UNIVERSITY OF CALIFORNIA

Santa Barbara

Decision-Making under Spatial Uncertainty in Downscaled Population Estimates: An Assessment of HIV Prevalence in Tanzania

A Thesis submitted in partial satisfaction of the requirements for the degree Master of Arts in Geography

by

Kevin Muriuki Mwenda

Committee in charge:

Professor Phaedon C. Kyriakidis, Co-Chair

Professor David López-Carr, Co-Chair

Professor Keith C. Clarke

June 2014

The thesis of Kevin Muriuki Mwenda is approved.

_____
Keith C. Clarke

_____
David López-Carr, Committee Co-Chair

_____
Phaedon C. Kyriakidis, Committee Co-Chair

May 2014

Decision-Making under Spatial Uncertainty

in Downscaled Population Estimates: An Assessment

of HIV Prevalence in Tanzania

by

Kevin Muriuki Mwenda

ACKNOWLEDGEMENTS

I would like to thank all the people who have provided me guidance and support throughout the process that has culminated in the writing of this thesis. Special thanks to the co-chair of my thesis committee, Dr. Phaedon Kyriakidis for grounding my interest in spatial statistics and challenging me along the way to always strive to solve a problem logically. Special thanks to my co-chair Dr. David López-Carr whose plethora of pragmatic ideas helped me keep the bigger picture of potential applications of this research in mind. Nothing can beat co-advisors who always make time to meet, share, discuss and break down complex concepts with such clarity. The lessons I have learnt working with Dr. Keith Clarke have enabled me to approach maps with a fresh perspective. A committee member who is readily available as a resource through-out the evolution of a graduate student is invaluable. I would also like to thank the staff members of the Geography Department at UCSB who work tirelessly behind the scenes to support the research and teaching efforts of faculty and graduate students. Furthermore, I would like to thank faculty and co-graduate students who have lent me an ear whenever I requested feedback.

Beyond the Geography Department, my heartfelt gratitude goes to Dr. Peter Kuhn and other research and graduate associates of the Broom Center for Demography who provided me with several opportunities to discuss my work with an inter-disciplinary audience.

Last but not least, I would like to thank my loving family who offered me a tremendous support during my pursuit of a graduate degree.

This thesis is dedicated to my late sister, Christine K. Mwenda.

ABSTRACT

Decision-Making under Spatial Uncertainty
in Downscaled Population Estimates: An Assessment
of HIV Prevalence in Tanzania
by

Kevin Muriuki Mwenda

Spatially explicit impact assessment analyses such as vulnerability studies often require spatially detailed population distribution as input. Downscaled population datasets otherwise known as disaggregated, gridded or fine resolution population datasets are becoming increasingly available at global and regional scales and are based on information from various sources with varying spatial and temporal resolutions as well as reliabilities. Uncertainty is endemic in such downscaled population estimates, particularly in developing countries yet it is hardly assessed. Consequently, decision-makers are potentially faced with a bias problem whereby uncertainties are masked and estimates are presented as unique or expected values even after being derived in a probabilistic context. This research explores how HIV prevalence in three districts of the United Republic of Tanzania might vary with the utilization of simulations of downscaled population estimates. In so doing, this study explores some scenarios in which HIV prevalence that corresponds to minimum expected cost of antiretroviral (ARV) treatment is estimated under three different decision-making attitudes, namely minimax regret, maximin and maximax, followed by a discussion of some

implications of any variation in best estimates of HIV prevalence corresponding to the least impact on ARV cost.

Our findings show that for effective decision analysis, rather than using coarse aggregated values such as census data at the district level, decision-makers may benefit from the application of multiple simulated spatial distributions of fine scale population along with the associated ARV cost estimates. From this distribution of ARV cost estimates, decision-makers could select best estimates based on an explication of risk attitude, thus avoiding unforeseeable consequences of underestimating or overestimating impact assessment outcomes of HIV prevalence.

ABBREVIATIONS AND ACRONYMS

AfriPop – Not an acronym. This is a downscaled population dataset created for countries within the continent of Africa

AIDS – Acquired Immunodeficiency Syndrome

AIS – AIDS Indicator Survey

ARV – Antiretroviral (treatment)

AsiaPop – Not an acronym. This is a downscaled population dataset created for countries within the continent of Asia

CIA – Central Intelligence Agency

DHS – Demographic and Health Surveys

GIS – Geographic Information Systems

GPW – Gridded Population of the World

HIV – Human Immunodeficiency Virus

Landscan – Not an acronym. This is a downscaled global population dataset

MODIS – Moderate-Resolution Imaging Spectroradiometer

PHC – Population and Housing Census

SLEUTH – Slope, Land-use, Exclusion, Urban extent, Transportation and Hill-shade

THMIS – Tanzania HIV/AIDS and Malaria Indicator Survey

UN – United Nations

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. Introduction

In most developing countries, the low resolution aggregated census data that are used to create the high resolution, disaggregated or downscaled population datasets show significant variation by year and by spatial resolution (Linard & Tatem, 2012). In addition, contemporary census data are not usually available for such countries and they therefore typically rely on census data that is around a decade old and collected at coarse administrative units (Tatem et al., 2011). The cumulative effect of using varying years of census data, aggregated intercensal growth rates and adjustments to estimate total population results in downscaled datasets have significant variations in estimated population size and spatial distribution may result (Tatem et al., 2011; Mondal & Tatem, 2012). The choice of such datasets that 'hide' uncertainty behind their input data, methods, and output estimates, may lead to significant discrepancies in vulnerability studies by decision-makers (e.g. researchers, policy-makers and/or agencies) who use the data (Thompson & Graham, 1996; Savage, 2009). For example, Mondal and Tatem (2012) compared two different downscaled population datasets and found differences exceeding 7.5 million people characterized as vulnerable to sea level rise and coastal flooding in Indonesia and Japan. Similarly, Tatem et al. (2012) reported a difference of over ten million people at risk of malaria when employing different gridded population datasets.

The 'vulnerable' populations in our study are adults infected by HIV/AIDS in the United Republic of Tanzania. Tanzania is one of the developing countries worst hit by the HIV/AIDS epidemic (TACAIDS, 2014). Tanzania was ranked 13th in the world based on the 2012 national adult prevalence rate of 5.1% (CIA, 2012a) which translates to 1.3 million adults (AllAfrica, 2012). The age bracket of 15-49 years is typically referred to as the

reproductive age group (DHS, 2013) and the people who are infected by HIV therein (HIV prevalence) are the main consumers of antiretroviral (ARV) drugs that the Tanzanian Government is contemplating manufacturing locally rather than importing. This move is predicted to cost each adult individual with HIV approximately 12,000 Tanzanian Shillings ($7.30 U.S. Dollars) for one dose of ARV treatment per month instead of approximately 16,000 Tanzanian Shillings ($10 U.S. Dollars) (AllAfrica, 2012). Given that these HIV prevalence and ARV cost values are derived from aggregate, not to mention uncertain, information, it begs the question – how might the cost of ARV treatment change with the utilization of simulated spatially distributed population estimates and what might the resulting implications of that change be?

In an attempt to answer the aforementioned question, we first review some downscaling methods used to disaggregate population datasets, then we introduce some concepts in decision theory within which we investigate how the cost of ARV treatment might change with the utilization of simulated spatially distributed population estimates in three districts in Tanzania. First, we determine and compare the district-level aggregate ARV costs with spatially distributed ARV costs derived from spatially distributed population of reproductive age. Next, we derive a distribution of spatially varying total ARV costs from multiple realizations of spatially distributed population. After the resulting implications of any discrepancy in ARV cost from the aforementioned three cost-determination procedures are discussed, three decision-making attitudes are employed to determine the "best" estimates of HIV prevalence that corresponds to the least expected loss or impact in terms of ARV costs by the decision-maker who is tasked with the acquisition of ARV drugs for the HIV prevalent population.

II.Background

A. Downscaling Methods: A Review

The integration of different population measurements obtained over different spatial units or zones is a major problem in spatial population studies. The varying geometrical characteristics of the aforementioned spatial zones warrant the need to redistribute or interpolate population measurements from coarse or low resolution source zones to fine resolution target zones using methods that are collectively known as population downscaling. Two main types of widely used downscaling methods are simple area weighting and dasymetric mapping. The simple area weighting method is one of the most-straight-forward interpolation methods to use because it is only concerned with the re-distribution of a population value into the overlap of target and source zones, without the need of any other extraneous or ancillary data. An example of contemporary downscaled population datasets that use simple area weighting for population modeling is Gridded Population of the World (GPW) version 2 and 3 (Linard & Tatem, 2012).The method is formulated by Fisher & Langford (1995) as

$$\hat{P}_t = \sum_{s=1}^{S} \frac{A_{ts}\, P_s}{A_s} = \sum_{s=1}^{S} A_{ts}\, D_s$$

where $\hat{P}_t$ is the population estimates at target zone *t*; A_{ts} is the overlapping area between target zone *t* and source zone *s*; P_s is the population of source zone *s;* A_s is the area of source zone *s; S* is the total number of source zones; and D_s is the mean population density of source zone *s*.

Dasymetric mapping was first introduced as a method of cartographic representation in which source zones are subdivided into smaller regions in order to ensure consistency of the

density of the variable to be mapped (McCleary, 1984). Examples of contemporary downscaled population datasets that use dasymetric modeling are AfriPop, AsiaPop, Landscan and the Gridded Population of the World (GPW) (Linard & Tatem, 2012). Essentially, each source zone is divided into two or more parts and population is redistributed therein based on a priori information or assumptions that those are the specific areas which are inhabited – a form of density measuring (Wright, 1936; Tobler, 1979). There are two main subtypes of dasymetric mapping: binary and multi-class (polycategorical) dasymetric methods.

In binary dasymetric mapping, the source zone is divided into two sub-regions, one of which is assigned a population from the original source zone, and the other considered uninhabited. The process is formulated by Fisher & Langford (1995) as

$$\hat{P}_t = \sum_{s=1}^{S} \frac{A_{tsp}\, P_s}{A_{sp}} = \sum_{s=1}^{S} A_{tsp}\, D_{sp}$$

where A_{tsp} is the overlapping area between target zone *t* and source zone *s,* with populated land cover having been identified; A_{sp} is the area of source zone *s,* with populated land cover having been identified*; S* is the total number of source zones; and D_{sp} is the dasymetric population density of the populated class in source zone *s*.

The multi-class (polycategorical) dasymetric method captures more realistic variations within populated target zones, for example using satellite imagery to derive information about population densities. The polycategorical dasymetric model is formulated as follows: (Yuan et al., 1997; Eicher & Brewer, 2001; Mennis, 2003; Langford, 2006)

$$\hat{P}_t = \sum_{s=1}^{S} \sum_{c=1}^{C} \frac{A_{tsc}\, P_s}{A_{sc}} = \sum_{s=1}^{S} \sum_{c=1}^{C} A_{tsc}\, D_{sc}$$

where A_{tsc} is the overlapping area between target zone *t* and source zone *s,* with populated land cover with class *c* having been identified; A_{sc} is the area of source zone *s,* with populated land cover with class *c* having been identified*; S* is the total number of source zones; and D_{sc} is the dasymetric population density of the populated class *c* in source zone *s*.

While dasymetric mapping generally incorporates more spatial information than simple area weighting, the main challenges of the former are objectively establishing a relationship between ancillary data and population distribution, and quantifying the inherent spatial uncertainty in the downscaled data based on the original data (Eicher & Brewer, 2001; Langford, 2006). For instance, the relationship between population and ancillary data such as land use and street data might be unknown, and the original data from census-designated polygons also lack precision.

B. Downscaled population Dataset: AfriPop

For our study, we utilize AfriPop, a downscaled population dataset created for each respective African country at a resolution of 3 arcseconds (~100m), using publicly available source data and mapping methods that are transparent (Linard & Tatem, 2002; Linard et al., 2012). The first AfriPop dataset was created for the East African region in 2009, using datasets and methods outlined in the main paper by Tatem et al. (2007). The main ancillary layers that went into creating AfriPop include settlement data, land cover data (e.g. rivers, roads etc.) and respective national census data (Tatem et al., 2004, 2005).

C. HIV/AIDS Surveys

The Demographic and Health Surveys (DHS) program is one of the world's leading surveying agencies that supports the collection, compilation and reporting of HIV/AIDS national statistics, among other health, nutrition and population information, for over 90 countries in the world (DHS, 2013). Specifically through the AIDS indicator Survey (AIS), the DHS collects demographic and behavioral aspects of HIV prevalence among adults of reproductive age (15-49 years of age). The AIS takes about 9 months to complete and the design typically consists of biomarkers, individual and household questionnaires, and geographic information in the form of sample points that are representative of household clusters in urban and rural areas (DHS, 2013).

The geo-referenced clusters are collected during fieldwork and have an accuracy of 15-20 meters. However, due to privacy protection purposes, the household cluster points (centroids) are randomly displaced[1] such that points located in urban areas contain between 0 and 2 kilometers of error while those in rural areas contain between 0 and 5 kilometers of error, with a further 1% of the rural points achieving a maximum of 10 kilometers (DHS, 2013). However, the displacement is geographically restricted such that the cluster points remain in their respective survey regions which, for surveys conducted after 2009, translate to the country's second administrative regions that are the district level for Tanzania (DHS, 2013).

[1] Detailed Information on GPS Data Collection Methods used by DHS can be found in the following link: http://dhsprogram.com/What-We-Do/GPS-Data-Collection.cfm

D. Decision Theory

There are several sources of uncertainty in spatial data resulting from data collection, processing and usage, or a combination thereof (Hunter & Beard, 1992; Hunter and Goodchild, 1997). This inherent uncertainty in spatial data introduces the question of how the decision-making process may be influenced by the nature of spatial dataset. For more than half a century, decision-making studies have been at the core economics, business, sociology and statistics, among others. The overlying goal of this interdisciplinary sub-field is to improve decision-making efficiency by optimizing results of decisions presumably made under rational conditions. Following the traditional concept of rational behavior, the goal is the pure intent of maximizing utility hewing to certain characteristics of human behavior that comprise rationality (Savage, 1954). However, studies have shown that rationality is instead bounded by the complexity of a certain decision, which in turn is partly determined by the uncertainty of the outcomes (Simon, 1955). In other words, rationality is determined by the potential consequences or risk factors of making a decision based on a probability of occurrence, hence the Bayesian approach to estimation that is one of the decision-making criteria (Ang & Tang, 1990).

There are three other main criteria or attitudes in decision theory. The first criterion is the minimax regret criterion or 'risk-neutral' attitude in which the decision-maker makes his/her decision based on regret or opportunity loss (Savage, 1954). For example, if a decision-maker knows that the actual outcome of not acquiring enough ARV drugs would be loss of life of the prevalent population due to HIV/AIDS, the best choice under the minimax regret attitude would be to pay to acquire a moderate amount of ARV drugs since it would be the least expensive option.

The second criterion is the maximin criterion or 'risk-averse' attitude in which the decision-maker still ignores probabilities and decides that he/she wants to picks the alternative that makes the minimal cost as high as possible, i.e., the one that maximizes his/her minimum cost (Wald, 1950). Therefore, the decision-maker would take the budget-conscious approach of putting the weight of his/her decision on the worst possible outcome of overspending on ARV drugs, by choosing an action with the greatest payoff in terms of costs saved. For example, if a decision-maker considers the outcome of acquiring more than enough ARV drugs worse than failing to buy enough ARV drugs to cater for the HIV prevalent population, then the best choice under the maximin attitude would be to pay for the least possible amount of ARV drugs since it would be the least expensive option.

The third criterion is the maximax criterion or 'risk-seeking' attitude in which the decision-maker ignores probabilities and decides that he/she wants to picks the alternative that makes the maximum cost as high as possible. Therefore, the decision-maker would adopt an optimistic approach of putting the weight of his/her decision on the best possible outcome of lives saved, by choosing an action with the greatest payoff in terms of ARV drugs acquired. For example, if a decision-maker considers the outcome of failing to buy enough ARV drugs worse than the outcome of acquiring more than enough ARV drugs, then the best choice under the maximax attitude would be to pay for the highest possible amount of ARV drugs such that ideally, all the HIV prevalent people would be catered for.

III. Study Area

A. Population

The United Republic of Tanzania is a country located in Eastern Africa between Longitude 29° and 41° east and Latitude 1° and 12° south. The country consists of a Mainland and set of Islands with a total land area of nearly 900,000 square kilometers (NBS, 2012b). The 2012 Population and Housing census (PHC) was conducted on 26th August 2012 by the Tanzania National Bureau of Statistics across 169 districts and the results showed that Tanzania had a population of 44,498,923 of which 97% reside on the mainland (NBS, 2012). This was the fifth census conducted since Tanzania became a republic in 1964. The census preceding the one in 2012 was conducted in 2002. The Tanzania 2012 census data together with the geographic boundary files were obtained from the NBS website, after which population data of reproductive age (15-49 years of age) were derived using district-level demographic data (NBS, 2012). Three districts of varying geographic and demographic characteristics were selected for analysis in order to test our models under diverse population and areal contexts. The first one is *Kinondoni* district, which is the largest of the three with an area of 537 km^2 and the most population gain in the country between 2010 and 2012 (~630,000 people). The second district is *Mbeya Urban* district which is roughly half the size of *Kinondoni* district but had the least population change in Tanzania between 2010 and 2012 (~500 people). The third district is *Mjini* district which is the smallest district in Tanzania (15 km^2) but has the highest population density of the three districts.

