Correspondence and
requests for materials
should be addressed to
A.W. (awesolow@
hsph.harvard.edu)

Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones

Amy Wesolowski[1,2,3], Gillian Stresman[4], Nathan Eagle[2,5], Jennifer Stevenson[6,7,8], Chrispin Owaga[7], Elizabeth Marube[7], Teun Bousema[4,9], Christopher Drakeley[4], Jonathan Cox[6] & Caroline O. Buckee[2,3]

[1]Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, USA, [2]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA, [3]Center for Communicable Disease Dynamics, Harvard School of Public Health, Boston, MA, USA, [4]Department of Immunology and Infection, London School of Hygiene and Tropical Medicine, London, UK, [5]College of Computer and Information Science, Northeastern University, Boston, MA, USA, [6]Department of Disease Control, London School of Hygiene and Tropical Medicine, London, UK, [7]Centre for Global Health Research, Kenya Medical Research Institute/ Centers for Disease Control and Prevention, Kisumu, Kenya, [8]Johns Hopkins Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA, [9]Radboud University Nijmegen, Nijmegen, 6525 GA, Netherlands.

Human travel impacts the spread of infectious diseases across spatial and temporal scales, with broad implications for the biological and social sciences. Individual data on travel patterns have been difficult to obtain, particularly in low-income countries. Travel survey data provide detailed demographic information, but sample sizes are often small and travel histories are hard to validate. Mobile phone records can provide vast quantities of spatio-temporal travel data but vary in spatial resolution and explicitly do not include individual information in order to protect the privacy of subscribers. Here we compare and contrast both sources of data over the same time period in a rural area of Kenya. Although both data sets are able to quantify broad travel patterns and distinguish regional differences in travel, each provides different insights that can be combined to form a more detailed picture of travel in low-income settings to understand the spread of infectious diseases.

Improvements in transportation infrastructure and increasing human mobility are enabling unprecedented connectivity between populations at both local and global scales, allowing for the rapid dissemination of pathogens[1–6]. Humans are able to introduce diseases into immunologically naïve populations through direct transmission or by introducing them into the environment[1,7–9], and travel plays a critical role in the spatial spread of influenza, polio, cholera, and dengue, as well as in the spatial spread of drug resistance among pathogens such as malaria[2,3,5,10–15]. Quantifying population travel dynamics is difficult, however, particularly in low-income countries where individual level data sets that include information about travel behavior are difficult to obtain and collect.

Traditionally, travel history questions from household surveys or from census data have provided the most comprehensive source of travel information[16]. During these surveys, which often include data on variables such as age, sex, income, household structure, health status, or ethnicity, for example, individuals are asked questions about their movement patterns. Surveys therefore provide insights into the demographic biases and motivations underlying movement patterns. However, these data sets often only sample a small subset of the population and may be subject to recall bias. Moreover, these questions are typically nested in larger surveys with disparate objectives that may impact their generalizability and oversample individuals of interest to the larger survey objective, for instance they may be part of country wide Malaria Indicator Surveys, questions asked during hospitalization, or household budget surveys[16–20]. The most common source of travel survey data in Africa is records from a national or micro-census, but these typically address only long-term changes in residence[21,22].

In contrast, anonymized mobile phone usage data have recently been shown to provide a valuable source of information on regular movement patterns on various spatial scales[14,15,23–25]. Call detail records (CDRs) store locational information for each subscriber when they make a call or send a SMS (Short Message Service), providing a detailed temporal and spatial picture of often millions of people. Due to privacy concerns and pre-paid plans, individual socio-demographic data about subscribers are unavailable to researchers. Analysis from previous work has shown that mobile phone ownership is biased towards wealthy, urban males,
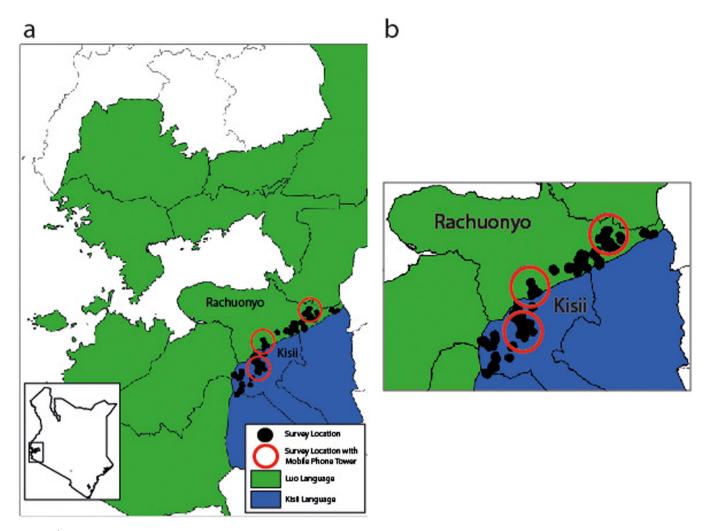
Figure 1 | The household survey locations within the study site in western Kenya. (a) Surveys were taken at households within western Kenya (larger map is highlighted in the inset, created using ArcGIS v10.1) with their locations mapped as black points. Households within 3 km of a mobile phone tower are outlined in red. Areas are colored by their dominant language with DhoLuo (Luo language) in green and Kisii in blue. In Rachuonyo district, the dominant language is DhoLuo whereas it is Kisii in Kisii district. (b) A zoomed image of the study site.

