

Research



Cite this article: Wesolowski A, Eagle N, Noor AM, Snow RW, Buckee CO. 2013 The impact of biases in mobile phone ownership on estimates of human mobility. *J R Soc Interface* 10: 20120986.
<http://dx.doi.org/10.1098/rsif.2012.0986>

Received: 28 November 2012

Accepted: 15 January 2013

Subject Areas:

biometrics, environmental science

Keywords:

mobile phone, human mobility,
socio-economic status

Author for correspondence:

Caroline O. Buckee
e-mail: cbuckee@hsph.harvard.edu

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2012.0986> or via <http://rsif.royalsocietypublishing.org>.

The impact of biases in mobile phone ownership on estimates of human mobility

Amy Wesolowski¹, Nathan Eagle^{2,3}, Abdisalan M. Noor^{4,5}, Robert W. Snow^{4,5} and Caroline O. Buckee^{3,6}

¹Department of Engineering and Public Policy, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15221, USA

²College of Computer and Information Science, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA

³Department of Epidemiology, and ⁶Center for Communicable Disease Dynamics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA

⁴Malaria Public Health Cluster, KEMRI-Wellcome Trust-University of Oxford Collaborative Programme, PO Box 43630-00100, Nairobi, Kenya

⁵Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, Churchill Hospital, University of Oxford, Old Road, Oxford OX3 7LJ, UK

Mobile phone data are increasingly being used to quantify the movements of human populations for a wide range of social, scientific and public health research. However, making population-level inferences using these data is complicated by differential ownership of phones among different demographic groups that may exhibit variable mobility. Here, we quantify the effects of ownership bias on mobility estimates by coupling two data sources from the same country during the same time frame. We analyse mobility patterns from one of the largest mobile phone datasets studied, representing the daily movements of nearly 15 million individuals in Kenya over the course of a year. We couple this analysis with the results from a survey of socioeconomic status, mobile phone ownership and usage patterns across the country, providing regional estimates of population distributions of income, reported airtime expenditure and actual airtime expenditure across the country. We match the two data sources and show that mobility estimates are surprisingly robust to the substantial biases in phone ownership across different geographical and socioeconomic groups.

1. Introduction

Increasing human mobility is rapidly changing the landscape of human societies in low-income nations. The temporal and spatial scope of this mobility ranges from the permanent migration of individuals into cities and surrounding slums from rural areas [1,2], the large-scale movement of individuals owing to conflict or drought [3–5] and the short-term journeys people take to health clinics or to visit family. Although an understanding of human mobility on these scales is critical to public policy and planning, the dynamics of human movement on a population level are almost impossible to capture with traditional travel diary and global positioning system (GPS) approaches [6,7]. Proxies for population movements including road networks, transportation accessibility maps and bank note dispersal data may not be representative of travel [6,8], and transportation-related proxies cannot capture non-motorized travel, an important component of human movement in low-income nations. As the global adoption of mobile phones continues to escalate, however, so does the opportunity to use call data records (CDRs) that identify the cell tower locations of calls and text messages to accurately quantify population mobility. Low-income nations are particularly exciting in this respect, as mobile phone adoption continues to rise. In 2008, there were 280 million

subscribers (approx. 30% of the entire population) in Africa, a number greater than the number of subscribers in North America [2].

Studies of human travel patterns using CDRs have ranged in detail and size. On one end of the spectrum, several studies have examined detailed behavioural data about a small number of individuals in the USA and Europe, generating an individual-level model of travel that includes reasons for travelling [9–11]. A mid-sized study by Gonzales *et al.* [12] analysed CDRs from 100 000 randomly chosen mobile phone users in Europe as well as detailed location data for around 200 subscribers, facilitating the development of statistical models of human mobility patterns on a larger scale. This study identified a power law governing mobility patterns in their data, which has widely been interpreted as reflecting universal rules of human movement [13–17]. On the other end of the spectrum, we have previously quantified the movement of nearly 15 million mobile phone subscribers in Kenya to estimate the impact of mobility on malaria transmission [18]. The scale of these datasets will inevitably increase as mobile phones continue their rapid diffusion into even the most rural and economically deprived areas. Studies of this kind will provide invaluable information on the dynamics of human populations that can be used to aid policy makers in decisions about transportation, health services or disaster management.

