TNM's competitors.

Author Contributions. Conceptualization, T.S., R.S., A.L., and M.M.; Methodology, T.S., D.G., M.M., B.K., S.A., and N.S.; Formal analysis, D.G., M.M., S.A., and N.S.; Data curation, M.M., S.A., C.M., and D.G.; Writing-original draft, D.G. and G.F.; Writing-review & editing, D.G., G.F., M.M., B.K., and T.S.; Supervision, A.L., T.S., C.M., and G.F.; Funding acquisition, T.S., G.F., and R.S.

Competing Interests. We declare no competing interests for this study. Neither the funder nor TNM (the data provider) had a role in designing the study, the data analytics, the report development for the ministry, or the writing of the manuscript.

Funding Statement. The funding source for this study is the United Nations Foundation Digital Impact Alliance, grant number UNF-20-1113.

References

- Africa Center for Strategic Studies** (2020) Coronavirus Spreads Through Africa. Available at <https://africacenter.org/spotlight/coronavirus-spreads-through-africa/>.
- Agence France Presse Staff** (2020) Malawians Flee “Chaotic” Isolation Sites After Repatriation. *Barrons*. Available at <https://www.barrons.com/news/malawi-virus-cases-and-suspects-flee-isolation-sites-after-repatriation-01590594303>.
- Ahmed N, Michelin RA, Xue W, Ruj S, Malaney R, Kanhere SS, Seneviratne A, Hu W, Janicke H and Jha SK** (2020) A survey of COVID-19 contact tracing apps. *IEEE Access* 8, 134577–134601. <https://doi.org/10.1109/ACCESS.2020.3010226>.
- Al Jazeera Staff** (2020) Malawi High Court Blocks Coronavirus Lockdown. *Al Jazeera*. Available at <https://www.aljazeera.com/news/2020/04/malawi-high-court-blocks-coronavirus-lockdown-200417184430403.html>.
- Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, Liu X, Wei L, Truelove SA, Zhang T, Gao W, Cheng C, Tang X, Wu X, Wu Y, Sun B, Huang S, Sun Y, Zhang J and Feng T** (2020) Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: A retrospective cohort study. *The Lancet Infectious Diseases* 20(8), 911–919. [https://doi.org/10.1016/S1473-3099\(20\)30287-5](https://doi.org/10.1016/S1473-3099(20)30287-5).
- Blondel VD, Decuyper A and Krings G** (2015) A survey of results on mobile phone datasets analysis. *EPJ Data Science* 4(1), 1–55. <https://doi.org/10.1140/epjds/s13688-015-0046-0>.
- Booth A, Reed AB, Ponzo S, Yassae A, Aral M, Plans D, Labrique A and Mohan D** (2021) Population risk factors for severe disease and mortality in COVID-19: A global systematic review and meta-analysis. *PLoS One* 16(3), e0247461. <https://doi.org/10.1371/journal.pone.0247461>.
- Carlitz RD and Makhura MN** (2020) Life under lockdown: Illustrating tradeoffs in South Africa’s response to COVID-19. *World Development* 137, 105168. <https://doi.org/10.1016/j.worlddev.2020.105168>.
- Chilunga Z** (2020) Mutharika Lays Out Malawi “Response Plan” on Coronavirus: Bans Gatherings of 100 People, Schools Closing. *Nyasa Times Malawi*. Available at <https://www.nyasatimes.com/mutharika-lays-out-malawi-response-plan-on-coronavirus-bans-gatherings-of-100-people-schools-closing/>.
- Chitete S** (2020) Malawi Top Court Annuls Presidential Election Results. *Al Jazeera*. Available at <https://www.aljazeera.com/news/2020/02/malawi-top-court-annuls-presidential-election-results-200203060112731.html>.
- Cooper/Smith** (2020a) Can We Do More to Scale COVID-19 Testing in Africa? *Medium*. Available at <https://medium.com/cooper-smith/can-we-do-more-to-scale-covid-19-testing-in-africa-6ee9eace51ca>.
- Cooper/Smith** (2020b) COVID-19 Cases in Kenya Are Rising Fast and We Don’t Know Why. *Medium*. Available at <https://medium.com/cooper-smith/covid-19-cases-in-kenya-are-rising-fast-and-we-dont-know-why-2bfa3bed23e0>.
- Cooper/Smith; DIAL; Infosys** (2019) Using Mobile Phone Data to Make Policy Decisions—A Study in How New Data Sources Optimized Health Facility Placement in Malawi. Available at https://static1.squarespace.com/static/548487dce4b08bf981fe60d5t/5df906a4635ee00d66cc082/1576601277654/DIAL-CooperSmith_UsingMNOData-Malawi_FINAL.pdf.
- Dong E, Du H and Gardner L** (2020) An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* 20(5), 533–534. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1).
- Endo A, Abbott S, Kucharski AJ and Funk S** (2020) Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Research* 5, 67. <https://doi.org/10.12688/wellcomeopenres.15842.3>.
- Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, Parker M, Bonsall D and Fraser C** (2020) Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 368(6491), eabb6936. <https://doi.org/10.1126/science.abb6936>.
- Fiko M** (2020) Malawi Confirms Three Cases of Coronavirus: President Mutharika Calls for Calm. *Malawi Nyasa Times*. Available at <https://www.nyasatimes.com/malawi-confirms-three-cases-of-coronavirus-president-mutharika-calls-for-calm/>.
- Gauravaram P** (2012) Security analysis of salt|password hashes. In *Proceedings of the 2012 International Conference on Advanced Computer Science Applications and Technologies, ACSAT 2012*, pp. 25–30. <https://doi.org/10.1109/ACSAT.2012.49>.
- GSMA** (2012) Telekom Networks Malawi (TNM) Ltd. Feasibility Study. Available at <https://www.gsma.com/mobilefordevelopment/resources/telekom-networks-malawi-tnm-ltd-malawi-feasibility-study/>.
- GSMA** (2014) GSMA Guidelines on the Protection of Privacy in the Use of Mobile Phone Data for Responding to the Ebola Outbreak. Available at https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2014/11/GSMA-Guidelines-on-protecting-privacy-in-the-use-of-mobile-phone-data-for-responding-to-the-Ebola-outbreak_October-2014.pdf.
- GSMA** (2018) Digital Identity Country Report—Malawi. Available at <https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/02/Digital-Identity-Country-Report.pdf>.
- Google** (2020) COVID-19 Community Mobility Reports, <https://www.google.com/covid19/mobility/>.
- Hoteit S, Chen G, Viana A, Fiore M and Viana AC** (2017) Spatio-temporal completion of call detail records for human mobility analysis. *HAL Open Archives*. Available at <https://hal.archives-ouvertes.fr/hal-01516717>.
- Jones KH, Daniels H, Heys S and Ford DV** (2018) Challenges and potential opportunities of mobile phone call detail records in health research: Review. *JMIR mHealth and uHealth* 6(7), e161. <https://doi.org/10.2196/mhealth.9974>.