despite remarkable levels of ownership across all income brackets in Kenya, for example[26]. Furthermore, phone sharing practices may hinder the use of mobile phone data to accurately capture individual level inferences about movement patterns[26]. Nevertheless, we have shown that these biases can be corrected for and are unlikely to impact the routes and relative volumes of travel between most populations[27].

We have previously quantified intra-national travel patterns from nearly 15 million mobile phone subscribers in Kenya on a range of spatial and temporal scales using mobile phone data, with a particular emphasis on the role of travel in the importation of malaria parasites across the country[15]. The volumes and direction of travel varied seasonally, and depended on both the origin and destination locations, with a large amount of travel occurring to and from the capital city, Nairobi. Here we compare a subset of these data with information from a detailed survey about travel from the same time and place, collected during cross-sectional surveys of 2,650 individuals in two districts in western Kenya. The travel survey was conducted as part of a study to characterize patterns of malaria transmission and risk factors for infection in an area of low malaria endemicity. We describe patterns of travel, highlight the differences and strengths in each data set and discuss how the data sets can be used in conjunction to enhance their utility.

Results

Travel history results from community surveys. Travel data was collected as part of a malariometric survey conducted in February 2009 and covered 2,650 (0.13% of the population in the survey sites N = 19,744) individuals in two districts: Kisii Central (formerly part of Kisii district) and Rachuonyo South (formerly part of Rachuonyo) (see Figure 1). Despite being predominantly rural, these districts have relatively high population densities (707 (Kisii) and 705 (Rachuonyo) individuals/km², total populations: 457,105 and 307,126 individuals from the national census) that characterize the regions surrounding Lake Victoria. The individuals included in the study are from the rural parts of both districts. Kisii is primarily made up of the Kisii ethnic group whereas Rachuonyo is primarily made up of the Luo ethnic group (see Figure 1). Travel surveys provided general socio-demographic information (see Table 1) that was used to analyze travel patterns stratified by age, gender, and other covariates.

One of the most striking findings in the travel survey was that the vast majority of people (90%, N = 2,388) reported that they had not made an overnight trip to another district within the last 3 months (see Table S1). More individuals within households in Rachuonyo reported traveling more often than those in Kisii (see Figure 2). When individuals did travel, they reported spending the majority of their time in neighboring districts or those including a major city, predominantly Nairobi (Figure 3, Table S2). The primary motivations for

Table 1 | Basic survey descriptive statistics. All percentages (sample size) do not necessarily add to 100 if the survey respondent did not answer the question

	Travel Survey
Number of individuals	2650
% Male	46 (1222)
% Female	53 (1398)
% Adults (15+)	45 (1194)
% Children (0–14)	55 (1456)
Survey Date	February, 2009
Number of households	776
Average household size	3.7

travel were either visiting family or friends (54%, N = 105) or attending a funeral (17%, N = 46) (see Tables S3–S4). Of those who traveled, most reported taking only a single trip that had most often occurred less than four weeks ago (64%, N = 125 see Table S5–S6, see Figure 4). Of adults (aged 15 or older) who had traveled, men were slightly more likely to have taken an overnight trip (males: 13% = 70/525, females 11% = 83/657, $x^2 = 2.3889$, p = 0.6646). The destinations for travel were primarily the same for both men and women, although men reported that they traveled to Nairobi more often than women (16% = 11/70 versus 7% = 6/83, $x^2 = 29$, p < 0.001) (see Table S7). Children (under 15 years of age) were less likely to travel than adults (3% (42/1318) of children had taken an overnight trip).