A prerequisite for these studies is an understanding of the demographics of mobile phone owners. It is generally assumed that mobile phones are sufficiently widespread that CDRs represent a true random sample of a population [19]. Because this is unlikely to be the case, particularly in low-income nations, interpretation of mobility estimates must account for bias introduced by heterogeneous mobile phone ownership. Recently, we showed that mobile phone owners in Rwanda and Kenya are not representative of the general population, being disproportionately male, educated and from larger households compared with the general population [20,21]. Travel diary and GPS studies have shown that people with the greatest extent, variety and frequency of travel are those with the fewest economic and social constraints, as well as those with access to transportation [22–24]. Urban individuals with higher occupational status, income and level of education appear to take more journeys per day, more social trips and travel greater distances [22]. The same factors contributing to the probability of an individual owning a mobile phone may also impact his or her travel patterns, therefore, leading to biased estimates of human mobility from CDRs. So far it has not been possible to analyse or account for bias in mobility estimates from mobile phone data because CDR data generally does not contain any information about the subscriber base, particularly in countries where pre-paid plans dominate. Additional sources of data on socioeconomic status and mobile phone usage and ownership are, therefore, needed to determine how representative mobility estimates from CDRs are for the general population.

Here, we quantify and adjust for these biases by taking advantage of two parallel datasets from Kenya from the same year. We analyse the daily movement patterns of nearly 15 million individuals from mobile phone CDRs over the course of a year, and match these to calculated regional airtime expenditures. We cross-validate these values with reported airtime expenditures from a randomized survey of nearly

33 000 individuals in the same year, and correlate them with distributions of reported income and frequency of mobile phone ownership in each district. By linking these two sources of data, we can determine which income brackets are under- or overrepresented in the mobile phone data, allowing us to adjust our mobility estimates. We show that despite the bias introduced by differential phone ownership, mobility patterns within particular regions are surprisingly robust across populations.

2. Results

We analysed the daily travel patterns of 14 816 521 (approx. 38% of the Kenyan population) individuals across Kenya from June 2008 to June 2009 using nearly 12 billion calls and text messages to estimate daily locations for each one (see §2). At the time of data collection, the incumbent mobile phone operator had 92 per cent of market share, so these data represent the vast majority of mobile phone owners in the country. We used a common measure of individual mobility, the radius of gyration, to examine how mobility patterns varied across the country on the district and population levels (see §2). This measure reflects both the frequency of travel and distance [12]. For each subscriber, we calculated a radius of gyration value over the year and then aggregated these population values to the district level. We combined individual mobility values from a previously described survey [21] (see §2) in order to analyse the relationship between mobility and income. The survey data included 32 748 individuals across the country in 2009, and included questions on mobile phone ownership, airtime expenditure, income and a variety of other social and economic characteristics.

Our estimates of individual mobility from CDRs (figure 1a) show that on a population level, individuals who spend more money on airtime travel further and more frequently than individuals who spend less ($F = 1.6752$, $p < 0.0001$, see the electronic supplementary material). As expected, the survey data indicate that airtime expenditure reflects monthly income (contingency coefficient of 0.47, $p < 0.001$), and is therefore a reasonable proxy for economic status (figure 1b). We have previously shown not only that mobile phone ownership is biased towards wealthier individuals, but also that substantial geographical heterogeneities exist in the levels of phone ownership across the country [21]. Together, these observations and previous results would suggest that CDRs are likely to result in overestimates of the mobility of a population and that these estimates may vary geographically.

In order to examine the extent to which ownership biases may impact mobility estimates, we matched reported and actual expenditure on airtime from both sources of data. For each district, we used the CDR to calculate the average mobility of individuals within discretized ranges of actual airtime expenditures. We could link these mobility estimates to the demographics of particular groups of people on a district level using self-reported monthly airtime expenditures. Figure 2a–c illustrates our approach. Across the majority of districts, consistent with the aggregated data shown in figure 1, individual mobility and airtime expenditure were positively correlated in the CDR data, as well as reported monthly income and expenditure on airtime from the survey (see the electronic supplementary material). This

