BMJ Global Health

Beyond HIV prevalence: identifying people living with HIV within underserved areas in South Africa

Hana Kim,[1,2] Frank Tanser,[3,4,5,6] Andrew Tomita,[7,8] Alain Vandormael,[9] Diego F Cuadros [1,2]

To cite: Kim H, Tanser F, Tomita A, *et al*. Beyond HIV prevalence: identifying people living with HIV within underserved areas in South Africa. *BMJ Global Health* 2021;**6**:e004089. doi:10.1136/bmjgh-2020-004089

Handling editor Sanni Yaya

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjgh-2020-004089).

Received 30 September 2020
Revised 10 March 2021
Accepted 12 March 2021

Check for updates

For numbered affiliations see end of article.

Correspondence to
Dr Diego F Cuadros; diego.cuadros@uc.edu

ABSTRACT

Introduction Despite progress towards the Joint United Nations Programme on HIV/AIDS 95-95-95 targets, South Africa is still suffering from one of the largest HIV epidemics globally. In this study, we generated high-resolution HIV prevalence maps and identified people living with HIV (PLHIV) in underserved areas to provide essential information for the optimal allocation of HIV-related services.

Methods The data come from the South Africa Demographic and Health Survey conducted in 2016 and spatial variables from other published literature. We produced high-resolution maps of HIV prevalence and underserved areas, defined as a greater than 30 min travel time to the nearest healthcare facility. Using these maps and the population density, we mapped PLHIV and the PLHIV within underserved areas for 30, 60 and 120 min thresholds.

Results There was substantial geographic variation in HIV prevalence, ranging from 1.4% to 24.2%, with a median of 11.5% for men, and from 2.1% to 48.1%, with a median of 20.6% for women. Gauteng province showed the highest density for both HIV prevalence and PLHIV. 80% of all areas in the country were identified as underserved areas (30 min threshold), which contained more than 16% and 20% of the total men and women living with HIV, respectively. KwaZulu-Natal province had the largest number of PLHIV in underserved areas (30 min threshold) and showed less than one healthcare facility per 1000 PLHIV.

Conclusion Our study showed extensive spatial variation of HIV prevalence and significant numbers of PLHIV in underserved areas in South Africa. Moreover, we identified locations where HIV-related services need to be intensified to reach the ~1.5 million PLHIV in underserved areas, particularly in KwaZulu-Natal province, with less than one healthcare facility per 1000 PLHIV.

Key questions

What is already known?

- ► South Africa suffers from one of the largest HIV epidemics, with more than 7.5 million people living with HIV (PLHIV) in 2018.
- ► The country achieved its first 90-90-90 target by diagnosing more than 90% of PLHIV.
- ► However, the other targets (the proportion of diagnosed people on antiretroviral therapy (ART) and the proportion of virally suppressed PLHIV on ART) are still behind, despite the aggressive expansion of ART coverage.

What are the new findings?

- ► South Africa showed extensive spatial variation of HIV-positive population for both genders.
- ► The largest PLHIV population is concentrated in the *Gauteng* province, with a density of 13 862 women per 5 km^2 and 15 484 men per 5 km^2.
- ► More than 80% of areas in the country are more than 30 min travel time from the nearest healthcare service.
- ► More than 16% of men living with HIV and 20% of women living with HIV reside in underserved areas where the nearest healthcare service cannot be reached within 30 min travel time.
- ► A total of 80% of PLHIV reside in only five provinces; *KwaZulu-Natal* province has the largest number of PLHIV within 30 min threshold of underserved areas containing more than 100 000 of PLHIV for each gender (149 197 for men and 174 543 for women) and shows less than one healthcare facility per 1000 PLHIV.

What do the new findings imply?

- ► There are still several barriers for HIV treatment service coverage that need to be overcome, particularly for PLHIV living in underserved areas where health services are not likely to be reached within threshold times.
- ► Efforts to allocate resources for HIV treatment need to be oriented towards the PLHIV located in underserved areas in provinces with low number of healthcare facilities per 1000 PLHIV.

INTRODUCTION

With the rapid achievement of the 90-90-90 strategy (90% of all people living with HIV (PLHIV) know their HIV status, 90% of all people diagnosed with HIV receive sustained antiretroviral therapy (ART) and 90% of all people receiving ART have viral suppression) to guide country-led and regional-led efforts to rapidly scale up HIV prevention and treatment services, the annual global number of new HIV infections had declined from 2.1 million in 2010 to 1.7 million (16%) by 2018.[1] This progress led to revised 95-95-95 targets and an acceleration of efforts to end the HIV/AIDS epidemic by 2030. In addition, the Joint United Nations Programme on HIV/AIDS (UNAIDS) announced the Fast-Track approach focusing on the 30 countries that account for most of the world's people newly infected with HIV, which includes South Africa.[2]

South Africa, with more than 7.5 million PLHIV,[1] has one of the largest HIV epidemics in the world. To reverse the tide of the epidemic, the country[3] aggressively scaled up ART coverage, which has led to improvements in HIV outcomes. In 2018, the annual number of new HIV infections declined to 240 000 people, which represented 0.49% of the uninfected population, compared with 0.64% in 2015.[1] South Africa has achieved a 90% of PLHIV diagnosed, 68% of those diagnosed on ART and 88% of those on ART virally suppressed in 2018,[4 5] and is potentially on its way to meeting the 90-90-90 targets by 2022.[6] Despite remarkable progress, HIV prevalence in the country remains persistently high with 18.9% of adults aged 15 and older infected with the virus, and more than 200 000 new infections occurring every year.[7 8] Although the proportion of PLHIV who know their HIV status has been steadily increasing, the proportion of PLHIV on ART is remarkably low despite the expansion of ART coverage. The likelihood of South Africa achieving its 95-95-95 targets and ending the HIV epidemic by 2030 remains far from certain.

In order to meet those targets, we have previously argued for prioritising resource allocation and accelerating HIV prevention and treatment services in certain high-risk areas in South Africa.[9 10] South Africa shows substantial spatial variation in the HIV epidemic with infections clustering in the eastern part of the country, between the *Limpopo* and *KwaZulu-Natal* provinces.[6 11] However, most past measures of HIV prevalence in the country are frequently available only for large geographical administrative units (ie, low spatial resolution), or they are only focused on localised epidemics.[7] These large-scale national or regional measures can obscure localised aspects of the HIV transmission process across the country. Moreover, these measures are estimated for the total population, but spatial disparities of the epidemic and their social and behavioural drivers can exist.[8] Microlevel analyses on the spatial structure of the HIV epidemic and the density of PLHIV in low-access areas could help facilitate the operation of HIV prevention and treatment services by targeting resource allocations to the areas. To maximise the achievement of the 95-95-95 targets by 2030, identifying areas where the burden of HIV is concentrated could play a key role in recognising vulnerable populations at high risk of infection, but also in prioritising the access to critical healthcare for PLHIV in these areas.[12–19]

Given this context, the main objective of this study was to generate high-resolution, spatial predictions of HIV epidemiological measures for men and women (HIV prevalence and density of PLHIV) in South Africa. We then used these spatial predictions to conduct a detailed spatial analysis of underserved areas from critical healthcare services to identify men and women living with HIV in areas with low access to HIV prevention and treatment services. This information would be of high relevance for assisting the national strategy to target appropriate areas and populations and achieve the UNAIDS targets in South Africa.[20]

METHODS

Study area and data sources

The main source of data for this study was the South Africa Demographic and Health Survey (SADHS) conducted in 2016,[21] which contains HIV serological biomarker information along with the corresponding spatial information of sample locations. The SADHS is a cross-sectional household survey designed to collect nationally representative

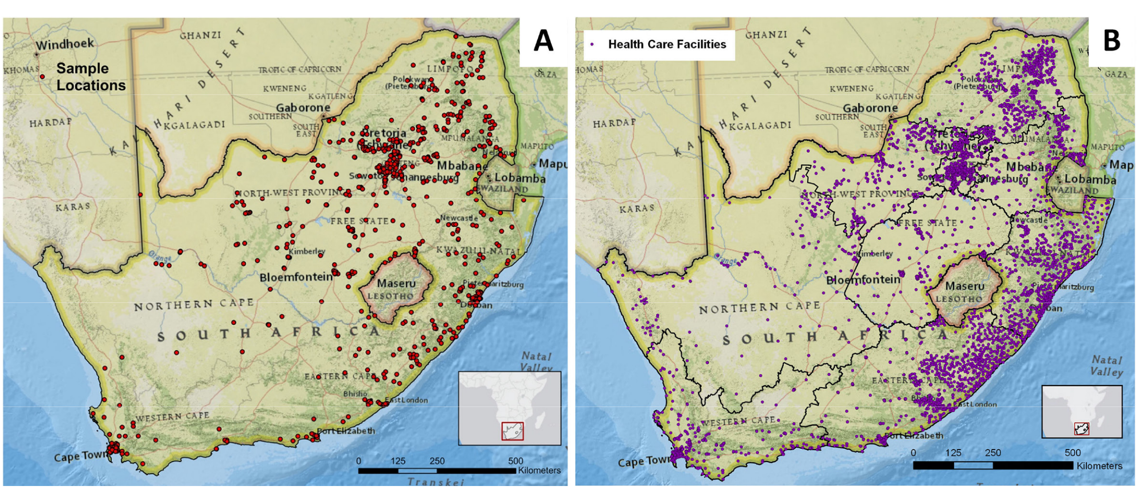


Figure 1 Sample locations. (A) Demographic and Health Survey sample locations for South Africa Demographic and Health Survey (SADHS) in 2016, and (B) healthcare facility locations. Map was created using ArcGIS by ESRI V.10.5 (http://www.esri.com).

data on population, health and socioeconomic parameters.[21] Individuals were enrolled in the SADHS via two-stage sampling procedure to select households. Men and women aged 15–49 in the selected households were eligible for the study. Of 12 132 participants (3618 men and 8514 women), 4862 individuals (2136 men and 2726 women) who interviewed and provided specimens for HIV testing were included. The SADHS sample locations are where the individuals participated the survey. The number of the sample location is 601 for men and 633 for women. The global positioning system was used to identify and record the geographical coordinates of each SADHS sample location (figure 1A).[22] Further details related to the population of the study can be found in the online supplemental materials.