Figure 1 shows some geographic and population characteristics of Tanzania and Figure 2 shows the three districts of interest.

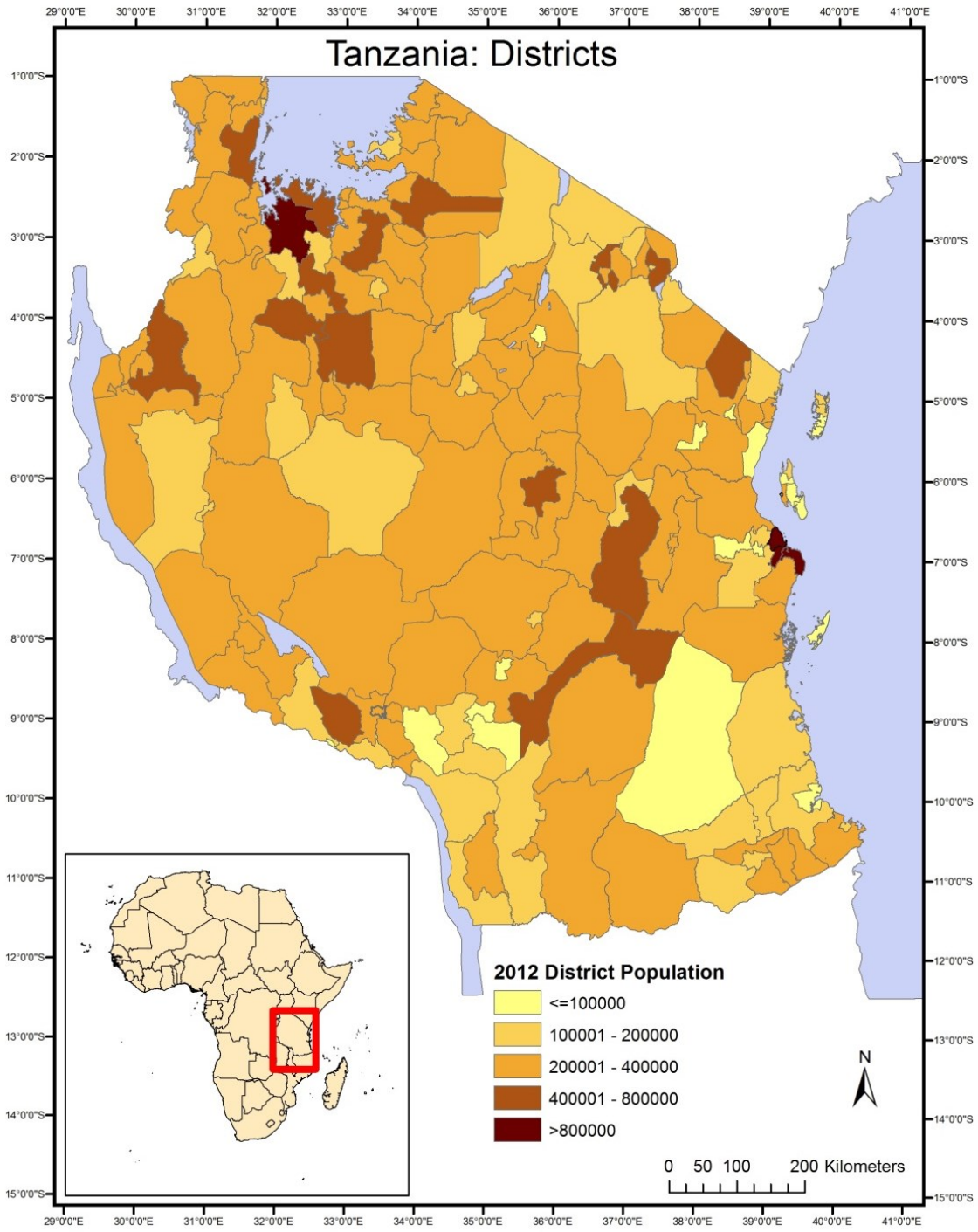

Figure 1. Geography and Population of Tanzania in 2012

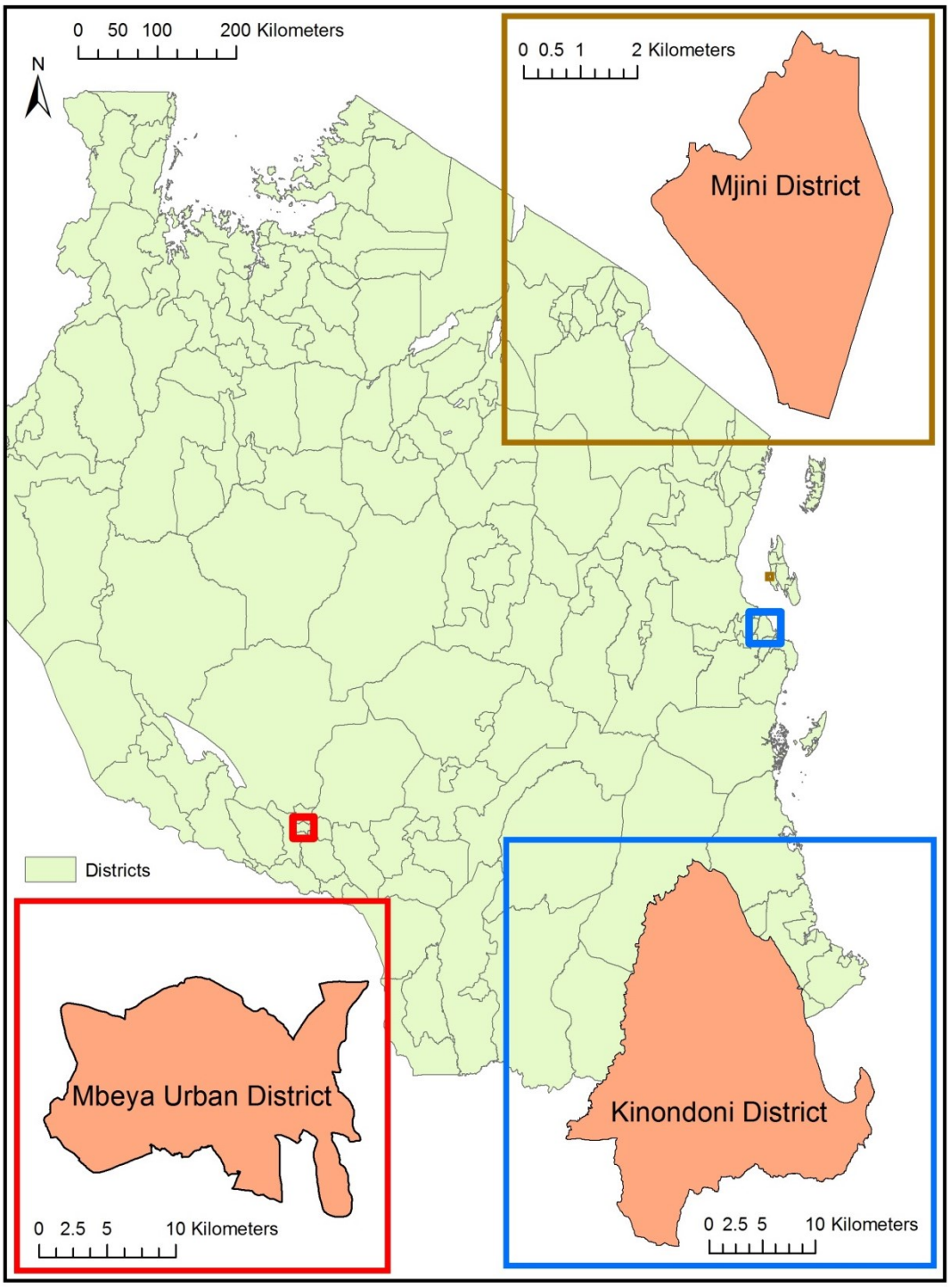

Figure 2. Study Areas

The Tanzania AfriPop 2010 dataset was obtained from the AfriPop website (AfriPop, 2013). The Tanzania AfriPop 2010 was created by downscaling 2002 national population census data to 100m grid and applying national intercensal growth rates provided by the United Nations (AfriPop, 2012) to create 2010 AfriPop national population estimates (AfriPop, 2013). We summed up the U.N. adjusted Afripop 2010 population dataset within each district to create 2010 aggregated population estimates. Figure 3 shows the AfriPop 2010 gridded population maps of Tanzania with the three districts of interest highlighted.

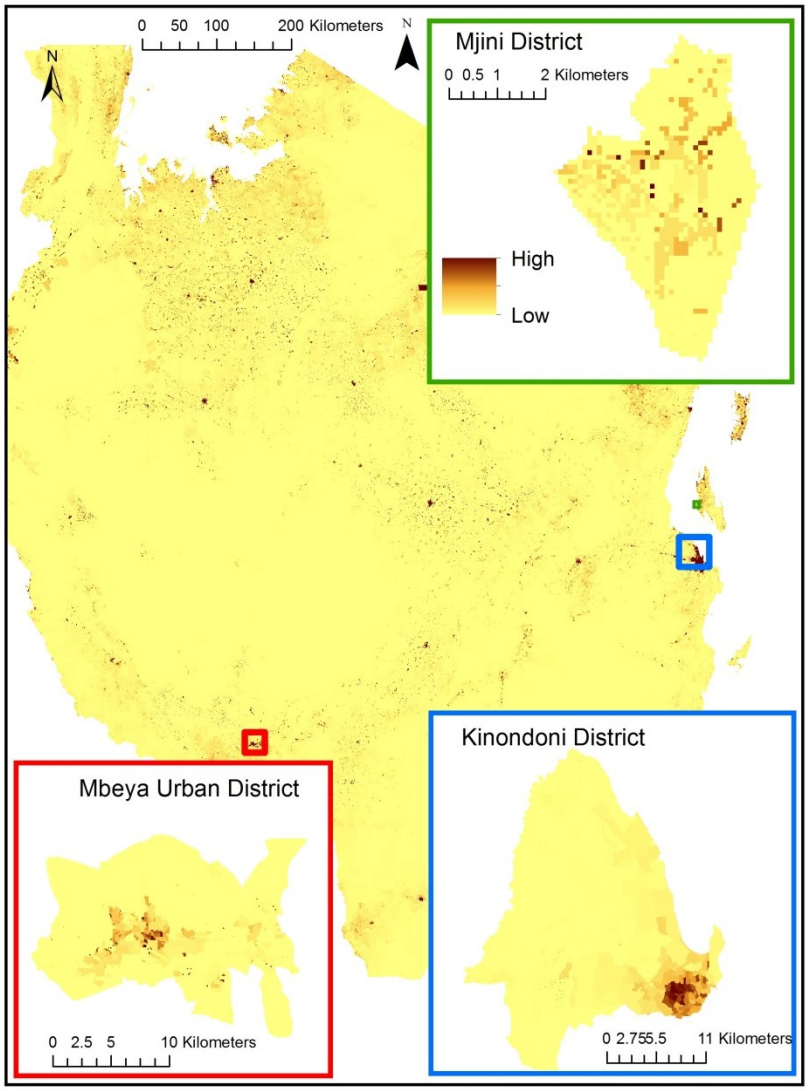

Figure 3. Tanzania AfriPop 2010 Population

B. HIV/AIDS Survey Data

The main survey used by NBS to collected recent information on HIV prevalence among Tanzania adults is collectively known as the 2011-12 Tanzania HIV/AIDS and Malaria Indicator Survey (THMIS) that is carried out as part of the DHS project in Tanzania (NBS, 2012b). The THMIS sampled 10,040 households containing 8352 men and 10,967 women between 15-49 years of age (DHS, 2013b). The response rate for men was 89% and that of women was 96% (NBS, 2012b). The 2011-12 THMIS cluster data were obtained from the DHS website (DHS 2013b). Since the district-level HIV prevalence rate data do not come readily available from the NBS website, we re-created the HIV prevalence rates at the district level using similar methodologies adopted by DHS whereby the average prevalence rate for each district was calculated in the proceeding steps. First, the total number of respondents with HIV in each household cluster was divided by the total number of respondents in each cluster and converted to a percentage value representing prevalence rate for each cluster. Finally, the average of all cluster prevalence rates within each district was obtained and that value assigned as the average HIV prevalence rate for each district. We confirmed the aforementioned district-level HIV prevalence rate by checking to see that collectively with other district level data, they averaged to the established national level of 5.1%[2] prevalence rate (TACAIDS, 2014). The population and HIV prevalence data for the three districts are summarized in table 1.

[2] Detailed Information on the Tanzania 2011-12 HIV/AIDS and Malaria Indicator Survey (THMIS) can be found in the following report: http://dhsprogram.com/pubs/pdf/SR196/SR196.pdf

District	Area (km^2)	Aggregated AfriPop (2010)	Census Population (2012)	Population Difference (2010-2012)	Population of Reproductive Age (2012) [%]	Average Prevalence Rate (%)
Kinondoni	537	1,142,170	1,775,049	+632,879	1,110,917 [62.6]	6.88
Mbeya Urban	253	305,901	305,319	-582	142,494 [46.7]	6.63
Mjini	15	243,515	223,033	-20,482	120,616 [54.1]	0.73

Table 1. Aggregate population and HIV Prevalence Information

IV. Methods

In order to investigate how the cost of ARV treatment might change with the utilization of simulated spatially distributed population estimates, we first employ the district-level aggregate census population values, the average HIV prevalence rates and the individual cost of ARV treatment to determine the aggregate ARV costs. Next, we compare the aforementioned aggregate or fixed costs to those obtained from applying spatially distributed prevalence rates to spatially distributed or gridded population values. Finally, we apply spatially distributed prevalence rates to multiple simulations of spatially distributed population estimates to obtain a distribution of varying total cost of ARV per district. The resulting implications of any change in ARV cost from the aforementioned procedures are measured by the adoption of three decision-making preferences to determine the best estimates of HIV prevalence that correspond to the least expected loss or impact in terms of costs of ARV acquisition by the decision-maker.

A. Calculating Aggregate ARV Costs from Aggregate Population Data

To create aggregate ARV costs we take the following steps. First, we adopt reported 2012 census population data from which we derive the aggregate population of reproductive age (15-49 years) at the district level. We then apply average prevalence rates to the population of reproductive age to obtain the HIV prevalence or the number of people with HIV/AIDS for each district. Finally, we apply the individual cost of ARV treatment to the HIV prevalence to obtain the aggregate total ARV cost at the district level.

Let $\bar{y}_k$ represent[3] the reported 2012 census population at the k^{th} district where, k = 1,…,K. In this study, K = 3. Furthermore, let f_k represent the proportion of census population of reproductive age (15 – 49 years) in each district k. The census population of reproductive age $\bar{z}_k$ at each district k is thus calculated as follows:

$$\bar{z}_k = f_k * \bar{y}_k$$

Let $\bar{r}_k$ represent the average HIV prevalence rate at each district k. The average prevalence ($\bar{v}_k$) is thus calculated as follows:

$$\bar{v}_k = \bar{r}_k * \bar{z}_k$$

Let c represent the proposed cost of a dose of ARV treatment for each adult individual infected with HIV ($7.30). The aggregate ARV cost ($\bar{c}_k$) for each district k is thus calculated as follows:

$$\bar{c}_k = c * \bar{v}_k$$

[3] In the methodology of this study and henceforth, we deviate from the symbols (e.g. $\hat{P}_t$) typically employed in downscaling literature since we are utilizing downscaled population values rather than conducting the population downscaling

B. Creating Spatially Distributed ARV Costs from Spatially Varying Population Data

To create spatially distributed ARV costs we take the following steps. First, we create spatially distributed or gridded population values on which we apply spatially distributed prevalence rates to determine spatially varying prevalence at the district level. We then apply the proposed cost of a dose of ARV treatment for each adult individual infected with HIV ($7.30) to the spatially varying prevalence rate to obtain the spatially varying total ARV cost at the district level.