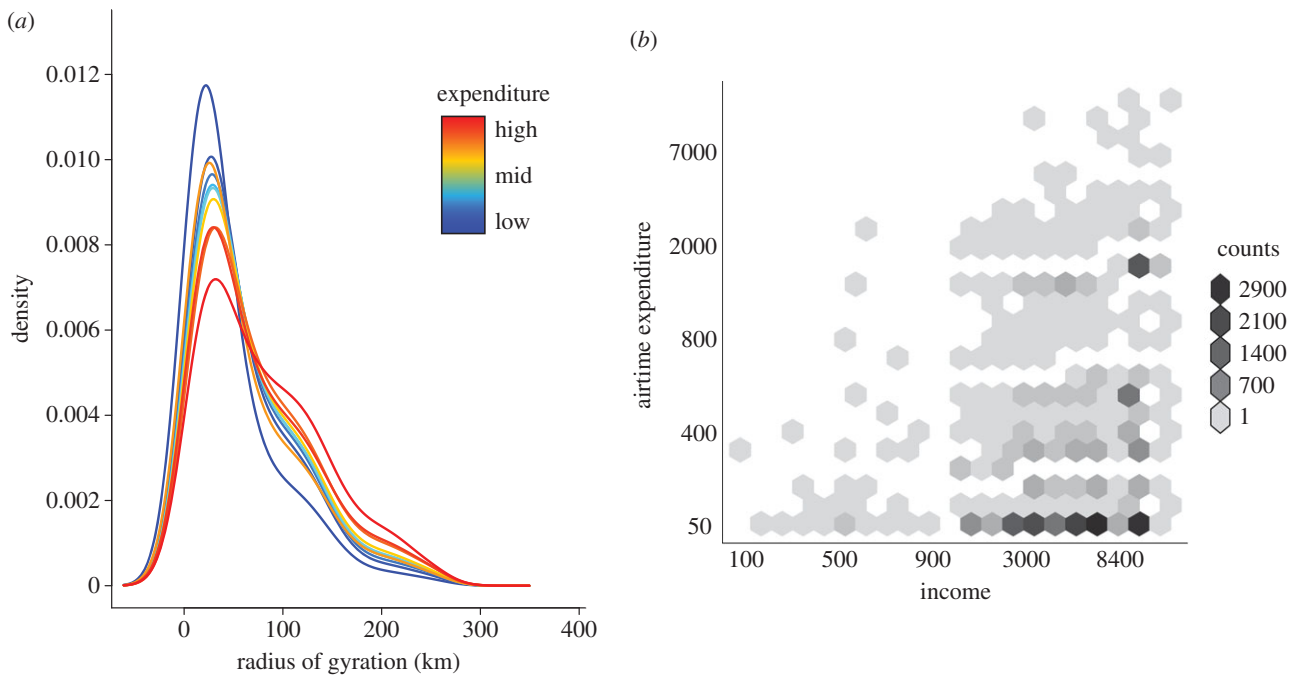


Figure 1. The relationship between monetary and movement variables. (a) Individuals were aggregated into percentiles of mobile phone expenditure calculated from the CDRs. Distributions of radius of gyration values are shown with high expenditure individuals travelling further (red) and low expenditure individuals travelling the least (blue). (b) The relationship between minimum monthly income and airtime expenditure for individuals. In general, airtime expenditure increased with income.

relationship enabled us to assign mobility estimates from the CDR to particular income brackets from the survey, and use the frequency of individuals within those income brackets (irrespective of phone ownership) to assess how skewed our mobility estimates are likely to be in each district (see §2 for more detail). Note this was only possible because some fraction of the population within every income bracket owned phones. This mapping enabled us to readjust observed mobility estimates to be representative of the underlying population even though we were only able to measure the mobility of mobile phone owners.

For each district, we compared the distribution of radius of gyration estimates from CDRs alone with the re-weighted distribution that accounted for differential mobile phone ownership (mean and median radius of gyration values shown in the electronic supplementary material, table S1). We compared the divergence between the two using a non-parametric method, a relative distribution, yielding an entropy measure that describes the overall similarity of the distributions, as well as estimates of which parts of the distribution were most skewed (see §2). Figure 3*a–c* illustrates hypothetical examples of distributions compared in this way. Figure 3*d* shows the entropy estimates and the direction of skew for all districts, highlighting how similar the CDR-only and re-weighted distributions were (mean entropy = 0.0011; see the electronic supplementary material, table S2). For comparison, we simulated two sets of data generated from a normal distribution with identical means and standard deviations and found similar values (mean entropy = 0.00252, see the electronic supplementary material).

