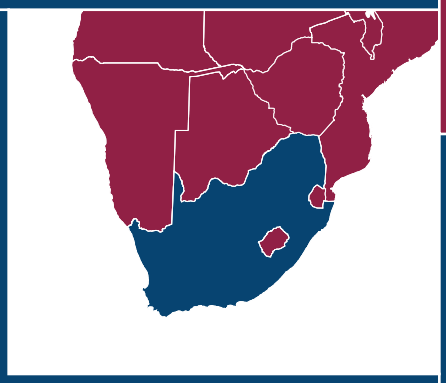
policy

August 2015

# HPP GeoHealth MAPPING



## *Using Geospatial Analysis to Understand the Local HIV Epidemic in KwaZulu-Natal Province and eThekwini Municipality in South Africa*

This publication was prepared by Ian Wanyeki, Benard Mitto, Lucy Nganga, Isabel Brodsky, Anita Datar, and Eurica Palmer of the Health Policy Project.

USAID FROM THE AMERICAN PEOPLE

PEPFAR

HEALTH POLICY PROJECT

Suggested citation: Wanyeki, I., B. Mitto, L. Nganga, I. Brodsky, A. Datar, and E. Palmer. 2015. *HPP GeoHealth Mapping: Using Geospatial Analysis to Understand the Local HIV Epidemic in KwaZulu-Natal Province and eThekwini Municipality in South Africa.* Washington, DC: Futures Group, Health Policy Project.

ISBN: 978-1-59560-107-0

The Health Policy Project is a five-year cooperative agreement funded by the U.S. Agency for International Development under Agreement No. AID-OAA-A-10-00067, beginning September 30, 2010. The project's HIV activities are supported by the U.S. President's Emergency Plan for AIDS Relief (PEPFAR). It is implemented by Futures Group, in collaboration with Plan International USA, Futures Institute, Partners in Population and Development, Africa Regional Office (PPD ARO), Population Reference Bureau (PRB), RTI International, and the White Ribbon Alliance for Safe Motherhood (WRA).

# HPP GeoHealth Mapping

*Using Geospatial Analysis to Understand the Local HIV Epidemic in KwaZulu-Natal Province and eThekwini Municipality in South Africa*

**AUGUST 2015**

This publication was prepared by Ian Wanyeki, Benard Mitto, Lucy Nganga, Isabel Brodsky, Anita Datar, and Eurica Palmer of the Health Policy Project.

The information provided in this document is not official U.S. Government information and does not necessarily represent the views or positions of the U.S. Agency for International Development.

# CONTENTS

**Background** ..... 1

**Objectives** ..... 1

**Methodology** ..... 2

Setting ..... 2
Data Sources ..... 2
Catchment Areas ..... 2
Inclusion and Exclusion Criteria ..... 2
Geocoding and Linking Programmatic Data to the Health Facilities ..... 3
Statistical Methods ..... 3

**Limitations, Potential Sources of Bias** ..... 6

**Timeline and Staff Time per Location of Interest** ..... 7

**Illustrative Examples** ..... 7

**Recommendations** ..... 7

**Annex A. The Provincial Monthly PMTCT Data Collection Form** ..... 8

**Annex B. The Provincial Monthly ART Data Collection Form** ..... 8

**Annex C. Illustrative Mapping Examples** ..... 9

**References** ..... 12

# BACKGROUND

Donors and development partners are now emphasizing the importance of taking a geographic approach to funding requests. For example, the Global Fund's New Funding Model encouraged countries to support funding requests with mapping and spatial analysis as part of the evidence base for Concept Note submission (i.e., funding requests).

In support of the Global Fund's New Funding Model, the President's Emergency Plan for AIDS Relief (PEPFAR) worked in coordination with the United States Agency for International Development (USAID) through the Health Policy Project (HPP) to provide technical assistance to select high-impact countries. This work was intended to strengthen these countries' capacity to use mapping and spatial analysis to identify the location of epidemic 'hot spots' at subnational levels to more effectively target the HIV response. Technical assistance included the use of mapping and spatial analysis to strengthen the evidence base of Concept Note submissions.

