Supporting Information for

Exploring the use of mobile phone data for national migration statistics

Shengjie Lai[1,2,3*], Elisabeth zu Erbach-Schoenberg[1,2], Carla Pezzulo[1], Nick W Ruktanonchai[1,2], Alessandro Sorichetta[1,2], Jessica Steele[1], Tracey Li[2], Claire A Dooley[1,2], Andrew J Tatem[1,2*]

[1]WorldPop, School of Geography and Environmental Science, University of Southampton, Southampton SO17 1BJ, United Kingdom

[2]Flowminder Foundation, SE-113 55 Stockholm, Sweden

[3]School of Public Health, Fudan University, Key Laboratory of Public Health Safety, Ministry of Education, 130 Dongan Road, Shanghai 200032, China

Shengjie Lai and Elisabeth zu Erbach-Schoenberg contributed equally to this work.

*Correspondence and requests for materials should be addressed to A.J.T (email: Andy.tatem@gmail.com) or S.L. (email: laishengjie@foxmail.com).

This PDF file includes:

SI Text

1.1. CDR Data Processing

Each call detail record (CDR) contained the location of the tower that the communication was routed through, giving us an approximate estimate of the user's location. Since we aimed for region level estimates, we mapped each entry to the region where the routing tower was contained (Mobile Telecommunications 2018). The daily location of a mobile phone user could be defined if the user made or received at least one communication (call or text) on that day. To reduce occasional travel or seasonal movement, e.g. travel over the Christmas holidays, infrequent mobile phone users with 30 days or less worth of defined daily locations (next section) for each year (12 months) were filtered out. This corresponded to 12% mobile phone users that were filtered out between October 2010 and September 2012 (Period 2011), and 20% users between October 2011 and September 2013 (Period 2012), respectively. The higher proportion of users filtered out in Period 2012 may be due to the constant increase of new SIM subscribers and shifting mobile phone usage from voice to data (Mobile Telecommunications 2012, 2014). We then calculated the most frequently observed region for each user and day as that user's daily location. Using these daily locations, we could define the residence as the most frequent daily location during a given year-long period (12 months) for each user. Finally, we derived migration flows of mobile users by comparing the places of residence at regional level across years for each user.

1.2. Comparing Different Time Lengths to Define Residence

As sufficient CDR data points are needed to accurately estimate the place of residence, different time windows of data were compared to define the residences of mobile phone users. We first investigated how many months of data were available for each mobile phone user in an example dataset from January to December 2012 that was used for migration estimate in Periods 2011 and 2012. A monthly location for each individual was calculated as the most frequent daily location on regional levels over each month if this user made at least one call or text in that month. We found that the most frequent case was that users have data available for all months of the year (Figure S2A), but it still showed a high proportion (15%) of infrequent users with only 1-month of data, which could introduce a strong bias for deriving migration flows compared to using yearly locations as residence. Therefore, to exclude very infrequent users, we used 31 days of defined daily locations as a cutoff for an individual to be included in the migration estimation. Moreover, most individuals with more than 1-month of data had monthly locations that were identical to yearly locations (Figure S2B), and the spatial differences in the mean percentage of monthly locations matching yearly locations were likely homogenous across the country, from the lowest percentage of 91% to the highest of 97% (Figure S2C).

Seasonal movements might lead to an individual temporarily residing at different locations, which would influence estimates of the place of residence in settings where only shorter periods (e.g. a month or two per year) of CDRs are available (Wesolowski et al., 2017). Figure S3A shows the percentage of locations using the random N months (ranged 1-11 months) of data that match the locations using a full year of data. As expected, the accuracy of estimating the usual residence increases with increasing length of period covered by the data. However, we observed the largest increase going from one month to two months of data used, likely due to the strong seasonal effect of individuals' temporary locations deviating from their usual places of residence during holiday periods and other times when short-term mobility is prevalent. Furthermore, Figure S3B shows the Z-score of the percentage of users whose monthly locations matched the yearly location. The negative deviation in December and January means fewer users being found at their usual residences in these holiday months, with a similar, but smaller effect being seen in May. This is part of the reason that censuses are generally timed to occur during a period with low seasonal population movement, e.g. August and September, and again it highlights the need to exclude infrequent mobile phone users to prevent the inclusion of short-term travel in longer-term migration estimation.

1.3. Phone Ownership

As mobile phone ownership is not homogenous across the population, we utilized data from the 2013 Namibia Demographic and Health Survey (DHS) (The Namibia Ministry of Health et al., 2014) to assess to what extent there is a possible exclusion of certain groups from the CDRs, and the potential of using the indicators to adjust CDR-based migration estimates and thus, to better match the census-based observations. Using a stratified two-stage cluster design, the 2013 Namibia DHS sampled 9,849 households that were distributed in 554 clusters across the country, with 269 in urban areas and 285 in rural areas (The Namibia Ministry of Health et al., 2014). This survey collected data for a wide range of indicators including demographic, health and socio-economic characteristics, and it also reported information on mobile phone ownership at household level. Moreover, households were ranked by the International Wealth Indicator (IWI) score, indicating to what extent the household possesses a basic set of assets. The IWI score is an asset-based wealth index used for measuring the economic situation of households in developing countries (Smits and Steendijk 2015). Households were then divided by five quintiles, based on the IWI ranking. Households falling within the upper two quintiles ('richer' and 'richest') were then classified as being wealthy, while households in the bottom three quintiles were defined as not being wealthy.

We first performed an exploratory bivariate analysis to compare the characteristics between households owning at least one mobile phone and households without mobile phones. Data were weighted using sampling weights and adjusted for the survey sampling design, and analyses were done using STATA 14 software. Table S2 shows that households owning at least one mobile phone are significantly different from households without mobile phones for almost all background characteristics considered in this analysis. In comparisons to households without mobile phones, those with mobiles are significantly more likely to be wealthy, to reside in urban areas and live in Khomas (22%). Households without mobile phones are more likely to be located in Kavango, Zambezi and Ohangwen regions. Moreover, households with one or more mobile phones are significantly more likely to have younger heads and a higher number of household members, whereas households without mobile phones tend to have more uneducated household heads and uneducated female and male residents (Table S2).

Furthermore, we performed a binary logistic regression model for the probability of households without mobile phones, by including all variables that resulted as being statistically significant in the bivariate analysis. In particular, we aimed to identify groups of households that had a high probability of not owning a mobile phone and the characteristics they share within a multiple regression analysis framework, after adjusting for the effects of other variables (Callegaro and Poggio 2004). Table S3 shows that there is a significant differential in the ownership of mobiles between households regarding wealth, age of the household head, household size, and education. Our findings also show that there is a significant ownership differential between regions in Namibia, confirming the results from the bivariate analysis. The odds of ownership of a mobile phone for households residing in most regions range between about 2 and 5 times greater than that in Kavango, meaning it may be necessary to take the regional mobile ownership bias into account in estimates of migration by CDRs for each region.

1.4. Model Covariates

We also collated potential migration-related demographic, socioeconomic, geographic and environmental variables for migration modelling as described in previous studies (Henry et al., 2003; Henry et al., 2004; Garcia et al., 2015; Wesolowski et al., 2015; Ruktanonchai et al., 2016; Sorichetta et al., 2016; Vobruba et al., 2016). An administrative unit boundary file at regional level matching the year of the census was obtained from the Global Administrative Areas Database (GADM 2018). Following previous studies (Garcia et al., 2015; Sorichetta et al., 2016), the shapefile was used to calculate variables that measure distance and contiguity between administrative units, respectively. Euclidean distance between geometric centroids is commonly used as a parameter in gravity models, where it represents the barriers to, as well as potential costs of, migration (Garcia et al., 2015; Sorichetta et al., 2016). To calculate possible environmental drivers of migration in models, moreover, high resolution monthly precipitation grids (30 seconds, ~1 km^2) were obtained from WorldClim version 2 (worldclim.org/version2). We then aggregated the precipitation data to obtain average annual precipitation (mm) by region as a proxy of push-pull factors such as agricultural productivity and the potential of floods and droughts.

A variety of demographic and socioeconomic variables known to be associated with migration flows were also collated from 2011 census data for each region in Namibia (Table S1). First, we included the populations in origin and destination (POP_i and POP_j) in 2010 and 2011. Given that the urbanization can be a significant pull factor for migrants (Lall et al., 2006; Garcia et al., 2015), then we included the percentage of population living in urban areas in origin and destination, denoted as $URBAN_i$ and $URBAN_j$, respectively.

Previous studies on migration also suggested that human migration is, at least in part, driven by economic opportunities and that different demographic characteristics, such as age, sex, educational attainment and marital status influence migration rates (Henry et al., 2004; Garcia et al., 2015; Sorichetta et al., 2016). Therefore, we also collated the following covariates from 2011 Namibia census data: *MALEPROP*, the proportion of males in the region; *AGE15_19*, the proportion of the population between 15 and 59 years old in the region; *ACTIVE*, the proportion of labour force participation in the population aged 15 and above by region; *LITERACY*, the percentage of literate individuals out of the population aged 15 and above; and *SINGLEPROP*, the proportion of unmarried people in the population aged 15 and above (Namibia Statistics Agency 2013). However, all of these variables show strong correlations (Pearson's *r* >0.5) with $URBAN_i$ and $URBAN_j$, in part because urban areas might offer more opportunities for improving these socioeconomic status (Lall et al., 2006). As the urbanization is related to socioeconomic development, we removed the demographic and socioeconomic variables that were highly correlated with the urbanization variables to avoid multicollinearity and overfitting in models.

SI Tables

Table S1. The summary of models.

Type	Model	Independent variables	
		Mobile phone data	**Other variables**
CDR-based linear model (CDRLM)	1	$CDR_{i,j}$ or $adjCDR_{i,j}$ or $CDR_{i,j} + PHONEPROP_i$	None
	2		$+ POP_i + POP_j + DIST_{i,j}$
	3[a]		$+ URBAN_i + URBAN_j + DIST_{i,j}$
	4[b]		$+ RAIN_i + RAIN_j + DIST_{i,j}$
Gravity-type spatial interaction model (GTSIM)	1	None	$\ln(POP_i) + \ln(POP_j) + DIST_{i,j}$
	2[a,b]		$URBAN_i + URBAN_j + DIST_{i,j}$
	3		$RAIN_i + RAIN_j + DIST_{i,j}$
CDR-based Gravity-type spatial interaction model (CGTSIM)	1[a,b]	$CDR_{i,j}$ or $adjCDR_{i,j}$ or $CDR_{i,j} + PHONEPROP_i$	$+ \ln(POP_i) + \ln(POP_j) + DIST_{i,j}$
	2		$+ URBAN_i + URBAN_j + DIST_{i,j}$
	3		$+ RAIN_i + RAIN_j + DIST_{i,j}$

Variable descriptions:

$CDR_{i,j}$: Number of migrating mobile phone users from origin i to destination j based on call detail records (CDRs).