Demographic and Health Surveys (DHS) has conducted nationally representative anonymous HIV testing to track HIV epidemics by collecting blood from representative samples of the population of women and men in a country since 2001.[21 23] Respondents who participated in the testing received educational materials and referrals for free voluntary counselling and testing. HIV serostatus was determined using the ELISA, based on a parallel testing algorithm. Further details related to the SADHS methodology, study design and data can be found in online supplemental materials and elsewhere.[21 23] This study follows the guidelines of the Strengthening the Reporting of Observational Studies in Epidemiology.[24]

Selection of spatial variables and HIV mapping

We selected variables from the SADHS data sets and other sources to investigate the association between socioeconomic and behavioural factors and the spatial structure of HIV in South Africa. Six socioeconomic and behavioural variables that have been previously associated with the risk of HIV prevalence were measured for men and five for women.[12 25] These included education level, poverty estimated using the DHS wealth index, condom use, number of lifetime sexual partners, if the participants had ever been previously tested for HIV and male circumcision (MC; male only). MC was a self-reported variable of circumcision status and thus did not consider which types of circumcision (eg, voluntary medical MC or traditional MC) were conducted. These variables were evaluated as percentages at each SADHS sample location.

We included seven environmental variables in the analysis: the normalised difference vegetation index (NDVI), night light, distance to main roads and dams, global human influence index (GH-I), friction map representing allocated fastest speed of travel based on the types of travel mode estimated within the pixel and population density. We quantified all variables at each DHS sample location. For example, we extracted the NDVI value at each SADHS sample location, and the distance from each sample to the closest main road was calculated and assigned at each sample location. The definition and descriptive statistics for each variable are presented in online supplemental table S1.

To generate disease maps of HIV prevalence in South Africa for each gender, we implemented a method previously developed to describe the spatial structure of HIV prevalence in several countries in sub-Saharan Africa (SSA) including Kenya, Malawi, Mozambique, Tanzania and Zimbabwe.[12 25] We used ordinary kriging mapping methods to generate continuous surface maps for each

Table 1 Variables in the final logistic regression models for men and women

Gender	Variable	Estimate	SE	P value	Moran's Index	P value
Women	Intercept	−1.008441	0.1343	<0.001	–	–
	Condom use	0.007380	0.0015	**<0.001**	0.04	0.08
	Poverty	0.005842	0.0012	**<0.001**	0.19	<0.001
	Lifetime number of sexual partners (<3)	−0.009180	0.0016	**<0.001**	0.07	0.00
	Friction	−24.314521	11.6500	**0.04**	–	–
	Distance to main road	−0.004377	0.0031	0.16	–	–
	NDVI	0.892539	0.1972	**<0.001**	–	–
Men	Intercept	−2.376504	0.3426	<0.001	–	–
	Ever been tested for HIV	0.006193	0.0028	**0.03**	0.04	0.06
	Poverty	0.007057	0.0019	**<0.001**	0.21	<0.001
	Lifetime number of sexual partners (<3)	−0.014912	0.0031	**<0.001**	0.04	0.07
	Male circumcision	−0.009507	0.0020	**<0.001**	0.11	<0.001
	GH-I	0.018049	0.0059	**<0.001**	–	–
	Distance to main road	−0.004983	0.0047	0.29	–	–
	NDVI	0.230944	0.2767	0.40	–	–

Bold values denote statistical significance at the p <0.05 level.
GH-I, global human influence index; NDVI, normalised difference vegetation index.

significant socioeconomic, demographic and behavioural factor identified in the multivariable regression model. The variables satisfied both criteria: (1) the bivariate logistic regression slope coefficient must have p value <0.1; and (2) the significance test for spatial autocorrelation must have p value <0.1. Then, we generated a map of HIV prevalence in raster format with 5 km grid resolution by substituting values from all continuous surface maps into the gender-specific multivariable logistic regression model for each gender. Finally, a 5 km × 5 km pixel resolution map of the density of women and men living with HIV was generated by combining the HIV prevalence map with the women and men density maps. Further details related to this methodology can be found in online supplemental materials.

Mapping HIV-underserved areas

To identify HIV-underserved areas, we used two data sets: the friction map from Malaria Atlas Project and the locations of public healthcare facilities in South Africa (figure 1B).[26] We employed travel time estimates, which offer the most accurate representation of the cost of travel, for measuring geographic accessibility and for identifying underserved areas.[27 28] We obtained all public healthcare facilities and private-not-for-profit sectors managed by government (eg, the Ministry of Health), local authorities, and faith-based organisation and non-governmental organisation to represent public health services.[26] A total of 4287 healthcare facilities were included in the analysis. We modelled travel time as a function of distance and travel speed. It can be conceptualised as the cost of movement using the cell dimensions (pixel resolution) and travel time assigned to each cell. In this model, travel occurs through cell to cell in both cardinal and diagonal directions, and a travel time was designated for each cell, representing the time required to traverse the cell estimated from the friction map. To identify underserved areas, we reclassified the accumulated travel time surface into a Boolean surface based on whether the cell was greater than three thresholds, 30, 60 and 120 min, from a healthcare location, then we converted the underserved areas from a raster grid to a vector data form (polygons) to identify the women

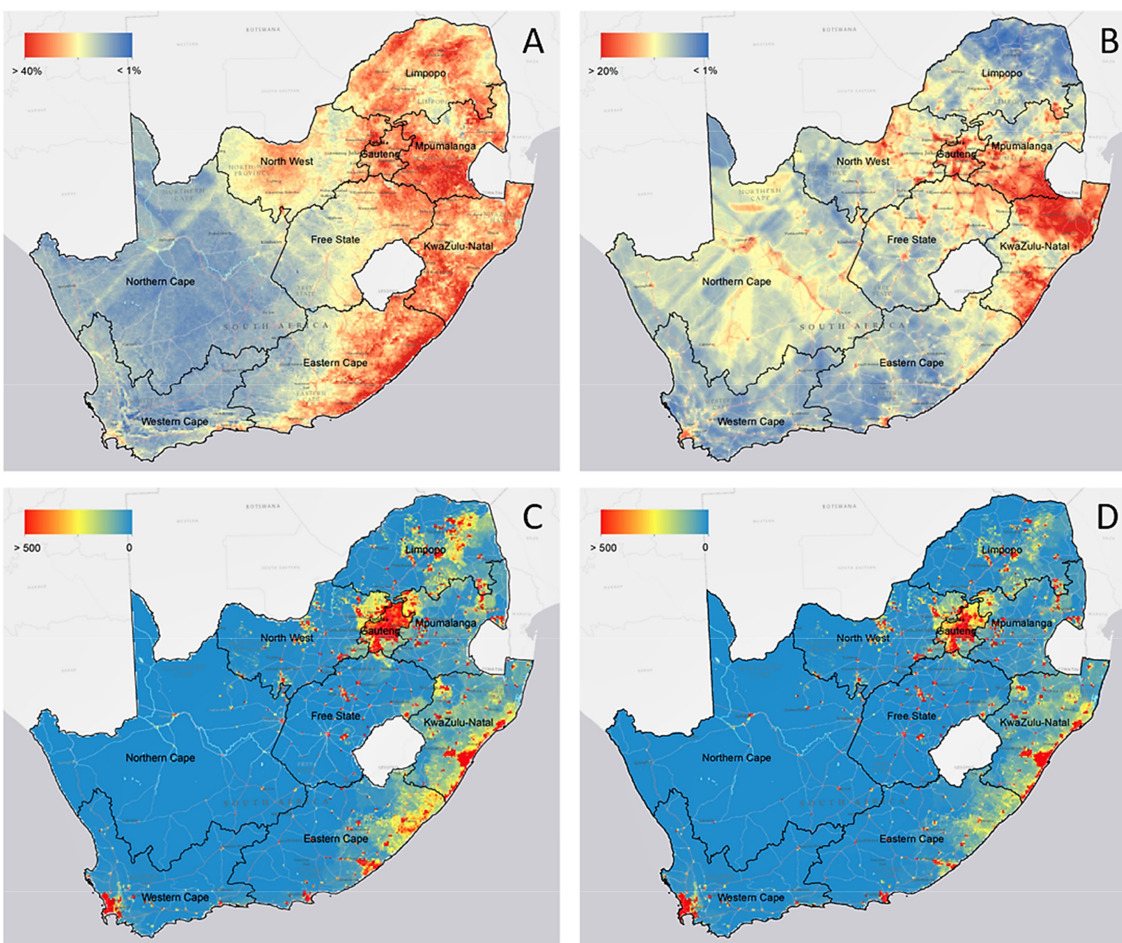


Figure 2 High-resolution maps of HIV prevalence and people living with HIV (PLHIV) in South Africa. High-resolution maps of HIV prevalence in South Africa for (A) men and (B) women in 2016; geographic dispersion of men (C) and women (D) living with HIV in South Africa. HIV prevalence for women is higher in the north-eastern part of the country from Limpopo to Eastern Cape province, whereas HIV prevalence for men is more concentrated in the mid-eastern part of the country among the Gauteng and KwaZulu-Natal provinces. The density of PLHIV for both genders shows similar spatial patterns, concentrating in Gauteng province. Maps were created using ArcGIS by ESRI V.10.5 (http://www.esri.com).

and men living with HIV in these areas. We used three travel times, that is, 30, 60 and 120 min, as thresholds for underserved areas in this study based on previous studies. Various travel times have been suggested as a threshold to access healthcare in South Africa and other countries in SSA.[15 29–31] A study of PLHIV in Uganda reports 30 min as a median travel time for PLHIV to access healthcare facility.[29] Another study for travel time to access primary healthcare in South Africa shows 81 min as a median travel time, and 65% of households travel more than 60 min to attend the primary healthcare.[15] In addition, other studies in African countries use 120 min as a threshold to access healthcare facilities.[30 31] Finally, we generated density maps of women living with HIV and men living with HIV in the underserved areas using the continuous surface maps of HIV generated in the previous step and the estimations for underserved areas. Further details related to this methodology, study design and data can be found in online supplemental materials.

Patient and public involvement

There were no patients involved in this research.

RESULTS

Spatial variable selection

The variables selected for inclusion in the final logistic regression models for men and women are summarised in table 1, and the results from bivariate logistic regression analysis and the corresponding spatial correlation analysis (Moran's Index) for each variable are presented in online supplemental table S2. For men, the final logistic model included ever been tested for HIV, poverty, lifetime number of sexual partners, MC, GH-I, distance to main roads and NDVI. For women, the final logistic model included condom use, lifetime number of sexual partners, friction, distance to main roads and NDVI. Continuous surface maps of each variable selected for men (online supplemental figure S1) and women (online supplemental figure S2) are included in online supplemental materials.