Toward this end, HPP supported the South Africa National AIDS Council (SANAC) and provincial health authorities in KwaZulu-Natal (KZN) on a proposal to map estimated HIV prevalence using routinely collected facility-level data and other related data in KZN province and the metro municipality of eThekwini (formerly known as Durban). These efforts were intended to help inform the country's Concept Note submission, scheduled for June 2015. In addition to providing technical assistance in support of the analysis, HPP documented the methodology (Judice and Datar, 2014)—adapted from Dr. Frank Tanser's work in Mpumalanga Province—so that SANAC might include a request to replicate this analysis in multiple municipalities throughout the country as part of the funding application to the Global Fund.

This report documents the methodology, data sources, and statistical methods used to map and analyze routine data to examine HIV variation at subnational levels.

# OBJECTIVES

The objective of this analysis is to use a geographic approach to examine spatial variations in HIV disease burden within KwaZulu-Natal Province, with a focus on eThekwini municipality. HIV burden is analyzed in three ways: 1) HIV positivity, 2) number of people living with HIV (PLHIV), and 3) antiretroviral therapy (ART) coverage gaps.

# METHODOLOGY

## Setting

The primary area of study was the province of KwaZulu-Natal and, specifically, eThekwini municipality.

## Data Sources

**Table 1: Indicators, Level of Aggregation and Data Source**

| Data | Key Indicators/Variables | Date | Level of Aggregation | Source |
|---|---|---|---|---|
| **PMTCT** | Annual HIV positivity among women* | 2012–2014 | Clinic | DHIS |
| **ART** | # Adults currently on ART* | December 2014 | Clinic | DHIS |
| **Administrative boundaries** | Catchment areas | 2012 | Province, district, municipality | StatsSA |
| **Gridded population** | Population 15–49 | 2013 | 1 $Km^2$ | LandScan |
| **Master facility list** | Latitude, longitude of health facilities | 2014 | Clinic | DHIS |

* Indicators collected from the provincial monthly PMTCT and ART data collection form (Appendix A and B), which are then entered in the electronic DHIS database.

## Catchment Areas

The boundary files from the Census 2012 were used. Estimates were aggregated to the district level, as this is where decision making occurs in KwaZulu-Natal; however, estimates can be aggregated up to any desired boundary level.

## Inclusion and Exclusion Criteria

The following criteria (tables 2 and 3) were used to determine the final number of prevention of mother-to-child transmission (PMTCT) and ART clinics for analysis.

**Table 2: Inclusion and Exclusion Criteria for PMTCT Clinics**

| | Criteria | KZN |
|---|---|---|
| **Inclusions** | PMTCT clinics that provided and reported on PMTCT services between 2012 and 2014 | **840** |
| **Exclusions** | < 50 women tested for HIV in 2014 | 232 |
| | Missing geo-coordinates | 48 |
| **Total PMTCT clinics included for analysis** | | **560** |

**Table 3: Inclusion and Exclusion Criteria for ART Clinics**

| | Criteria | KZN |
|---|---|---|
| **Inclusions** | ART clinics that provided and reported on ART services in 2014 | **804** |
| **Exclusions** | Missing geo-coordinates | 175 |
| **Total ART clinics included for analysis** | | **629** |

## Geocoding and Linking Programmatic Data to the Health Facilities

An attempt was made to obtain geocoded program data from the KwaZulu-Natal Department of Health; however, program data was provided without the geocodes of clinics. MEASURE Evaluation Strategic Information for South Africa (SIFSA) provided a geocoded master facility list (MFL) that contained a list of clinics in KwaZulu-Natal. Programmatic PMTCT and ART data were linked to this MFL and matched for facility name, district, and province. This enabled latitude and longitude coordinates of health facilities to be linked to program data. Facilities with missing geo-coordinates were excluded from the analysis.