1. Creating Spatially Distributed Population Estimates

The gridded AfriPop population estimates in 2010 are used as a surrogate for creating gridded population estimates in 2012. Let $\hat{x}_i$ represent AfriPop population gridded at the i^{th} pixel contained in district k, where $i = 1,\ldots, N_k$ pixels. A population change factor (δ_k) depicting a change of aggregate population from 2010 to 2012 is calculated using the following equation:

$$\delta_k = \frac{\bar{y}_k}{\bar{\hat{x}}_k}$$

where $\bar{\hat{x}}_k$ represents aggregate AfriPop population estimates in each district k in 2010 and is calculated as follows:

$$\bar{\hat{x}}_k = \sum_{i=1}^{N_k} \hat{x}_i$$

The population change factor (δ_k) is then applied to AfriPop gridded estimates $\hat{x}_i$ to obtain gridded population 2012 estimates ($\hat{y}_i$) as follows:

$$\hat{y}_i = \delta_k * \hat{x}_i$$

Finally, the gridded population estimates of reproductive age in 2012 ($\hat{z}_i$) are obtained using the equation below:

$$\hat{z}_i = f_k * \hat{y}_i$$

2. Creating Spatially Distributed Prevalence Values

Spatially distributed HIV prevalence values are created by disaggregating the average HIV prevalence rate to the grid level then applying the spatially varying HIV prevalence rates to each adjusted simulation of spatially varying population of reproductive age, $\hat{z}_i$. The average HIV prevalence rate for each district is disaggregated as follows. First, we create buffers around each DHS cluster centroid, specifying a radius of 2km for urban clusters and 5km for rural clusters. These specifications are consistent with the threshold of random displacement of cluster centroids by the DHS (2013b) coupled with the premise that the prevalence rate (r_d)from each household cluster (d) would be applicable to all population grids falling within each buffer. Furthermore, for population grids that fall within two or more buffers, $r_d = \max(r_d)$ and those that do not fall within any buffer, $r_d = 0$. Therefore, the spatially varying prevalence ($\hat{v}_i$) is calculated as

$$\hat{v}_i = r_d \hat{z}_i$$

The spatially varying prevalence at each pixel is then summed up to obtain the total spatially varying prevalence ($\bar{\hat{v}}_k$) as follows:

$$\bar{\hat{v}}_k = \sum_{i=1}^{N_k} \hat{v}_i$$

3. Calculating Spatially Distributed Total ARV Costs

The spatially varying prevalence ($\hat{v}_i$) is multiplied by c to obtain the spatially varying cost ($\hat{c}_i$) of ARV treatment at the pixel level, as follows:

$$\hat{c}_i = c * \hat{v}_i$$

Consequently, the ARV cost ($\hat{c}_i$) at each pixel is then summed up to obtain the total spatially varying cost ($\bar{\hat{c}}_k$) as follows:

$$\bar{\hat{c}}_k = \sum_{i=1}^{N_k} \hat{c}_i$$

C. Creating Varying ARV Costs from Simulated Spatially Varying Population Data

To create varying ARV costs we take the following steps. First, we conduct multiple non-homogenous Poisson simulations of gridded population of reproductive age (z_i) and then we sum up z_i for each realization. Next, we adjust each realization by increasing and reducing population values of candidate pixels such that when summed up, they match the estimated census 'target' population of reproductive age ($\bar{z}_k$) at the district level. We then apply spatially varying prevalence rates (r_d) to each realization of adjusted population of reproductive age at the pixel level then summing them up to obtain a distribution of varying aggregate prevalence for each district. Finally, we apply c to the varying prevalence to obtain varying total ARV cost at the district level.

1. Non-homogenous Poisson Simulation of Population of Reproductive Age

We employ the Poisson process to simulate multiple realizations of the gridded population of reproductive age since the latter is considered an outcome of a counting process depicting a population that has occurred in a particular space and time. Specifically, the non-homogenous Poisson process is used to simulate 100 independent realizations of the gridded population of reproductive age (z_i) for each district, based on the presumption that each $\hat{z}_i$ represents the respective mean observed or expected value (λ_i) and is also independent of each other within a single realization. We choose to conduct 100 realizations, a number that is computationally efficient for our purposes. The non-homogenous Poisson simulation of gridded population of reproductive age (z_i) is represented as

$$z_i \sim Pois\ (\hat{z}_i)$$

with a cumulative distribution function (CDF) that can be expressed as follows:

$$F\ (z_i|\ \lambda_i) = \ e^{-\lambda_i} \frac{{\lambda_i}^{z_i}}{(z_i)!}$$

where λ_i represents the local intensity of the Poisson distribution (Rubinstein & Kroese, 2007). In other words, λ_i is the mean observed or expected value at each pixel and is equivalent to $\hat{z}_i$ in this study.

2. Adjustment of Simulated Population of Reproductive Age

The sum of each realization from the non-homogenous Poisson simulation is calculated as

$$\bar{z}_k = \sum_{i=1}^{N_k} z_i$$

where $\bar{z}_k$ represents the sum of each Poisson realization of gridded population of reproductive age. The gridded population realizations ($\bar{z}_k$) are then adjusted in order to match the spatial pattern of the AfriPop dataset while summing up to the estimated census 'target' population of reproductive age ($\bar{z}_k{}^t$) that is previously defined as the aggregate census population $\bar{z}_k$ for each district k.

$$\bar{z}_k{}^t = f_k * \bar{y}_k$$

We set up the spatial population adjustment algorithm to make changes to realizations in which $\bar{z}_k \neq \bar{z}_k{}^t$. For realizations in which $\bar{z}_k > \bar{z}_k{}^t$, a threshold ϕ is selected as the median of z_i for which changes would only be made randomly to pixels for which $z_i > \phi$. Consequently, any randomly selected pixels would have their respective populations decreased by a randomly selected net migration rate ϖ where $m_l \leq \varpi \leq m_h$ until $\bar{z}_k = \bar{z}_k{}^t$. Net migration rate refers to the difference between the number of people entering and leaving a specific country over a given period of time that is typically a year (CIA, 2012b). If more people leave the country than enters it over a specific time frame, then the process is known as a net emigration. In our study, the net migration rate m_l represents a "low" national net emigration rate of 0.29 migrants per 1000 people as estimated in 2012 (CIA, 2012b) and the net migration rate m_h represents a "high" national net emigration rate of 0.6 migrants per 1000 people as estimated in 2013 (IOM, 2014). Each randomly selected net

migration rate ϖ is converted into a ratio on a 0-1 scale and used to decrease z_i by a factor of $(1 - \varpi)$ as follows:

$$z_i{}' = (1 - \varpi) * (z_i)$$

In the equation above, $z_i{}'$ is the adjusted simulated gridded population of reproductive age z_i. During the adjustment process, adjusted simulated values $z_i{}'$ replace the original z_i and the adjustment is considered complete when

$$\sum_{i=1}^{N_k} z_i = \bar{z}_k = \bar{z}_k{}^t$$

The algorithm then checks if there is any adjusted realization in which

$$\sum_{i=1}^{N_k} z_i < \bar{z}_k{}^t$$

If found, the algorithm increases $\sum_{i=1}^{N_k} z_i$ by $[\bar{z}_k{}^t - \bar{z}_k]$ to match $\bar{z}_k{}^t$.

For realizations in which $\bar{z}_k < \bar{z}_k{}^t$, a similar threshold ϕ was selected as the median of z_i for which changes would only be made randomly to pixels in which $z_i > \phi_i$. Consequently, any randomly selected pixels would have their respective populations increased by a randomly selected urbanization rate ψ where $b_l \leq \psi \leq b_h$ until $\bar{z}_k = \bar{z}_k{}^t$. Urbanization rate refers to the average rate of growth of the size of an urban population over a specific period in time for a given country (UN Population Division, 2013). The urbanization rate b_l represents a "low" annual urbanization rate of 4.2%, estimated for the period between 2005-2010 (UN Population Division, 2013) and the urbanization rate b_h represents a "high" annual rate of 4.77%, estimated for the period between 2010-2015 (CIA, 2012c).

Each randomly selected urbanization rate ψ is converted into a ratio on a 0-1 scale and used to increase z_i by a factor of $(1 + \psi)$ as follows:

$$z_i' = (1 + \psi) * (z_i)$$

During the adjustment process, adjusted simulated values z_i' replace the original z_i and the adjustment is considered complete when

$$\sum_{i=1}^{N_k} z_i = \bar{z}_k = \bar{z}_k{}^t$$

The algorithm then checks if there is any adjusted realization in which

$$\sum_{i=1}^{N_k} z_i > \bar{z}_k{}^t$$

If found, the algorithm decreases $\sum_{i=1}^{N_k} z_i$ by $[\bar{z}_k{}^t - \bar{z}_k]$ to match $\bar{z}_k{}^t$.

The adjustment of simulated population is summarized in Figure 4 and 5.

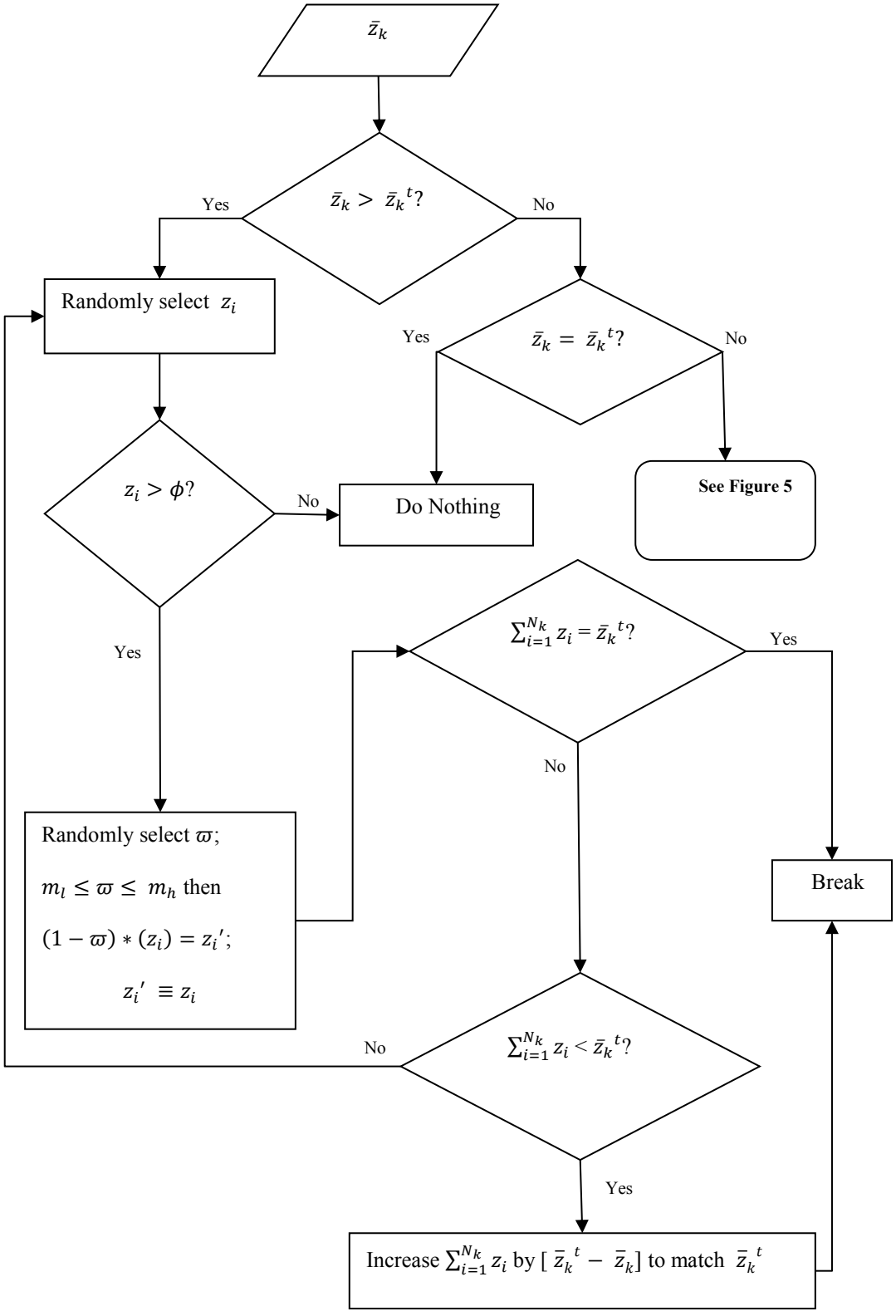

Figure 4. Spatial Adjustment Flowchart for Decreasing Population

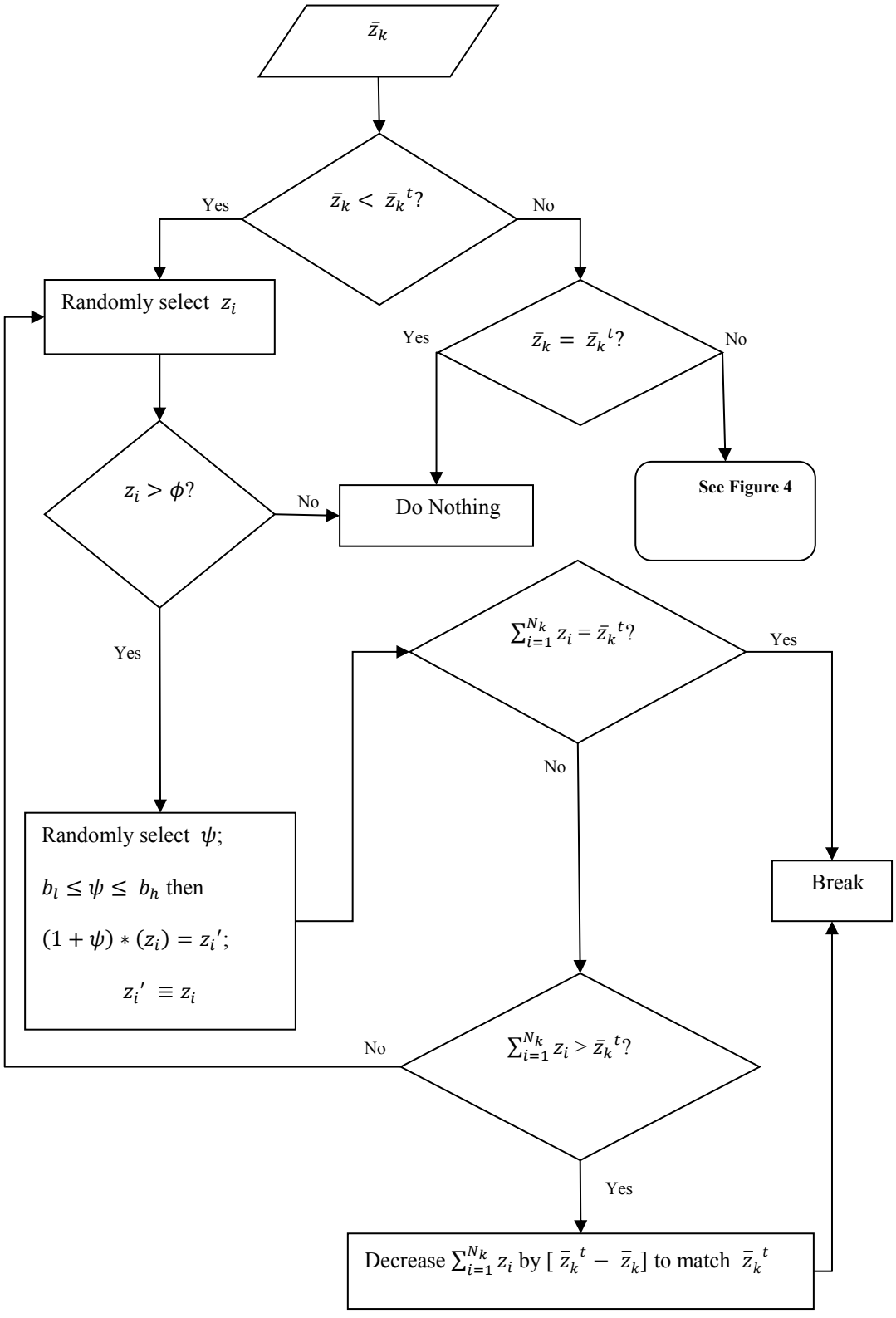

Figure 5. Spatial Adjustment Flowchart for Increasing Population

3. Calculating Varying Prevalence and ARV Cost from Simulated Population of Reproductive Age

The gridded prevalence values (v_i) for each district are obtained by applying spatially varying prevalence rates (r_d) to each realization of adjusted population of reproductive age (z_i) at the pixel level, as follows:

$$v_i = r_d z_i$$

These gridded prevalence values v_i are then summed up across all realizations to obtain a distribution of varying aggregate prevalence values ($\tilde{v}_k$) for each district *k*, as follows:

$$\tilde{v}_k = \sum_{i=1}^{N_k} v_i$$

Finally, we apply c to the varying aggregate prevalence values $\tilde{v}_k$ to obtain varying total ARV cost $\tilde{c}_k$ at the district level, as follows:

$$\tilde{c}_k = c * \tilde{v}_k$$

D. Estimation of Utility Functions for Antiretroviral Costs

The estimation of utility or loss functions for ARV cost by a decision-maker is based on the adjusted gridded population of reproductive age (z_i) established after the non-homogenous Poisson simulation process and consequently used to calculate $\tilde{c}_k$. Assuming that a decision-maker has a particular choice of $\tilde{c}_k$ and that the remaining values could be possible 'actual' values $\widehat{\tilde{c}_k}$ this would imply an estimation error which would be depicted by a loss (Ang & Tang, 1990). The goal for a decision-maker then would be to minimize the expected loss associated with an estimation error which is in turn informed by the decision-maker's risk attitude. In our study, $\tilde{c}_k$ is the estimator of a cost parameter whose 'actual'

value is dictated by the distribution $f(\hat{\tilde{c}}_k)$. Therefore, the expected loss resulting from the error in estimation is calculated as

$$L = \int_{-\infty}^{\infty} g(\hat{\tilde{c}}_k, \tilde{c}_k) f(\hat{\tilde{c}}_k) d\hat{\tilde{c}}_k$$

where $g(\hat{\tilde{c}}_k, \tilde{c}_k)$ is the loss function from the estimation error that is $(\hat{\tilde{c}}_k - \tilde{c}_k)$. For the estimator $\tilde{c}_k$ to minimize the expected loss, it should satisfy the following equation:

$$\frac{dL}{d\tilde{c}_k} = \frac{d}{d\tilde{c}_k}\left[\int_{-\infty}^{\infty} g(\hat{\tilde{c}}_k, \tilde{c}_k) f(\hat{\tilde{c}}_k) d\hat{\tilde{c}}_k\right] = 0$$

1. Testing Cost Scenarios in Varying Risk Tolerance

Few studies have suggested loss functions for minimizing ARV cost. One study explored various algorithms for minimizing and maximizing non-linear loss functions for ARV cost and efficacy, respectively (Oyelami & Ogidi, 2013). However, the aforementioned study was conducted from a theoretical perspective without use of empirical HIV prevalence and ARV cost data. With little reference to inform our loss functions, we simply assume the loss function $g(\hat{\tilde{c}}_k, \tilde{c}_k)$ to be linear and is formulated as

$$g(\hat{\tilde{c}}_k, \tilde{c}_k) = \begin{cases} c(\hat{\tilde{c}}_k - \tilde{c}_k); & \hat{\tilde{c}}_k \leq \tilde{c}_k \\ c(\hat{\tilde{c}}_k - \tilde{c}_k); & \hat{\tilde{c}}_k > \tilde{c}_k \end{cases}$$

where c is a previously defined cost constant that could be preceded with a variable that would depend on the approach to decision making.