Less than half (47%, N = 366) of households reported having a mobile phone. Mobile phone ownership (on a household level) was positively correlated with the likelihood of reporting having traveled (see Table 2). The percentage of households where at least one person reported traveling was 60% (83/138) in households with a mobile phone versus 40% (55/138) ($x^2 = 10.72$, p = 0.001) for those without a mobile phone.

Mobile phone data analysis. We analyzed CDRs using methods previously described[15] (see Materials and Methods), identifying 34,861 subscribers (4.6% of the total population in these districts assuming each subscriber is an individual) in the region (see Materials and Methods). Briefly, cell tower locations were assigned to districts, demarcated by political boundaries. Using a daily time series of tower locations over the course of the data set, subscribers whose most used mobile phone tower was within 3 km, the typical service range of the study site, were considered (see Materials and Methods). During the three-month study time period corresponding to the travel survey, movement between districts was quantified.

In contrast to the travel survey, we inferred from the CDRs that the vast majority of mobile phone subscribers had spent at least one night outside Kisii and Rachuonyo districts during the time frame of the survey (61% from Kisii, 95% from Rachuonyo, in total 27,668 subscribers, see Table 3). As observed in the survey data, subscribers from Rachuonyo traveled more than those from Kisii, possibly related to the geographic distribution of the Luo ethnic group. We excluded travel between Kisii and Rachuonyo because many cell towers lie on the border between the two districts, making it difficult to separate travelers within this sub-region. Half of subscribers traveled for at least 2 days away from Kisii and Rachuonyo to other districts (36% from Kisii, 63% from Rachuonyo, 17,560 subscribers, see Table 3). Thus, we estimate that between 17,560 (two days or more) and 27,668 (one night or more) subscribers traveled to other districts during the study time frame. Including travel lasting at least one night, subscribers took a total of 13,860 trips. These trips were often short with 65% lasting less than three days (see Table S8).

Comparing travel between data sources. Given the wide divergence in terms of the magnitude of travel between the two data sets, we calculated an adjustment to compare the two data sets (see Table S9). The survey sites had a collective population of 19,744 individuals when accounting for the total enumerated population for the areas that represented the survey clusters. Using the survey data, we estimated that between 2,500 and 11,500 mobile phone subscribers were located in the study site at the time, with the range determined by the estimated number of subscribers per household (see Supplementary Information and Table 4). This value is up to one order of magnitude less than the number of mobile phone subscriber IDs we have included in the analysis, indicating that i) we may be capturing subscribers who reside in neighboring areas in our CDR analysis, ii) individuals own multiple mobile phones or SIM cards, and/or iii) estimates from the two data sources are extremely different.

We cannot address this last option, but it seems unlikely that on average each individual owned 5 SIM cards. Furthermore, even if we assume that all mobile phone subscribers were adult men, since they represent the most mobile demographic group, at most 16% of men reported traveling away from their home district in the survey. This would correspond to 200 to 1,800 mobile phone subscribers within the study site traveling (see Supplementary Information, Tables S10–S11). Since this value is orders of magnitude less than measured number of trips by mobile phone subscribers (approximately 28,000, see Table 4), the two sources of data remain markedly different in their estimates of the number of travelers, although both were able to identify the main districts where people travel.
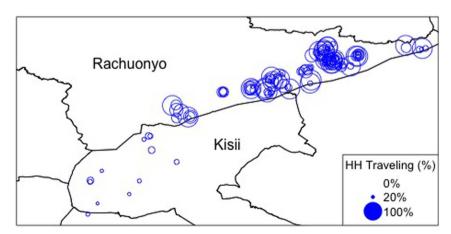

Figure 2 | **The percentage of individuals within a household who reported traveling.** From the travel survey data, the percentage of individuals per household who reported traveling was quantified. Households within Rachuonyo traveled much more than those in Kisii (t = −7.401, df = 410.141, p-value < 0.001). This map was created using ArcGIS v10.1.
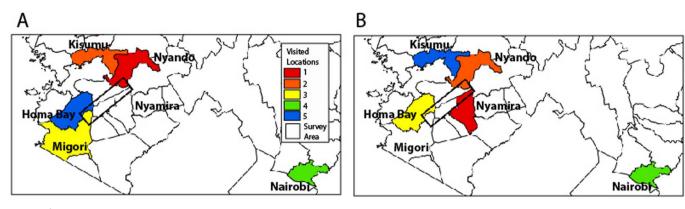
3

Figure 3 | The locations of the most commonly visited districts. From the (A) survey data and (B) mobile phone data, the five most commonly visited districts are colored by their rank with the survey area outlined in black. The most common districts visited were Nyamira, Nyando, Homa Bay, Nairobi, Kisumu, and Migori, also primarily nearby districts and those including major population centers (Kisumu and Nairobi) (in descending order). This did vary slightly between Kisii and Rachuonyo. Amongst subscribers in Kisii the districts most commonly visited were: Nyamira, Nairobi, Gucha, and Migori whereas those in Rachuonyo commonly visited Nyamira, Nyando, Homa Bay, and Kisumu. The map was created using ArcGIS v10.1.