560 (92%) of the 608 health facilities offering PMTCT services selected for analysis were geocoded, while 629 (78%) of the health facilities offering ART services were geocoded. The final dataset, after exclusions, resulted in 560 PMTCT sites and 629 ART sites.

## Statistical Methods

### *HIV positivity*

Annual trends in HIV positivity for the years 2012, 2013, and 2014 were graphed at both the provincial level and the eThekwini municipality level.

Annual HIV positivity at each PMTCT clinic was calculated as

*Antenatal client HIV 1st test HIV positive / Antenatal client HIV 1st test * 100*

### *Confidence intervals*

The 95% confidence interval around this estimate was calculated as

*HIV positivity% ± 1.96 * std error*

### *HIV prevalence estimates*

We calculated an estimation of HIV prevalence at each clinic (see below) through adjustment with HIV prevalence values from the 2012 national HIV population-based survey. In the 2012 National HIV Human Sciences Research Council (HSRC) survey, HIV prevalence among those ages 15–49 was 14.5 percent for men, 23.2 percent for women, and 18.8 percent overall. We created an adjustment factor (Gouws et al., 2008) by dividing overall prevalence by the female prevalence.

*(HRSC 2012 overall HIV prevalence / HRSC 2012 female prevalence) = 18.8/23.2 = 0.81*

We then used this adjustment factor to convert HIV positivity among pregnant women to an estimate of prevalence in the general population, as show below:

*% HIV-positive females from PMTCT * 0.81*

For example, if an HIV positivity of 13.8 percent was observed among pregnant women, this would result in an estimated prevalence of 11.2 percent in the general population.

*13.8 * 0.81 = 11.2%*

We used HIV prevalence for the 15–49 age group from the 2012 HRSC survey, as it is expected that women attending PMTCT clinics will be within this age range.

### *Interpolation*

Interpolation can be defined as the estimation of attribute values at unsampled points from measurements made at surrounding sites (sampled points). In this example, the sample points are the PMTCT clinics and the attribute value is estimated HIV prevalence. Interpolation enables us to obtain estimates of HIV prevalence at a fine level of aggregation (1 $km^2$); with these estimates, we can then calculate estimates of PLHIV at the same resolution. This enables us to drill down or aggregate these estimates up to any desired level of aggregation and examine HIV burden.

### *Selection of an appropriate interpolation method and parameters*

Multiple interpolation methods exist and different methods have been used to estimate HIV prevalence from point data (HIV Monitoring Consortium, 2014; Burgert, 2014). These methods vary in both complexity and prediction accuracy. Our aim was to select the most appropriate method that minimizes the prediction error and provides estimated errors of prediction. We formally evaluated several interpolation methods, along with various input parameters, prior to selecting an appropriate method for the KwaZulu-Natal dataset. Two validation processes were used, as described below.

- Partitioned holdback: This involves holding back 10 percent of data points and using the rest of the data to create an interpolated surface. The predicted estimates are compared with the actual values in the sample. We selected the model that resulted in the best prediction (the lowest root mean square error).
- Leave-one-out cross validation method: This involves holding back one data point and comparing this with the predicted estimate. This process is done iteratively for all data points. The predicted estimates are then compared with the actual values in the sample. We selected the model that resulted in the best prediction (the lowest root mean square error).

Universal Kriging was eventually selected to provide the best predicted HIV surface. Kriging utilizes a geostatistical technique in predicting values, and assumes that distance or direction between sample points reflects spatial correlation that can be used to explain variation in the surface.

### *Interpolation of HIV prevalence*

All data was first projected to Cape_UTM_Zone_35S, which is a good projection for areas such as KwaZulu-Natal that are located within the UTM grid between 28°E to 30°E and has units in meters. Adjusted HIV prevalence rates for the PMTCT clinics in 2014 were then interpolated using Universal Kriging.