In a minimax regret attitude, the assumption is that the risk-neutral decision-maker would consider both overestimation and underestimation of HIV prevalence as having similar impact (cost) and as such would adopt the following symmetrical utility function.

$$g(\hat{\tilde{c}}_k, \tilde{c}_k) = \begin{cases} c(\hat{\tilde{c}}_k - \tilde{c}_k); & \hat{\tilde{c}}_k \leq \tilde{c}_k \\ c(\hat{\tilde{c}}_k - \tilde{c}_k); & \hat{\tilde{c}}_k > \tilde{c}_k \end{cases}$$

In a maximin attitude, the assumption is that the budget-conscious decision-maker would be hesitant to spend more money on ARV acquisition than they need to and therefore would not mind catering to a lower HIV prevalence than the risk-neutral approach. To achieve such an outcome, the decision maker would most likely consider overestimation of HIV prevalence as having a greater impact (cost) than underestimation.

For our study, we defined overestimation of HIV prevalence as having 10 times of an impact (cost) than underestimation, as shown in the following asymmetrical utility function.

$$g(\hat{\tilde{c}}_k, \tilde{c}_k) = \begin{cases} c(\hat{\tilde{c}}_k - \tilde{c}_k); & \hat{\tilde{c}}_k \leq \tilde{c}_k \\ 10(\hat{\tilde{c}}_k - \tilde{c}_k); & \hat{\tilde{c}}_k > \tilde{c}_k \end{cases}$$

The coefficient value of 10 was chosen in order to create an asymmetrical linear loss function that presents a significant distinction between minimized values obtained from overestimating and underestimating population of reproductive age.

In a *Maximax* attitude, the assumption is that the 'risk-seeking' decision-maker would be willing to spend more money on ARV acquisition and potentially benefit a higher HIV prevalence than the risk-neutral approach. To achieve such an outcome, the decision maker would most likely consider underestimation of HIV prevalence as having a greater impact (cost) than overestimation. For our study, we defined underestimation of HIV prevalence as having 10 times of an impact (cost) than overestimation, as shown in the following asymmetrical utility function.

$$g(\hat{\tilde{c}}_k, \tilde{c}_k) = \begin{cases} 10c(\hat{\tilde{c}}_k - \tilde{c}_k); & \hat{\tilde{c}}_k \leq \tilde{c}_k \\ c(\hat{\tilde{c}}_k - \tilde{c}_k); & \hat{\tilde{c}}_k > \tilde{c}_k \end{cases}$$

In the aforementioned three decision-making attitudes, we minimized the loss functions to obtain the best HIV prevalence estimates and associated least expected impact (loss) in terms of the cost of ARV acquisition for each district.

V. Results

A. Aggregate ARV Costs from Aggregate Population Data

Aggregate ARV costs $(\bar{c}_k)$ are calculated from aggregate census population $(\bar{y}_k)$ using procedures described in part A of the *Methods* section. The results are shown in table 2.

District (K)	Census Population $(\bar{y}_k)$	Proportion of Reproductive Age (f_k)	Census Population of Reproductive Age $(\bar{z}_k)$	Aggregate HIV prevalence rate $(\bar{r}_k)$ [%]	Average HIV prevalence $(\bar{v}_k)$	Aggregate ARV cost $(\bar{c}_k)$ [$]
Kinondoni	1,775,049	0.63	1,110,917	6.88	76,431	557,947
Mbeya Urban	305,319	0.47	142,494	6.63	9,447	68,963
Mjini	223,033	0.54	120,616	0.73	881	6,431

Table 2. Aggregate Population and ARV Cost

B. Spatially Distributed ARV Costs from Spatially Varying Population Data

Empirical census population was discretized in the procedures outlined in part B of the *Methods* section for the respective number of grids in each district, as specified by the "surrogate" AfriPop population data. The discretized population of reproductive age ($\hat{z}_i$) values were obtained by applying the population change factor (δ_k) to the AfriPop gridded estimates ($\hat{x}_i$) to obtain gridded census population ($\hat{y}_i$), then applying the proportion of population of reproductive age (f_k) in each district *k*. The results of the discretization process are shown in table 3.

District (K)	Population Change Factor (δ_k)	Proportion of Reproductive Age (f_k)	Sum of Gridded Census Population ($\bar{y}_k$)	Sum of Gridded Census Population of Reproductive Age ($\bar{z}_k{}^t$)	Number of discretized points (N)
Kinondoni	1.03	0.63	1,775,049	1,110,917	53,332
Mbeya Urban	0.99	0.47	305,319	142,494	25,257
Mjini	0.92	0.54	223,033	120,616	1,511

Table 3. Results of Population Discretization

The discretized population of reproductive age in *Kinondoni* district is shown in figure 6. The legend limits have been re-scaled to a maximum of 500 to best show the pixel-level population since most of the pixels contain relatively few people, as shown in the left histogram in figure 7. The right histogram in figure 7 shows the same population distribution capped at 100 people.

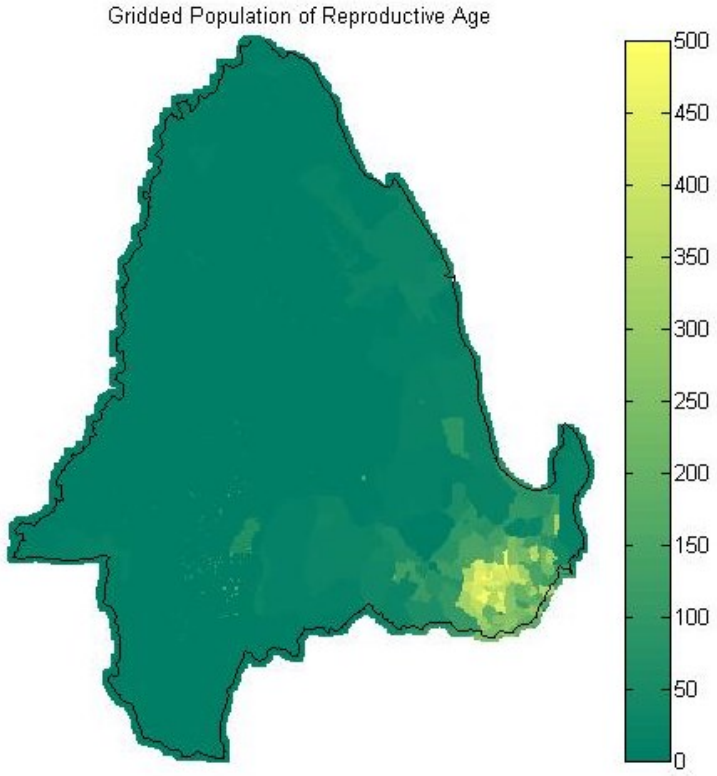

Figure 6. Discretized Population of Reproductive Age in Kinondoni District

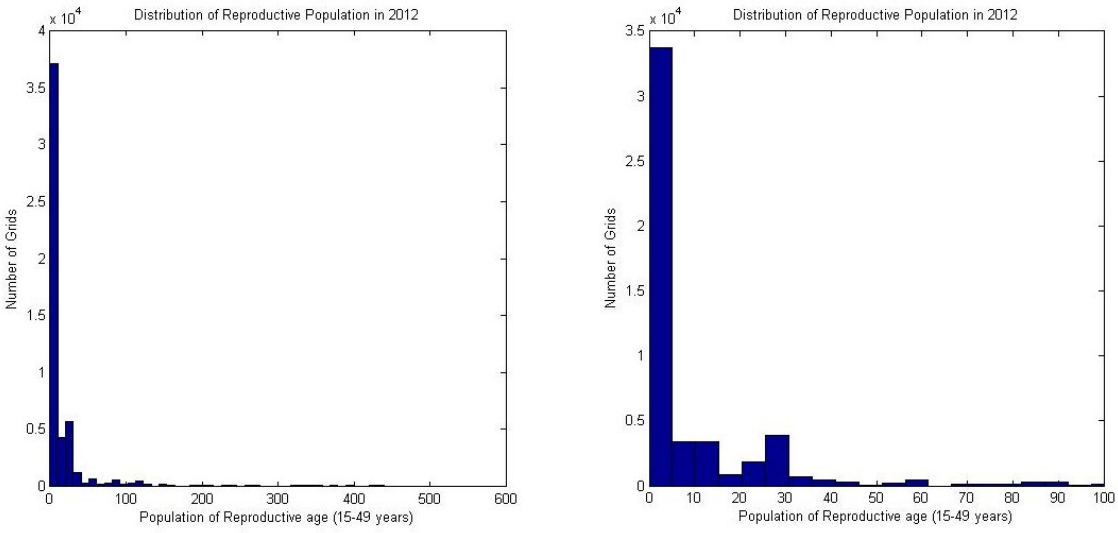

Figure 7. Histogram of Discretized Population of Reproductive Age in Kinondoni District

The discretized population of reproductive age in *Mbeya Urban* district is shown in figure 8. The legend limits have also been scaled to a maximum of 50 to best show the pixel-level population variation since most of the pixels contain relatively few people, as shown the left histogram in figure 9. The right histogram in figure 9 shows the same population distribution capped at 100 people.

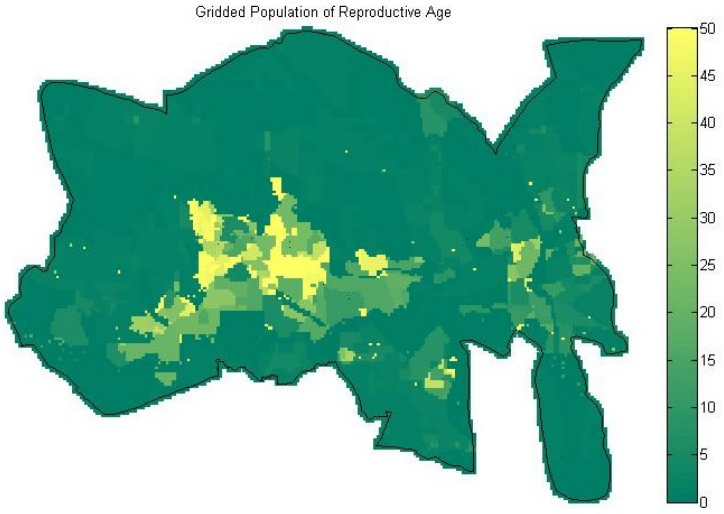

Figure 8. Discretized Population of Reproductive Age in Mbeya Urban District

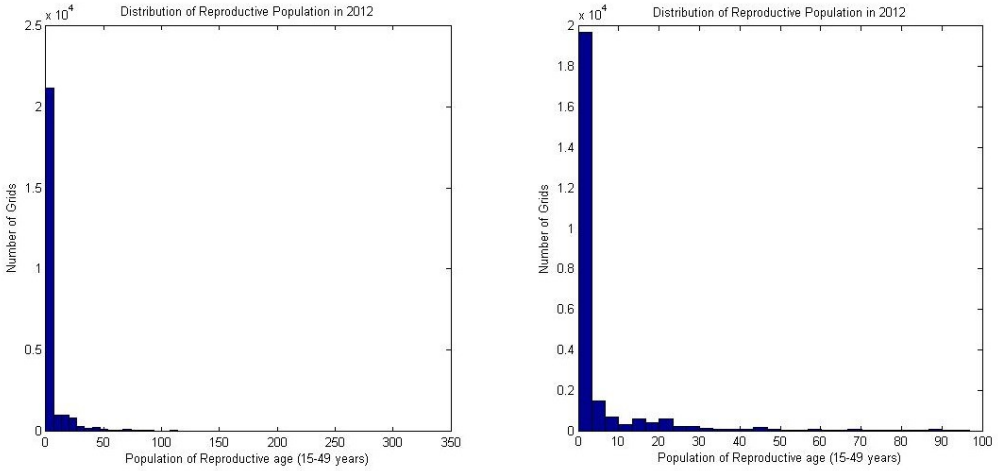

Figure 9. Histogram of Discretized Population of Reproductive Age in Mbeya Urban District

The discretized population of reproductive age in *Mjini* district is shown in figure 10. The legend limits have also been scaled to a maximum of 500 to best show the pixel-level population variation since most of the pixels contain less than 200 people, as shown the left histogram in figure 11. The right histogram in figure 11 shows the same population distribution capped at 200 people.

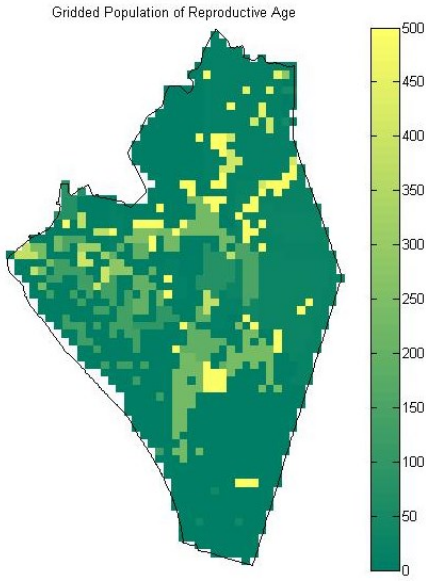

Figure 10. Discretized Population of Reproductive Age in Mjini District

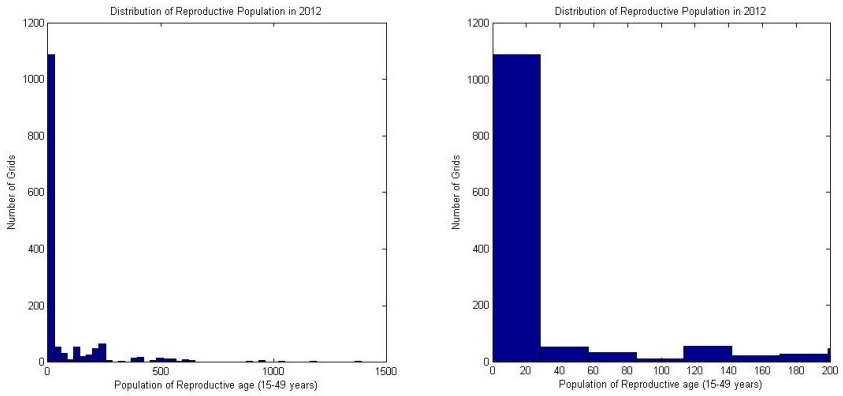

Figure 11. Histogram of Discretized Population of Reproductive Age in Mjini District

Figure 12 – 14 show the spatially varying prevalence rates (r_d) for each district that were determined using steps outlined in part B2 in the *Methods* section.