We next compared the percentage of individuals taking between one and 60 trips from each data set. In general, individuals from the survey data took a fewer number of trips than the mobile phone data would suggest (see Figure 4). Possible reasons for the discrepancies between the two data sets include recall bias or misreporting in the travel surveys, differences in the populations represented in each data set, and mobile phone sharing practices. We hypothesize that the first is highly likely, and although the last two are possible, they cannot account for the entirety of the difference[9,27,28]. It is likely, therefore, that all three of these contribute to varying degrees and actual travel falls somewhere between the two estimates.

Impact of travel estimates on predictions about malaria exchange. One of the most important reasons to quantify human mobility is in the assessment of the spread of disease in the region, including malaria. Previously, we quantified malaria (*Plasmodium falciparium*) importation within Kenya using mobile phone data[15] and spatial *P. falciparium* ($PfPR_{2-10}$) prevalence data from the Malaria Atlas Project[30]. Using a simplified metric that does not require as detailed data as in[15], we used a measure of malaria exchange (as opposed to malaria importation) that utilizes population-weighted travel as well as prevalence data (see Materials and Methods and Supplementary Information)[29]. In particular, this metric does not require information on the duration of travel since it is unavailable in the survey data. This measure describes the estimated exchange of malaria parasites adjusted based on the prevalence data between two locations. It almost certainly overestimates the impact of travel, since we use the higher parasite rate found in children age 2–10 years old, but illustrates the possible range of importation of parasites to and from the region.

For travelers from Kisii and Rachuonyo, the mobile phone data produces total malaria exchange estimates an order of magnitude greater than the survey data, in this case comparing the total number of travelers from both data sets (see Tables 5, S12). Both data sets predict that the amount of malaria being brought into Rachuonyo is much greater than into Kisii, and were both able to identify the major routes. Mobile phone data predicted that malaria exchange occurs between nearly all districts. However, the community survey data suggest that malaria parasites are likely to predominantly come from a few districts (see Supplementary Information). These findings have important implications for targeted surveillance in the region, since the overall volume and locations contributing to malaria exchange may be a more important consideration for control programs than travel surveys would indicate. We propose that while travel surveys provide important information about motivations for travel, the type of people who are traveling, and identify the main travel destinations, they are also likely to under-estimate the volume and range of mobility (see Supplementary Information).
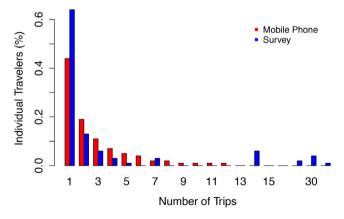

Figure 4 | The number of trips taken by individuals from each data source. The distribution of the number of trips (between 1–60 trips) taken by individuals who traveled from the mobile phone data (red) and the survey data (blue) is shown. In the survey, individuals rarely reported taking more than one trip, whereas in the mobile phone data multiple trips were measured from a substantial number of subscribers (KS statistic: 0.7947, p = 0.0005).

Table 2 | The percentage of adults within a household who travel versus the percentage of those households who own a mobile phone. For households (HH) where 0–100% of the adults in the HH have traveled, the percentage of those HH who own a mobile phone. In general, the households where a higher percentage of adults have traveled are more likely to own a mobile phone (Chi-Squared = 15.49, df = 6, p-value = 0.0168)

Percentage of adults within a HH who have traveled	Percentage of those HH who own a mobile phone
0	45 (277)
20	75 (3)
25	100 (2)
33	50 (5)
50	64 (37)
67	100 (3)
100	56 (29)

Table 3 | The basic travel statistics from the mobile phone data

	Kisii	Rachuonyo
Number of mobile phone towers within study site	6	9
Number of subscribers	16,196	18,665
Number of travelers, trips lasting at least 1 day	61% (N = 9,880)	95% (N = 17,732)
Number of travelers, trips lasting at least 2 days	36% (N = 5,830)	63% (N = 11,759)