### *Calculation of people living with HIV*

To obtain estimates of people living with HIV, we multiplied the projected 2013 surface of adult population ages 15–49 by the interpolated prevalence surfaces to obtain numbers of PLHIV per $km^2$. This information was then aggregated up to the desired level of aggregation.

### *ART coverage*

To calculate estimated ART coverage gaps as of December 2014, we calculated the proportion of adults ages 15–49 on ART divided by the estimated number of PLHIV in the same age group. We multiplied this by 100 to obtain the percentage of patients on ART by district.

### *Hot spot analysis*

Hot spot analysis can be used to identify the locations of statistically significant hot spots (locations where high values cluster together) and cold spots (locations where low values cluster together) in a dataset. For HIV burden analysis, hot spot analysis can be used to identify the locations of PMTCT clinics that have a cluster of HIV positivity values that are either statistically higher (hot spots) or lower (cold spots) than the average. To be a statistically significant hot spot, a feature will have a high value and be surrounded by other features with high values. This information can be used for resource allocation. We used the spatial autocorrelation (Moran's I) statistic iteratively to determine if clustering existed, and then to determine the optimal distance scale for analysis. Using the optimal parameters, we used the Getis-Ord GI* static (Getis and Ord, 1992) to identify statistically significant hot and cold spots at the ≥ 90% confidence level.

# LIMITATIONS, POTENTIAL SOURCES OF BIAS

### *HIV prevalence estimates*

Using data from women attending ANC clinics to obtain estimates of HIV prevalence may be limited, as it only samples pregnant women attending health services and excludes women who are not pregnant or sexually active, as well as women who are pregnant but do not attend ANC. This data also excludes men (Burgert, 2014). The proportion of pregnant women attending ANC who were not tested for HIV likely impacts the ability to adequately estimate HIV prevalence in this methodology. Excluding PMTCT clinics with a low sample size (fewer than 50 women tested for HIV in 2014) may have limited the geographic representativeness of the study. The reporting rate of PMTCT service data in 2014 was also likely to impact the results.

### *Geographical representativeness*

All PMTCT clinics in the district with geocodes were selected (excluding those with low sample size as outlined in table 2); in this way, issues around geographical representativeness were minimized.

### *Data quality*

There may be data quality issues. These could occur at several stages in the process, from quality of HIV testing to transcription or transmission of results. We were unable to assess the quality of data, as this was beyond our scope of work.

### *Data of patients on ART*

Programmatic data of people ages 15–49 currently on ART is likely to be overestimated. This could be a result of clinics lacking the ability to accurately keep track of patients on ART, or of patients who appear to be active on ART but who may have actually died or become lost to follow-up.

### *Assumptions on where patients get tested and where they receive treatment*

This methodology assumes that patients receive HIV testing or ART services at the clinic closest to where they live; however, this may not necessary be true. For urban areas like eThekwini—in which many people work but do not reside—this assumption could be a limitation of the analysis.

## TIMELINE AND STAFF TIME PER LOCATION OF INTEREST

| Activity | Month 1 | Month 2 | Month 3 | Month 4 | Month 5 | Month 6 |
|---|---|---|---|---|---|---|
| Proposal review and approval | | | | | | |
| Identification and collection of data sources | | | | | | |
| Data cleaning and geocoding (if necessary) | | | | | | |
| Data analysis and mapping | | | | | | |
| Dissemination and validation of analyses and maps | | | | | | |
| Methodology write-up | | | | | | |

Staff time: 2.5 full-time equivalents (FTEs) for duration of analysis, per location of interest (i.e., municipality, district, etc.).

- 1 spatial epidemiologist
- ½ time spatial analyst
- ½ time epidemiologist
- ½ time manager/consultant

## ILLUSTRATIVE EXAMPLES

(See Annex C for illustrative maps.)