HIV disease mapping

Figure 2A,B illustrates the spatial distribution of HIV prevalence estimated using the final multivariable logistic regression models for men and women, respectively. These maps show the distribution of the proportion of individuals who are HIV positive among the entire adult population (men aged 15–59 and women aged 15–49). Pixel-level HIV prevalence for men ranged from 1.4% to 24.2% per 5 km^2, with a median of 11.5% per 5 km^2, and from 2.1% to 48.1% per 5 km^2, with a median of 20.6% per 5 km^2 for women. The spatial structure of HIV prevalence showed some differences between the distribution of the disease for men and women. HIV prevalence for women was higher in the north-eastern part of the country, particularly in *Limpopo* province, and in the south-eastern areas along with the coastline in *KwaZulu-Natal* and *Eastern Cape* provinces. Conversely, HIV prevalence for men was more concentrated in the mid-eastern part of the country among the *Gauteng* and *KwaZulu-Natal* provinces.

Figure 2C,D illustrates the geographical distribution of the estimated number of men living with HIV and women living with HIV, respectively. Using the spatial maps, we estimated a total of 3 367 810 men living with HIV aged 15–59 and 3 861 228 women living with HIV aged 15–49 in 2016 in South Africa. The highest density of PLHIV for both genders shows in *Gauteng*, and the lowest density

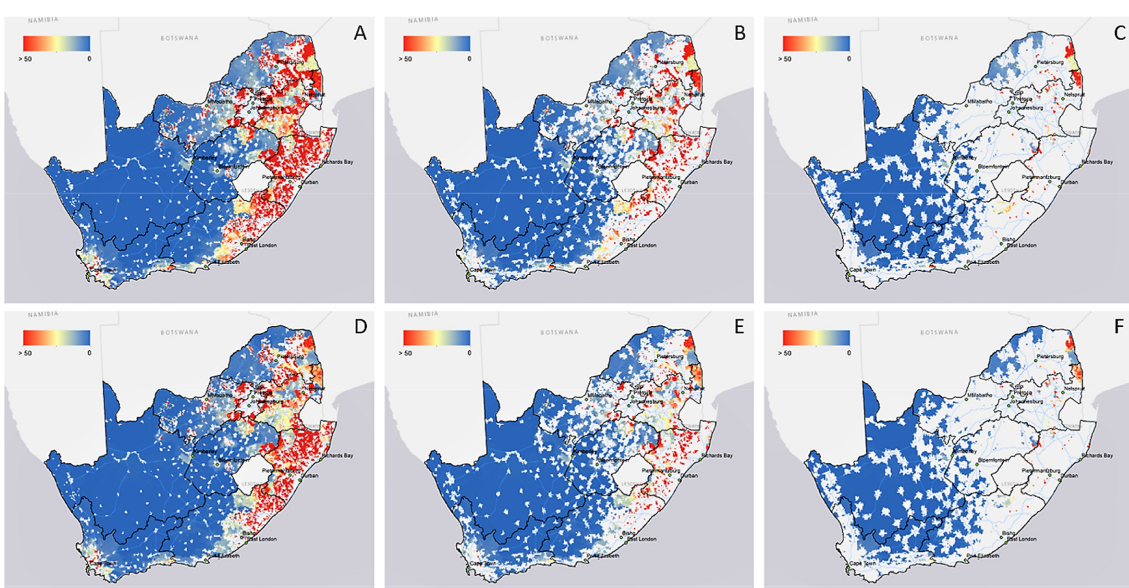


Figure 3 People living with HIV (PLHIV) within underserved areas (30, 60 and 120 min thresholds) in South Africa. Estimated men living with HIV within 30 min threshold (A), 60 min threshold (B), and 120 min threshold (C) underserved areas; estimated women living with HIV within 30 min threshold (D), 60 min threshold (E), and 120 min threshold (F) underserved areas. Maps were created using ArcGIS by ESRI V.10.5 (http://www.esri.com).

shows in *Northern Cape*. The density of PLHIV ranged from 0 per 5 km^2 in *Northern Cape* for both genders to a maximum of 13 862 per 5 km^2 for women and 15 484 per 5 km^2 for men in *Gauteng*. Specifically, men living with HIV concentrate in the urban areas, whereas women living with HIV were distributed both in urban and the periphery areas, covering the north-eastern and south-eastern parts of the country.

PLHIV within underserved areas

Figure 3A,D illustrates the population density of men living with HIV and women living with HIV within the underserved areas with 30 min threshold, respectively. The underserved areas based on 30 min travel time from the nearest healthcare facility across the country covered about 80% of the total territory (≈991 000 km^2). About 16.6% (≈559 581) of men living with HIV aged 15–59 lived outside of a 30 min minimum travel time from the nearest hospital, whereas 20.4% (≈787 022) of women living with HIV aged 15–49 lived in those underserved areas, with a similar spatial pattern of PLHIV within the underserved areas for both genders. Most part of *Gauteng* province and coastal areas in *KwaZulu-Natal* province including the city of Durban were areas where healthcare services can be reached within 30 min. Likewise, figure 3B,E shows the population density of men living with HIV and women living with HIV within the underserved areas with 60 min threshold, respectively. These underserved areas covered 62.1% of the total territory of the country, and 6.4% (≈216 419) of men living with HIV were located outside of the 60 min travel time to the nearest hospital threshold, whereas 7.9% (≈305 074) of women living with HIV lived in those underserved areas. Lastly, 37.1% of the total area of the country was located outside of a 120 min threshold from the nearest healthcare facility (figure 3C,F), and 1.6% (≈53 737) of men living with HIV lived outside of a 120 min travel time to the nearest hospital, whereas 1.9% (≈73 144) of women living with HIV were located in these underserved areas.

Table 2 summarises the estimated number and percentage of PLHIV within the underserved areas by province. Around 80% of PLHIV were living in five provinces, including *Gauteng, KwaZulu-Natal, Eastern Cape, Limpopo* and *Mpumalanga*. Moreover, 50% of the PLHIV were located in *Gauteng* and *KwaZulu-Natal* provinces (46.66% for women living with HIV and 51.27% for men living with HIV). Although *North West* province contained a lower percentage of PLHIV (7.65% for women living with HIV and 7.58% for men living with HIV of the total PLHIV in the country), this province had the highest proportion of PLHIV within underserved areas with

Table 2 The estimated number of PLHIV living in the underserved areas, and the number of health facilities per 1000 PLHIV for both men and women by province level

Province	PLHIV (% total)	PLHIV 30 min (% total)	PLHIV 60 min (% total)	PLHIV 120 min (% total)	Healthcare facilities per 1000 PLHIV
Women					
Eastern Cape	495 547 (12.83)	133 628 (16.98)	39 720 (13.02)	8285 (11.33)	1.83
Free state	189 708 (4.91)	45 899 (5.83)	26 297 (8.62)	4040 (5.52)	1.48
Gauteng	1 004 439 (26.01)	42 783 (5.44)	8479 (2.78)	0 (0.00)	0.45
KwaZulu-Natal	797 416 (20.65)	174 543 (22.18)	60 188 (19.73)	7593 (10.38)	0.93
Limpopo	431 823 (11.18)	133 555 (16.97)	58 690 (19.24)	24 100 (32.95)	1.24
Mpumalanga	376 178 (9.74)	126 764 (16.11)	57 593 (18.88)	12 456 (17.03)	0.94
North West	295 370 (7.65)	75 913 (9.65)	31 200 (10.23)	4175 (5.71)	1.21
Northern Cape	50 903 (1.32)	12 502 (1.59)	10 471 (3.43)	7097 (9.70)	4.58
Western Cape	219 776 (5.69)	14 429 (1.83)	12 430 (4.07)	5393 (7.37)	2.01
Men					
Eastern Cape	328 618 (9.76)	77 660 (13.89)	24 537 (11.34)	5872 (10.93)	2.76
Free state	181 787 (5.40)	34 339 (6.14)	19 569 (9.04)	3130 (5.83)	1.54
Gauteng	972 130 (28.87)	28 689 (5.13)	5499 (2.54)	0 (0.00)	0.47
KwaZulu-Natal	754 290 (22.40)	149 197 (26.68)	49 547 (22.90)	5948 (11.07)	0.99
Limpopo	250 860 (7.45)	69 400 (12.41)	29 579 (13.67)	12 125 (22.57)	2.13
Mpumalanga	290 587 (8.63)	89 993 (16.09)	41 027 (18.96)	9659 (17.98)	1.21
North West	255 133 (7.58)	56 360 (10.08)	22 196 (10.26)	2811 (5.23)	1.40
Northern Cape	65 048 (1.93)	14 327 (2.56)	12 182 (5.63)	8604 (16.01)	3.58
Western Cape	268 245 (7.97)	39 181 (7.01)	12 266 (5.67)	5578 (10.38)	1.65

PLHIV, people living with HIV.

30 and 60 min thresholds. Moreover, *Gauteng* province contained the largest proportion of PLHIV of the country for both genders, 26.01% for women living with HIV and 28.87% for men living with HIV. This province showed low number of PLHIV located within the underserved areas, with less than 5% of PLHIV in the underserved areas with 120 min threshold, and also had the lowest numbers of healthcare facilities per 1000 PLHIV among all provinces (0.45 for women living with HIV and 0.47 for men living with HIV). Conversely, *Northern Cape* and *Western Cape* provinces showed the lowest percentage of PLHIV of the entire country (1.32% and 5.69% for women living with HIV, and 1.93% and 7.97% for men living with HIV, respectively) and the highest numbers of healthcare facilities per 1000 PLHIV (4.58 and 2.01 for women living with HIV, and 3.58 and 1.65 for men living with HIV, respectively). However, these provinces contained a large number of PLHIV within underserved areas with 120 min threshold (9.70% and 7.37% for women living with HIV and 16.01% and 10.38% for men living with HIV).