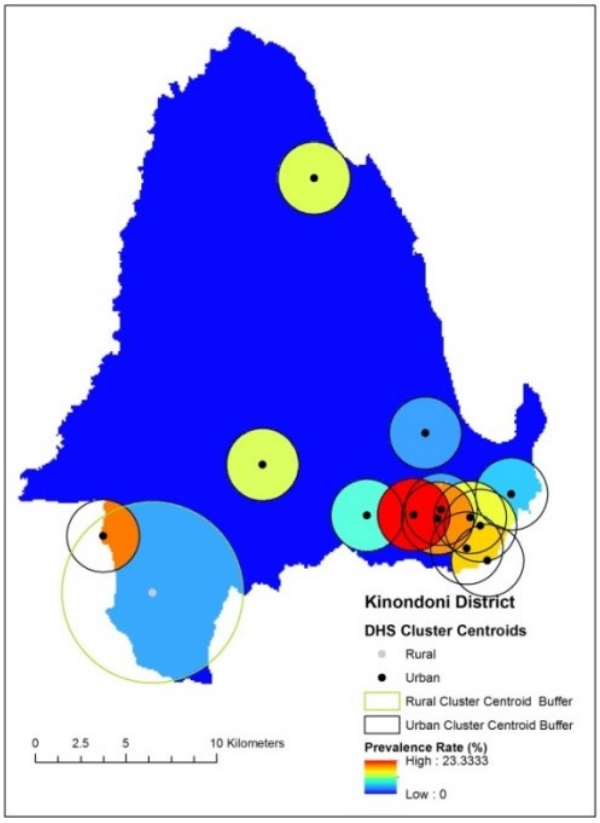

Figure 12. Map of Spatially Varying HIV Prevalence Rates in Kinondoni District

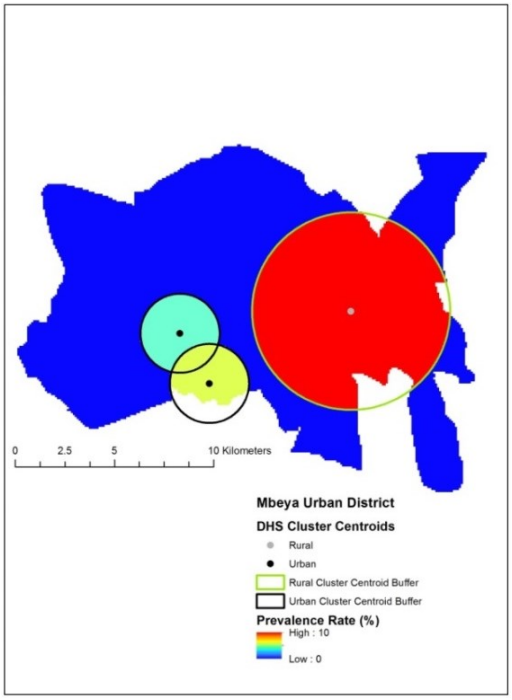

Figure 13. Map of Spatially Varying HIV Prevalence Rates in Mbeya Urban District

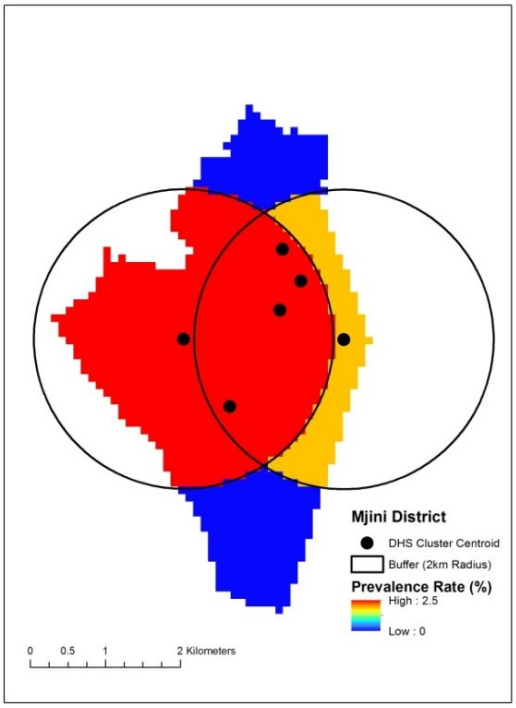

Figure 14. Map of Spatially Varying HIV Prevalence Rates in Mjini District

Spatially distributed total ARV costs ($\bar{\hat{c}}_k$) are calculated from the spatially varying gridded population estimates ($\hat{z}_i$) and spatially varying prevalence rates (r_d)using procedures described in part B2 and B3 in the *Methods* section. The results are shown in table 4.

District (K)	Sum of Gridded Census Population of Reproductive Age ($\bar{z}_k{}^t$)	Average of Spatially Varying prevalence rate (r_d/N) [%]	Sum of Spatially Varying prevalence ($\bar{\hat{v}}_k$)	Aggregate of Spatially Varying ARV cost ($\bar{\hat{c}}_k$) [$]
Kinondoni	1,110,917	1.65	18,330	133,810
Mbeya Urban	142,494	3.29	4,688	34,223
Mjini	120,616	1.78	2,147	15,673

Table 4. Spatially Varying Population Estimates and ARV Cost

C. Varying ARV Costs from Simulated Spatially Varying Population Data

1. Poisson Simulation of Population of Reproductive Age

For each district, the non-homogenous Poisson process was used to simulate 100 realizations of the gridded census population of reproductive age (z_i). Figure 15 shows a graph of 100 realizations of population values from the non-homogenous Poisson simulation of gridded census population of reproductive age in *Kinondoni* District. Population maps of some realizations are shown in figure 16.

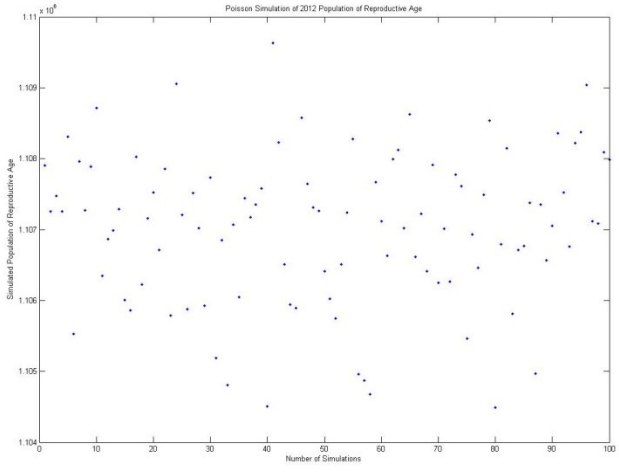

Figure 15. 100 Realizations of Gridded Census Population of Reproductive Age in Kinondoni District

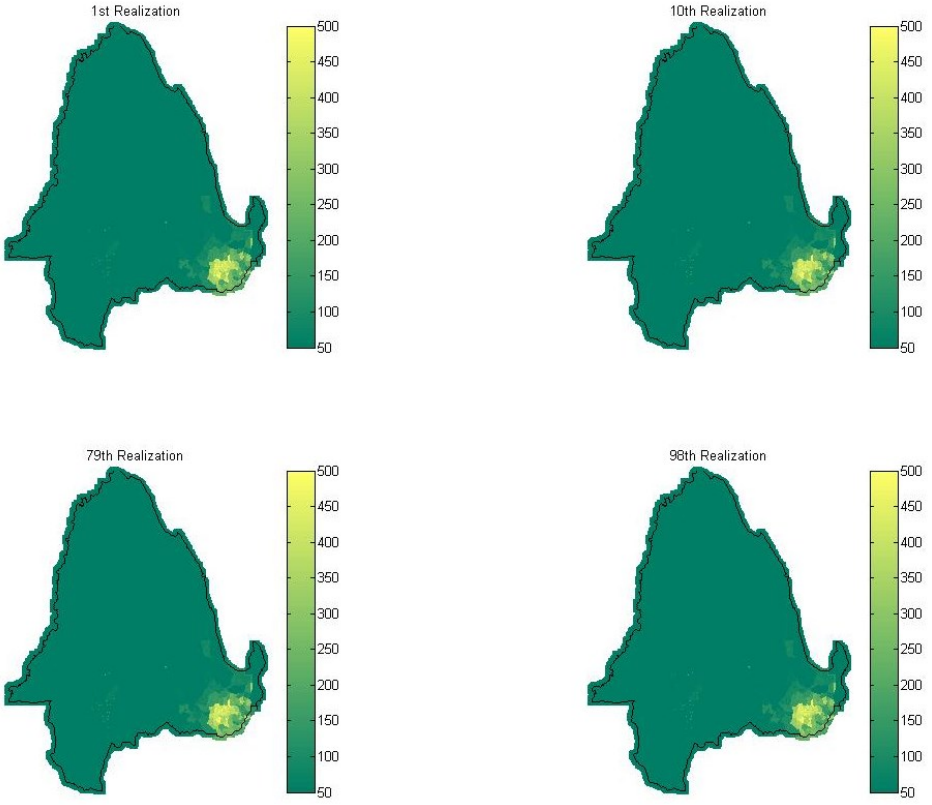

Figure 16. 1st, 10th, 79th and 98th Realizations of Simulated Gridded Census Population of Reproductive Age in Kinondoni District

Figure 17 shows a graph of 100 realizations of population values from the non-homogenous Poisson simulation of gridded census population of reproductive age in *Mbeya Urban* District. Population maps of some realizations are shown in figure 18.

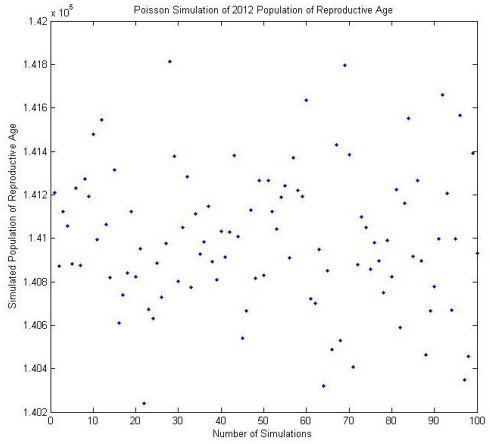

Figure 17. 100 Realizations of Gridded Census Population of Reproductive Age in Mbeya Urban District

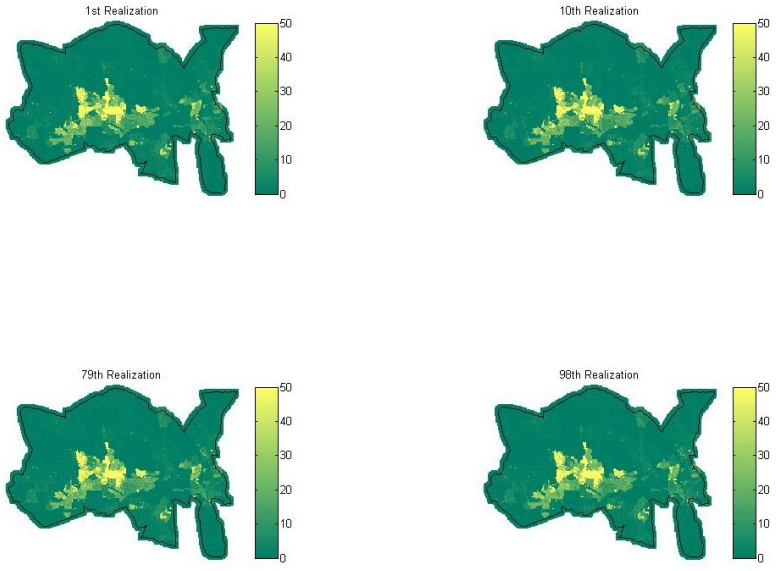

Figure 18. 1st, 10th, 79th and 98th Realizations of Simulated Gridded Census Population of Reproductive Age in Mbeya Urban District

Figure 19 shows a graph of 100 realizations of population values from the non-homogenous Poisson simulation of gridded census population of reproductive age in *Mjini* District. Population maps of some realizations are shown in figure 20.

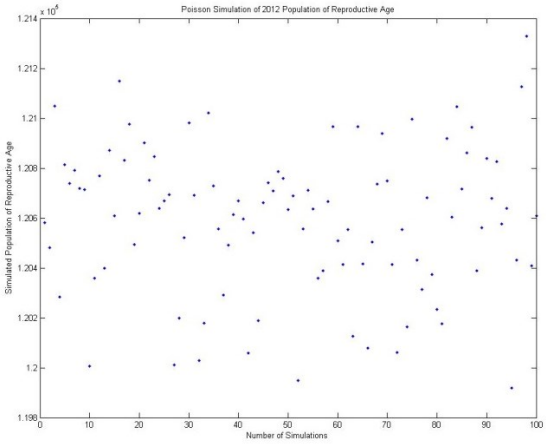

Figure 19. 100 Realizations of Gridded Census Population of Reproductive Age in Mjini District

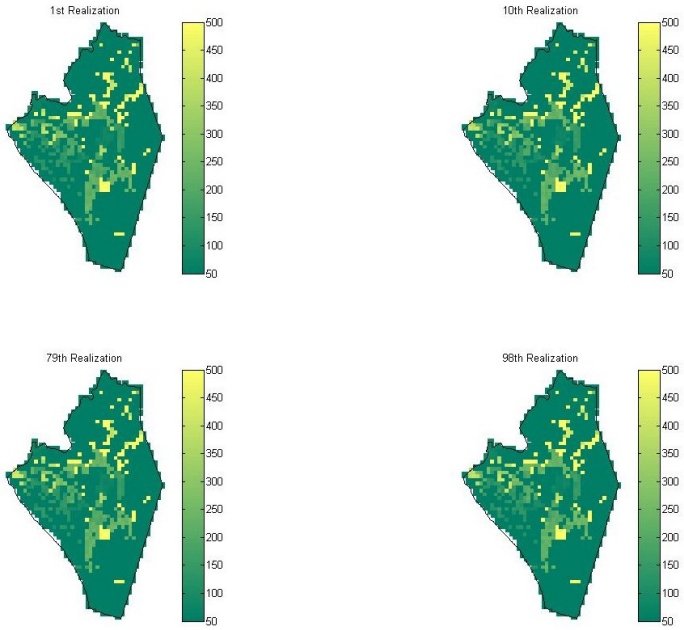

Figure 20. 1st, 10th, 79th and 98th Realizations of Simulated Gridded Census Population of Reproductive Age in Mjini District

2. Adjusted Realizations of Population of Reproductive Age

For each of the three districts, the population realizations were adjusted to spatially match the target census population of reproductive age as stipulated in the flowcharts in figure 4 and 5 (the threshold population of reproductive age (ϕ) for each district is shown in table 5 below). The population realizations were matched the target census population of reproductive age ($\bar{z}_k{}^t$).

District (K)	Threshold Population of Reproductive Age (ϕ)
Kinondoni	1
Mbeya Urban	1
Mjini	8

Table 5. Threshold Population of Reproductive Age

Figure 21 shows a graph of 100 realizations of population values from the non-homogenous Poisson simulation of discretized population in *Kinondoni* District, adjusted to the 'target' census population of reproductive age, a value of 1,110,917. Population maps of some adjusted realizations with histograms are shown in figure 22 and 23, respectively. The x-axis on the histograms in figure 23 is capped at 100 people to best display the subtleties in the pixels with low population.

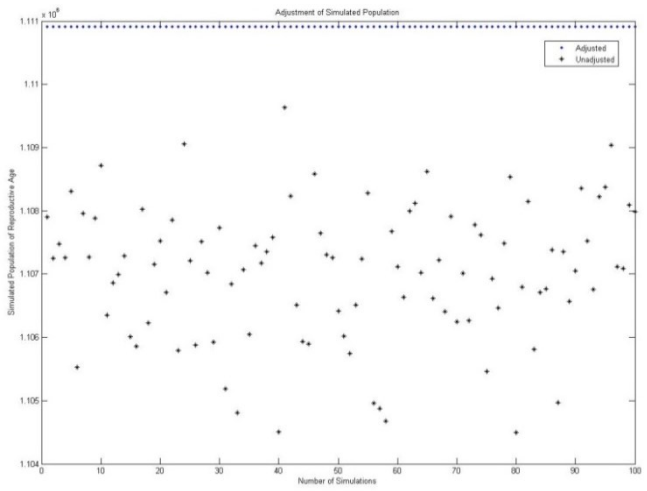

Figure 21. Adjusted Realizations of Discretized Census Population of Reproductive Age in Kinondoni District

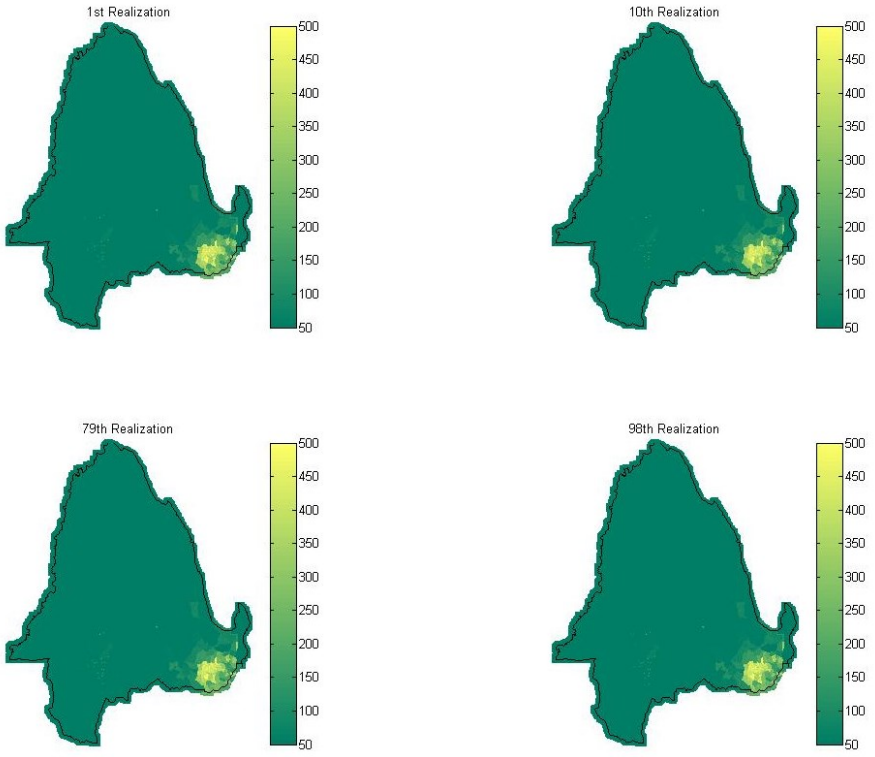

Figure 22. 1st, 10th, 79th and 98th Adjusted Realizations of Simulated Gridded Census Population of Reproductive Age in Kinondoni District