## RECOMMENDATIONS

Understanding geographic variation at the most granular level can help identify 'hot spots within hot spots' and target the HIV response, resulting in allocative efficiencies. The use of routinely available facility-level PMTCT data, which is both low-cost and timely, allows decisionmakers to monitor and manage key indicators, including HIV positivity, the estimated number of PLHIV, and ART coverage. Using maps to visualize these data allows decisionmakers to explore potential drivers of disease transmission, understand why some facilities experience a higher volume of HIV-positive patients than others, identify which sites are on track to meet targets, and identify where additional resources might be needed. The HPP team recommends replicating this analysis in multiple epidemic hot spots throughout South Africa so that SANAC and partners may identify where HIV disease burden is greatest and thereby effectively target new and existing resources.

# ANNEX A. THE PROVINCIAL MONTHLY PMTCT DATA COLLECTION FORM

**Provincial Data Management and GIS Services**
**Monthly PMTCT Data Collection Form**

health
Department:
Health
PROVINCE OF KWAZULU-NATAL

Health Facility: _____ Date: (dd/mm/yy) _____/_____/_____

| ANTENATAL CARE CLINIC | Number | Comment |
|---|---|---|
| Antenatal client Pre - Test counselled for HIV | | |
| Antenatal client HIV 1st test | | |
| Antenatal client HIV 1st test positive | | |
| Antenatal client HIV 1st test negative | | |

# ANNEX B. THE PROVINCIAL MONTHLY ART DATA COLLECTION FORM

**Provincial Data Management and GIS Services**
**Monthly ART Form**

health
Department:
Health
PROVINCE OF KWAZULU-NATAL

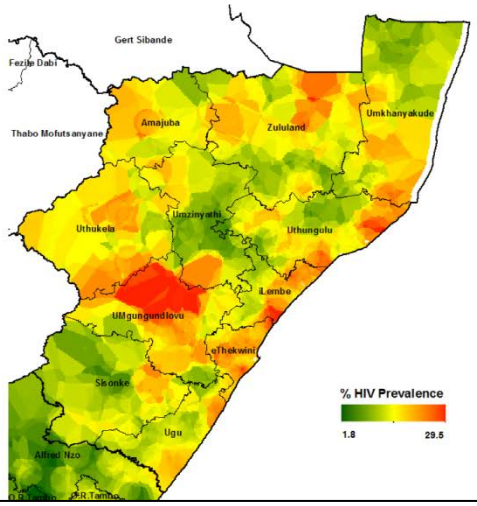
Health Facility: _____ Date: (dd/mm/yy) _____/_____/_____

| Pre - ART | Number | Comment |
|---|---|---|
| Pre-art assessment first visit - Male | | |
| Pre-art assessment first visit - Female | | |
| Pre-art assessment first visit - Paediatrics | | |
| HIV positive new patient screened for TB | | |
| HIV positive new patient diagnosed with TB | | |
| HIV positive client eligible for IPT | | |
| HIV positive patients started on Co-trimoxazole prophylaxis | | |
| TB/HIV co-infected client - total | | |
| TB/HIV co-infected client initiated on CPT | | |
| HIV positive client initiated on IPT | | |
| Patients who develop active TB whilst taking INH Preventive Therapy (IPT) | | |
| HIV Positive patients who completed 6 months of INH preventive therapy (IPT) | | |
| Newly diagnosed HIV positive patient (excluding antenatal client) with Blood drawn for CD4 | | |
| Child under 1 year eligible for ART during this month | | |
| Child 1- 5 years eligible for ART during this month | | |
| HIV positive adult patient eligible for ART | | |
| HIV positive child under 15 years eligible for ART | | |
| HIV patients medically eligible for ART on waiting list | | |
| **ART** | **Number** | **Comment** |
| Child under 1 year started on ART during this month - naïve | | |
| Child 1 -5 years started on ART during this month - naïve | | |
| Child under 15 years started on ART during this month - naive | | |
| Adult started on ART during this month - naive | | |
| Child <15 remaining on ART at the end of the month - Total | | |
| Adult remaining on ART at the end of the month - Total | | |