DISCUSSION

HIV prevalence for both men and women showed remarkable geographical variation in South Africa, with the burden of infection concentrated in the eastern part of the country, among the provinces of *Gauteng, KwaZulu-Natal, Eastern Cape, Limpopo* and *Mpumalanga,* and with high concentration of PLHIV identified near urban settings of the eastern part of the country, including Johannesburg and Durban. For example, *Gauteng* shows the highest percentage of the average HIV prevalence for women, with 30.6% per 5 km^2 and ranges from 18% to a maximum of 42% per 5 km^2. For men, the average HIV prevalence in *Gauteng* shows 14.7% per 5 km^2 and ranges from 8% to a maximum of 24% per 5 km^2. We identified gender differences in the distribution of the HIV epidemic, with the HIV prevalence for women more concentrated in the northern-eastern and southern-eastern parts of the country compared with the concentration of men in the eastern part of South Africa. Moreover, PLHIV who are unlikely to access healthcare services within at least a 60 min time threshold are located in eastern part of the country among the *Limpopo* and *KwaZulu-Natal* provinces, and a large number of PLHIV near *Kruger National Park* across *Limpopo* and *Mpumalanga,* in the vicinity villages of Phalaborwa and Shangaan, near Hazyview town, are unlikely to access health services within a 120 min travel time.

Our results suggest that variables such as MC and poverty are key drivers of the spatial variation of HIV prevalence in South Africa. MC showed a significant association with the spatial structure of HIV for men. With the effectiveness of MC to reduce HIV infections among men,[32] the South Africa government implemented the 2012–2016 National Strategic Plan (NSP) to circumcise 80% of HIV-negative men aged 15–49 by 2015 (SANAC, 2011).[33] As a result, South Africa achieved a significant increase of MC from 25% in 2013 to 57% in 2016, but with marked spatial variation among provinces, with 35% MC prevalence in *Northern Cape* compared with 86% in *Limpopo*.[21] The government launched the NSP 2017–2022 with a new target of 3 million circumcised men by 2021.[34] Besides the direct impact of MC on the reduction of HIV acquisition, MC could also generate secondary benefits by affecting the sexual behaviour of circumcised men.[35] For example, circumcised men tend to be concerned more about partner's HIV status and to use condoms more frequently compared with non-circumcised men in South Africa.[35] Likewise, poverty was also a significant explanatory variable associated with the spatial structure of HIV for both men and women. This result is consistent with a previous study showing that poorer women are more likely to have burden of HIV infection and have engaged in risky sexual behaviour especially for young women, such as earlier sexual debut, having had multiple sexual partners in the year, lower chance of condom use and having traded sex, compared with women in more affluent households.[36 37] For both genders, poverty reduces the chance of discussing safe-sex practices with their partner.[36]

According to the UNAIDS Global Report in 2019, South Africa is one of the countries showing the great decline in new HIV infections with more than 50% of reduction in AIDS-related deaths between 2010 and 2018.[1] The South Africa government launched a series of strategic plans as a response to HIV and updated the NSP four times (2000–2005, 2007–2011, 2012–2016, 2017–2022) to expand the coverage of prevention and treatment guidelines towards the ultimate goal of the elimination of HIV by 2030. South Africa has also increased domestic funds for HIV by about $650 million between 2010 and 2018, and now 78% of the total HIV funding resources in the country are domestic.[1] Despite these efforts, South Africa has a 19.1% HIV prevalence among the adult population aged 15–49 years.[1] One of the approaches included in the NSP 2017–2022 is the identification of priority areas and key vulnerable populations as a strategy for optimal resource allocation, and to build an environment that maximises coverage of HIV prevention and treatment services. As discussed previously, since spatial distance from residence to healthcare facilities for PLHIV is a major barrier to receiving adequate healthcare, allocating resources and controlling HIV prevention and treatment services are critical strategies to overcome the spatial barrier. Here, we localised HIV-underserved areas that could hinder South Africa in meeting UNAIDS targets and identified areas in need of appropriate HIV prevention and treatment services. Our analysis identified the areas where the travel time takes more than 30, 60 and 120 min to the nearest healthcare facility and indicated that more than 550 000 men living with HIV (17% among the total men living with HIV) and 750 000 women living with HIV (20% among the total women living with HIV) are living in areas where travel time to the nearest facility took at least 30 min.

The highest density of PLHIV in South Africa was concentrated in *Gauteng* province (around 1 million (26%) persons) with a maximum of 13 862 women per 5 km^2 and 15 484 men per 5 km^2. Since the borders of this province covered a small area, the numbers of PLHIV living in underserved areas rapidly decreased to zero. However, our results suggest that this province has the lowest number of healthcare facilities per 1000 PLHIV (0.45 for women and 0.47 for men), and although this province is characterised by easy-to-access health services with a short travel time, the absolute number of healthcare facilities may not satisfy the demand generated by the large numbers of PLHIV who need HIV treatment. Moreover, according to a recent analysis, *Gauteng* showed the slowest progress towards the 90-90-90 targets between 2003 and 2015 due to budget allocation.[38] *Gauteng* and *Mpumalanga* had the lowest amount of HIV expenditure per HIV-positive individuals whereas *Northern Cape* and *Free State* had the highest amount of HIV expenditure.[38]

KwaZulu-Natal province had the second largest numbers of PLHIV for men and women (more than 700 000 PLHIV; 21%) with more than 10% of these PLHIV living in underserved areas. Similar to *Gauteng,* the province had lower numbers of healthcare facilities per 1000 PLHIV (0.99 for men and 0.93 for women). With the NSP 2017–2022 for HIV, the province implemented the Provincial Growth and Development Strategy 2035 along with the 90-90-90 strategy and set a new goal to eliminate HIV infections by 2035.[39] As a result, the province achieved a significant decline of HIV infection by 27% between 2012 and 2016.[39] The province classified districts into three categories (ie, high triple burden of HIV, tuberculosis (TB) and sexually transmitted infections (STI); dual high burden of HIV; high-burden districts of STI) to efficiently treat districts with limited resources and expenditure based on importance towards eliminating of HIV, TB and STI.[39] According to the *KwaZulu-Natal* Provincial Multi-Sectoral HIV, TB and STI Implementation Plan 2017–2022, acceleration, intensification, inclusivity, participation and efficiency in the provision of services will be hallmark of the interventions and activities of this plan, given the dire situation of the epidemics in *KwaZulu-Natal* and for the fact that it is considered the epicentre of the epidemic in the country. Towards this end, the province adopted the ‘Focus for Impact’ methodology to assist localities with the greatest burden to targeting these settings with interventions and activities. The essence of the methodology is to deliver results that can provide the greatest impact with limited resources and highest value for money. Results have shown that five of the province’s 11 districts, namely (1) *eThekwini*; (2) *iLembe*; (3) *uThukela*; (4) *uMgungundlovu*; and (5) *Ugu*, have a high triple burden of HIV, TB and STIs. Four of the 11 districts have a dual high burden as follows: (1) *King Cetshwayo* and (2) *Zululand* (high burden of HIV and TB), (3) *Harry Gwala* (high HIV and STI burden), and (4) *uMkhanyakude* (high TB and STI burden). *Amajuba,* on the other hand, has a high STI burden.[39] As the priority group, the high triple burden of HIV districts, including *eThekwini, iLembe, uThukela, uMgungundlovu* and *Ugu*, is located in the southern part of the province near the city of Durban.[39] The results of our analysis show similar spatial patterns with this policy in terms of PLHIV. However, we found that the areas with higher needs of HIV prevention and treatment services are located in the northern part of the province, which is classified as a dual high-burden district since the districts contain a large number of PLHIV who are not likely to access the nearest health facility within a certain travel time.

Lastly, areas near Kruger National Park in *Limpopo* and *Mpumalanga* provinces had the largest number of PLHIV who needed at least 120 min to access the nearest healthcare facility. *Limpopo* and *Mpumalanga* provinces implemented the provincial strategic plan 2012–2016 with goals based on the NSP 2017–2022.[40,41] However, since the two provinces are exposed to truck routes across borders with Zimbabwe, Mozambique and Zambia, these areas could be exposed to high movement of population that might increase the risk of HIV, and thus the intensification programmes for accessible HIV prevention and treatment for PLHIV within the areas might need to be implemented.[40,41]

We identified spatial variations of HIV prevalence and PLHIV at a high-resolution level for both genders and discovered underserved areas that healthcare services cannot reach at the high-resolution level. However, several study limitations could have affected our results. First, some of the variables included in the study could have been affected by inherent biases in the data due to the multiple logistical difficulties in conducting SADHS, such as variability in response rates to HIV testing. SADHS was conducted based on individuals who were interviewed and participated in HIV testing. The sample weighting procedures were applied to correct for differential non-responses, considering provincial characteristics, residential place and gender. However, the generalisability of the prevalence estimates can be unclear.[42,43] For example, some high-risk subpopulations, such as women sex workers, injection drug users, men who have sex with men and mobile individuals, could have been missed.[12] Since South Africa shares borders with other countries including Mozambique, Zambia, Namibia, Eswatini and Lesotho, migration and mobile populations could affect our findings. The methodology used to calculate HIV prevalence and PLHIV is not able to capture temporal changes due to migration, and the sources of data did not include information regarding mobile individuals which could affect estimations near the country borders.[12] Moreover, MC prevalence was estimated from self-reported status of circumcision, and thus does not discriminate among different types of MC like traditional or medical circumcision. However, since our analysis focused on the general relationship between circumcised men and the spatial structure of HIV prevalence, we expect that this generalisation does not affect the main findings of our study.[44 45] Lastly, the displacement process of the SADHS sampling data point via global positioning system might cause potential spatial bias.[12] This process could have an impact on the precision of the continuous surface of HIV prevalence and PLHIV by a few kilometres.

CONCLUSION

Although South Africa achieved the first of the UNAIDS 90-90-90 targets, a large number of PLHIV are still in need for treatment, and precise geographic targeting of HIV prevention and treatment services might facilitate the country to meet the second of the UNAIDS 90-90-90 targets. Using spatial epidemiological approaches to investigate HIV prevalence and underserved areas from healthcare facilities, we identified more than 1 million of PLHIV in areas with the lack of access to healthcare within 30 min travel time. These findings can shed light on areas that have been aggregated in the province or national levels to represent HIV prevalence to support a cost-effective decision process for decision makers by geographically targeting certain areas. Therefore, special attention should be given to the development and implementation of tailored HIV prevention and treatment programme, especially in areas identified as underserved areas for PLHIV. The statistical modelling and high-resolution maps of HIV prevalence and PLHIV provide valuable information to support decision-making process to achieve the national and international goals for the elimination of HIV by increasing the effective HIV prevention and treatment services in those areas with the greatest need.