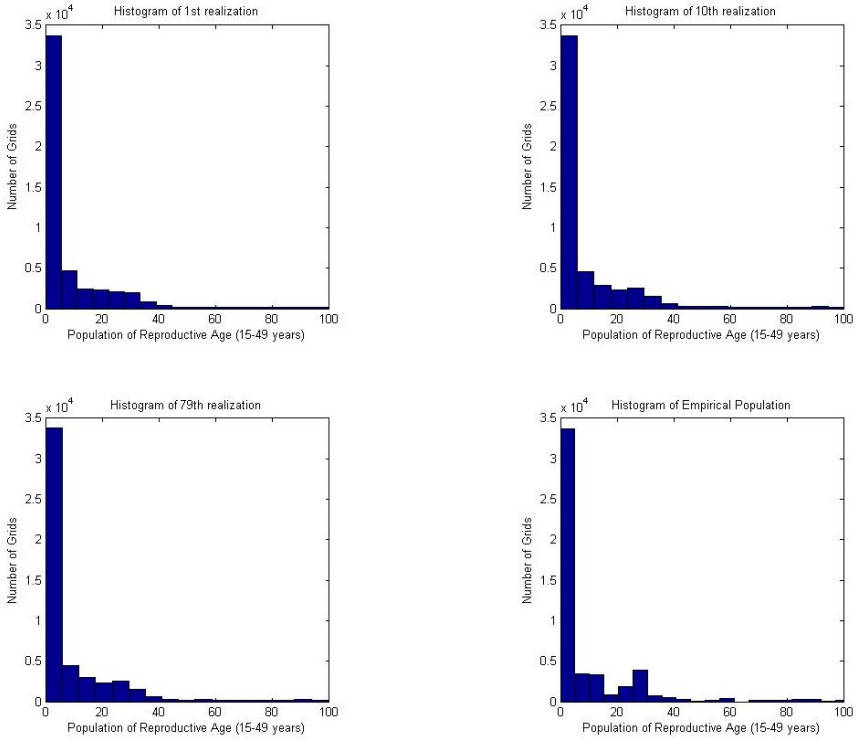

Figure 23. Histogram of 1st, 10th and 79th Adjusted Realizations with Histogram of Empirical Gridded Census Population of Reproductive Age (Bottom Right) in Kinondoni District

Figure 24 shows a graph of 100 realizations of population values from the non-homogenous Poisson simulation of discretized population in *Mbeya Urban* District, adjusted to the 'target' census population of reproductive age, a value of 142,494. Population maps of some adjusted realizations with histograms are shown in figure 25 and 26, respectively. The x-axis on the histograms in figure 26 is capped at 50 people to best display the subtleties in the pixels with low population.

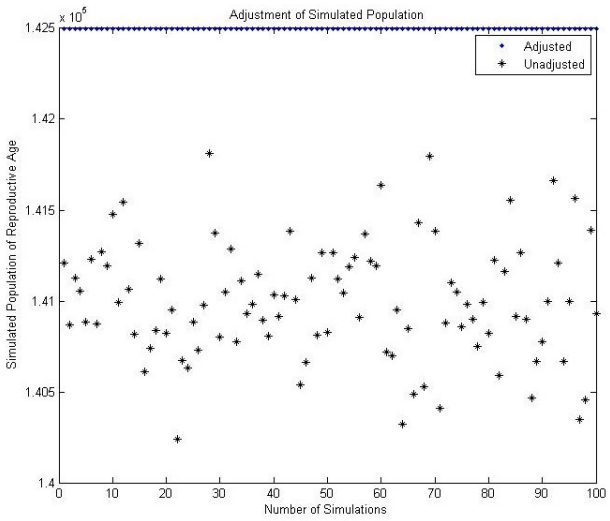

Figure 24. Adjusted Realizations of Discretized Census Population of Reproductive Age in Mbeya Urban District

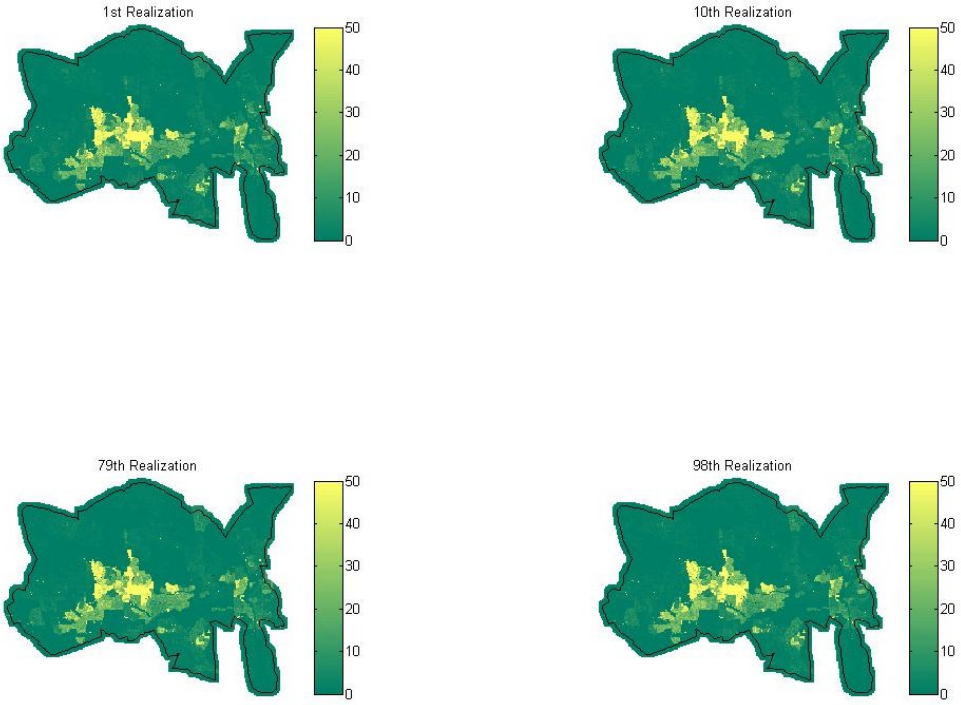

Figure 25. 1st, 10th, 79th and 98th Adjusted Realizations of Simulated Gridded Census Population of Reproductive Age in Mbeya Urban District

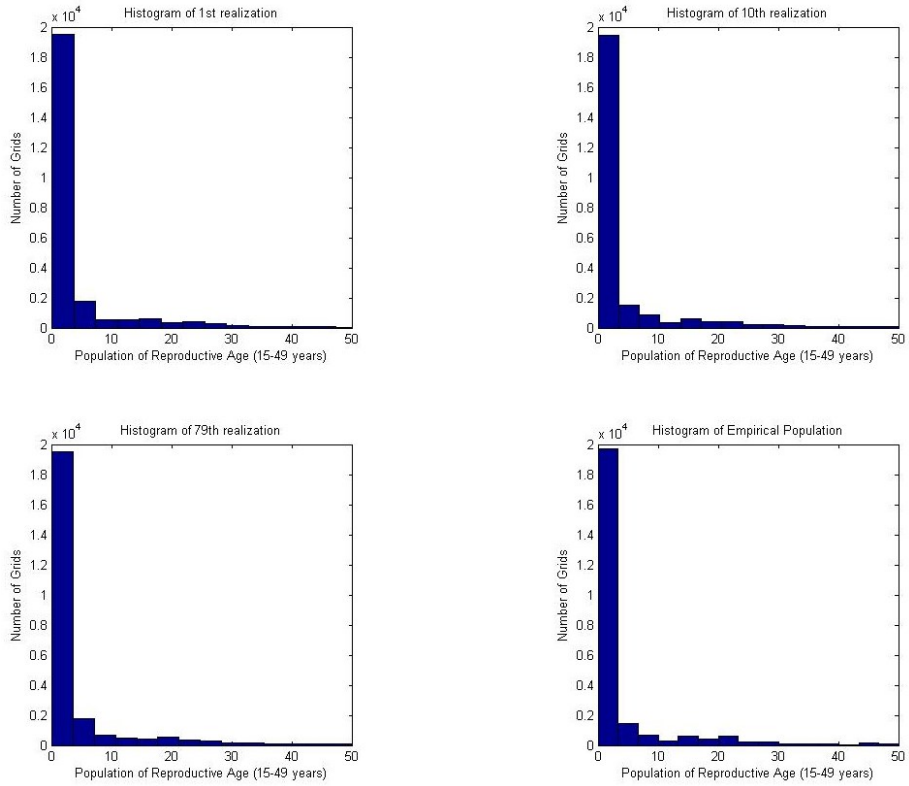

Figure 26. Histogram of 1st, 10th and 79th Adjusted Realizations with Histogram of Empirical Gridded Census Population of Reproductive Age (Bottom Right) in Mbeya Urban District

Figure 27 shows a graph of 100 realizations of population values from the non-homogenous Poisson simulation of discretized population in *Mbeya Urban* District, adjusted to the 'target' census population of reproductive age, a value of 120,616. Population maps of some adjusted realizations with histograms are shown in figure 28 and 29, respectively. The x-axis on the histograms in figure 29 is capped at 200 people to best display the subtleties in the pixels with low population.

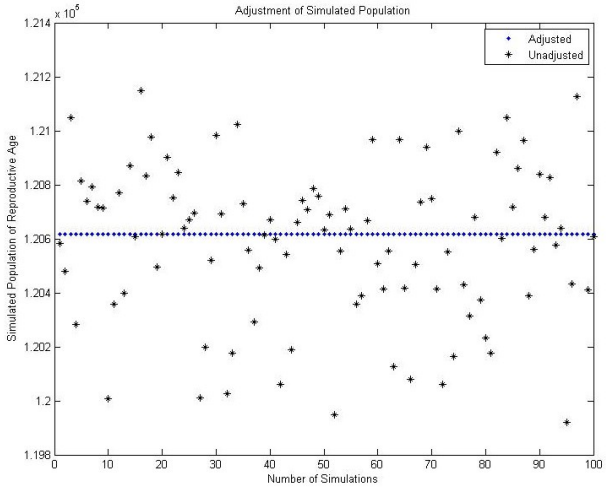

Figure 27. Adjusted Realizations of Discretized Census Population of Reproductive Age in Mjini District

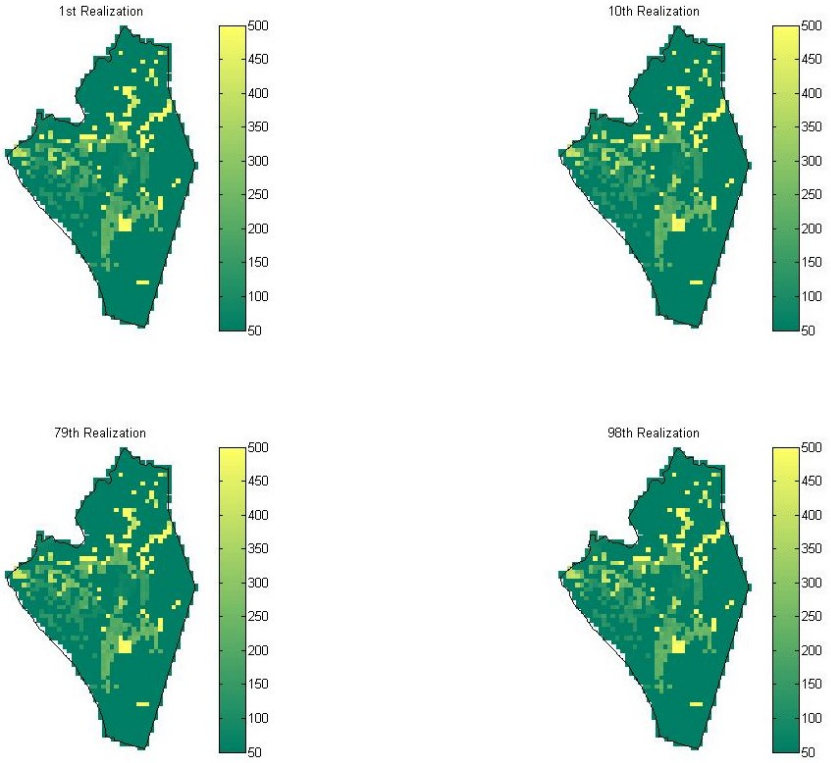

Figure 28. 1st, 10th, 79th and 98th Adjusted Realizations of Simulated Gridded Census Population of Reproductive Age in Mjini District

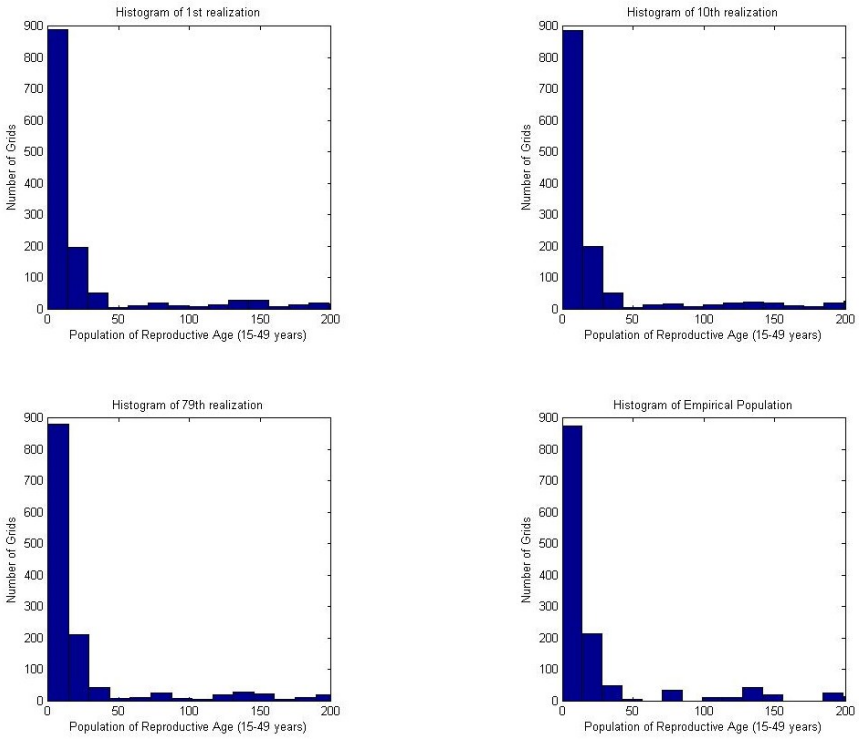

Figure 29. Histogram of 1st, 10th and 79th Adjusted Realizations with Histogram of Empirical Gridded Census Population of Reproductive Age (Bottom Right) in Mjini District

3. Varying HIV Prevalence

For each district, varying HIV prevalence values ($\tilde{v}_k$) were created by applying spatially varying HIV prevalence rates to each of the adjusted realizations of population of reproductive age. Figure 30 – 32 show the histograms varying HIV prevalence derived from adjusted realizations of gridded population of reproductive age.

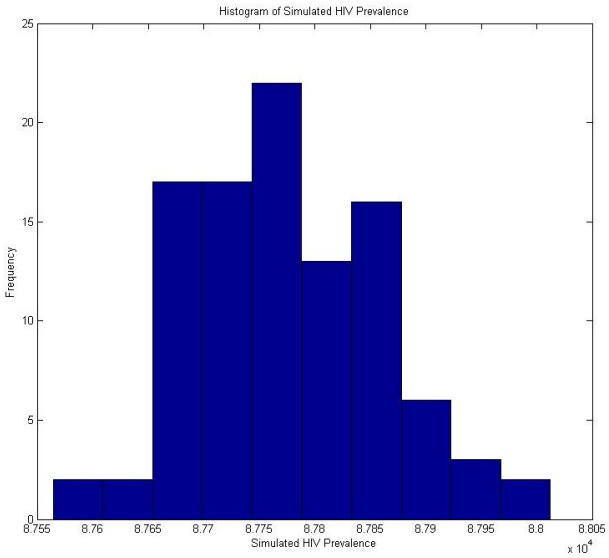

Figure 30. Histogram of Distribution of HIV Prevalence in Kinondoni District

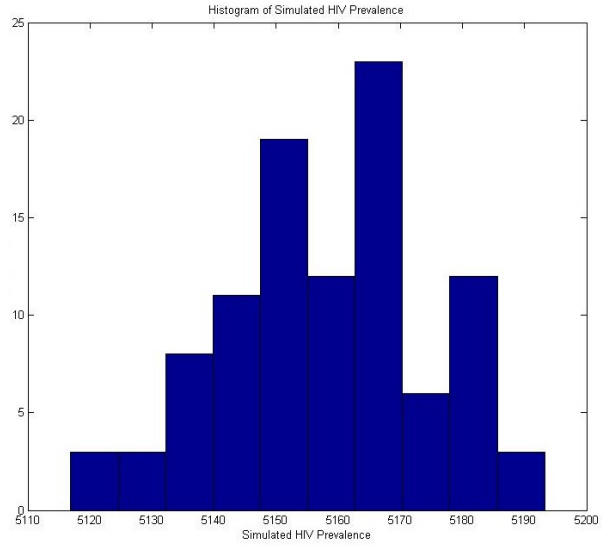

Figure 31. Histogram of Distribution of HIV Prevalence in Mbeya Urban District

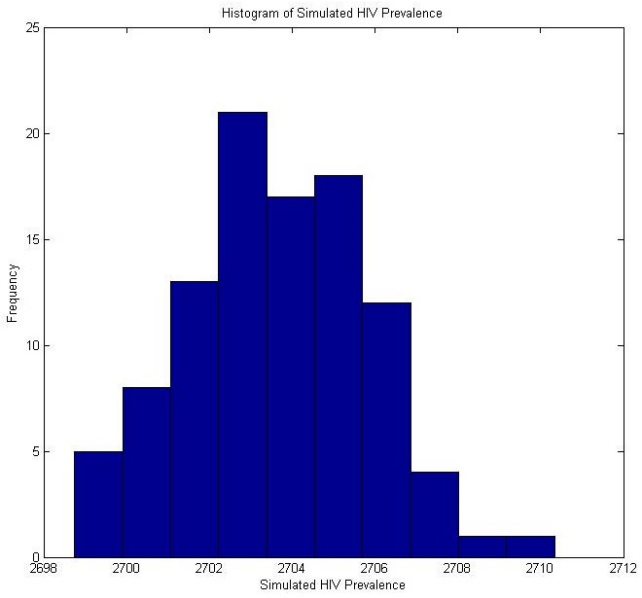

Figure 32. Histogram of Distribution of HIV Prevalence in Mjini District

Table 6 shows some statistics of the varying HIV prevalence values ($\tilde{v}_k$) for each district.