$PHONEPROP_i$: Proportion of the population owning mobile phones in origin region i.

$adjCDR_{i,j}$: $CDR_{i,j}$ divided by $PHONEPROP_i$.

POP_i and POP_j : Population of origin i and destination j.

$URBAN_i$ and $URBAN_j$: Proportion of population living in urban areas.

$RAIN_i$ and $RAIN_j$: Annual average precipitation (mm).

$DIST_{i,j}$: Euclidean distance between centroids of origin i and destination j.

[a] Optimal model of each model family for regions except Zambezi, based on the lowest root-mean-square error (RMSE) (Fig. S8).

[b] Optimal model of each model family for all regions, based on the lowest RMSE (Fig. S8).

Table S2. Demographic and socio-economic characteristics of households by mobile phone ownership.

	With MP % (N=8,589)	Without MP % (N=1,257)	Total % (N=9,846)	p value*
Gender of household head				0.159
Male	55.7	58.6	56.1	
Female	44.3	41.4	43.9	
Mean age of household head, yrs (SE)	45.5 (0.32)	50.7 (0.76)		<0.001**
Mean number of household members, yrs (SE)	4.4 (0.05)	3.6 (0.12)		<0.001**
Number of under-five children in households				0.019
0	56.2	62.7	57.0	
1-3	41.9	36.1	41.3	
≥4	1.8	1.2	1.8	
Ratio of under 5 yrs in household members (SE)	0.120 (0.002)	0.113 (0.006)		0.257
Household wealth (International Wealth Index, IWI)				<0.001
Non-wealthy	52.4	91.4	56.8	
Wealthy	47.5	8.5	43.1	
Household residence				<0.001
Urban	55.8	22.3	52	
Rural	44.2	77.7	48	
Region				<0.001
Zambezi (Caprivi)	4.8	10.6	5.5	
Erongo	10.1	4.2	9.5	
Hardap	3.8	4.4	3.9	
Karas	4.3	3	4.1	
Kavango	6.2	17.6	7.5	
Khomas	22.1	7.5	20.4	
Kunene	3.1	7.3	3.6	
Ohangwena	8.8	11.5	9.1	
Omaheke	3.2	4.9	3.4	
Omusati	9.7	9.4	9.6	
Oshana	8.8	6	8.4	
Oshikoto	8.3	8.3	8.3	
Otjozondjupa	6.8	5.4	6.6	
Education attainment of household head* **				<0.001
No education	13	36.3	15.7	
Incomplete primary	21.4	29.8	22.4	
Complete primary	5.5	8.3	5.8	
Incomplete secondary	30.6	18.1	29.2	
Complete secondary	16.3	4.9	15	
Higher	12.6	2.2	11.4	
Unknown	0.6	0.4	0.6	
Education attainment of female household population (aged 6+)^				<0.001

	With MP % (N=8,589)	Without MP % (N=1,257)	Total % (N=9,846)	p value*
No education	10.6	28.9	12.2	
Incomplete primary	32	42.2	32.9	
Complete primary	5.5	6.3	5.6	
Incomplete secondary	32.4	18	31.1	
Complete secondary	11.9	3.1	11.1	
Higher	7.4	1.4	6.8	
Unknown	0.2	0.1	0.2	
Education attainment of male household population (aged 6+)^^				<0.001
No education	12	28	13.6	
Incomplete primary	35.7	41.2	36.2	
Complete primary	5.3	6.8	5.4	
Incomplete secondary	28	19	27.1	
Complete secondary	11.2	3.6	10.4	
Higher	7.5	1.2	6.8	
Unknown	0.3	0.3	0.3	

MP: mobile phone. The data were obtained from the Namibia 2013 Demographic and Health Survey (DHS) (dhsprogram.com/pubs/pdf/fr298/fr298.pdf). The ownership was defined as the presence of at least one mobile phone in the household. Percentages were produced using sampling weights and adjusted for the survey design. *Pearson*'s Chi-square test and t-test (where specified) were run to test the association between variables. Where indicated, variables were constructed using the household member dataset.

* *p* values were produced using Pearson's Chi-square test, adjusted for the survey design.

** Calculated using a t-test for equality of means.

*** Information available for total population, N=9,796.

^ calculated for de facto female household members aged 6 and over by highest level of schooling attended or completed (N=18,230).

^^ calculated for de facto male household members aged 6 and over by highest level of schooling attended or completed (N=16,153).

Table S3. Estimates of factors for household without mobile phone (N=9,842).

	Coefficient	SE	Odds Ratio (95% CI)
Region (Ref.: Kavango)			
Zambezi (Caprivi)	-0.46	0.219	0.63 (0.41; 0.97)*
Erongo	-1.12	0.267	0.33 (0.19; 0.55)*
Hardap	-0.44	0.224	0.64 (0.41; 1.00)
Karas	-1.06	0.333	0.35 (0.18; 0.67)*
Khomas	-0.92	0.232	0.40 (0.25; 0.63)*
Kunene	-0.38	0.226	0.69 (0.44; 1.07)
Ohangwena	-1.09	0.264	0.34 (0.20; 0.56)*
Omaheke	-0.83	0.236	0.44 (0.28; 0.70)*
Omusati	-1.61	0.232	0.20 (0.13; 0.31)*
Oshana	-1.35	0.424	0.26 (0.11; 0.60)*
Oshikoto	-1.30	0.223	0.27 (0.18; 0.42)*
Otjozondjupa	-0.99	0.276	0.37 (0.22; 0.64)*
Place of residence (Ref.: Urban)			
Rural	0.80	0.137	2.24 (1.71; 2.93)*
Household wealth quintiles IWI (Ref.: Wealthy)			
Not wealthy	1.50	0.138	4.47 (3.41; 5.86)*
Household has at least one member with completed secondary or higher education (Ref.: Yes)			
No	0.84	0.12	2.32 (1.84; 2.94)*
Age of household head	0.02	0.003	1.02 (1.01; 1.02)*
Number of household members	-0.19	0.021	0.83 (0.79; 0.86)*
Intercept	-3.50	0.238	0.03 (0.02; 0.05)*

Note: A logistic regression model was used, and the estimates of coefficients, odds ratios and 95% confidence intervals (CI) are presented. “Svy” command in STATA was used to account for weighting and sampling design. * *p* value < 0.05.

SI Figures

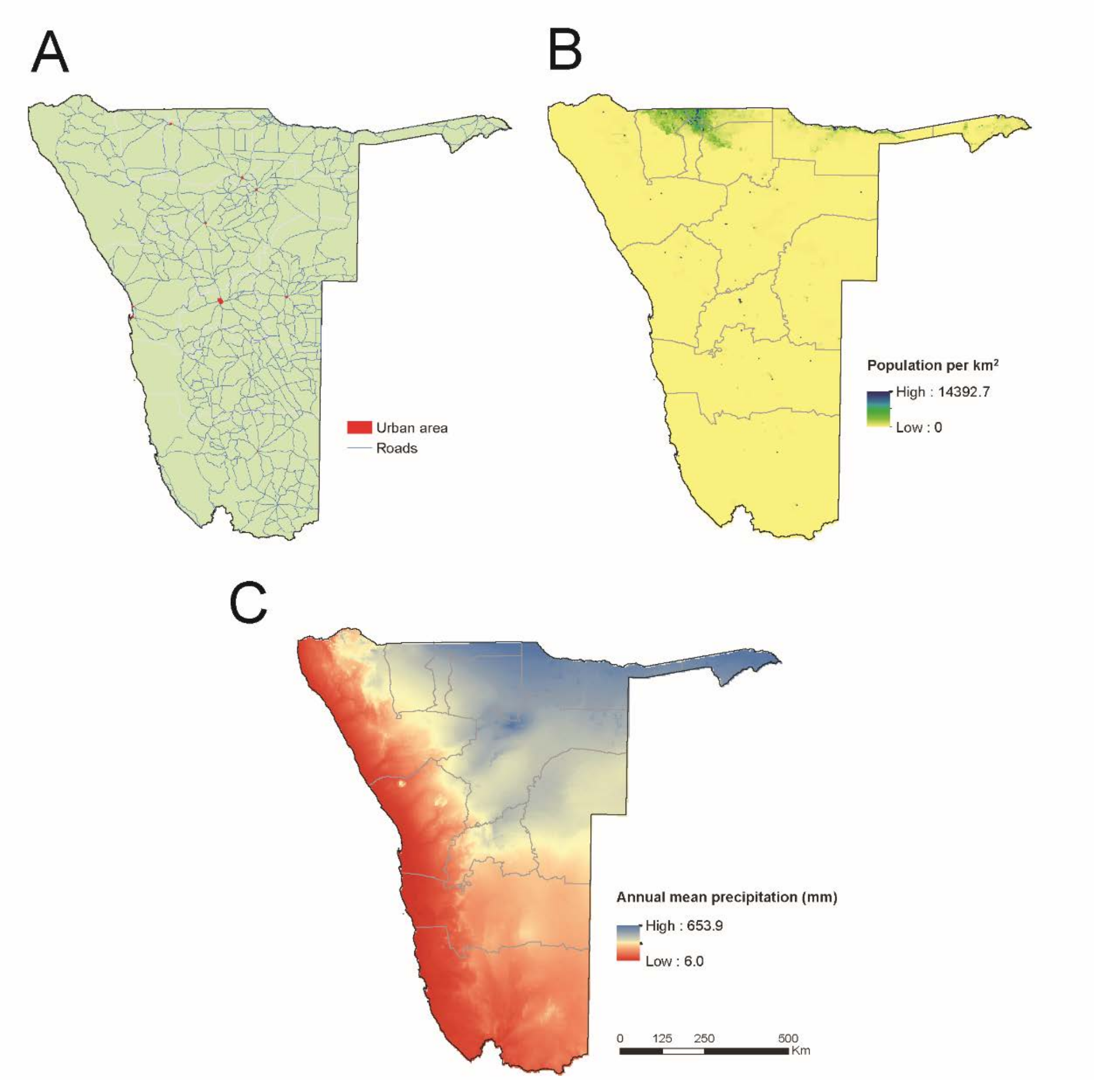

Fig. S1. Urban area and road networks (*A*), population density (*B*), and annual precipitation (*C*) in Namibia. The data on urban areas were obtained from the Natural Earth (www.naturalearthdata.com/downloads/10m-cultural-vectors/10m-urban-area/) (Schneider et al., 2003), and the road networks were obtained from DIVA-GIS (www.diva-gis.org), the population density data were downloaded from WorldPop (www.worldpop.org), and the average annual precipitation for 1970-2000 were obtained from WorldCllim (worldclim.org/version2).