Author affiliations

[1]Department of Geography and Geographic Information Science, University of Cincinnati, Cincinnati, Ohio, USA
[2]Health Geography and Disease Modeling Laboratory, University of Cincinnati, Cincinnati, Ohio, USA
[3]Lincoln International Institute for Rural Health, University of Lincoln, Lincoln, UK
[4]Africa Health Research Institute, KwaZulu-Natal, South Africa
[5]School of Nursing and Public Health, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa
[6]Department of Population Health, London School of Hygiene and Tropical Medicine, London, UK
[7]Centre for Rural Health, School of Nursing and Public Health, University of KwaZulu-Natal, Durban, South Africa
[8]KwaZulu-Natal Research Innovation and Sequencing Platform, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa
[9]Heidelberg Institute of Global Health (HIGH), Medical Faculty, Heidelberg University, Heidelberg, Germany

Acknowledgements The authors thank Measure Demographic and Health Surveys (Measure DHS) for releasing these national surveys in the service of science, and the US Agency for International Development and other donors supporting these initiatives.

Contributors Conceptualisation: HK, FT, AT, AV, DFC. Methodology: HK, DFC. Software: HK. Validation: HK, FT, AT, AV, DFC. Data curation: HK. Writing—original draft preparation: HK, DFC. Writing—review and editing: FT, AT, AV.

Funding This work was supported by two National Institutes of Health (NIH) grants (R01HD084233 and R01AI124389). Andrew Tomita was funded by the UKRI GCRF (MR/T029803/1).

Disclaimer The funder had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Map disclaimer The depiction of boundaries on this map does not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or of its authorities. This map is provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Procedures and questionnaires for standard Demographic and Health Surveys have been reviewed and approved by the ICF International Institutional Review Board (IRB). The ICF International IRB ensures that the survey complies with the US Department of Health and Human Services regulations for the protection of human subjects, while the host country IRB ensures that the survey complies with laws and norms of the nation (http://dhsprogram.com/What-We-Do/Protecting-the-Privacy-of-DHS-Survey-Respondents.cfm%23sthash.0t3N7n5m.dpuf).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open-access repository. The data that support the findings of this study are available from the Demographic and Health Surveys (http://www.measuredhs.com), but restrictions apply to the availability of these data, which were used under licence for the current study and so are not publicly available. However, data are available from the authors on reasonable request and with the permission of Demographic and Health Surveys. We sought and were granted permission to use the core data set for this analysis by Measure DHS.

ORCID iD

Diego F Cuadros http://orcid.org/0000-0001-7060-4203

REFERENCES

1 UNAIDS. *Global HIV & AIDS statistics: 2019 fact sheet*, 2019.
2 UNAIDS. *Understanding fast-track: accelerating action to end the AIDS epidemic by 2030*, 2015.
3 Johnson LF, May MT, Dorrington RE, *et al*. Estimating the impact of antiretroviral treatment on adult mortality trends in South Africa: a mathematical modelling study. *PLoS Med* 2017;14:e1002468.
4 (CESAR) CfSAaR. *2018 global AIDS monitoring report: analysis of current status and progress towards targets*, 2018.
5 UNAIDS. AIDSinfo, 2019. Available: http://aidsinfo.unaids.org [Accessed 15 Jan 2021].

6 Kleinschmidt I, Pettifor A, Morris N, *et al*. Geographic distribution of human immunodeficiency virus in South Africa. *Am J Trop Med Hyg* 2007;77:1163–9.
7 Cuadros DF, Sartorius B, Hall C, *et al*. Capturing the spatial variability of HIV epidemics in South Africa and Tanzania using routine healthcare facility data. *Int J Health Geogr* 2018;17:27.
8 Palk L, Blower S. Geographic variation in sexual behavior can explain geospatial heterogeneity in the severity of the HIV epidemic in Malawi. *BMC Med* 2018;16:22.
9 Tanser F, Bärnighausen T, Dobra A, *et al*. Identifying 'corridors of HIV transmission' in a severely affected rural South African population: a case for a shift toward targeted prevention strategies. *Int J Epidemiol* 2018;47:537–49.
10 Tomita A, Vandormael A, Bärnighausen T, *et al*. Sociobehavioral and community predictors of unsuppressed HIV viral load: multilevel results from a hyperendemic rural South African population. *AIDS* 2019;33:559.
11 Gutreuter S, Igumbor E, Wabiri N, *et al*. Improving estimates of district HIV prevalence and burden in South Africa using small area estimation techniques. *PLoS One* 2019;14:e0212445.
12 Cuadros DF, Li J, Mukandavire Z, *et al*. Towards UNAIDS fast-track goals: targeting priority geographic areas for HIV prevention and care in Zimbabwe. *AIDS* 2019;33:305–14.
13 Anderson S-J, Cherutich P, Kilonzo N, *et al*. Maximising the effect of combination HIV prevention through prioritisation of the people and places in greatest need: a modelling study. *Lancet* 2014;384:249–56.
14 Grantham KL, Kerr CC, Wilson DP. Local responses to local epidemics for national impact need advanced spatially explicit tools. *AIDS* 2016;30:1481–2.
15 Tanser F, Gijsbertsen B, Herbst K. Modelling and understanding primary health care accessibility and utilization in rural South Africa: an exploration using a geographical information system. *Soc Sci Med* 2006;63:691–705.
16 Vandormael A, Cuadros D, Kim H-Y, *et al*. The state of the HIV epidemic in rural KwaZulu-Natal, South Africa: a novel application of disease metrics to assess trajectories and highlight areas for intervention. *Int J Epidemiol* 2020;49:666–75.
17 Gerberry DJ, Wagner BG, Garcia-Lerma JG, *et al*. Using geospatial modelling to optimize the rollout of antiretroviral-based pre-exposure HIV interventions in sub-Saharan Africa. *Nat Commun* 2014;5:5454.
18 Barankanira E, Molinari N, Niyongabo T, *et al*. Spatial analysis of HIV infection and associated individual characteristics in Burundi: indications for effective prevention. *BMC Public Health* 2016;16:118.
19 Meyer-Rath G, McGillen JB, Cuadros DF, *et al*. Targeting the right interventions to the right people and places: the role of geospatial analysis in HIV program planning. *AIDS* 2018;32:957–63.
20 PEPFAR SA. *South Africa country operational plan (COP) 2018 strategic direction summary*, 2018.
21 Council SAMR, Health NDo, Africa SS. *South Africa demographic and health survey 2016: NDoH, STATs SA, SAMRC, and ICF Pretoria, South Africa and Rockville*, 2019.
22 ESRI. *ArcGIS 10.x program*. Redlands, California, 2004.
23 Mishra V, Vaessen M, Boerma JT, *et al*. Hiv testing in national population-based surveys: experience from the demographic and health surveys. *Bull World Health Organ* 2006;84:537–45.
24 von Elm E, Altman DG, Egger M, *et al*. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Epidemiology* 2007;18:800–4.
25 Cuadros DF, Li J, Branscum AJ, *et al*. Mapping the spatial variability of HIV infection in sub-Saharan Africa: effective information for localized HIV prevention and control. *Sci Rep* 2017;7:9093.
26 Maina J, Ouma PO, Macharia PM, *et al*. A spatial database of health facilities managed by the public health sector in sub Saharan Africa. *Sci Data* 2019;6:134.
27 Delamater PL, Messina JP, Shortridge AM, *et al*. Measuring geographic access to health care: raster and network-based methods. *Int J Health Geogr* 2012;11:15.
28 Martin D, Wrigley H, Barnett S, *et al*. Increasing the sophistication of access measurement in a rural healthcare study. *Health Place* 2002;8:3–13.
29 Akullian AN, Mukose A, Levine GA, *et al*. People living with HIV travel farther to access healthcare: a population-based geographic analysis from rural Uganda. *J Int AIDS Soc* 2016;19:20171.
30 Ouma PO, Maina J, Thuranira PN, *et al*. Access to emergency hospital care provided by the public sector in sub-Saharan Africa in 2015: a geocoded inventory and spatial analysis. *Lancet Glob Health* 2018;6:e342–50.
31 Hulland EN, Wiens KE, Shirude S, *et al*. Travel time to health facilities in areas of outbreak potential: maps for guiding local preparedness and response. *BMC Med* 2019;17:1–16.
32 Auvert B, Taljaard D, Lagarde E, *et al*. Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 trial. *PLoS Med* 2005;2:e298.
33 South African National AIDS Council (SANAC). Statement on the meeting of the South African National Aids Council 12 august, 2011. Available: http://www.sanac.org.za/files/uploaded/886_Plenary%20Press%20Statement%2012Aug11.pdf
34 Council SANA. *Let Our Actions Count: South Africa's National Strategic Plan for HIV*, 2017.
35 Mukudu H, Dietrich J, Otwombe K, *et al*. Voluntary medical male circumcision (VMMC) for prevention of heterosexual transmission of HIV and risk compensation in adult males in Soweto: findings from a programmatic setting. *PLoS One* 2019;14:e0213571.
36 Kelly C, Hulme C, Farragher T, *et al*. Are differences in travel time or distance to healthcare for adults in global North countries associated with an impact on health outcomes? A systematic review. *BMJ Open* 2016;6:e013059.
37 Bunyasi EW, Coetzee DJ. Relationship between socioeconomic status and HIV infection: findings from a survey in the free state and Western Cape provinces of South Africa. *BMJ Open* 2017;7:e016232.
38 Johnson LF, Dorrington RE, Moolla H. Progress towards the 2020 targets for HIV diagnosis and antiretroviral treatment in South Africa. *South Afr J HIV Med* 2017;18:694–8.
39 Directorate K-NOotPHA. *The KwaZulu-Natal provincial Multi-Sectoral HIV, TB and STIs implementation plan 2017-2022*, 2017.
40 council LPA. *Annual progress report 2015/16: provincial strategic plan 2012-2016*, 2017.
41 Council MPA. *Mpumalanga provincial implementation plan for HIV, TB, and STIs 2017-2022*, 2017.
42 WHO. *South Africa: HIV country profile, 2016*, 2017.
43 National Department of H, Icf. *South Africa demographic and health survey 2016*. Pretoria: National Department of Health - NDoH - ICF, 2019.
44 Kim H, Branscum A, Miller FD, *et al*. Geospatial assessment of the voluntary medical male circumcision programme in Tanzania, 2011–2016. *BMJ Glob Health* 2019;4:e001922.
45 Cuadros DF, Branscum AJ, Miller FD, *et al*. Are Geographical "Cold Spots" of Male Circumcision Driving Differential HIV Dynamics in Tanzania? *Front Public Health* 2015;3:218.