As predicted by previous work, even though the distributions were similar, using CDR alone tended to overestimate the mobility of populations (in 73% of the districts, see figure 3*d* and the electronic supplementary material). Note that the adjustment here is based on our ability to link

expenditure between datasets, and correlate income with expenditure in the survey. Although income is probably an important indicator of travel patterns, other metrics that we were not able to measure may have led to different adjustments. Despite this caveat, we are cautiously optimistic that CDRs from mobile phones offer a reasonable estimate of mobility patterns of individuals across a wide range of incomes, despite considerable heterogeneity in mobile phone ownership.

3. Discussion

This work represents one of the first attempts to quantify the biased nature of mobile phone ownership and to analyse the impact of these biases on mobility estimates derived from CDRs. Given the potential value of CDR datasets and the increasing prevalence of mobile phone ownership globally, understanding these relationships is critical for future studies of human movement using this approach. A direct link between particular survey respondents and their CDRs would breach privacy policies of most mobile phone operators, but travel surveys that incorporate questions about mobile phone usage would greatly increase the utility of CDR datasets in the future. Without the ability to link socio-economic survey data with CDR data, an understanding of biased ownership and the effect it has on call data analytics, such as radius of gyration, would not be possible.

CDR datasets have already been applied to better understand human movement worldwide [9,18,25–27]. The scale of CDR studies of human mobility will inevitably increase over the coming decades, providing a powerful tool for policy makers and scientists for analysing millions of individual movement patterns. CDR studies are able to estimate the overall population mobility for a specific region, often an entire country, on a scale unobtainable

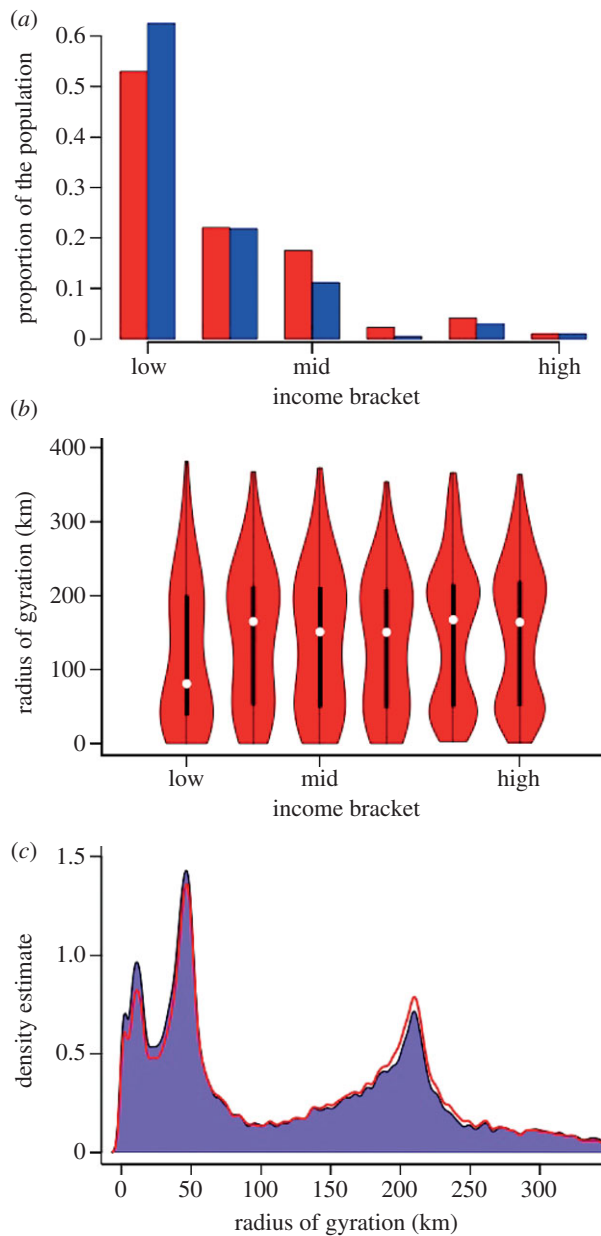


Figure 2. A schematic describing the method to readjust radius of gyration values based on the true population distribution for a sample district, Malindi. (a) For each income bracket, the number of mobile phone owners (red) versus non-owners (blue). (b) For each income bracket, the range of radius of gyration estimates by individuals within the income bracket (matched via mobile phone expenditure) is shown using a violin plot. Income brackets were defined based on percentiles of reported income for individuals in each district. (c) The kernel density of radius of gyration estimates for the unadjusted (red, i.e. only measured radius of gyration values from mobile phone owners) versus adjusted (violet, i.e. after taking into account the true population falling within each income bracket) are shown.