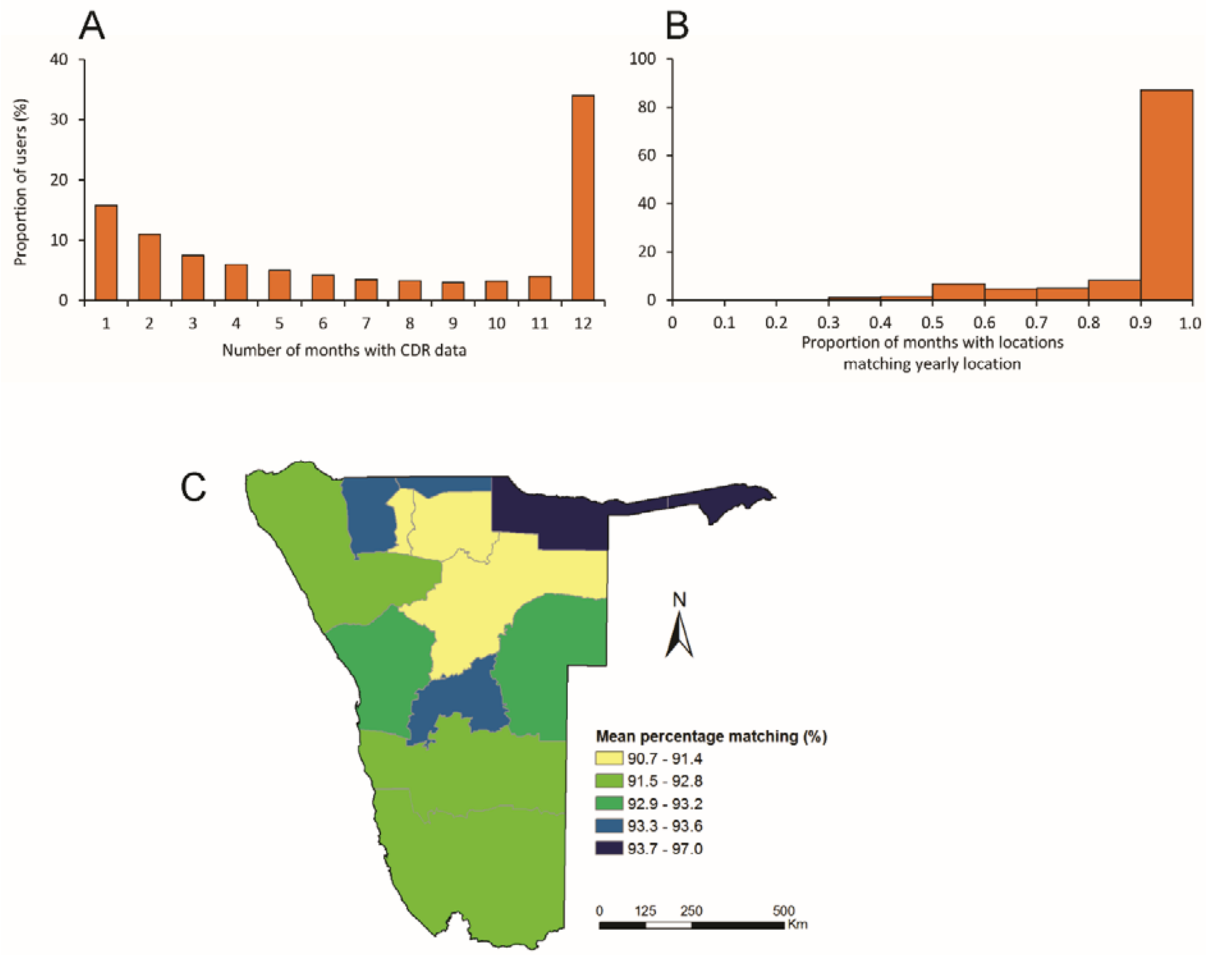

Fig. S2. Comparing CDR-derived monthly and yearly locations of subscribers in 2012. (*A*) Users with monthly CDR data for defining locations. (*B*) Users with monthly locations matching yearly locations, excluding users with only 1-month data. (*C*) Percentage of monthly locations matching yearly locations by region. The monthly/yearly location was defined as the most frequent location at regional level across the whole month/year.

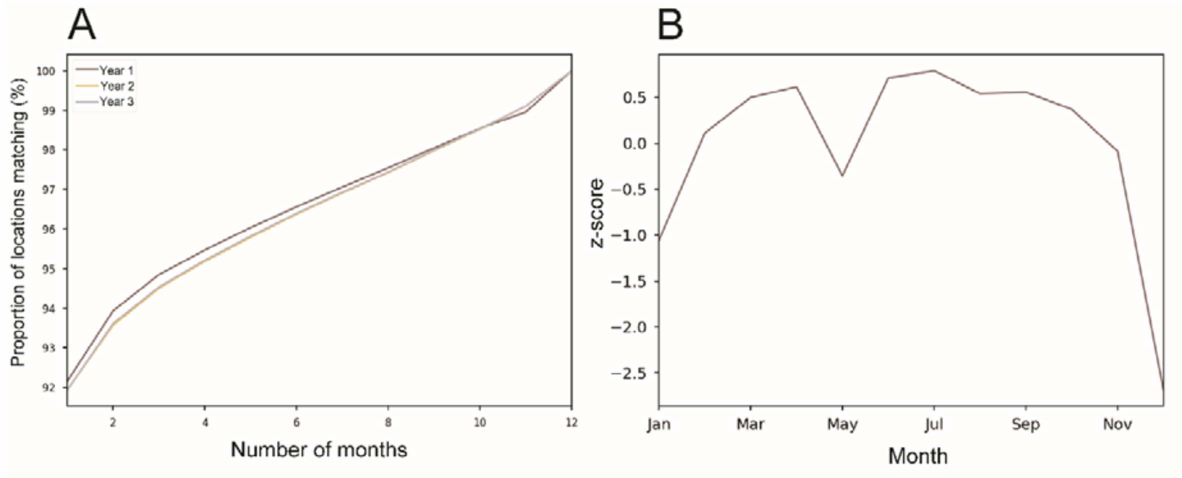

Fig. S3. Comparing the ability of periods of different lengths to define residential location. (*A*) Proportion of monthly locations using different period data and matching the yearly location in three periods: Year 1 (October 2010 – September 2011), Year 2 (October 2011 – September 2012), and Year 3 (October 2012 – September 2013). The proportions in Year 1 and Year 2 are almost identical. (*B*) The Z-score of the percentages of users with monthly locations matching yearly location. The monthly/yearly location was defined as the most frequent location of a phone user over the course of the corresponding period.

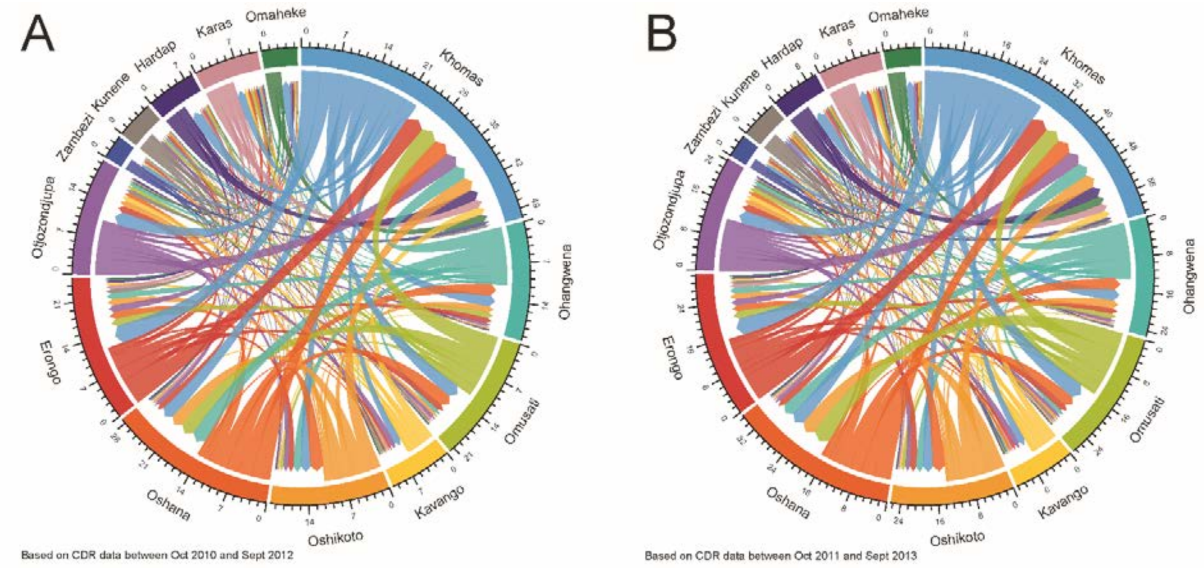

Fig. S4. CDR-derived migrating phone users between regions of Namibia in (*A*) 2011 and (*B*) 2012. The origins and destinations of migrants are each assigned a colour and represented by the circle's segments. The direction of the flow is encoded by both the origin region's colour and a gap between the flow and the destination region's segment. The volume of movement is indicated by the width of flow. Because the flow width is nonlinearly adapted to the curvature, it corresponds to the flow size only at the beginning and end points. Tick marks on the circle segments show the number of migrants (inflows and outflows) in thousands.