Discussion

Overall, the community survey provided a snapshot of travel behavior for 2,650 individuals. The volume of travel reported from the surveys was considerably lower than that captured by mobile phone data. It is possible that mobile phone subscribers were simply not captured by the survey, since working age men are often absent during community surveys. Other possible reasons for under-reporting of travel include recall problems of interviewed individuals; details about trips taken may be forgotten or when trips were taken not accurately reported. Lack of knowledge of, or recent changes in administrative boundaries may also result in underreporting of travel. Surveys are challenging to conduct on a large scale and it is not feasible to sample the majority of residents within even small geographic areas. Cross-sectional surveys can only collect travel data for each individual at one point in time and therefore do not provide a dynamic picture of overall movement patterns. For example, Nyamira district was once part of Kisii district and this may have caused confusion in the travel survey that would not be observed in the mobile phone data.

Mobile phone data enables researchers to estimate travel patterns for a large sample of the population over time, but can only provide an estimate of travel for mobile phone subscribers and is limited by mobile phone tower density. Community surveys are able to compliment mobile phone data by approximating travel patterns of non-subscribers. Here we used anonymized CDRs where every subscriber is assigned a unique ID. Subscriber IDs may not reflect individuals due to phone sharing and/or multiple SIM card ownership[26]. Subscribers also represent a biased sample of the general population, with ownership more prevalent among more educated, urban, males[26]. However, based on the results from the travel survey, it appears that those households that do not own a mobile phone are also less likely to travel, so bias of estimates due to skewed mobile phone ownership may not be as large as previously thought.

Interestingly, it appears that in this setting ethnicity influences travel behavior. From both data sets, we observed that those living in Rachuonyo travel more than those in Kisii (see Figure 2). Rachuonyo is predominantly Luo whereas Kisii is predominantly Kisii[31]. The large geographic coverage of the Luo ethnic group (see Figure 1) may go some way to explain this. The main reasons for travel given during the surveys were to visit family and friends or attend a funeral, both of which are more likely to have strong ethnic influences. However, at present we can only suggest this as a possible explanation. Aside from ethnicity, road access and travel times to other districts may also impact travel and we suggest that this should be investigated in future work.

Quantifying human travel patterns can have broad applications in epidemiology, particularly the spatial spread of infectious diseases. Being able to accurately parameterize movement patterns will be invaluable in identifying areas that are at risk of re- or continued importation of disease, which has major implications for control and elimination programs. Here we compared travel survey questions with mobile phone data over the same time period in western Kenya. We found that the survey data produces lower estimates of travel, although it did provide demographic information about travelers and motivations for travel. Mobile phone data can give a refined, spatio-temporal description of travel patterns, although it lacks information about subscribers, is often difficult to obtain, and as more providers become available such comprehensive estimates as presented here become even more challenging to achieve. In the case of malaria exchange via travel within these districts, although the volume of exchange differs by data source, both surveys were able to identify the same areas where the majority of exchange is likely to originate. In conjunction, these two data sources can be used to form a quantitative and qualitative description of travel within rural Kenya.

Methods

Travel survey data. A malariometric community survey was conducted using a cluster design in the highland districts Kisii Central and Rachuonyo South (referred to as Kisii and Rachuonyo in this paper), Nyanza province, western Kenya. For the survey, 23 enumeration areas (EA) (administrative areas with approximately 100 households or 500 residents) were randomly selected. Each EA was enumerated and mapped and 12–15 households were randomly selected for the survey.

The cross sectional surveys took place during February 2009. During this survey, individual informed consent was sought from all residents of the compound above the age of 6 months by signature or thumbprint accompanied by the signature of an independent witness. Consent for children under the age of 18 was provided by a parent/guardian, and children between 14 and 17 years also provided written assent by signature or thumbprint accompanied by the signature of an independent witness. Individuals between 15 and 18 years of age who were pregnant, married, or a parent were considered "mature minors" according to national policy and were able to

Table 4 | A comparison between the two data sets. For both data sets, the type of travel data available and scale (spatial and population) available. In general, the survey data is able to provide a coarser picture of travel, although refined socio-demographic data about travelers. The mobile phone data can only provide estimates on subscriber travel and is not able to provide any socio-demographic information about travelers. In order to compare between both data sets, we estimated the number of subscribers and the number of subscribers who have traveled from the survey data (see Materials and Methods). In comparison to the actual values quantified using the mobile phone data, the survey data produces estimates an order of magnitude less than the observed quantities

	Survey Data	Mobile Phone Data
Number of trips taken by individuals	Yes	Yes, for subscribers
Primary travel destination	Yes, district level	Yes, mobile phone tower
All destinations visited during traveling	No, only the primary destination	Yes, mobile phone tower
Duration of travel	No	Yes
Socio-demographic information about travelers	Yes	No
Spatial scale	District level	Mobile phone tower (~3 km)