District	$\tilde{v}_k$ Maximum	$\tilde{v}_k$ Mean	$\tilde{v}_k$ Median	$\tilde{v}_k$ Minimum	$\tilde{v}_k$ Range
Kinondoni	88,012	87,776	87,774	87,564	448
Mbeya Urban	5193.2	5158.2	5157.9	5116.9	76.3
Mjini	2710.3	2703.6	2703.7	2698.7	11.6

Table 6. Summary Statistics of Varying HIV Prevalence

4. Varying ARV Costs

For each district, total cost of ARV was calculated by multiplying the cost of one dose of ARV treatment per HIV patient per month (c) with the spatially distributed HIV prevalence ($\tilde{v}_k$). Figures 33 – 35 show the varying total cost of ARV for each of the districts.

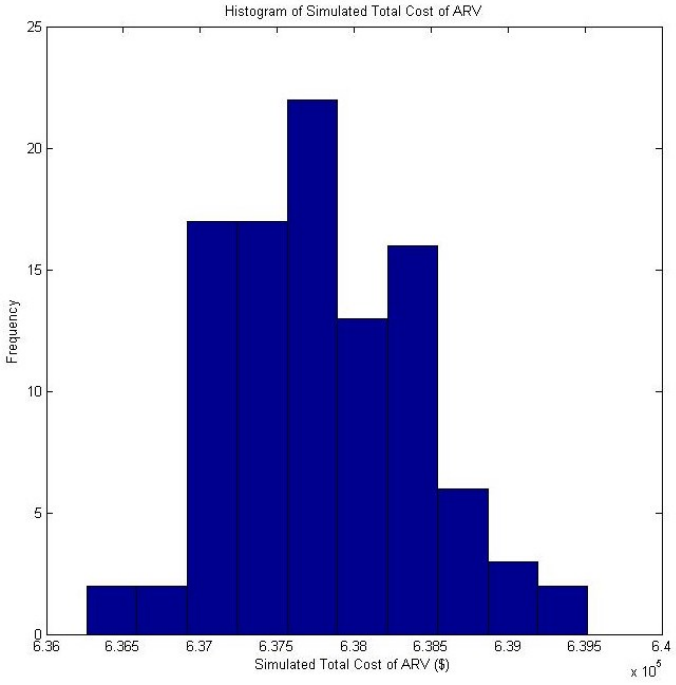

Figure 33. Histogram of Distribution of Total Cost of ARV in Kinondoni District

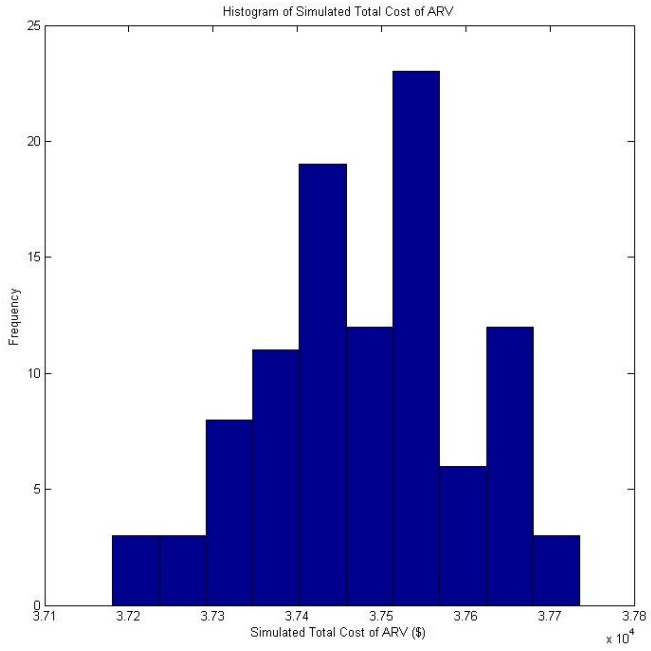

Figure 34. Histogram of Distribution of Total Cost of ARV in Mbeya Urban District

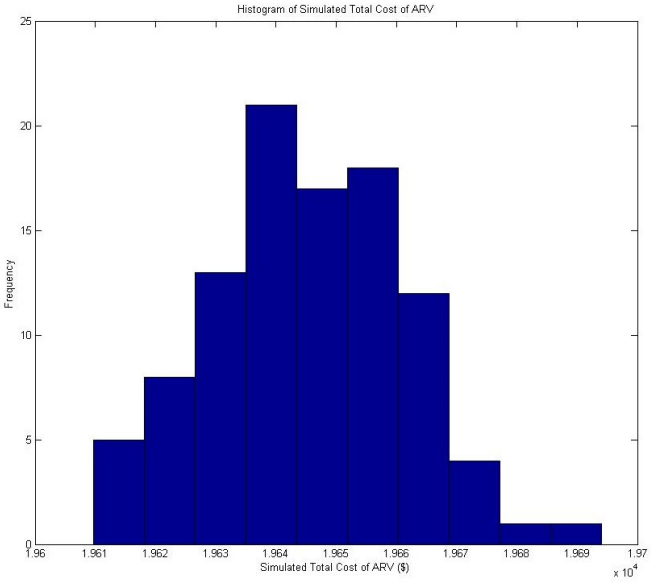

Figure 35. Histogram of Distribution of Total Cost of ARV in Mjini District

Table 7 shows some statistics of the varying ARV total cost for each district.

District	$\tilde{c}_k$Max ($)	$\tilde{c}_k$Mean ($)	$\tilde{c}_k$Median ($)	$\tilde{c}_k$Min($)	$\tilde{c}_k$Range ($)
Kinondoni	639,520	637,810	637,790	636,260	3255
Mbeya Urban	37,735	37,481	37,479	37,181	554.5
Mjini	19,694	19,645	19,645	19,610	84.3

Table 7.Summary Statistics of Varying ARV Costs

D. Loss Functions for Antiretroviral Costs

The varying ARV costs ($\tilde{c}_k$) are then used to create and minimize loss functions under three decision-making attitudes or varying risk tolerance, namely minimax regret, maximin and maximax attitudes. For each of the three districts, linear loss functions depicting the error in estimation of HIV prevalence were created and minimized to obtain the best estimates of HIV prevalence values that correspond to the least impact on ARV cost under three decision-making attitudes. The linear loss functions of the minimax regret, maximin and maximax attitudes, respectively, are represented on the graphs on the top row in figures 36 – 39 and the corresponding graphs of minimized loss values are shown on the bottom row of the same figures. Summary statistics from the each district's minimized expected loss values are shown in tables 8 – 10.

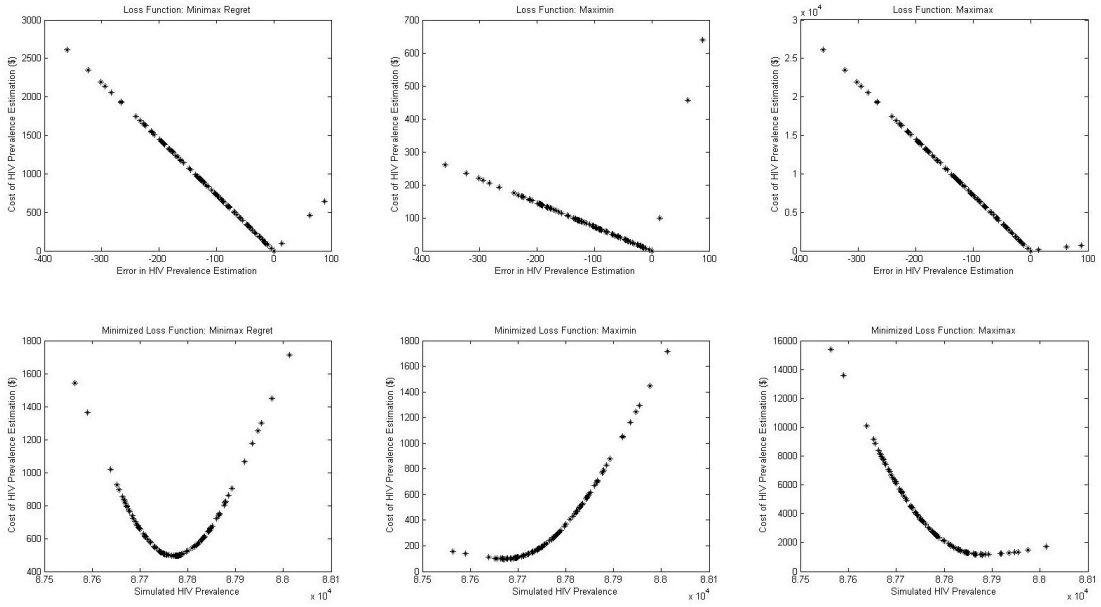

Figure 36. Linear Loss Functions and Expected Loss Values for Kinondoni District

Attitude	Best HIV Prevalence Estimate	Impact on ARV Cost ($)
Minimax Regret	87,774	494.6
Maximin	87,672	98.3
Maximax	87,880	1167.1

Table 8.Kinondoni District: Best HIV Prevalence Estimates and Impact on ARV Cost under Three Decision-Making Attitudes

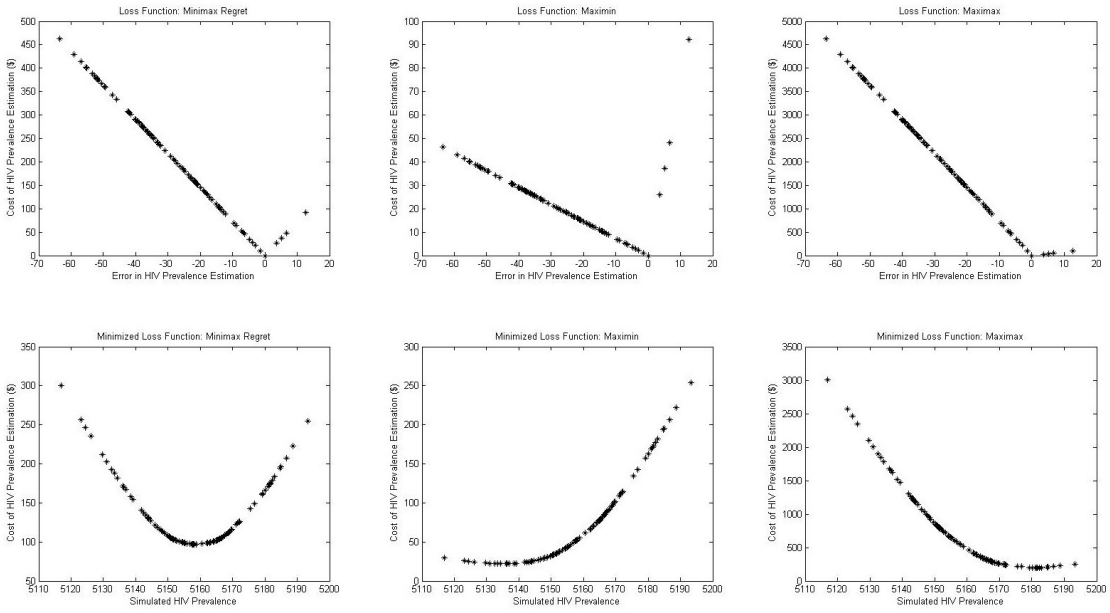

Figure 37. Linear Loss Functions and Expected Loss Values for Mbeya Urban District

Attitude	Best HIV Prevalence Estimates	Impact on ARV Cost ($)
Minimax Regret	5158	97.2
Maximin	5136	22
Maximax	5181.4	196

Table 9.Mbeya Urban District: Best HIV Prevalence Estimates and Impact on ARV Cost under Three Decision-Making Attitudes

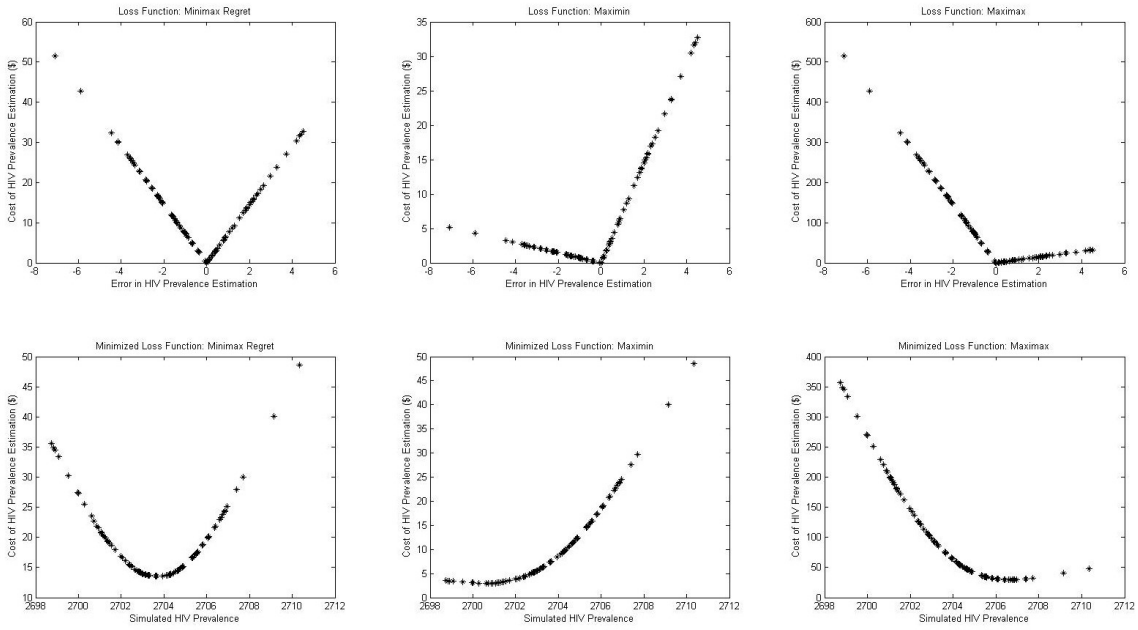

Figure 38. Linear Loss Functions and Expected Loss Values for Mjini District

Attitude	Best HIV Prevalence Estimates	Impact on ARV Cost ($)
Minimax Regret	2703.7	13.6
Maximin	2700.7	3
Maximax	2706.7	29.4

Table 10.Mjini District: Best HIV Prevalence Estimates and Impact on ARV Cost under Three Decision-Making Attitudes

VI. Discussion

A. Aggregate vs Spatially Varying Prevalence Values

In all the three districts in this study, the aggregate prevalence values ($\bar{v}_k$) did not fall within the distribution of the spatially varying values ($\tilde{v}_k$) derived from simulated population of reproductive age, as shown in table 11.

District (K)	Average HIV prevalence ($\bar{v}_k$)	Sum of Spatially Varying prevalence ($\bar{\tilde{v}}_k$)	$\tilde{v}_k$ Maximum	$\tilde{v}_k$ Mean	$\tilde{v}_k$ Median	$\tilde{v}_k$ Minimum
Kinondoni	76,431	18,330	88,012	87,776	87,774	87,564
Mbeya Urban	9,447	4,688	5193.2	5158.2	5157.9	5116.9
Mjini	881	2,147	2710.3	2703.6	2703.7	2698.7

Table 11.HIV Prevalence Statistics

For *Kinondoni* and *Mjini* districts, the spatially varying prevalence was at least 11,133 people (1% of respective gridded census population of reproductive age) and 1,817 people (1.5% of respective gridded census population of reproductive age) more than the aggregate prevalence, respectively. For *Mbeya Urban* district, the spatially varying prevalence was at least 4253 people (3% of respective gridded census population of reproductive age) less than the aggregate prevalence.