from sources such as travel surveys. However, they are unlikely to replace traditional travel survey studies that provide much richer insights into the socio-economic characteristics of travellers and motivations for travelling, for example. In particular, it is difficult to determine causation with CDR data, such as the effect of a changing landscape on travel patterns, whereas individual travel surveys can better address these questions.

Surprisingly, the bias in mobile phone ownership did not greatly skew estimates of human mobility. Using an adjustment scheme to correct for differential ownership throughout

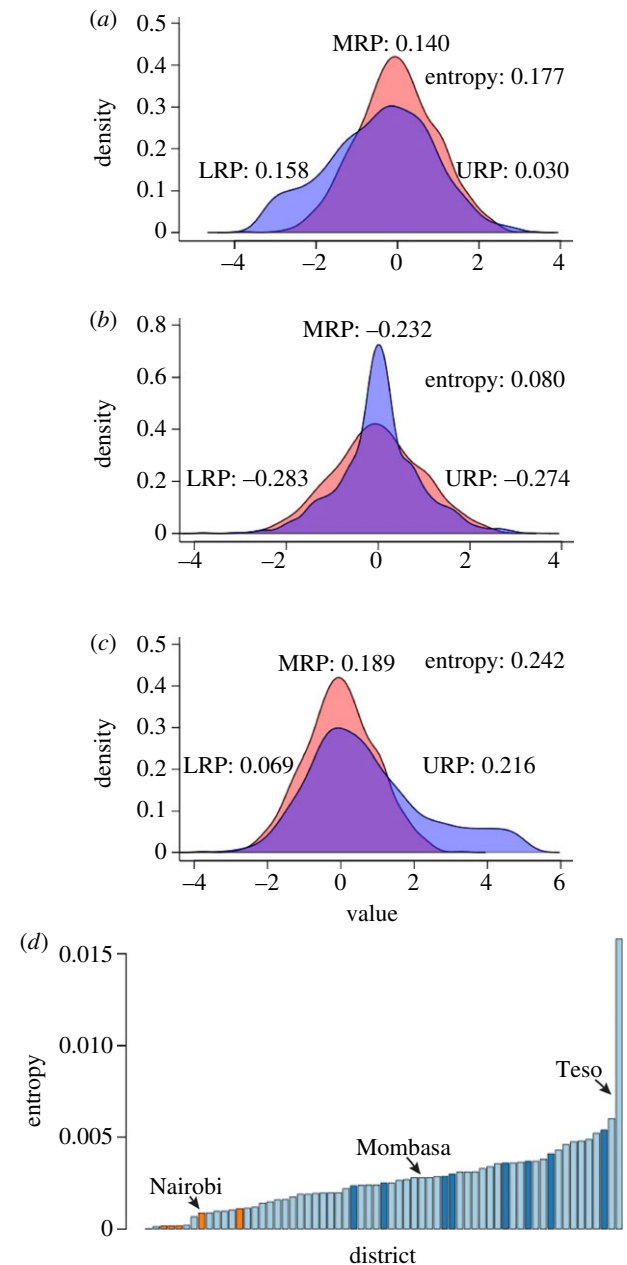


Figure 3. The statistical methodology used to assess the influence of biased ownership on mobility estimates. Density estimates for synthetic data that has been skewed towards (a) the lower tail, (b) the median, and (c) the upper tail (see the electronic supplementary material for further details). For each set of distributions, we have overlaid the skewed synthetic data over original synthetic data with lower, median, and upper relative polarization indices. (d) The effect of ownership bias on mobility estimates on each district in Kenya. Each district is ranked according to the entropy measure and then coloured by the direction of the skew (LRP, blue; MRP, orange; URP, dark blue).

the country, we were able to account for the unobserved movement patterns of non-owners in each district. Although we believe income is a reasonable variable to use for this type of approach, we could have used a range of variables that correlate with mobile phone ownership and usage, and these may have slightly different results. The distributions of mobility estimates within very different regions of Kenya were surprisingly robust to biases in mobile phone ownership even with large differences in ownership [21]. Although our results must be confirmed in other settings, we are optimistic that CDRs give reasonable estimates of the distributions of individual human mobility.