	Estimated Value – Survey Data	**Estimated Value – Mobile Phone Data**
Total Population of Survey Site	19,744	
Number of Subscribers	2,500–11,500	35,000
Number of Subscribers who Travel	200–1,800	28,000

Table 5 | The population weighted malaria-travel metric estimated importation from the survey data and mobile phone data in both study areas

Data Source, Location	Population Weighted Malaria-Travel Metric	Top destination district for malaria-travel metric
Survey, Kisii	0.00014	Butere/Mumias
Survey, Rachuonyo	0.015	Nyando
Mobile Phone Data, Kisii	0.055	Nyamira
Mobile Phone Data, Rachuonyo	0.25	Nyamira

consent for themselves. The household was interviewed to assess household wealth indices and use of anti-malarial measures. All consenting individuals above the age of 6 months were tested for malaria and anemia. Individuals were asked basic travel questions about themselves and their children, specifically whether they had made any overnight trips to another district, the total number of overnight trips made, when they came back from their journey, and the reason for traveling (see Table 6).

Mobile phone data. Call data records (CDR) from June 2008 till June 2009 for 14,816,521 subscribers within Kenya were obtained from all months except for February 2009. For each entry in the CDR, the sender, receiver, date, and location of the call (or SMS) was recorded by the leading mobile phone provider. In total, subscribers sent and received approximately 12 billion calls and SMS geolocated at one of 12,502 mobile phone towers. For each subscriber, we approximated their daily location based on the location of the mobile phone tower that serviced the majority of their calls (or SMS) or the tower that serviced their most recent call (or SMS) if no call was made. For this analysis, we aggregated tower locations to districts based on the location of the mobile phone tower. We only considered subscriber IDs where the majority of their calls were serviced by mobile phone towers within the service area (3 km) and the district of each study site to conservatively only consider travel by subscribers whose primary mobile phone tower location was in the study site. At the time of data collection, this was the standard service area for mobile phone towers. This data was then restricted to include the sets of subscribers that overlap with the area of the community survey, one set in Kisii and two in Rachuonyo (see Figure 1). In total, we considered the data generated from 16,196 (based at 6 mobile phone towers) and 18,665 (9 mobile phone towers) subscribers in Kisii and Rachuonyo respectively (see Table 3).

We only considered travel that crossed district boundaries outside of the study area and not local movement within the study site (i.e. no travel between Kisii and Rachuonyo). Although the study site spans a district border, climate and topography are similar and we wanted to assess the extent of travel to areas where disease transmission would be markedly different. Also, there were a number of mobile phone towers along the borders of these districts making differentiating travel between the two locations more difficult. To match the time period of the survey, which specifically asked about travel in the previous 3 months, we only considered travel that occurred between the start of November 2008 till the end of January 2009. The mobile phone data describe the movements to the entire country of approximately 35,000 subscribers who primarily call from one of 15 mobile phone towers. No other demographic information is available from mobile phone data.

Comparing between the two sources of travel data. To compare between the mobile phone and survey data, we estimated the number of subscribers using the survey data and calculated a range for the number of trips taken by these subscribers. To estimate the number of mobile phone subscribers in the study area from the survey data, we used the number of individuals in the study area (~20,000), number of households (776), percentage of households with a mobile phone (47% reported in the survey), as well as the average number of individuals per sleeping structure (3.7). We did not know the number of subscribers per household but assumed a range of 1–4 subscribers per household to produce a range of ~2,500–11,500 mobile phone subscribers in the study area from the survey data.

To estimate the number of subscribers who have traveled using only the survey data, we considered a range of the percentage of subscribers who have traveled. At the low end, 8% of individuals have reported traveling results in 200–920 subscribers who have traveled. At the high end, adult males living with a household with a mobile phone were the demographic group with the highest percentage of travelers (16%). This value would imply that between 400–1,800 subscribers have traveled. As

reported in the results section, these estimates are at least an order of magnitude lower than the measured values from the mobile phone data.