Supplementary Materials for

Beyond HIV prevalence: Identifying people living with HIV within underserved areas in South Africa

Hana Kim, Frank Tanser, Andrew Tomita, Alain Vandormael, Diego F. Cuadros

Reprints or correspondence: Diego F. Cuadros, PhD, E-mail: diego.cuadros@uc.edu.

Supplementary methods

Definition of the total population for the study

We selected population based on the HIV testing. The total population of our analysis are 4,862 individuals (2,136 men and 2,726 women) among 12,132 participants of 2016 South Africa DHS (SADHS; 3,618 men and 8,514 women) who had the HIV testing results either *yes* or *no* and had their sample location. The 2016 SADHS report mentions that their total population of respondents are 12,132 (3,618 men and 8,514 women), and 58% of women and 44% of men were interviewed and provided specimens for HIV testing.[1]

In terms of SADHS sample location, the "SADHS locations" means the location of each primary sample units (PSU) in this paper. In the most DHS surveys, the groupings of households that participated in the survey are geo-referenced, and the locations of their participating are called sample locations or clusters. DHS survey used a two-stage sampling frame to select households and reduce sampling errors. A sampling frame is a complete list of all sampling unites that covers entire target population. The PSU is the sampling unit for the first stage of selection, and typically corresponded to census enumeration areas (EAs). 2016 SADHS used the Statistics South Africa Master Sample Frame (MSF), which was created using Census 2011 EAs. The frame contains information about the geographic type (urban, rural, or farm) and the estimated number of residential dwelling units (DUs) in each PSU. In rural areas, a PSU is a natural village, or a segment of a large village, or a group of small villages whereas a PSU is a street or a city block in urban areas. The 2016 SADHS followed a stratified two-stage sample design with a probability proportional to size sampling of PSUs at the first stage and systematic sampling of DUs at the second stage. The Census 2011 DU count was used as the PSU measure of size. In the first stage, a total of 750 PSUs were selected from the 26 sampling strata, yielding 468 selected PSUs in urban areas, 224 PSUs in rural areas, and 58 PSUs in farm areas. As most DHS surveys use a fixed take of household per PSU about 25-30 households, a fixed number of 20 DUs per PSU were selected with systematic selection in the second stage of selection for 2016

Kim H, *et al. BMJ Global Health* 2021; 6:e004089. doi: 10.1136/bmjgh-2020-004089

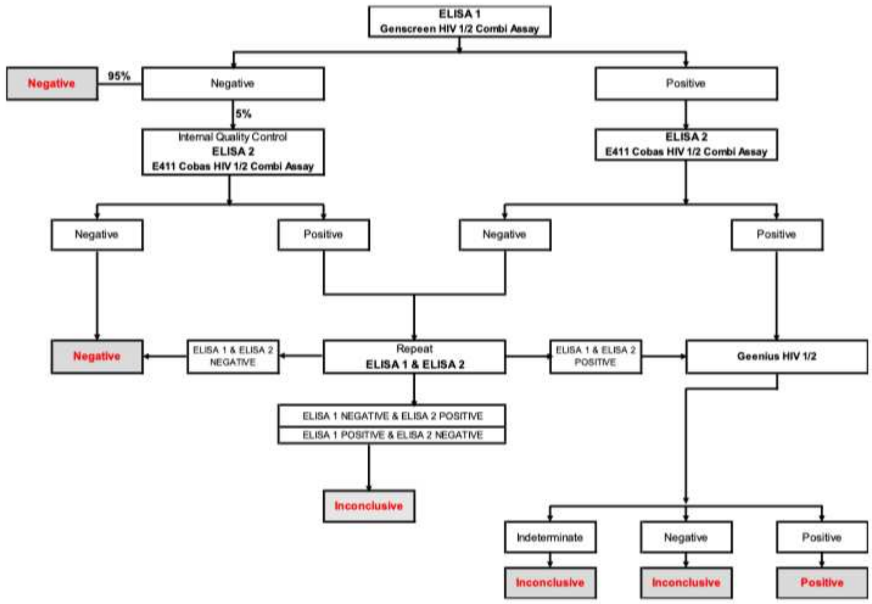
SADHS. As such, the sample locations represent 80% and 84% for men and women, respectively, not 25%. Additionally, SADHS includes sampling weights at household and individual levels in order to restore representativeness of the sample and to prevent bias due to over- and under-sampling depend on region size.[1] Thus, among the total population for 2016 SADHS, 2,136 men and 2,726 women who interviewed HIV testing at the sample locations (601 for men and 633 for women) are the population of this study.

In terms of the HIV testing, the 2016 SADHS incorporates the HIV testing biomarkers.[1] For each biomarker test for which an individual was eligible, the respondent was required to provide written consent before the test proceed. The SADHS provides brochure including information of HIV and the national AIDS hotline number and nearby facilities for HIV testing and counselling.

The HIV testing algorithm (figure1) called for testing all samples with an enzyme-linked immunosorbent assay (ELISA). All samples that tested positive on the ELISA 1 were subjected to a second ELISA (ELISA 2). Similar to samples that tested positive on the ELISA 1, 5% of the samples that tested negative on the ELISA 1 were also subjected to the ELISA 2 for internal quality control, while the other 95% were recorded as negative. All the concordant negative results on the ELISA 1 and ELISA 2 were recorded as negative. If the results on the ELISA 1 and ELISA 2 were mismatched, the two ELISAs were repeated. If the results remained discordant, the specimen was classified as inconclusive. All the concordant positive results on the ELISA 1 and ELISA 2 were subjected to a third assay. When both the ELISA 1 and ELISA 2 were positive, the sample was classified as positive if the confirmatory rapid test was positive, and inconclusive if the confirmatory rapid test was negative or indeterminate. For the detailed description of HIV testing and ELISA test, please see the 2016 SADHS reports.[1]

Kim H, *et al. BMJ Global Health* 2021; 6:e004089. doi: 10.1136/bmjgh-2020-004089

Diagram S1. HIV testing algorithm



Definition of socio-economic and behavioral variables

To access the association between socioeconomic and behavioral factors, and the prevalence of HIV infection, we selected a variety of socio-economic and behavioral attributes from the South Africa Demographic and Health Survey (SADHS) dataset. Six variables measured for men and five for women included: education level, poverty, condom use, number of lifetime sexual partners, if the participants had ever been tested for HIV, and male circumcision (men only).

Socioeconomic and behavioral variables were quantified as percentages at each SADHS sample location for men and women separately. Specifically, level of education is individual's highest educational level. SADHS standardized the variable into *No education, Primary, Secondary, and Higher.*[2] We classified individuals with secondary or higher education as a higher education and calculated the percent of individuals at each DHS sample location. We used DHS wealth index to identify poverty variable. DHS wealth index is an ordinal variable that describes standard of living as determined by material possessions. The index is calculated based on the questions about wealth status, such as an ownership of a television and car, housing construction materials, type of drinking water source, and access to sanitation facilities.[2 3] This index were used to classify individuals in five categories: *poorest, poorer, middle, richer, and richest.* We classified *poorest* and *poorer* as a poverty and calculated the percent of those individuals at each DHS sample location.

Kim H, *et al. BMJ Global Health* 2021; 6:e004089. doi: 10.1136/bmjgh-2020-004089

Behavioral variables, such as condom use, ever been tested for HIV, and male circumcision, are from the binary questions. The three questions for the behavioral variables are: 1) Did you use condom during last sex with most recent partner?, 2) Have you ever been tested for HIV?, 3) Are you circumcised? (men only). The binary, *yes or no*, answer to the questions was used as the measure of individuals, and we calculated the percent of individuals who answered *yes* to the question. The number of lifetime sexual partners is continuous variable in DHS, and we calculated the percent of individuals whose lifetime sexual partners are less than three at each DHS sample location.

Definition of environmental variables

Environmental variables included the normalized difference vegetation index (NDVI), distance to main roads, distance to the nearest dams, night light, global human influence index (GH-I), friction map, and population density. NDVI serves as a proxy for degree of urbanization at a sampled location, and it has been shown to be associated with the potential risk of coinfection with diseases transmitted by parasites, such as malaria, and could lead to increased HIV transmission efficiency in SSA.[4 5] Night light represents dynamics of urban settlements and related economic activities since urban areas are expected to have higher night light intensity than rural areas. We collected NDVI and night light datasets from the National Aeronautics and Space Administration (NASA) Earth Observatory Group.[6] Those datasets were used to extract the corresponding variables at each DHS sample location. Distance to main roads was derived from a road networks dataset (in vector format) from the World Bank.[7] Distance to dams was derived from a dam dataset from NASA. The dams were geospatially referenced and assigned to polygons depicting reservoir outlines with multiple attributes[8] A raster layer of distance to main roads and dams were calculated as Euclidean distance for each pixel to the nearest main road and dams, respectively. We further resampled these raster layers to 5X5km spatial resolution with the same projection (WGS_1984_UTM_Zone_35S). GH-I provides an updated information of anthropogenic impacts on the environment based on global data of human population pressure, human land use and infrastructure (built-up areas, land use), and human access (coastlines, railroads, and roads).[9] Population density was obtained from the WorldPop database.[10] Population count data with detailed satellite-derived settlement extents was resampled to 100m X 100m resolution is used to generate population distribution. Estimates of population that we used in the analysis were summarized in raster image with a resolution of 100m x 100m pixel. For consistency with the resolution of the maps generated previously, we reduced the resolution of the raster file to 5km x 5km pixel. We included the friction map from Malaria Atlas project (MAP) to represent accessibility and mobility in the model.[11] The friction map is expressed in units of minutes required to travel on meter, representing allocated fastest speed of travel based on the types of travel mode estimated within the pixel.[12] It used mainly the most updated data until 2015 for road, railroad and water travel distances from the Open Street Maps (OMS) and Google roads projects. The datasets include roads, railways, rivers, lakes, oceans, topographic conditions (e.g., slope and elevation), landcover types, and national borders.[12] We overlaid the friction map with the map of study to link the sample locations with corresponding values.