# ANNEX C. ILLUSTRATIVE MAPPING EXAMPLES

KwaZulu-Natal Analysis

## Estimated HIV Prevalence

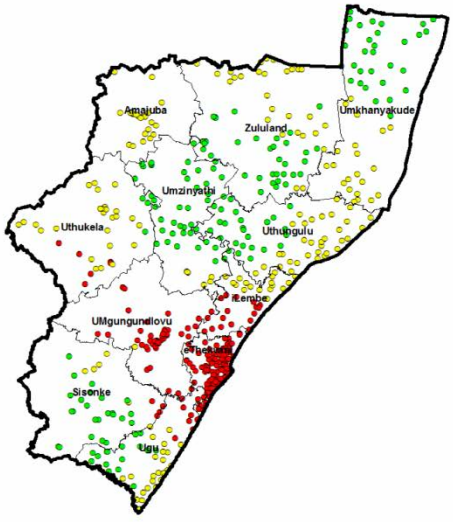



Using the same methodology of HIV prevalence estimation from HIV positivity at PMTCT sites and interpolation, HPP created a map of estimated HIV prevalence in KwaZulu-Natal.

The map shows major hotspot areas in uMgungundolvu, eThekwini, iLembe, uThungulu, and Zululand.

KwaZulu-Natal Analysis

## HIV Positivity at Health Facilities




HPP conducted a hotspot analysis to identify the locations of statistically significant "hot" and "cold" spots in the province.

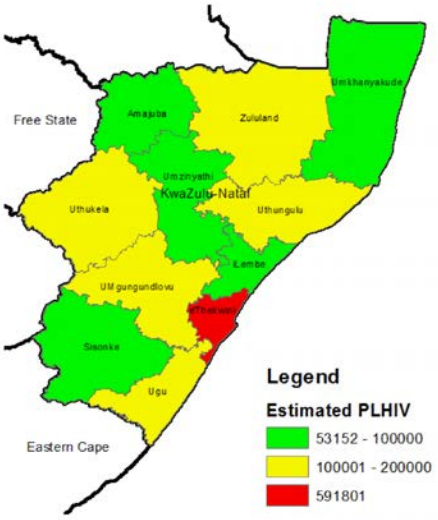
Hot spots have a high HIV positivity and surrounded by other facilities with high values. Cold spots have a low HIV positivity and are surrounded by other facilities with a low value.

For HIV burden analysis, this map can be used to identify the locations of PMTCT clinics that identify a significantly high number of HIV positive cases.

uMgungundlovu, eThekwini, and iLembe have almost all of the hot spots within the province. Note that these three districts also showed high HIV prevalence in the previous map.

KwaZulu-Natal Analysis

## Estimated PLHIV by District



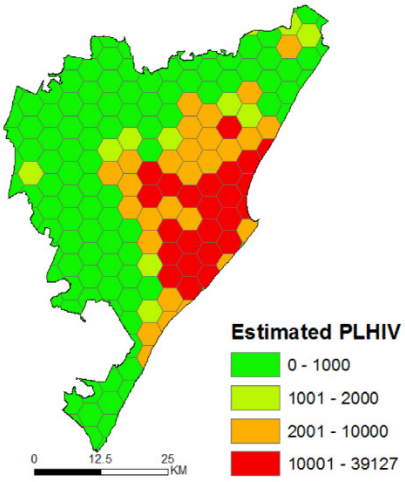

The second method HPP used to estimate disease burden was number of PLHIV by district. This is calculated by multiplying estimated HIV prevalence by the total population.

When examining disease burden through this measure, the only district that shows up as a "hot spot" is eThekwini. Note that uMgungundlovu is in yellow and iLembe is in green.

Examining PLHIV takes into account population density, so urban areas will have higher numbers than rural areas even if the prevalence is similar in both locations. This has important implications for resource allocation and access to services.