Quantifying malaria exchange. To further compare both data sets, we quantified a malaria (*P. falciparum*) exchange metric using each set of travel data along with malaria endemicity data. Spatially explicit quantitative malaria endemicity estimates were obtained from the Malaria Atlas Project[29]. *P. falciparum* malaria endemicity data were obtained from the MAP (www.map.ox.ac.uk/) as measured by the parasite rate in the 2–10 age group ($PfPR_{2-10}$)[29]. This measure is an overestimate on the parasite rate since we are quantifying travel by adults, who generally have lower rates of parasite carriage. We use prevalence in children to avoid complex adjustments for patterns of prevalence by age, which vary with transmission intensity and are not straightforward to measure since many semi-immune individuals have sub-patent infections. Our estimates therefore represent an upper limit, and are intended to reflect the potential range and extent of spatial spread of malaria.

We calculated population rescaled travel from Rachuonyo and Kisii to other districts using the mobile phone and census data. For the mobile phone data, the population in each district's coverage area was the number of subscribers (18,665 and 16,196 in Rachuonyo and Kisii) whereas in the survey data it was the total number of individuals surveyed (1,297 and 1,352 in Rachuonyo and Kisii). From the survey data, we separated individuals by their district study site and considered travel to other districts.

In previous work, we utilized the mobile phone data to quantify the role of travel for malaria importation within Kenya[15]. However, due to the coarseness of the travel survey data and inability to describe the duration and exact destinations for all trips reported in the survey data, we choose to use a simplified malaria-travel metric that describes malaria exchange between locations[30] (see Tables 4, 5, Supplementary Information for further discussion). This metric, *Pfm*, is based on travel between individuals from the study sites (i) to all other districts (j) is defined as:

$$Pfm_{i,j} = \frac{PfPR_i * PfPR_j}{PfPR_i + PfPR_j} m_{i,j}$$

where $m_{i,j}$ is the population weighted travel to other districts.

Statistical analysis. The proportion of people traveling to another district was calculated for both datasets and summary values compared. Data from the travel survey data were analyzed to estimate the conditional probabilities of travel outside the district to provide insight on the demographics of travelers. Statistical and spatial analyses were carried out using R statistical analysis software (R v3.0.1, The R Foundation for Statistical Computing).

Geographic analysis. Mapping shown in Figures 1–3 was carried out by one of the co-authors using ArcGIS v10.1.

Ethical considerations. The community surveys were conducted and approved by the ethical committees of the London School of Hygiene and Tropical Medicine (LSHTM) and the Kenya Medical Research Institute (KEMRI) under protocol number SSC1802. Call data records were provided by the leading mobile phone provider to one of the co-authors of the paper. All received records were anonymized and could not be linked to individual users. The de-identified mobile phone records analysis was approved as not human subjects researchers by Harvard University IRB.

1. Balcan, D. *et al.* Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21484–21489, DOI:10.1073/pnas.0906910106 (2009).
2. Ferguson, N. M. *et al.* Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437**, 209–214, DOI:10.1038/nature04017 (2005).
3. Longini, I. M., Jr. *et al.* Containing pandemic influenza at the source. *Science* **309**, 1083–1087, DOI:10.1126/science.1115717 (2005).
4. Morens, D. M., Folkers, G. K. & Fauci, A. S. The challenge of emerging and re-emerging infectious diseases. *Nature* **430**, 242–249, DOI:10.1038/nature02759 (2004).
5. Kilpatrick, A. M. *et al.* Predicting the global spread of H5N1 avian influenza. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19368–19373, DOI:10.1073/pnas.0609227103 (2006).
6. Funk, S., Salathe, M. & Jansen, V. A. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface* **7**, 1247–1256, DOI:10.1098/rsif.2010.0142 (2010).

Table 6 | A brief outline of the travel questions asked in the travel survey

Travel Question
Have you made any overnight trips to another district in the last 3 months?
How many overnight trips have you made in the last 3 months?
Where did you spend the majority of time during this trip?
When did you get back?
What was your reason for traveling?