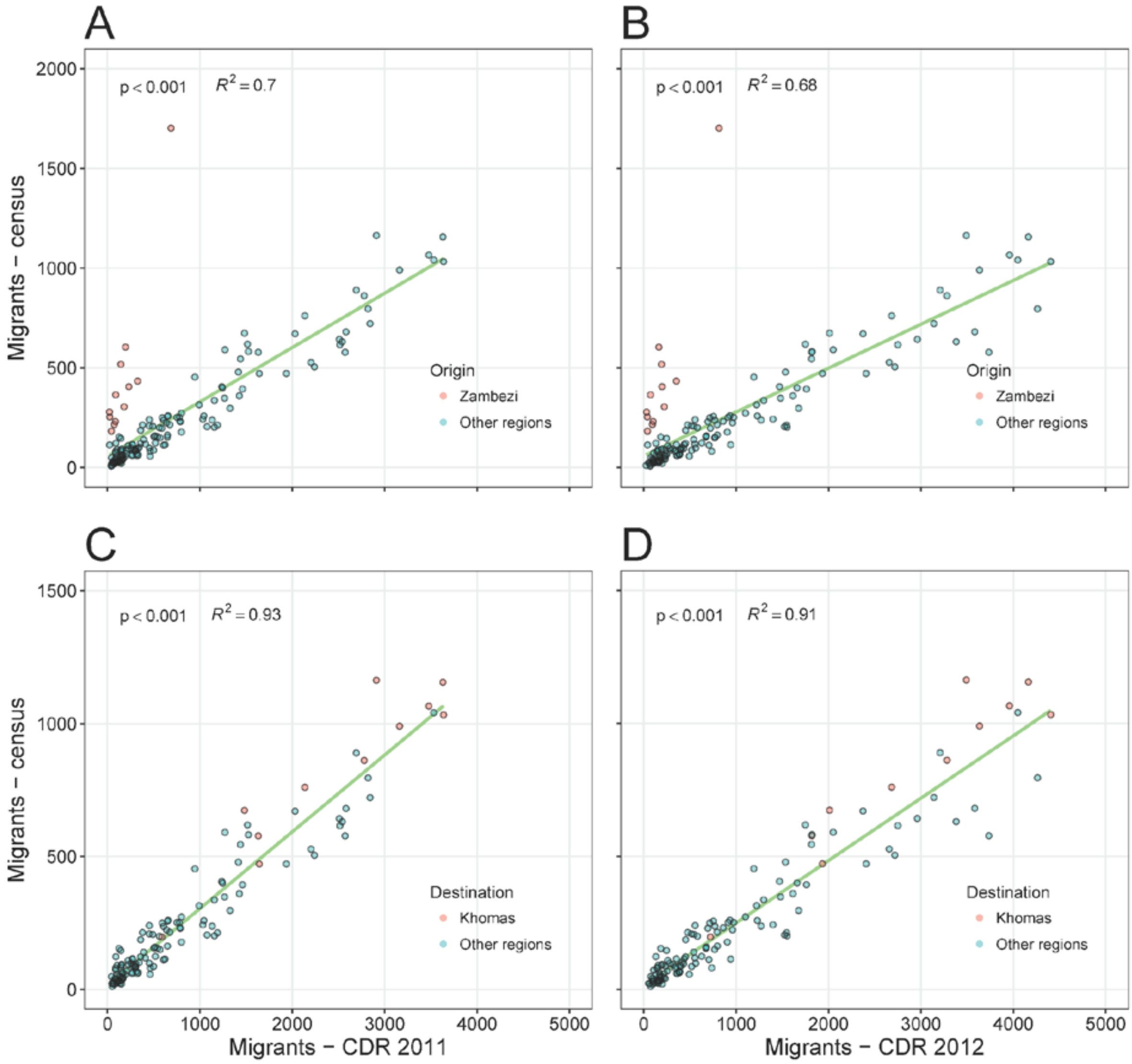

Fig. S5. Relation between census-based migration data and CDR-derived migrating mobile phone users at regional level, with (*A*) and (*B*) for all 13 regions of Namibia, and (*C*) and (*D)* for 12 regions except Zambezi. The green solid lines represent linear regression fit, with *p* values and R-squared (R^2) values presented.

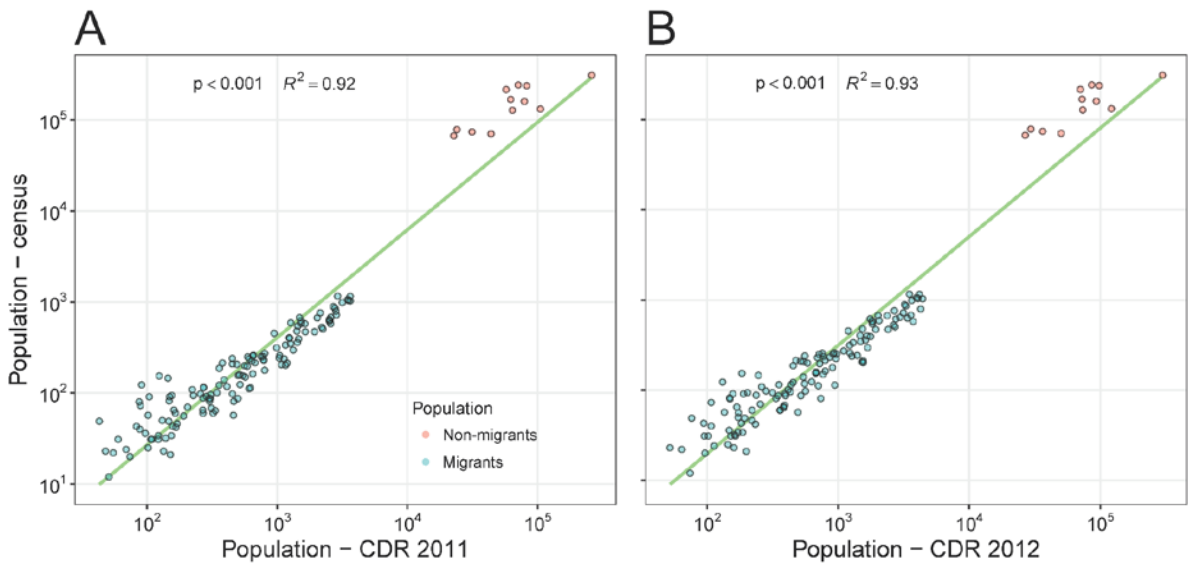

Fig. S6. Relations between 2011 census-derived population and CDR-derived population of mobile phone users in Periods 2011 (*A*) and 2012 (*B*) between regions of Namibia except Zambezi. The blue dots represent the population migrating from one region to another region, and the red dots represent the residents staying in the same region. The green solid lines represent linear regression fit with *p* and R^2 values.

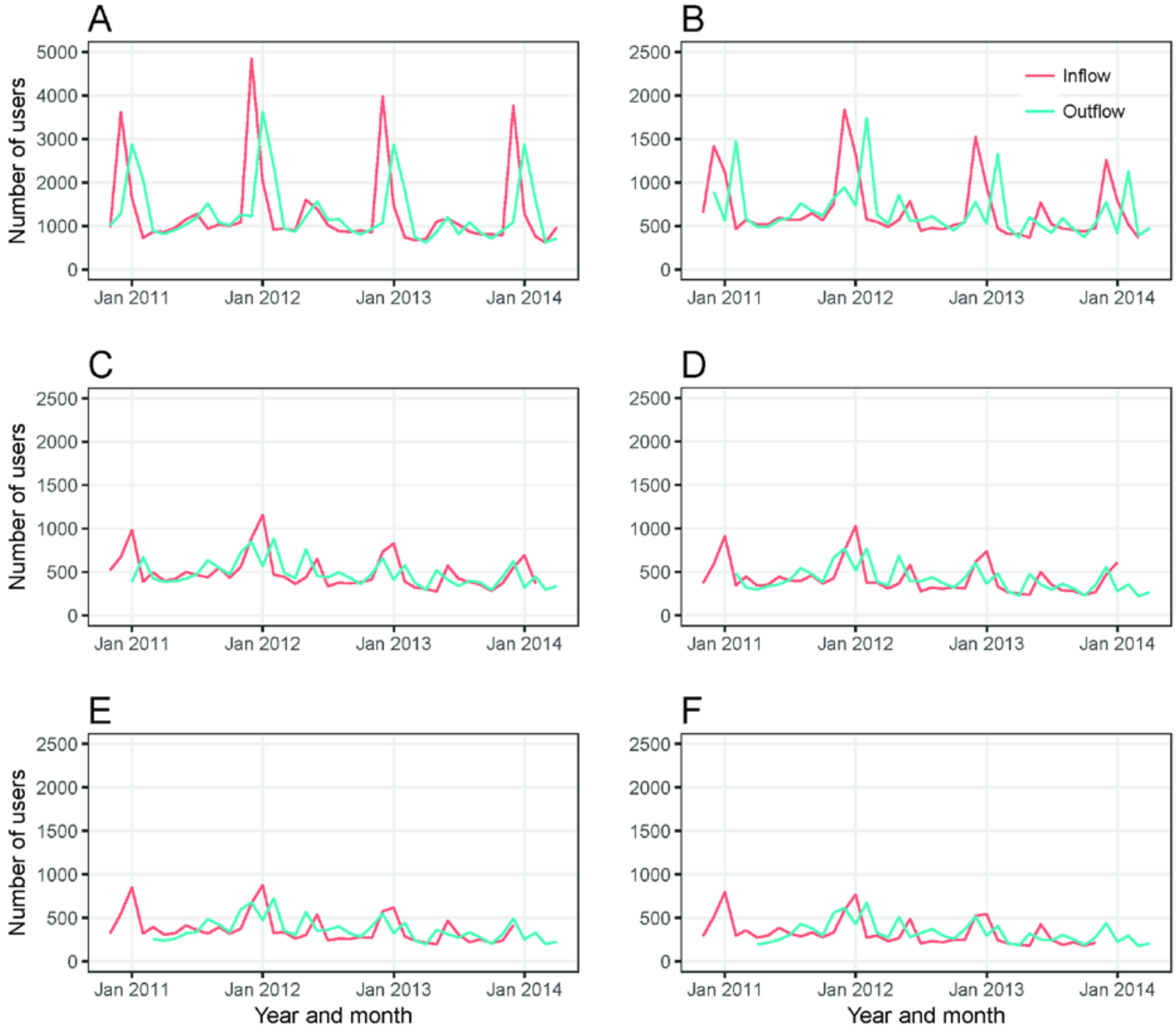

Fig. S7. CDR-derived inflow and outflow of Zambezi with usual residences defined by different lengths of periods, from one month (A) to six months (E). The usual residence at regional level was defined as the most frequent location of a mobile phone user over the course of corresponding period.