- Jones KH, Daniels H, Heys S and Ford DV** (2019) Toward an ethically founded framework for the use of mobile phone call detail records in health research. *JMIR mHealth and uHealth* 7(3), e11969. <https://doi.org/10.2196/11969>.
- Joseph A** (2020). Contact Tracing May Help Avoid Another Lockdown. Can It Work in the U.S.? *STAT*. Available at <https://www.statnews.com/2020/05/29/contact-tracing-can-it-help-avoid-more-lockdowns/>.
- Kondowe R** (2020). Malawi's Do-Over Election Fraught with Tensions, Covid Fears. *Quartz Africa*. Available at <https://qz.com/africa/1872200/malawis-do-over-election-fraught-with-tensions-covid-fears/>.
- Li J, Huang DQ, Zou B, Yang H, Hui WZ, Rui F, Yee NTS, Liu C, Nerurkar SN, Kai JCY, Teng MLP, Li X, Zeng H, Borghi JA, Henry L, Cheung R and Nguyen MH** (2021) Epidemiology of COVID-19: A systematic review and meta-analysis of clinical characteristics, risk factors, and outcomes. *Journal of Medical Virology* 93(3), 1449–1458. <https://doi.org/10.1002/jmv.26424>.
- Marron O, Thomas G, Burdon Bailey JL, Mayer D, Grossman PO, Lohr F, Gibson AD, Gamble L, Chikungwa P, Chulu J, Handel IG, De C Bronsvoot BM, Mellanby RJ and Mazeri S** (2020) Factors associated with mobile phone ownership and potential use for rabies vaccination campaigns in southern Malawi. *Infectious Diseases of Poverty* 9(1), 62. <https://doi.org/10.1186/s40249-020-00677-4>.
- Masina L** (2020) Malawi COVID-19 Cases Rise as Citizens Return from South Africa. *Voice of America*. Available at <https://www.voanews.com/africa/malawi-covid-19-cases-rise-citizens-return-south-africa>.
- Ministry of Health—Malawi** (2021) COVID-19—National Information Dashboard. Available at <https://covid19.health.gov.mw/>.
- Munzert S, Selb P, Gohdes A, Stoetzer LF and Lowe W** (2021) Tracking and promoting the usage of a COVID-19 contact tracing app. *Nature Human Behaviour* 5(2), 247–255. <https://doi.org/10.1038/s41562-020-01044-x>.
- National Statistics Office** (2018) *Malawi Population and Housing Census Main Report*.
- Nordling L** (2020) The pandemic appears to have spared Africa so far. Scientists are struggling to explain why. *Science*. <https://doi.org/10.1126/science.abe2825>.
- Nyasulu JCY, Munthali RJ, Nyondo-Mipando AL, Pandya H, Nyirenda L, Nyasulu PS and Manda S** (2021) COVID-19 pandemic in Malawi: Did public social-political events gatherings contribute to its first wave local transmission? *International Journal of Infectious Diseases* 106, 269–275. <https://doi.org/10.1016/j.ijid.2021.03.055>.
- Porter G** (2002) Living in a walking world: Rural mobility and social equity issues in sub-Saharan Africa. *World Development* 30(2), 285–300. [https://doi.org/10.1016/S0305-750X\(01\)00106-1](https://doi.org/10.1016/S0305-750X(01)00106-1).
- Salyer SJ, Maeda J, Sembuche S, Kebede Y, Tshangela A, Moussif M, Ihekweazu C, Mayet N, Abate E, Ouma AO and Nkengasong J** (2021) The first and second waves of the COVID-19 pandemic in Africa: A cross-sectional study. *The Lancet* 397(10281), 1265–1275. [https://doi.org/10.1016/S0140-6736\(21\)00632-2](https://doi.org/10.1016/S0140-6736(21)00632-2).