KwaZulu-Natal Analysis

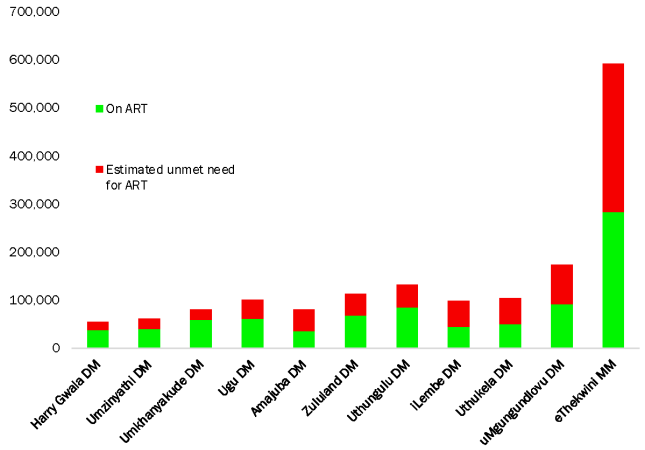
## Estimated PLHIV in eThekwini




HPP drilled down to examine estimated PLHIV within eThekwini. 16 $km^2$ hexagons were created as artificial sub-district boundaries to further understand the spatial distribution of PLHIV living in a concentrated area with high disease burden.

KwaZulu-Natal Analysis

## ART and Estimated Unmet Need by District, December 2014

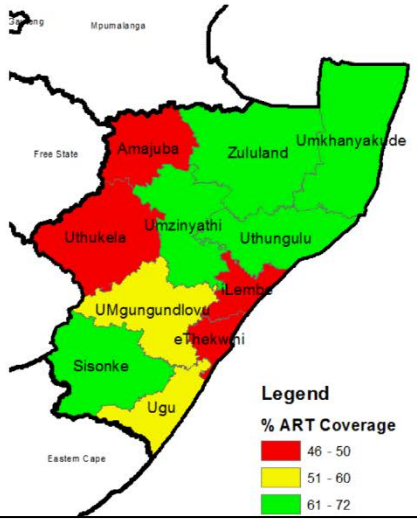



The number of people on ART can be compared to the total number of PLHIV in each district to calculate and estimated unmet need for ART. The limitation to this estimate is that it does not take into account South Africa's ART eligibility guidelines.

Understanding unmet need for ART can have implications on decisions with regards to resource allocation.

KwaZulu-Natal Analysis

## Estimated ART Coverage




This same information can be portrayed in a map to further understand geographic variations in ART coverage. eThekwini and iLembe both have very low ART coverage, as do uThukela and Amajuba.

# REFERENCES

Burgert, C.R. 2014. *Spatial Interpolation with Demographic and Health Survey Data: Key Considerations*. DHS Spatial Analysis Reports No. 9. Rockville, MD: ICF International.

Getis, A. and J.K. Ord. 1992. "The Analysis of Spatial Association by Use of Distance Statistics." *Geographical Analysis* 24(3): 190–206.

Gouws, E., V. Mishra, and T.B. Fowler. 2008. "Comparison of Adult HIV Prevalence from National Population-based Surveys and Antenatal Clinic Surveillance in Countries with Generalised Epidemics: Implications for Calibrating Surveillance Data." *Sexually Transmitted Infections* 84(Suppl 1): i17–i23.

HIV Monitoring Consortium. 2014. *Technical Consultation on Methods for Generating Sub-National Estimates of HIV Epidemiology to Support Country Programme Planning and Evaluation*. Nairobi, Kenya: HIV Monitoring Consortium, UNAIDS Reference Group of Estimates, UNAIDS Hotspot Taskforce.

Judice, N. and A. Datar. 2014. "Demand for and Uses of Geospatial Mapping in HIV Programs: Documenting the Experiences of End-Users in Mpumalanga Province, South Africa." Washington, DC: Futures Group, Health Policy Project.

For more information, contact:

Health Policy Project
Futures Group
1331 Pennsylvania Ave NW, Suite 600
Washington, DC 20004
Tel: (202) 775-9680
Fax: (202) 775-9694
Email: policyinfo@futuresgroup.com
www.healthpolicyproject.com