7. Prothero, R. M. Disease and mobility: a neglected factor in epidemiology. *Int. J. Epidemiol.* **6**, 259–267 (1977).
8. Meloni, S. *et al.* Modeling human mobility responses to the large-scale spreading of infectious diseases. *Sci. Rep.* **1**, 62, DOI:10.1038/srep00062 (2011).
9. Stoddard, S. T. *et al.* The role of human movement in the transmission of vector-borne pathogens. *PLoS Negl. Trop. Dis.* **3**, DOI:10.1371/journal.pntd.0000481 (2009).
10. Kimman, T. G. & Boot, H. The polio eradication effort has been a great success--let's finish it and replace it with something even better. *The Lancet Infect. Dis.* **6**, 675–678, DOI:10.1016/S1473-3099(06)70603-X (2006).
11. Kyle, J. L. & Harris, E. Global spread and persistence of dengue. *Annu. Rev. Microbiol.* **62**, 71–92, DOI:10.1146/annurev.micro.62.081307.163005 (2008).
12. Bajardi, P. *et al.* Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PloS ONE* **6**, DOI:10.1371/journal.pone.0016591 (2011).
13. Gaudart, J. *et al.* Spatio-temporal dynamics of cholera during the first year of the epidemic in Haiti. *PLoS Negl. Trop. Dis.* **7**, DOI:10.1371/journal.pntd.0002145 (2013).
14. Le Menach, A. *et al.* Travel risk, malaria importation and malaria transmission in Zanzibar. *Sci. Rep.* **1**, 93, DOI:10.1038/srep00093 (2011).
15. Wesolowski, A. *et al.* Quantifying the impact of human mobility on malaria. *Science* **338**, 267–270, DOI:10.1126/science.1223467 (2012).
16. Santos, A., McGuckin, N., Nakamoto, H. Y., Gray, D. & Liss, S. *Summary of travel trends: 2009 national household travel survey. National Technical Information Service: Trends in travel behavior,* http://nhts.ornl.gov/2009/pub/Compendium_2013.pdf (2011) Date of Access: 01/07/2013.
17. Buliung, R. N. & Remmel, T. K. Open source, spatial analysis, and activity-travel behaviour research: capabilities of the aspace package. *J. Geogr Syst* **10**, 191–216 (2008).
18. Craig, M. H., Kleinschmidt, I., Le Sueur, D. & Sharp, B. L. Exploring 30 years of malaria case data in KwaZulu-Natal, South Africa: part II. The impact of non-climatic factors. *Trop. Med. Intl. Health* **9**, 1258–1266, DOI:10.1111/j.1365-3156.2004.01341.x (2004).
19. Noor, A. M. *et al.* Establishing the extent of malaria transmission and challenges facing pre-elimination in the Republic of Djibouti. *BMC Infect. Dis.* **11**, 121, DOI:10.1186/1471-2334-11-121 (2011).
20. Schlich, R. & Axhausen, K. W. Habitual travel behaviour: evidence from a six-week travel diary. *Transp.* **30**, 13–36 (2003).
21. Wesolowski, A. *et al.* The use of census migration data to approximate human movement patterns across temporal scales. *PloS ONE* **8**, DOI:10.1371/journal.pone.0052971 (2013).
22. Pindolia, D. K. *et al.* Human movement data for malaria control and elimination strategic planning. *Malar. J.* **11**, 205, DOI:10.1186/1475-2875-11-205 (2012).
23. Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A. L. Understanding individual human mobility patterns. *Nature* **453**, 779–782, DOI:10.1038/nature06958 (2008).
24. Song, C., Qu, Z., Blumm, N. & Barabasi, A. L. Limits of predictability in human mobility. *Science* **327**, 1018–1021, DOI:10.1126/science.1177170 (2010).
25. Simini, F., Gonzalez, M. C., Maritan, A. & Barabasi, A. L. A universal model for mobility and migration patterns. *Nature* **484**, 96–100, DOI:10.1038/nature10856 (2012).
26. Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W. & Buckee, C. O. Heterogeneous mobile phone ownership and usage patterns in Kenya. *PloS ONE* **7**, DOI:10.1371/journal.pone.0035319 (2012).
27. Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W. & Buckee, C. O. The impact of biases in mobile phone ownership on estimates of human mobility. *J. R. Soc. Interface* **10**, DOI:10.1098/rsif.2012.0986 (2013).
28. Clarke, M., Dix, M. & Jones, P. Error and uncertainty in travel surveys. *Transp.* **10**, 105–126 (1981).
29. Noor, A. M. *et al.* The risks of malaria infection in Kenya in 2009. *BMC Infect. Dis.* **9**, 180, DOI:10.1186/1471-2334-9-180 (2009).
30. Tatem, A. J. & Smith, D. L. International population movements and regional Plasmodium falciparum malaria elimination strategies. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12222–12227, DOI:10.1073/pnas.1002971107 (2010).
31. Morgan, W. T. W. The ethnic geography of Kenya on the eve of independence: the 1962 census. *Erdkunde* **54**, 76–87 (2000).

Author contributions

A.W. designed the research, performed the analysis of data, and wrote the manuscript. C.D., J.C. and J.S. designed and implemented the field travel surveys. A.W., G.S., J.S., N.E., T.B., E.M., C.O., C.D., J.C. and C.O.B. all contributed to designing the research and manuscript preparation. All authors reviewed the manuscript.

Additional information