4. Material and methods

4.1. Mobile phone data

We used mobile phone communication logs from the incumbent mobile phone provider in Kenya. Our data represents all communications, both calls and text messages, by subscribers from June 2008 to June 2009. Unique hashed IDs were associated with one of 12 502 mobile phone towers' latitude and longitude coordinates for the caller, allowing us to map the location of the user at that time (see the electronic supplementary material, figure S1a). The annual radius of gyration was calculated for each individual. This measure takes into account the range and frequency of travel, with higher values corresponding to more journeys and larger distances travelled. Radius of gyration has been used in several studies of human mobility [12–14] as follows:

$$r_g^a(t) = \sqrt{\frac{1}{n_c^a(t)} \sum_{i=1}^{n_c^a(t)} (\vec{r}_i^a - \vec{r}_{cm}^a)^2},$$

where \vec{r}_i^a represents the $i = 1, \dots, n_c^a(t)$ tower locations recorded for subscriber a at time t , and $\vec{r}_{cm}^a = 1/n_c^a \sum_{i=1}^{n_c^a} \vec{r}_i^a$ represents the user's centre of mass trajectory. Note we have analysed the observed positive relationship between call volume and radius of gyration, and through simulations have shown that it increases the variance on our estimates (see the electronic supplementary material, figure S2).

4.2. Survey data

In 2009, the Financial Sector Deepening Kenya (FSDK) conducted a survey asking 32 748 individuals located at 646 communities several questions about mobile phone usage, ownership and monthly expenditure on airtime, as well as detailed demographic questions concerning income, education level and housing type [21] (see the electronic supplementary material, figure S1b). Cluster stratified probability sampling based on National Sample Survey and Evaluation Programme (NASSEP IV provided by the National Bureau of Statistics) ensured representative populations were included in the survey. First level selection (cluster level) yielded a representative set at the national, provincial and urbanization levels in each province. The Kenyan National Bureau of Statistics (KNBS) determined how many clusters should be selected for each province, with clusters being randomly selected from a list in the sampling frame for each region to ensure urban regions were adequately represented. Second level selection (household level) of households aimed for 10 households within each cluster based on standard sample size calculations. Finally, third level selection (individual level) of individuals aged 16+ years was performed using a standard Kish grid (available in the original survey at <http://www.fsdkkenya.org>).

4.3. Correlating expenditure with top-up denominations

The FSDK survey asked individuals their mobile phone monthly expenditure. To match these reported values with the CDR, we used a monthly airtime expenditure calculated from each individual's call volume, assuming the tariff at the time of our dataset (3 KSH per minute, 1 USD ~ 75 KSH). This calculation allowed us to estimate radius of gyration values for different brackets of monthly expenditure in each district. Figure 2 shows a schematic of the approach. While the correlation between monthly airtime expenditures and income or other variables that are measurable for individuals who do not own phones may not be exact, they provide a mechanism for adjusting our mobility estimates that would be possible for CDR studies.

The relationship between monthly income and radius of gyration for each district was slightly positive (mean correlation

coefficient = 0.19). The strength of this relationship varied by district and was adjusted by population per district using a non-parametric method to account for this fact (see the electronic supplementary material). Using the population adjustments, for each district, we constructed a weighted kernel density estimation of the population adjusted and non-population adjusted radius of gyration values. Following the method of Hancock & Morris [28], the comparison sample y_1, y_2, \dots, y_n refers to the non-population adjusted radius of gyration values, and the reference distribution x_1, x_2, \dots, x_m refers to the population adjusted radius of gyration values. The comparison and reference distributions have cumulative distribution functions (CDFs; probability density functions (PDFs)) denoted $F(f)$ and $F_0(f_0)$, respectively. We construct relative data $r_i = F_0(y_i)$ with relative CDF of $G(r) = F(F_0^{-1}(r))$ and relative density $g(r) = \frac{f(F_0^{-1}(r))}{f_0(F_0^{-1}(r))}$. The relative data provides a non-parametric method for a full comparative distributional analysis that is a complete summary of the comparisons between the comparison and reference distributions. Regardless of the shape of the raw data, it can be compared with the adjusted data in a systematic way. We can use the relative distribution to measure the divergence, skew and location of skew of the comparison distribution to the reference distribution.