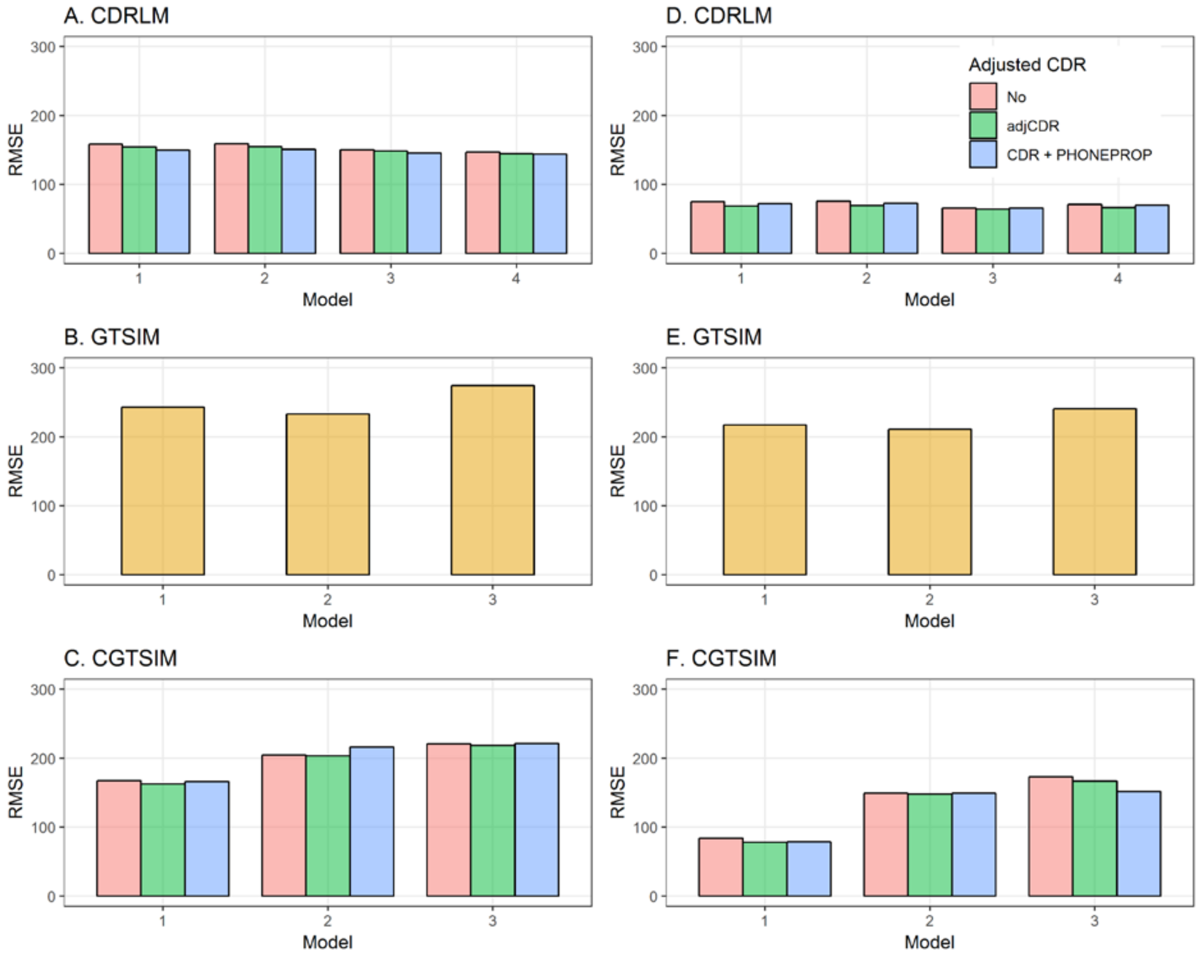

Fig. S8. The root-mean-square error (RMSE) of the CDR-based linear models (CDRLMs), gravity-type spatial interaction models (GTSIMs) and CDR-based gravity-type spatial interaction models (CGTSIMs). The results in (*A*), (*B*), and (*C*) are the RMSE of models tested for all regions, while results in (*D*), (*E*), and (*F*) are the RMSE of models tested for regions except Zambezi. The variables included in these models are detailed in Table S1.

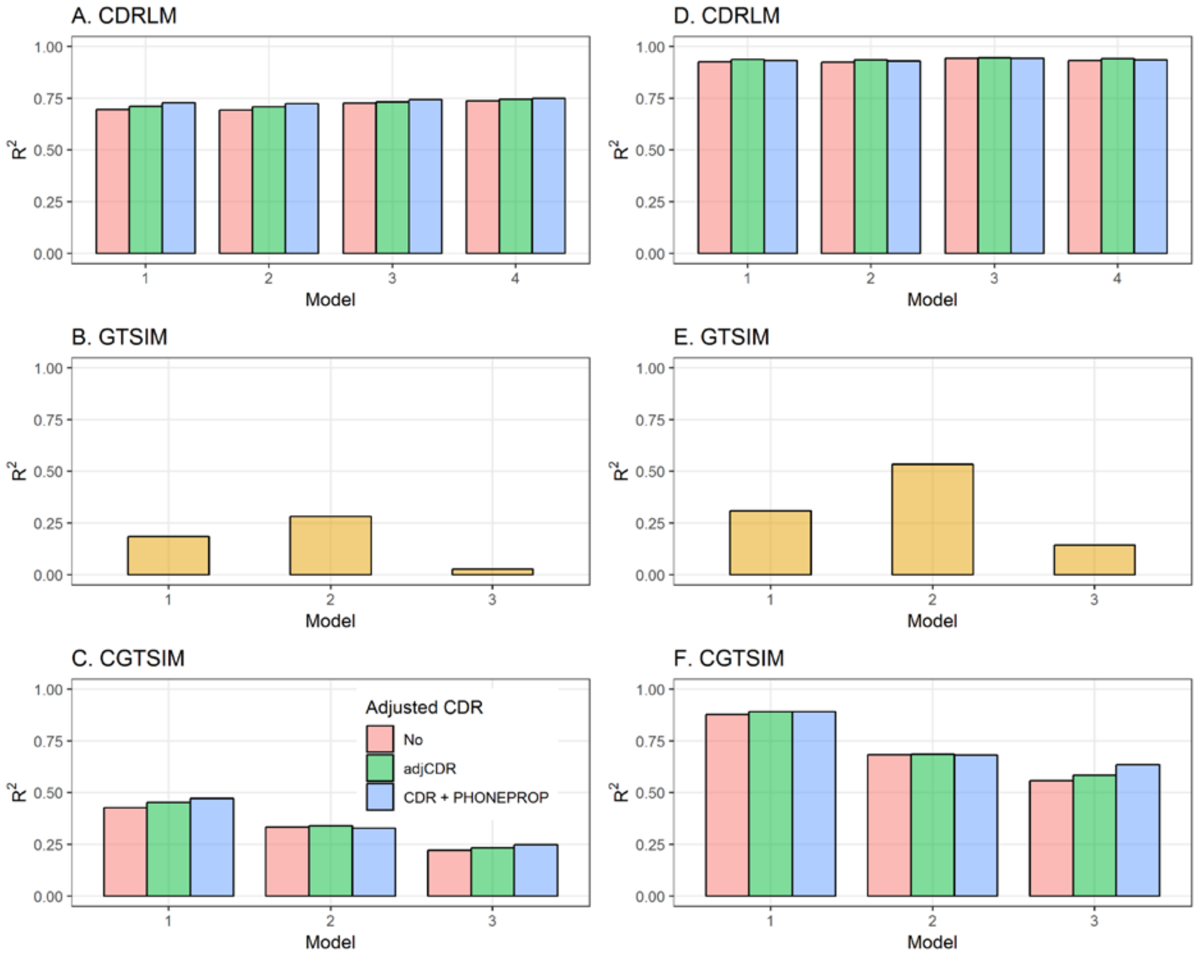

Fig. S9. The R^2 of CDRLMs, GTSIMs and CGTSIMs. The results in (*A*), (*B*), and (*C*) are the R^2 of models tested for all regions, while results in (*D*), (*E*), and (*F*) are the R^2 of models tested for regions except Zambezi. The variables included in these models are detailed in Table S1.