For selection of variables, first we conducted preliminary bivariate logistic regression analyses to assess the existence of associations between each variable and HIV prevalence for each gender. Second, since we were interested in variables that included spatial information, we explored the randomness of the geographical distribution (spatial structure) of each variable using Global Moran's Index for spatial autocorrelation. We selected variables that satisfied both criteria: 1) the bivariate logistic regression slope coefficient must have *p-value* < 0.1; and 2) the significance test for spatial autocorrelation must have *p-value* < 0.1. Then, we used multivariable logistic regression models, one for women and another for men, to examine the associations between each selected variables and HIV prevalence adjusting for the other variables. For this model, the dependent variable, y_i, is the percentage of HIV seropositive individuals for women and men at the sample location *i,* respectively. The HIV prevalence at sample location i was modeled as $y_i \sim binomial\ (N_i, p_i)$, where N_i is the total number of individuals sampled at sample location i and p_i is the HIV prevalence at sample location *i.* The multivariable regression model for the HIV prevalence is

$$logit\ (p_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki},$$

where X_{ji} is the value of predictor variable *j* at sample location i and β_i is the corresponding regression coefficient.

Supplementary Figure 1 represents the cofactors in the final logit model for men, and the cofactors are poverty, lifetime number of sexual partners more than two, ever been HIV tested, GH-I, and male circumcision. Each variable shows the spatial variance in the country. For example, the high percentage of poverty tend to focus on the northern and eastern part of the country across the *Limpopo, KwaZulu-Natal*, and *Eastern Cape* Provinces. In contrast, the southeastern part of the country has very low poverty percentages. Supplementary Figure 2 shows the cofactors in the final logit model for women. The cofactors are poverty, condom use, lifetime number of sexual partners more than two, NDVI, and friction. These cofactors show spatial variance. The poverty and the high number of lifetime sexual partners which are also selected for men show similar spatial patterns with men, showing *KwaZulu-Natal* province has the large areas of high prevalence.

Mapping predictor variables and HIV prevalence

We used ordinary kriging mapping methods, which is one of the most widely used interpolation methods, to generate continuous surface maps for each significant socioeconomic, demographic, and behavioral factor identified in the multivariable regression model. This method interpolates the values of variables at unmeasured locations by estimating a variogram of weighted averages of the data.[13] Using this method, we generated continuous surface maps of each selected variable with 5km x 5km pixel resolution. We generated an HIV prevalence map for each gender by substituting values from all continuous surface maps into the gender-specific multivariable logistic regression model using Map Algebra and the inverse logistic equation:

$$HIV\ Prevalence = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k}}{1+\ e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots++\beta_k X_k}}$$

We used this method to generate a map of HIV prevalence in raster format with 5-kilometer grid resolution for each gender, implemented in the software ArcGIS version 10.5. Finally, a 5km x 5km pixel resolution map of the density of women and men living with HIV was generated by combining the HIV prevalence map with the women and men density maps. We obtained estimates of population densities from the WorldPop dataset as a raster image with a resolution of 100m x 100 m pixel. For consistency with the resolution of the maps generated previously, we reduced the resolution of the raster image to 5km x 5km pixel. Since the survey dataset restricted the respondents' age from 15-49 for women and 15-59 for men, we estimated the population for each gender to represent precise estimates of people living with HIV. We obtained the ratio of women aged 15-49 and men aged 15-59 among the total population in 2016 from *PopulationPyramid* dataset [14] and combined the ratio with the population density to generate the estimates of study population, women aged 15-49 and men 15-59.

Definition of variables for underserved areas

To measure underserved areas from health care facilities, we used two datasets: the friction map that we previously addressed for from MAP and the locations of health care facilities in South Africa.[15] We obtained all public health care facilities and private-not-for-profit sectors managed by government (e.g., the Ministry of Health, MoH), local authorities, and faith-based (FBO) and non-governmental organization (NGO) to represent public health services. Those facilities serve expanded immunization programs, health data surveillance, and receive government funding to provide services to the general population. Facilities that provided services for special groups and for private were excluded from the analysis. The main source of the locations of health care facilities is MoH, and other relevant sources such as other government agencies, website of FBOs and NGOs, or personal contact were used if MoH had inadequate facility listing information.[15] A total of 4,287 health care facilities were included in the analysis.

Mapping HIV underserved areas

We modeled travel time as a function of distance and travel speed. It can be conceptualized as the cost of movement using the cell dimensions (pixel resolution) and travel time assigned to each cell. In this model, travel occurs through cell to cell in both cardinal and diagonal directions, and a travel time was designated for each cell, representing the time required to traverse the cell estimated from the friction map from MAP. For example, if we travel from cell A to contiguous cell B in cardinal directions, travel time from cell A to cell B (T_{AB}) would be calculated as following,

$$T_{AB} = \left(T_A * \frac{d}{2}\right) + \left(T_B * \frac{d}{2}\right),$$

where T_i is the travel time assigned to the cell i, and d is the distance between cell centers, which is equal to the cell resolution. The movement consists in two steps because half of each cell is traversed with each step. In case of diagonal direction movement, travel time from cell A to cell C, the program could choose either a direct route between the two cells or a route traversing another cell, which is located at a right angle. The movement of direct route between the two points would be calculated using the Pythagorean theorem to adjust the distance term as following,

$$T_{AC} = \left(T_A * \frac{\frac{\sqrt{2}}{2} * d}{2}\right) + \left(T_B * \frac{\frac{\sqrt{2}}{2} * d}{2}\right).$$

Travel time or cost for traversing each cell was calculated using the cell length and travel time assigned to each cell. An accumulated cost surface was created wherein cell values represented the total travel time from the cell to the nearest hospital location (i.e., least cost path for each cell). To identify underserved areas, we reclassified the accumulated travel time surface into a Boolean surface based on whether the cell was greater than thresholds (30, 60, and 120 minutes) from a health care location. Then, we filtered the grid representing underserved areas to remove any groups of less than three contiguous cells. We conducted the filtering process in an effort to remove single cells and very small areas generally inside the catchment areas. After the filtering process, we converted the underserved areas from a raster grid to a vector data form (polygons) to identify the women and men living with HIV in these areas.

We combined the continuous surface maps of women and men living with HIV generated in the previous step with the maps of underserved areas to generate density maps of women living with HIV and men living with HIV in the underserved areas, and calculate the number of women living with HIV and men living with HIV in these areas.

Supplementary Tables

Supplementary Table S1. Descriptive statistics of socio-behavioral variables included in the analysis.

Gender		Variable	Mean	Standard Deviation
Men	**Socio-behavioral**	*Secondary or higher education (%)*	80.06	27.61
		Poverty (poorer or poorest) (%)	43.67	42.28
		Condom use (%)	50.01	36.84
		Lifetime number of sexual partners more than two (%)	19.65	30.27
		Ever been tested for HIV (%)	72.55	30.50
		Male circumcision (%)	54.44	38.22
	Environmental	*NDVI*	0.05	0.24
		Night light	25.69	22.19
		GH-I	31.96	13.77
		Distance to main roads (km)	11.79	17.17
		Distance to dams (km)	34.79	31.14
		Population density (/km^2)	1,049.33	1,368.41
		Friction (min/m)	0.0063	0.0047
Women	**Socio-behavioral**	*Secondary or higher education (%)*	87.17	22.17
		Poverty (poorer or poorest) (%)	40.73	42.00
		Condom use (%)	43.14	34.89
		Lifetime number of sexual partners more than two (%)	46.37	34.36
		Ever been tested for HIV (%)	84.92	23.15
	Environmental	*NDVI*	0.04	0.24

Kim H, *et al. BMJ Global Health* 2021; 6:e004089. doi: 10.1136/bmjgh-2020-004089

	Night light	26.48	22.46
	GH-I	32.23	13.77
	Distance to main roads (km)	11.66	17.08
	Distance to dams (km)	34.77	31.24
	Population density (/km^2)	1,057.90	1,340.49
	Friction (min/m)	0.0062	0.0046

Kim H, *et al. BMJ Global Health* 2021; 6:e004089. doi: 10.1136/bmjgh-2020-004089

Supplementary Table S2. Bivariate logistic regression analysis and Moran's I analysis for cofactor selection

Gender	Variable	Estimate	*P*	Moran's *I*	*P*
Men	*Secondary or higher education (%)*	-0.0095	<0.001	0.00	0.85
	Poverty (poorer or poorest) (%)	0.0026	0.09	0.21	<0.001
	Condom use (%)	0.0007	0.73	0.00	0.82
	Lifetime number of sexual partners more than two (%)	-0.0160	<0.001	0.04	0.07
	Ever been tested for HIV (%)	0.0079	<0.001	0.04	0.06
	Male circumcision (%)	-0.0063	<0.001	0.11	<0.001
	NDVI	0.4692	0.06	-	-
	Night light	0.0030	0.30		
	GH-I	0.0155	<0.001	-	-
	Distance to main roads (km)	-0.0116	0.01	-	-
	Distance to dams (km)	0.0000	0.30	-	-
	Population density (/km^2)	0.0001	0.17	-	-
	Friction (min/m)	-13.0737	0.34	-	-
Women	*Secondary or higher education (%)*	-0.0072	<0.001	0.02	0.29
	Poverty (poorer or poorest) (%)	0.0045	<0.001	0.19	<0.001
	Condom use (%)	0.0083	<0.001	0.04	0.08
	Lifetime number of sexual partners more than two (%)	-0.0085	<0.001	0.07	0.00
	Ever been tested for HIV (%)	0.0108	<0.001	0.02	0.24
	NDVI	0.9206	<0.001	-	-
	Night light	-0.0005	0.81	-	-
	GH-I	0.0001	0.97	-	-