To measure the divergence between the two distributions, we calculated the Kullback–Leibler divergence between the relative and comparison distributions. K–L divergence measures the difference between two probability distributions and quantifies the expected number of extra bits required to code samples from the reference distribution when using a code based on the comparison distribution rather than a code based on the reference distribution. It is defined as

$$D(F; F_0) = \int_{-\infty}^{\infty} \log\left(\frac{f(x)}{f_0(x)}\right) dF(x) = \int_0^1 \log(g(r)) g(r) dr.$$

To measure polarization, i.e. if the comparison distribution is skewed towards the middle or tail values, we calculate the median relative polarization index of Y relative to Y_0 .

$$\text{MRP}(F; F_0) = 4 \int_0^1 \left|r - \frac{1}{2}\right| g_0^A(x) dx - 1.$$

Positive values of median relative polarization (MRP) represent more polarization (increases in the tails of the distribution), negative values represent less polarization (convergence towards the centre of the distribution), and a zero value represents no differences in distributional shape. We can also define lower and upper polarization indices to quantify the contributions above and below the median of the relative distribution.

$$\text{LRP}(F; F_0) = 8 \int_0^{1/2} \left|r - \frac{1}{2}\right| g_m(x) dx - 1$$

and

$$\text{URP}(F; F_0) = 8 \int_{1/2}^1 \left|r - \frac{1}{2}\right| g_m(x) dx - 1.$$

The authors thank Dr Cosma Shalizi for his help with the statistical methodology. A.P.W. was supported by the National Science Foundation Graduate Research Fellowship programme (no. 0750271). A.M.N. was supported by the Wellcome Trust as an Intermediary Research Fellow (no. 095127). R.W.S. was supported by the Wellcome Trust as Principal Research Fellow (no. 079080). A.M.N. and R.W.S. also acknowledge programmatic support provided by the Wellcome Trust Major Overseas Programme grant to the KEMRI/Wellcome Trust Programme (no. 092654). C.O.B. was supported by the Models of Infectious Disease Agent Study programme (cooperative agreement 1U54GM088558). The authors thank the Financial Sector Deepening Kenya for access to the data. The content is solely the responsibility of

the authors and does not necessarily represent the official views of the National Institute Of General Medical Sciences or the National Institutes of Health. Mobile phone data were provided by an anonymous service

provider in Kenya and is not available for distribution. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