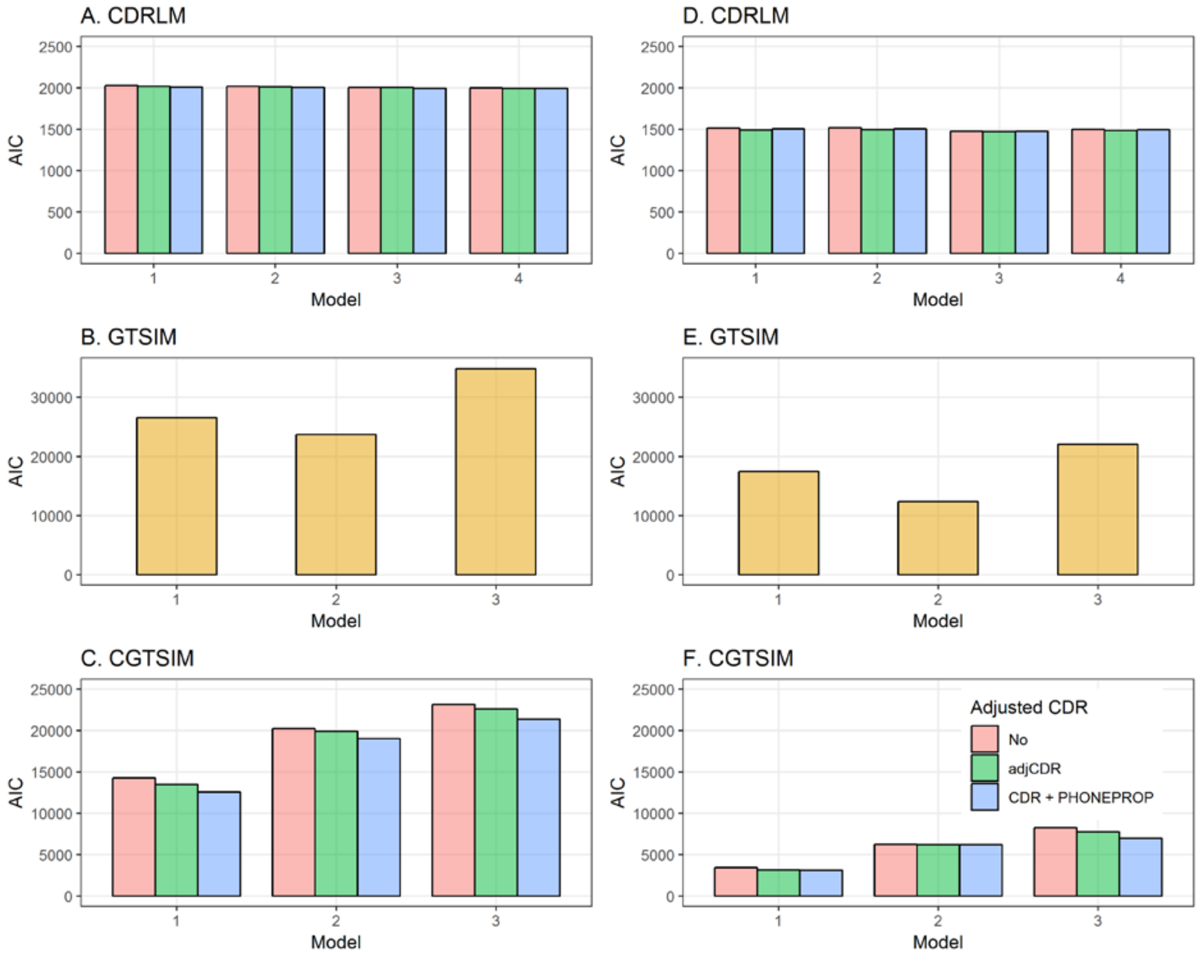

Fig. S10. The Akaike Information Criterion (AIC) for CDRLMs, GTSIMs and CGTSIMs. The results in (*A*), (*B*), and (*C*) are the AIC of models tested for all regions, while results in (*D*), (*E*), and (*F*) are the AIC of models tested for regions except Zambezi. The variables included in these models are detailed in Table S1.

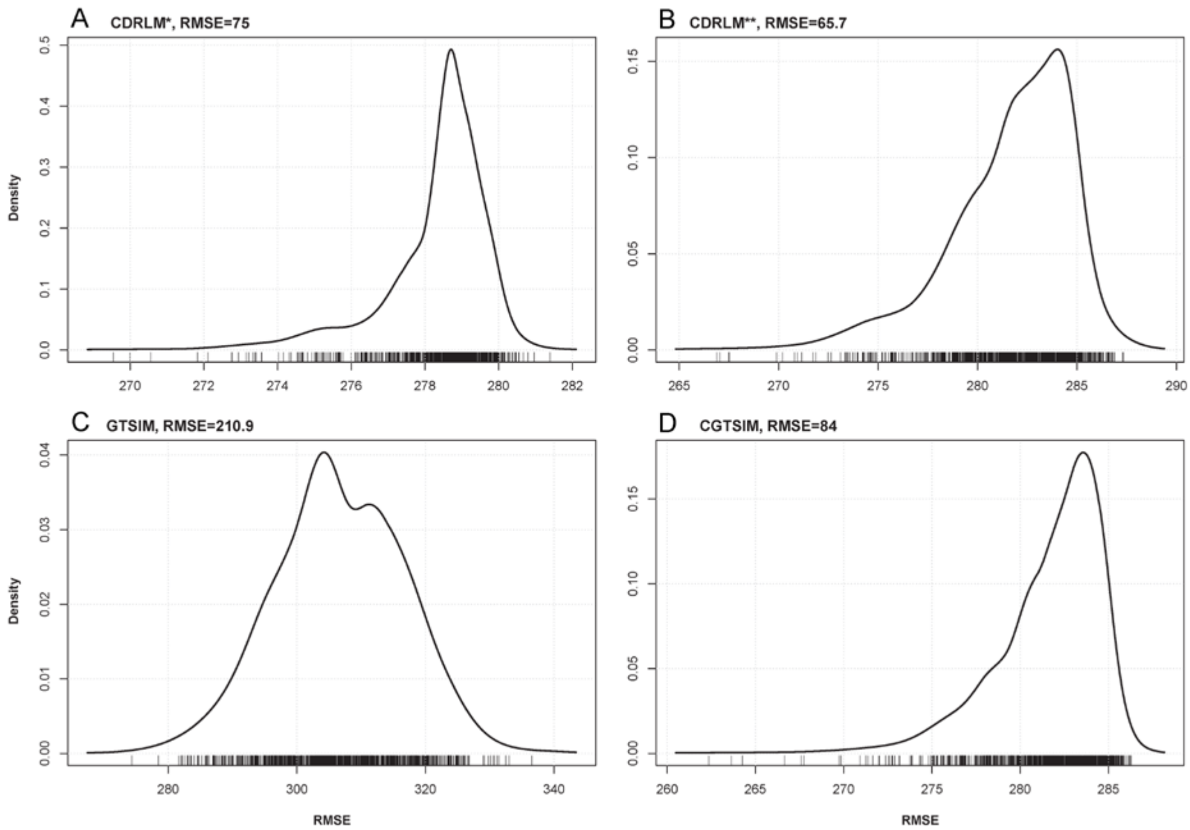

Fig. S11. The distribution of RMSE of models tested by shuffled census-derived migration data through 1000 simulations. (*A*) CDRLM only using a variable of unadjusted CDRs, and (*B*), (*C*) and (*D*) results of optimal CDRLM, GTSIM and CGTSIM, respectively, using unadjusted CDR data and other variables (Fig S8 and Table S1). The RMSE of each model fitted by real, unshuffled census data is given in the title of each graph. The Zambezi region as an outliner is excluded in the dataset.

* The model #1 of CDRLM.

** The model #3 of CDRLM.

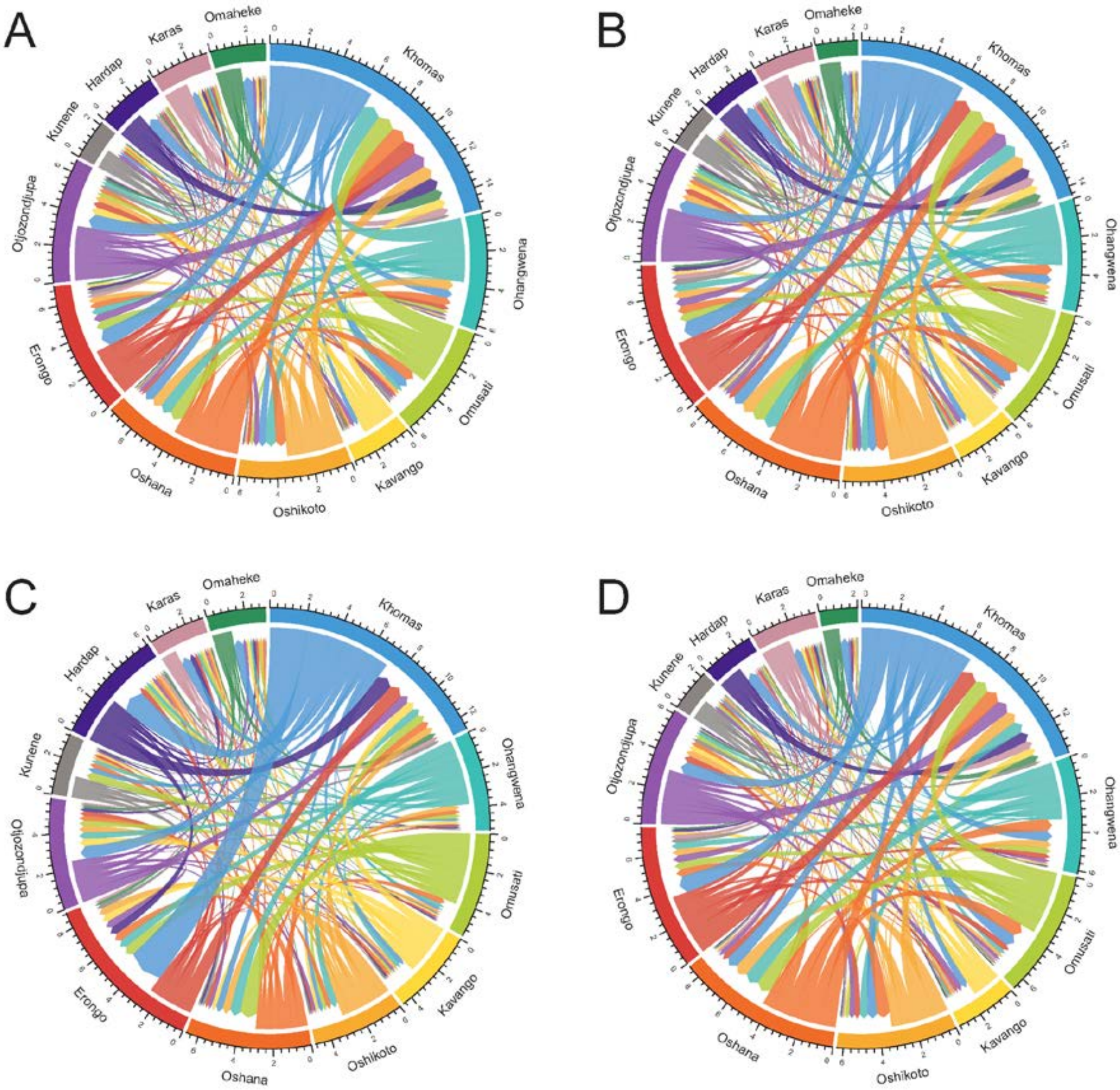

Fig. S12. Comparing census-derived and model-estimated internal migration between regions in Namibia. (*A*) Migration flows derived from the 2011 Namibia Population and Housing Census data. (*B*) Estimates made by the CDRLM. (*C*) Estimates made by GTSIM. (*D*) Estimates made by the CGTSIM. Tick marks on the circle segments show the number of migrants (inflows and outflows) in thousands. The Zambezi region as an outliner is excluded, and the estimates of optimal models with the lowest RMSE using unadjusted CDR data (Fig S8 and Table S1) are presented here.