- Sneppen K, Nielsen BF, Taylor RJ and Simonsen L** (2021) Overdispersion in COVID-19 increases the effectiveness of limiting nonrepetitive contacts for transmission control. *Proceedings of the National Academy of Sciences of the United States of America* 118(14). <https://doi.org/10.1073/pnas.2016623118>.
- Steinfeld C, Wyche S, Chivasa H and Cai T** (2015) *The Mobile Divide Revisited: Mobile Phone Use by Smallholder Farmers in Malawi*. ACM International Conference Proceeding Series, 15. New York: Association for Computing Machinery. <https://doi.org/10.1145/2737856.2738022>.
- Stevens M, Bursztein E, Karpman P, Albertini A and Markov Y** (2017) The first collision for full SHA-1. In Katz J and Shacham H (eds), *Advances in Cryptology—CRYPTO 2017*. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10401. Cham: Springer, pp. 570–596. https://doi.org/10.1007/978-3-319-63688-7_19.
- The Apple** (2020) citation is: Apple Maps COVID-19 Mobility Trends Report, <https://covid19.apple.com/mobility>.
- UNICEF** (2021) *Malawi COVID-19 Situation Report, 13 January 2021*.
- University of Oxford** (2020) Coronavirus Government Response Tracker, *Blavatnik School of Government*. Available at <https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker>.
- WHO Africa** (2021) New COVID-19 Variants Fuelling Africa's Second Wave. Available at <https://www.afro.who.int/news/new-covid-19-variants-fuelling-africas-second-wave>.
- Wikipedia** (2021a) COVID-19 Apps. Available at https://en.wikipedia.org/wiki/COVID-19_apps.
- Wikipedia** (2021b) Timeline of First Confirmed Case by Country. Available at https://en.wikipedia.org/wiki/COVID-19_pandemic_by_country_and_territory#Timeline_of_first_confirmed_case_by_country.
- WorldPop** (2020) Unconstrained 1km Population Counts. Available at <https://www.worldpop.org/project/categories?id=3>.
- Zeller M, Gangavarapu K, Anderson C, Smither AR, Vanchiere JA, Rose R, Dudas G, Snyder DJ, Watts A, Matteson NL, Robles-Sikisaka R, Marshall M, Feehan AK, Sabino-Santos G, Bell-Kareem A, Hughes LD, Alkuzweny M, Snarski P, Garcia-Diaz J, Scott RS, Melnik LI, Klitting R, McGraw M, P Belda-Ferre, DeHoff P, Sathe S, Marotz C, Grubaugh N, Nolan DJ, Drouin AC, Genemaras KJ, Chao K, Topol S, Spencer E, Nicholson L, Aigner S, Yeo GW, Farnaes L, Hobbs CA, Laurent LC, Knight R, Hodcroft EB, Khan K, Fusco DN, Cooper VS, Lemey P, Gardner L, Lamers SL, Kamil JP, Garry RF, Suchard MA, Andersen KG** (2021) Emergence of an Early SARS-CoV-2 Epidemic in the United States. *MedRxiv* 32. <https://doi.org/10.1101/2021.02.05.21251235>.

Cite this article: Green D, Moszczynski M, Asbah S, Morgan C, Klyn B, Foutry G, Ndira S, Selman N, Monawe M, Likaka A, Sibande R and Smith T (2021). Using mobile phone data for epidemic response in low resource settings—A case study of COVID-19 in Malawi. *Data & Policy*, 3: e19. doi:10.1017/dap.2021.14