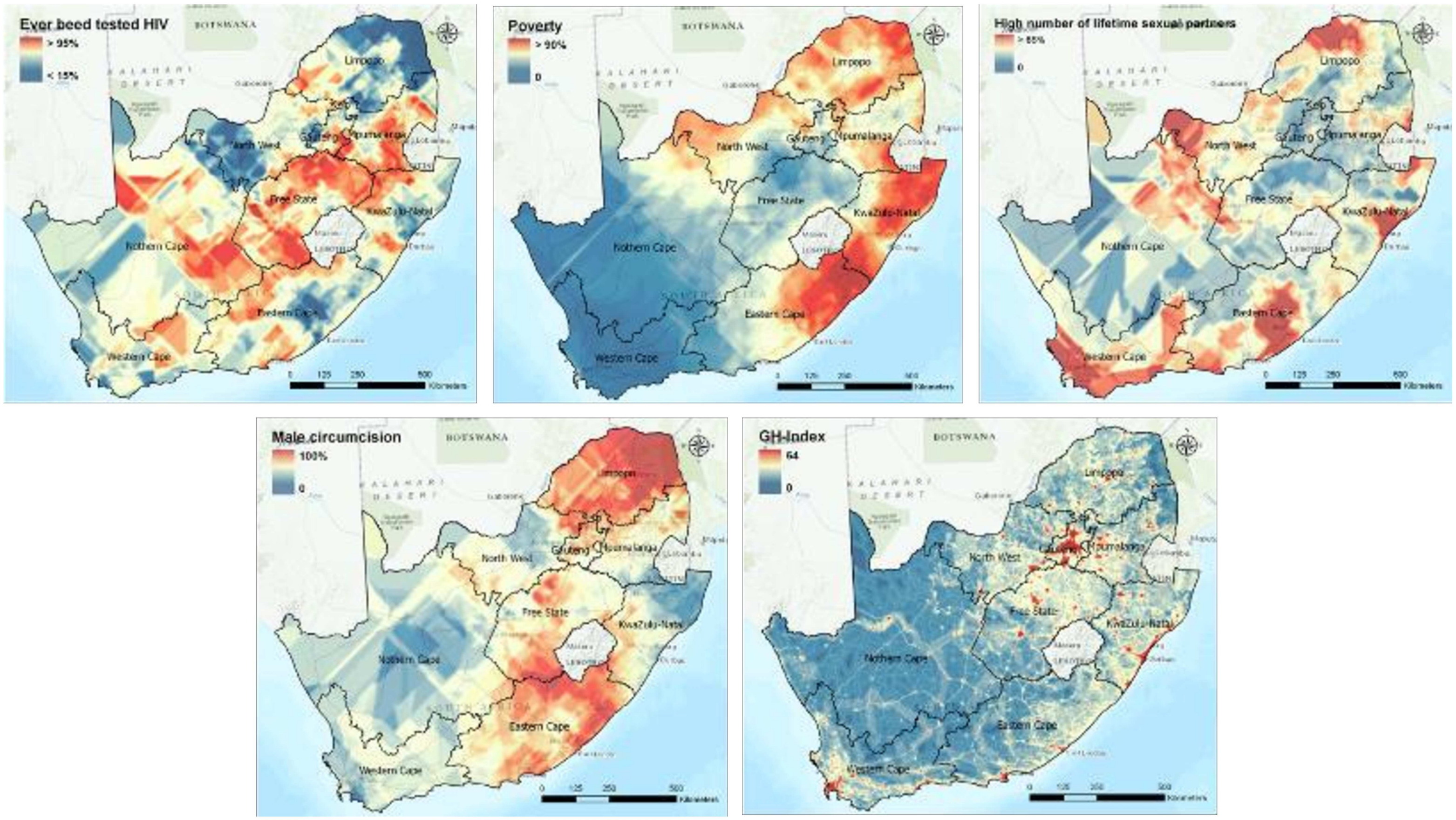
Kim H, *et al. BMJ Global Health* 2021; 6:e004089. doi: 10.1136/bmjgh-2020-004089

	Distance to main roads (km)	-0.0058	0.03	-	-
	Distance to dams (km)	0.0000	0.21	-	-
	Population density (/km^2)	0.0000	0.97	-	-
	Friction (min/m)	-16.0918	0.09	-	-

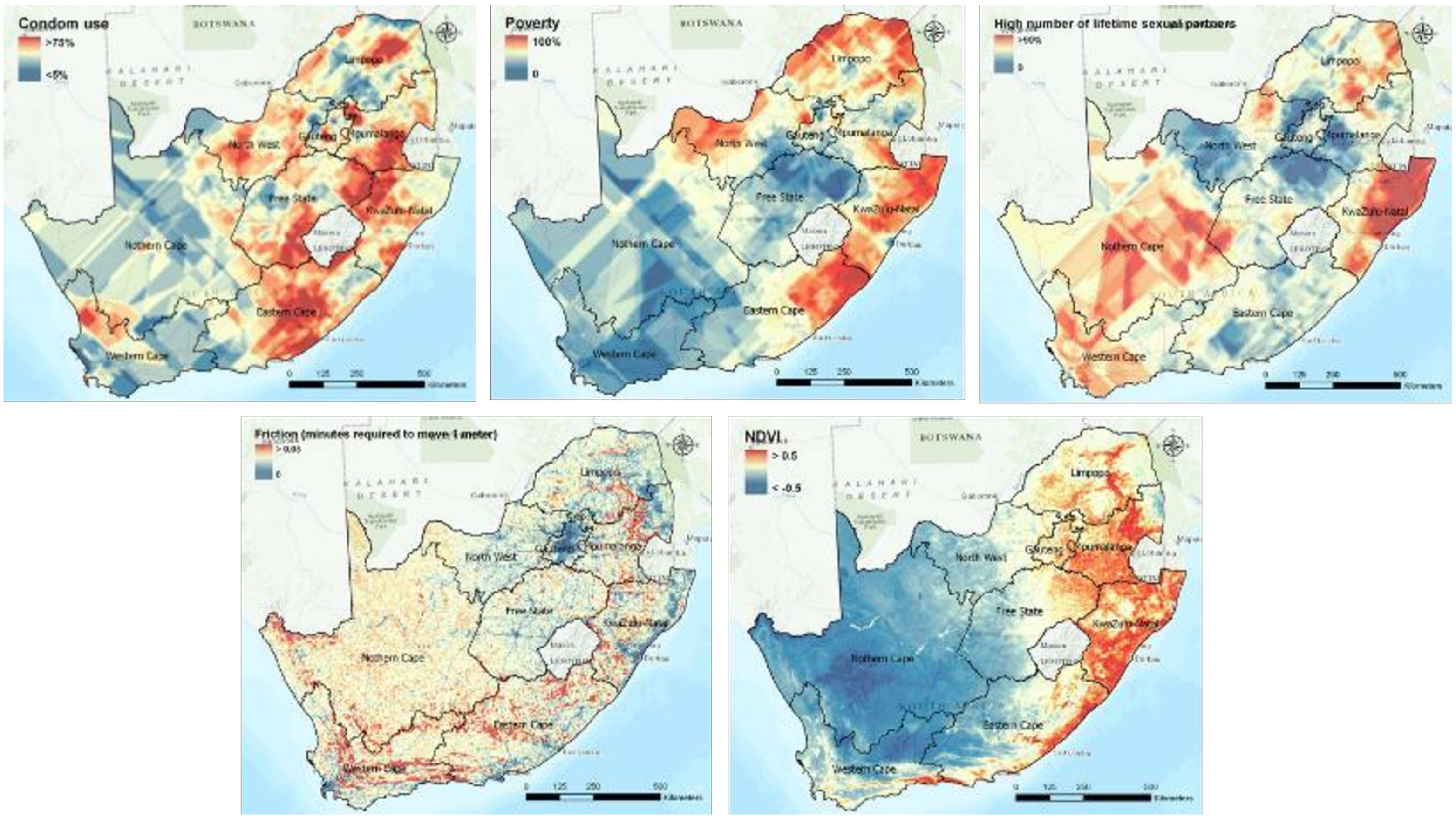
Kim H, *et al. BMJ Global Health* 2021; 6:e004089. doi: 10.1136/bmjgh-2020-004089

Supplementary Figures

Supplementary Figure S1. Continuous surface maps for the cofactors in the final model for men. Maps were created using ArcGIS by ESRI version 10.5 (http://www.esri.com).



Supplementary Figure S2. Continuous surface maps for the cofactors in the final model for women. Maps were created using ArcGIS by ESRI version 10.5 (http://www.esri.com).



References

1. National Department of H, Icf. South Africa Demographic and Health Survey 2016. Pretoria: National Department of Health - NDoH - ICF, 2019.
2. ICF. Demographic and Health Surveys Standard Recode Manual for DHS7. Rockville, Maryland, U.S.A.:ICF: The Demographic and Health Surveys Program, 2018.
3. Rutstein SO, Johnson K. The DHS wealth index. DHS Comparative Reports No 6. Calverton, Maryland, USA: ORC Macro, 2004.
4. Abu-Raddad LJ, Patnaik P, Kublin JG. Dual infection with HIV and malaria fuels the spread of both diseases in sub-Saharan Africa. *Science* 2006;314(5805):1603-06.
5. Cuadros DF, Branscum AJ, Crowley PH. HIV–malaria co-infection: effects of malaria on the prevalence of HIV in East sub-Saharan Africa. *International journal of epidemiology* 2011;40(4):931-39.
6. NASA, Observatory E. Measuring Vegetation (NDVI & EVI) [Available from: http://earthobservatory.nasa.gov/Features/ MeasuringVegetation/.
7. Group WB. Underlying Data compiled by Africon for AICD study 2009 [Available from: https://datacatalog.worldbank.org/dataset/south-africa-roads.
8. Lehner B, Reidy Liermann C, Revenga C, et al. Global Reservoir and Dam Database, Version 1 (GRanDv1): Dams, Revision 01. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), 2011.
9. Wildlife Conservation Society - WCS, Center for International Earth Science Information Network - CIESIN - Columbia University. Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Influence Index (HII) Dataset (Geographic). Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), 2005.
10. Linard C, Gilbert M, Snow RW, et al. Population Distribution, Settlement Patterns and Accessibility across Africa in 2010. *PLOS ONE* 2012;7(2):e31743. doi: 10.1371/journal.pone.0031743
11. Project MA. Malaria Atlas Project (MAP) [cited 2019. Available from: https://map.ox.ac.uk/.
12. Weiss DJ, Nelson A, Gibson HS, et al. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* 2018;553(7688):333-36. doi: 10.1038/nature25181
13. Wackernagel H. Ordinary Kriging. In: Multivariate Geostatistics. Berlin, Heidelberg: Springer 1995.
14. PopulationPyramid. Population Pyramids of the World: South Africa 2016 [Available from: https://www.populationpyramid.net/south-africa/2016/.
15. Maina J, Ouma PO, Macharia PM, et al. A spatial database of health facilities managed by the public health sector in sub Saharan Africa. *Scientific Data* 2019;6(1):134. doi: 10.1038/s41597-019-0142-2