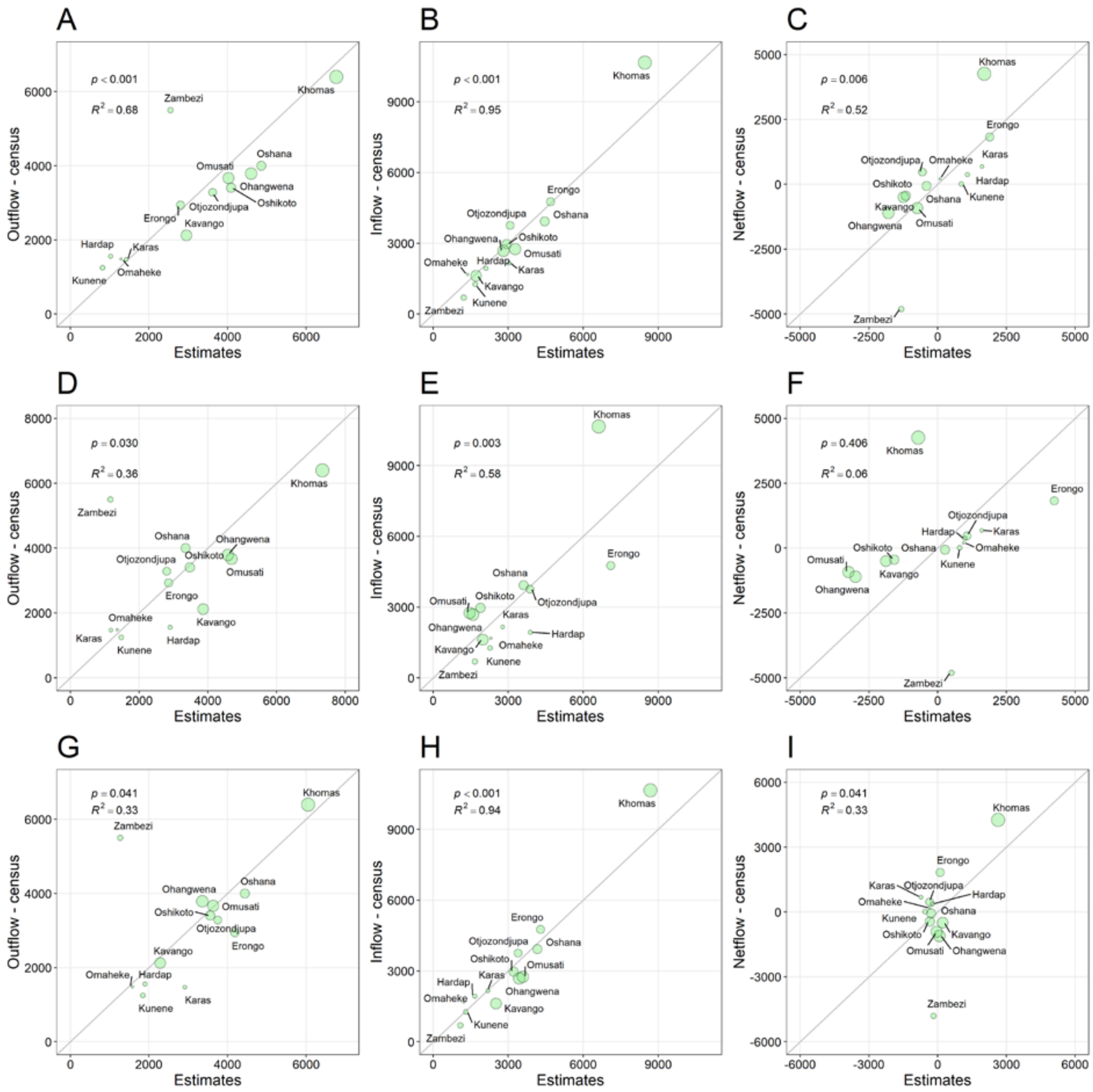

Fig. S13. Comparing regional outflow, inflow and net migration between 2011 census data and estimates made by models for all regions in Namibia. The estimates of optimal models with the lowest RMSE using unadjusted CDR data (Fig S8 and Table S1) are presented, with CDRLM in (*A*), (*B*) and (*C*), GTSIM in (*D*), (*E*) and (*F*), and CGTSIM in (*G*), (*H*) and (*I*).

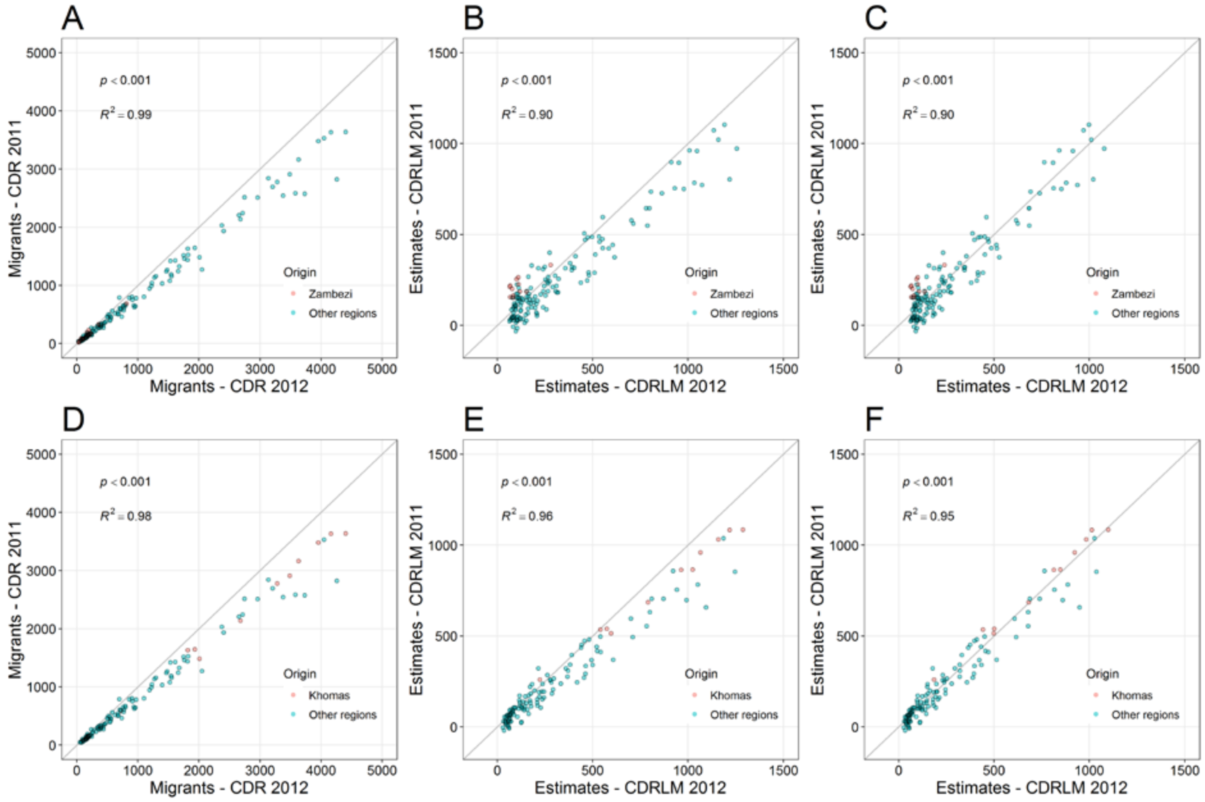

Fig. S14. CDR-derived migrating users and estimates by CDRLM for Periods 2011 and 2012. The graphs in (*A*), (*B*) and (*C*) were plotted for all regions, while Zambezi as an outliner was excluded in (*D*), (*E*) and (*F*). CDR-derived number of migrating mobile phone users are presented in (*A*) and (*D*), respectively. Estimates made by CDRLM are showed in (*B*) and (*E*), and results of CDRLM using CDR data adjusted for the effect of increasing users across years are presented in (*C*) and (*F*). The fitted CDRLMs using only CDRs for 2011 were used to predict migration in 2012.