- Hanson S. 2005 Perspective on the geographic stability and mobility of people in cities. *Proc. Natl Acad. Sci. USA* **43**, 15 301–15 306. (doi:10.1073/pnas.0507309102)
- Wesolowski A, Eagle N. 2010 Parameterizing the dynamics of slums. In *AAAI Spring Symp. on Artificial Intelligence for Development*, March. Stanford, CA: Stanford University.
- Bharti N, Tatem AJ, Ferrari MJ, Grais RF, Djibo A, Grenfell BT. 2011 Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science* **9**, 1424–1427. (doi:10.1126/science.1210554)
- Gray CL, Mueller V. 2012 Natural disasters and population mobility in Bangladesh. *Proc. Natl Acad. Sci. USA* **109**, 6000–6005. (doi:10.1073/pnas.1115944109)
- Ezra M, Kiros G-E. 2001 Rural out-migration in the drought prone areas of Ethiopia: a multilevel analysis. *Int. Migr. Rev.* **35**, 749–771. (doi:10.1111/j.1747-7379.2001.tb00039.x)
- Brockmann D, Hufnagel L, Geisel T. 2006 The scaling laws of human travel. *Nature* **439**, 462–465. (doi:10.1038/nature04292)
- Funk S, Salathé M, Jansen VA. 2010 Modeling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface* **7**, 1247–1256. (doi:10.1098/rsif.20120.0142)
- Tatem AJ, Smith DL. 2010 International population movements and regional *Plasmodium falciparum* malaria elimination strategies. *Proc. Natl Acad. Sci. USA* **107**, 12 222–12 227. (doi:10.1073/pnas.1002971107)
- Eagle N, Pentland A. 2006 Reality mining: sensing complex social systems. *Pers. Ubiquit. Comput.* **10**, 255–268. (doi:10.1007/s00779-005-0046-3)
- Yang Y, Atkinson P, Ettema D. 2007 Individual space–time activity-based modeling of infectious disease transmission within a city. *J. R. Soc. Interface* **5**, 759–772. (doi:10.1098/rsif.2007.1218)
- Demirbas M, Bayir MA, Eagle N. 2009 Discovering spatiotemporal mobility profiles of cellphone users. In *World of Wireless, Mobile and Multimedia Networks and Workshops*, 2009. WoWMoM 2009. IEEE International Symposium. New York, NY: IEEE. (doi:10.1109/WOWMOM.2009.5282489)
- Gonzales M, Hidalgo C, Barabasi AL. 2008 Understanding individual human mobility patterns. *Nature* **453**, 779–782. (doi:10.1038/nature06958)
- Song Q, Blumm N, Barabasi AL. 2010 Limits of predictability in human mobility. *Science* **327**, 1018–1021. (doi:10.1126/science.1177170)
- Barabási AL. 2009 Scale-free networks: a decade and beyond. *Science* **325**, 412–413. (doi:10.1126/science.1173299)
- Wang P, González MC. 2009 Understanding spatial connectivity of individuals with non-uniform population density. *Phil. Trans. R. Soc. A* **367**, 3321–3329. (doi:10.1098/rsta.2009.0089)
- Balcan D, Colizza V, Gonc B, Hu H. 2009 Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl Acad. Sci. USA* **106**, 21 484–21 489. (doi:10.1073/pnas.0906910106)
- Merler S, Ajelli M. 2010 The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proc. R. Soc. A* **277**, 557–565. (doi:10.1098/rspb.2009.1605)
- Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, Buckee CO. 2012 The impact of human mobility on malaria. *Science* **338**, 267–270. (doi:10.1126/science.1223467)
- James MJ, Versteeg M. 2007 Mobile phones in Africa: how much do we really know? *Soc. Indicators Res.* **84**, 117–126. (doi:10.1007/s11205-006-9079-x)
- Blumenstock JE, Eagle N. 2010 Mobile divides: gender, socioeconomic status, and mobile phone use in Rwanda. In *Proc. 4th International Conference on Information and Communication Technologies and Development*, London, UK, 13–16 December, 2010. New York, NY: ACM.
- Wesolowski A, Eagle N, Noor AM, Snow RW, Buckee CO. 2012 Heterogeneous mobile phone ownership and usage patterns in Kenya. *PLoS ONE* **7**, e35319. (doi:10.1371/journal.pone.0035319)
- Hanson S, Hanson P. 1981 The travel-activity patterns of urban residents: dimensions and relationships to sociodemographic characteristics. *Stud. Choice ConstraintsHum. Spat. Behav.* **57**, 332–347.
- Popoola KO, Faborode TG. 2011 Effect of socio-economic status on household all purpose travel patterns of men and women in Ibadan, Oyo State, Nigeria. In *ECAS: 4th Eur Conf on African Studies*, 15–18 June, Uppsala, Sweden. Uppsala, Sweden: The Nordic Africa Institute.
- Camlin CS, Hosegood V, Newell M-L, McGrath N, Barnighausen T, Snow RC. 2010 Gender, migration and HIV in rural KwaZulu-Natal, South Africa. *PLoS ONE* **5**, e11539. (doi:10.1371/journal.pone.0011539)
- Le Menach A, Tatem AJ, Cohen JM, Hay SI, Randell H, Patil AP, Smith DL. 2011 Travel risk, malaria importation and malaria transmission in Zanzibar. *Sci. Rep.* **1**, 93. (doi:10.1038/srep00093)
- Lu X, Bengtsson L, Holme P. 2012 Predictability of population displacement after the 2010 Haiti earthquake. *Proc. Natl Acad. Sci. USA* **109**, 11 472–11 473. (doi:10.1073/pnas.1203882109)
- Phithakkitnukoon S, Smoreda Z, Olivier P. 2012 Socio-geography of human mobility: a study using longitudinal mobile phone data. *PLoS ONE* **7**, e39253. (doi:10.1371/journal.pone.0039253)
- Handcock MS, Morris M. 1998 Relative distribution methods. *Am. Sociol. Assoc.* **28**, 53–97. (doi:10.2307/270964)