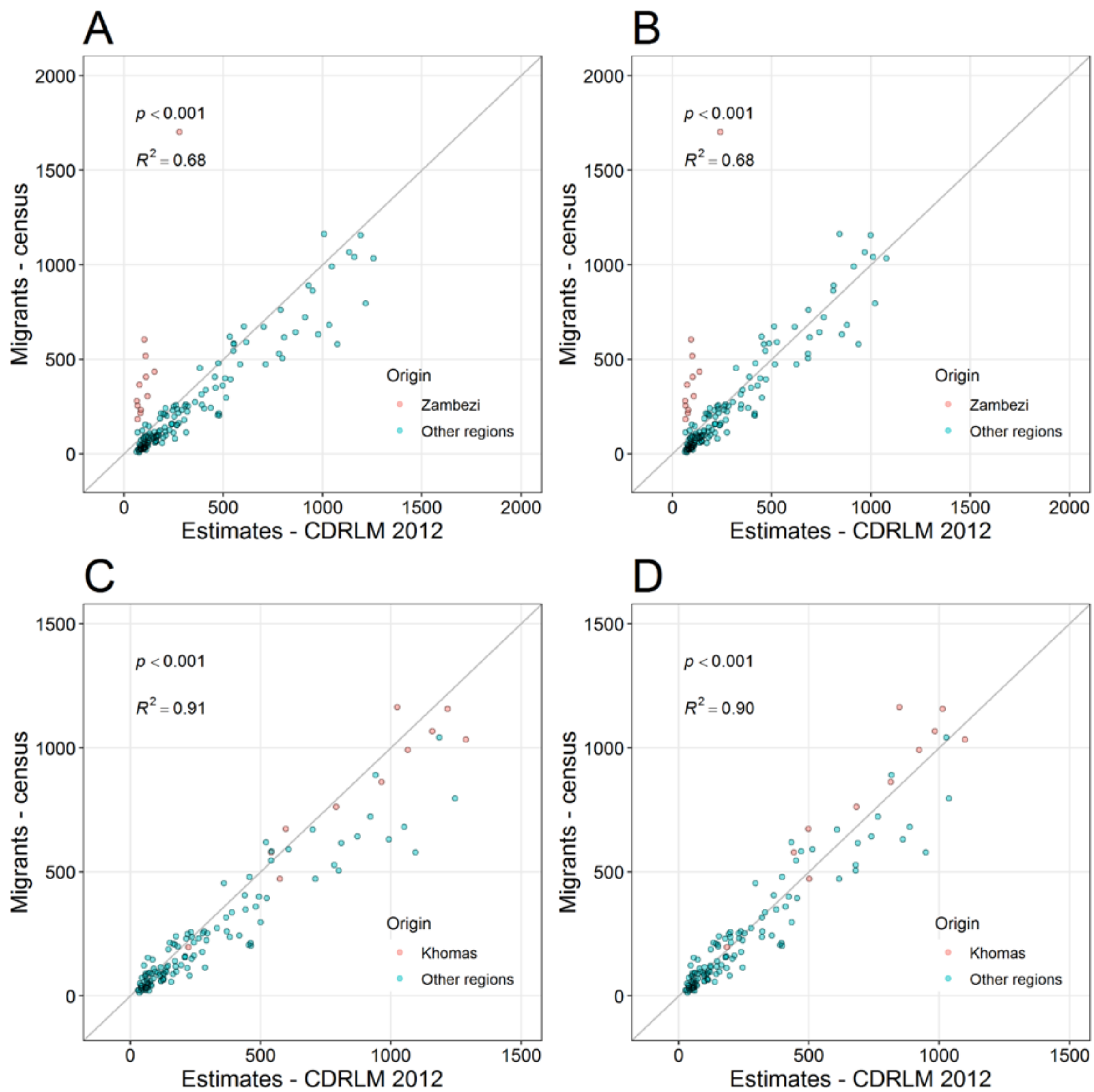

Fig. S15. Comparing census-derived migration in 2011 and estimates made by the CDRLM for 2012. (*A*) Estimates using unadjusted 2012 CDR data for all regions. (*B*) Estimates made using 2012 CDR-derived migrating user data adjusted for the effect of increasing users across years. (*C*) Estimates using unadjusted 2012 CDR data for regions excluding Zambezi. (*D*) Estimates using 2012 CDR data adjusted for the effect of the increasing users across years in regions excluding Zambezi. The fitted CDRLMs using only CDRs for 2011 were used to predict the migration in 2012 with corresponding CDR data.

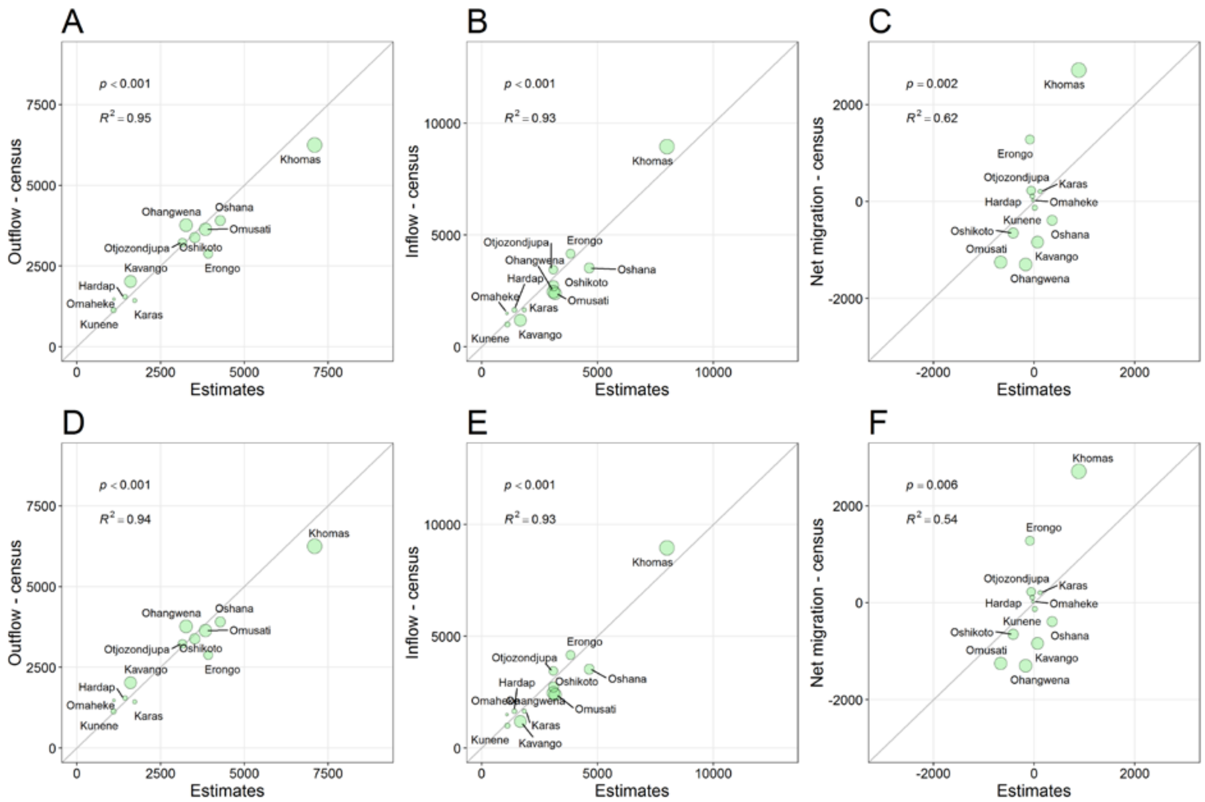

Fig. S16. Comparing regional outflows, inflows and net migration between census data in 2011 and estimates made by CDRLM in 2012. The predictions of CDRLM using unadjusted 2012 CDR data are presented in (*A*), (*B*) and (*C*), while (*D*), (*E*) and (*F*) show the results of CDRLM using CDR data adjusted to offset the effect of increasing number of users across years. The statistics of migration were obtained from the 2011 Namibia Population and Housing Census data. The Zambezi region as an outliner is excluded in these graphs, and the fitted CDRLMs using only CDRs for 2011 were used to predict the migration in 2012 with corresponding CDR data.

Reference

Callegaro M and Poggio T (2004) Where can I call you? The "Mobile (Phone) Revolution" and its impact on survey research and coverage error: a discussion of the Italian case.

GADM (2018) *Database of Global Administrative Areas (version 3.4)*. Available from: http://www.gadm.org [Accessed 24 April 2018]

Garcia A J, Pindolia D K, Lopiano K K and Tatem A J (2015) Modeling internal migration flows in sub-Saharan Africa using census microdata. *Migration Studies*; **3**(1): 89-110.

Henry S, Boyle P and Lambin E F (2003) Modelling inter-provincial migration in Burkina Faso, West Africa: the role of socio-demographic and environmental factors. *Applied Geography*; **23**(2-3): 115-136.

Henry S, Piche V, Ouedraogo D and Lambin E F (2004) Descriptive analysis of the individual migratory pathways according to environmental typologies. *Population and Environment*; **25**(5): 397-422.

Lall S V, Selod H and World Bank. (2006) *Rural-urban migration in developing countries a survey of theoretical predictions and empirical findings*. Washington, D.C.: World Bank, Available from: http://econ.worldbank.org/external/default/main?pagePK=64165259&theSitePK=469382&piPK=64165421&menuPK=64166093&entityID=000016406_20060505110833

Mobile Telecommunications (2012) *MTC Annual Report 2012*. Available from: http://www.mtc.com.na/sites/annual-reports/2012/Annual_Report_2012.pdf [Accessed 21 November 2017]

Mobile Telecommunications (2014) *MTC Annual Report 2014*. Available from: http://www.mtc.com.na/sites/annual-reports/2014/pdf/MTCAnnualReport2014.pdf [Accessed 23 November 2017]

Mobile Telecommunications (2018) *Coverage*. Available from: http://www.mtc.com.na/coverage [Accessed 3 June 2018]

Namibia Statistics Agency (2013) *Namibia 2011 Population & Housing Census Main Report*. Available from: https://cms.my.na/assets/documents/p19dmn58guram30ttun89rdrp1.pdf [Accessed 22 November 2017]

Ruktanonchai N W, Bhavnani D, Sorichetta A, Bengtsson L, Carter K H, Cordoba R C, Le Menach A, Lu X, Wetter E, Erbach-Schoenberg E Z and Tatem A J (2016) Census-derived migration data as a tool for informing malaria elimination policy. *Malaria Journal*; **15**

Schneider A, Friedl M A, McIver D K and Woodcock C E (2003) Mapping Urban Areas by Fusing Multiple Sources of Coarse Resolution Remotely Sensed Data. *Photogrammetric Engineering & Remote Sensing*; **69**(12): 1377-1386.

Smits J and Steendijk R (2015) The international wealth index (IWI). *Social Indicators Research*; **122**(1): 65-85.

Sorichetta A, Bird T J, Ruktanonchai N W, Erbach-Schoenberg E Z, Pezzulo C, Tejedor N, Waldock I C, Sadler J D, Garcia A J, Sedda L and Tatem A J (2016) Mapping internal connectivity through human migration in malaria endemic countries. *Scientific Data*; **3**

The Namibia Ministry of Health, Social Services - MoHSS/Namibia and ICF International (2014) Namibia Demographic and Health Survey 2013. Windhoek, Namibia: MoHSS/Namibia and ICF International

Vobruba T, Korner A and Breitenecker F (2016) Modelling, Analysis and Simulation of a Spatial Interaction Model. *Ifac Papersonline*; **49**(29): 221-225.

Wesolowski A, O'Meara W P, Eagle N, Tatem A J and Buckee C O (2015) Evaluating Spatial Interaction Models for Regional Mobility in Sub-Saharan Africa. *Plos Computational Biology*; **11**(7)

Wesolowski A, Zu Erbach-Schoenberg E, Tatem A J, Lourenco C, Viboud C, Charu V, Eagle N, Engo-Monsen K, Qureshi T, Buckee C O and Metcalf C J E (2017) Multinational patterns of seasonal asymmetry in human movement influence infectious disease dynamics. *Nature Communications*; **8**(1): 2069.