In all the three districts in this study, the sum of spatially varying prevalence ($\bar{\tilde{v}}_k$) values also did not fall within the distribution of the spatially varying values ($\tilde{v}_k$). However, the aggregates of spatially varying prevalence ($\bar{\tilde{v}}_k$) values were closer to the $\tilde{v}_k$ distribution in *Mbeya Urban* and *Mjini* districts by at least 428 and 551 people, respectively. In *Kinondoni* district, the aggregates of spatially varying prevalence was at least 69,234 people less than the spatially varying values ($\tilde{v}_k$) because of the relatively high difference between the aggregate prevalence rate and the average of spatially varying prevalence rate as shown in table 12.

District	Aggregate HIV prevalence rate ($\bar{r}_k$) [%]	Average of Spatially Varying prevalence rate (r_d/N) [%]
Kinondoni	6.88	1.65
Mbeya Urban	6.63	3.29
Mjini	0.73	1.78

Table 12.Aggregate vs. Spatially Varying HIV Prevalence Rates

While the disaggregation of HIV prevalence rates from the average district values to the spatially varying values may help a decision-maker to avoid spatial analysis bias, assigning DHS cluster centroid values to discretized points within the urban and rural buffers may lead to a significant discrepancy between the aggregate prevalence rate and the average of spatially varying prevalence rate, as is the case in *Kinondoni* district. Therefore, a decision-maker would want to be cautious of which prevalence rate values to employ, keeping in mind that using the average of spatially varying prevalence rates instead of the aggregate prevalence rate may underestimate HIV prevalence in relatively large districts such as *Kinondoni* and *Mbeya Urban* while overestimating HIV prevalence in smaller districts such

as *Mjini*, as shown in this study. Another consideration that might affect the average of spatially varying prevalence rates is how prevalence rates are assigned to grids that are contained within the intersection of two or more DHS centroid buffers. In our study, the grids in the intersection of two or more buffers were assigned the higher prevalence rate value: thus, it is likely that the difference between the average of spatially varying prevalence rates and aggregate prevalence rate would change somewhat had we considered the alternative approach of averaging the values of the prevalence rates contained within the intersection of buffers. That said, since our goal in this study was to explicate the differences between using aggregate values and spatially varying values in the light of possible decision-making attitudes, we considered the somewhat extreme but realistic possibility of employing the maximum r_d value as the spatially varying prevalence rate to study the decision-making implication thereof.

B. Aggregate vs Spatially Varying Prevalence ARV Costs

In all the three districts in this study, the aggregate ARV costs ($\bar{c}_k$) did not fall within the distribution of the spatially varying values ($\tilde{c}_k$) derived from simulated population of reproductive age, as shown in table 13.

District (K)	Aggregate ARV cost ($\bar{c}_k$) [$]	Aggregate of Spatially Varying ARV cost ($\bar{\tilde{c}}_k$) [$]	$\tilde{c}_k$Max ($)	$\tilde{c}_k$Mean ($)	$\tilde{c}_k$Median ($)	$\tilde{c}_k$Min($)
Kinondoni	557,947	133,810	639,520	637,810	637,790	636,260
Mbeya Urban	68,963	34,223	37,735	37,481	37,479	37,181
Mjini	6,431	15,673	19,694	19,645	19,645	19,610

Table 13.ARV Cost Statistics

The aforementioned discrepancies translate to a difference of at least $78,313 and $13,179 between spatially varying ARV cost values ($\tilde{c}_k$) and aggregate ARV cost ($\bar{c}_k$) for *Kinondoni* and *Mjini* districts, and aggregate ARV cost of at least $31,228 more than the spatially varying cost in *Mbeya Urban* district. In all the three districts in this study, the sum of spatially varying ARV cost ($\bar{\tilde{c}}_k$) also did not fall within the distribution of the spatially varying ARV cost ($\tilde{c}_k$). However, the aggregates of spatially varying ARV cost ($\bar{\tilde{c}}_k$) were closer to the $\tilde{c}_k$ distribution in *Mbeya Urban* and *Mjini* districts by at least $2958 and $3937, respectively. In *Kinondoni* district, the aggregates of spatially varying ARV cost was at least $502,450 less than the spatially varying ARV cost ($\tilde{c}_k$).

The aforementioned cost differences can be better put into perspective if one was to take into account how much the decision-maker, say the Tanzanian government, would be trying to save the consumer by manufacturing ARV drugs locally at approximately $7.30 rather than importing at approximately $10. Considering the aggregate HIV prevalence and cost of

one dose of ARV treatment per month per person, the decision-maker would be aiming to save approximately $200,000 for *Kinondoni* district, $25,000 for *Mbeya Urban* district, $2000 for *Mjini* district and approximately $3.5 million for the entire country of Tanzania. Considering such potential aggregate savings, the discrepancy between aggregate ARV cost ($\bar{c}_k$) and aggregate of spatially varying ARV cost ($\bar{\tilde{c}}_k$) translates to approximately two times the intended potential savings in *Kinondoni* district, exceeds intended potential savings in *Mbeya Urban* by 36% and is four times the intended potential savings in *Mjini* district. A viable conclusion then is that if a decision-maker utilizes the aggregate ARV cost ($\bar{c}_k$) instead of the aggregate of spatially varying ARV cost ($\bar{\tilde{c}}_k$), then it is unlikely that they will realize the full potential savings from locally manufacturing rather than importing ARV drugs.

For comparison, the discrepancy between the aggregate of spatially varying ARV cost ($\bar{\tilde{c}}_k$) and the spatially varying ARV cost ($\tilde{c}_k$) is approximately double the intended potential savings in *Kinondoni* district, at least 11% of intended potential savings in *Mbeya Urban* and exceeds intended potential savings in *Mjini* district by 65%. Therefore, if a decision-maker utilizes the spatially varying ARV cost ($\tilde{c}_k$) instead of the aggregate of spatially varying ARV cost ($\bar{\tilde{c}}_k$), then they would be relatively closer to realizing the full potential savings from locally manufacturing rather than importing ARV drugs.

C. Adjustment of Realizations of Population of Reproductive Age

During the adjustment process, the threshold population of reproductive age (ϕ_i) chosen was equivalent to the median gridded population of reproductive age. While the choice of the latter may have been subjective, it was computationally sufficient enough for our adjustment algorithm to make changes to enough pixels as to achieve the required spatial pattern. The choice of threshold population was informed by the percentiles of population of reproductive age shown in table 14.

District	Percentiles (%)					
	10	25	50	75	90	100
Kinondoni	0	0	1	13	36	512
Mbeya Urban	0	0	1	3	17	334
Mjini	0	1	8	38	239	1419

Table 14.Percentiles of Gridded Population of Reproductive Age

In all the three districts, the gridded population values of reproductive age were relatively low across most grids. For instance, in *Kinondoni*, the largest district in this study, 90% of the population in the 53,332 grids contained 36 people of reproductive age or lower. In *Mbeya Urban* and *Mjini* districts, 90% of the population in their grids contained 17 and 239 people of reproductive age or lower, respectively.

All in all, the adjusted realizations resembled the original gridded dataset relatively well across all districts as shown in the maps and associated histograms in figures 39 – 41. As expected, in all the three districts, the relatively higher population differences are observed in pixels that contained more people of reproductive age to begin with.

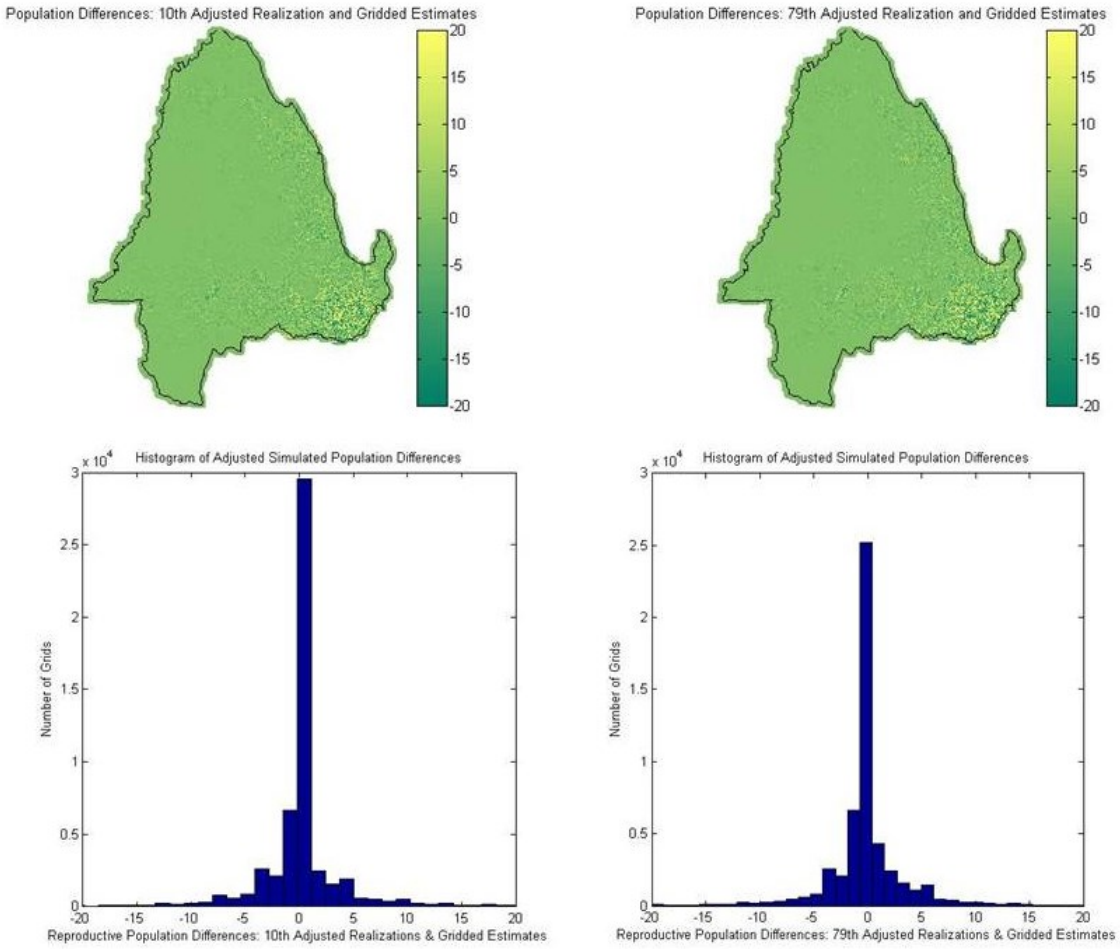

Figure 39. Kinondoni District: Maps and Histograms showing Population Differences between some Adjusted Realizations and Gridded Population of Reproductive Age

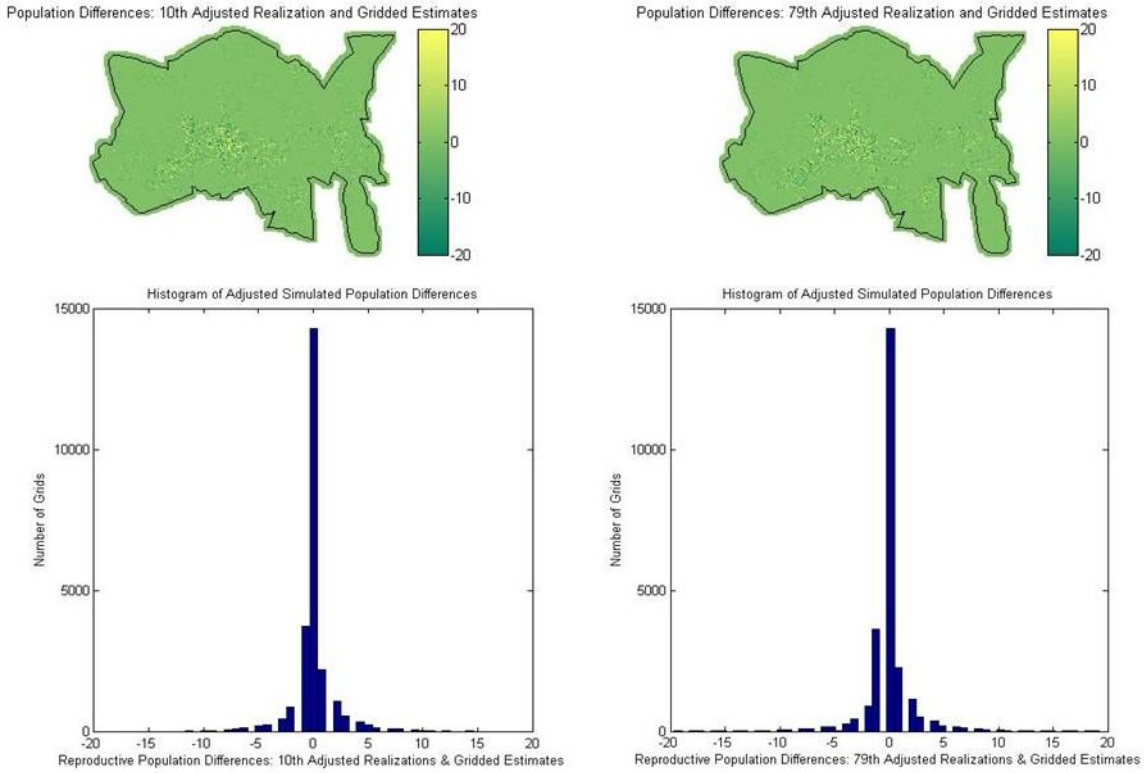

Figure 40. Mbeya Urban District: Maps and Histograms showing Population Differences between some Adjusted Realizations and Gridded Population of Reproductive Age

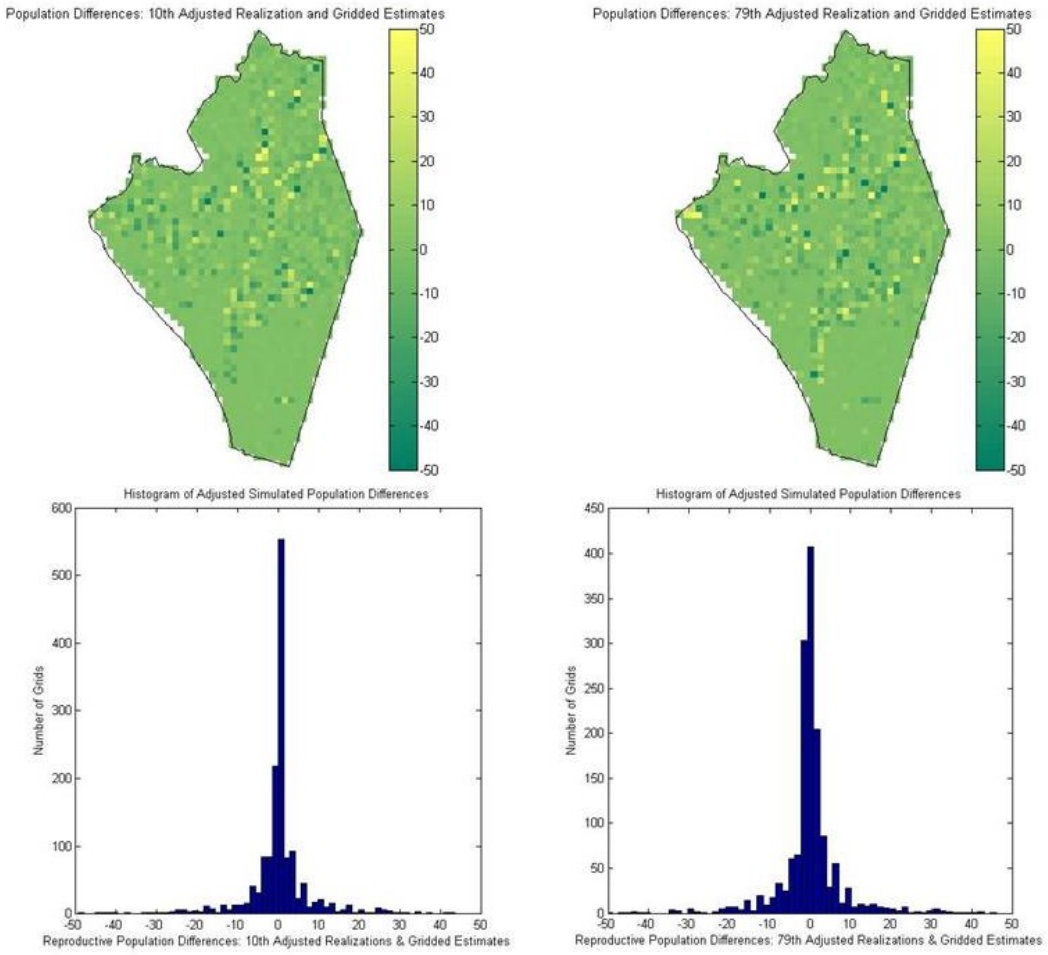

Figure 41. Mjini District: Maps and Histograms showing Population Differences between some Adjusted Realizations and Gridded Population of Reproductive Age

Moreover, the adjusted realizations resembled each other relatively well across all districts as shown in the maps and associated histograms in figures 42 – 44. As expected, the relatively higher population differences are observed in pixels that contained more people of reproductive age to begin with.

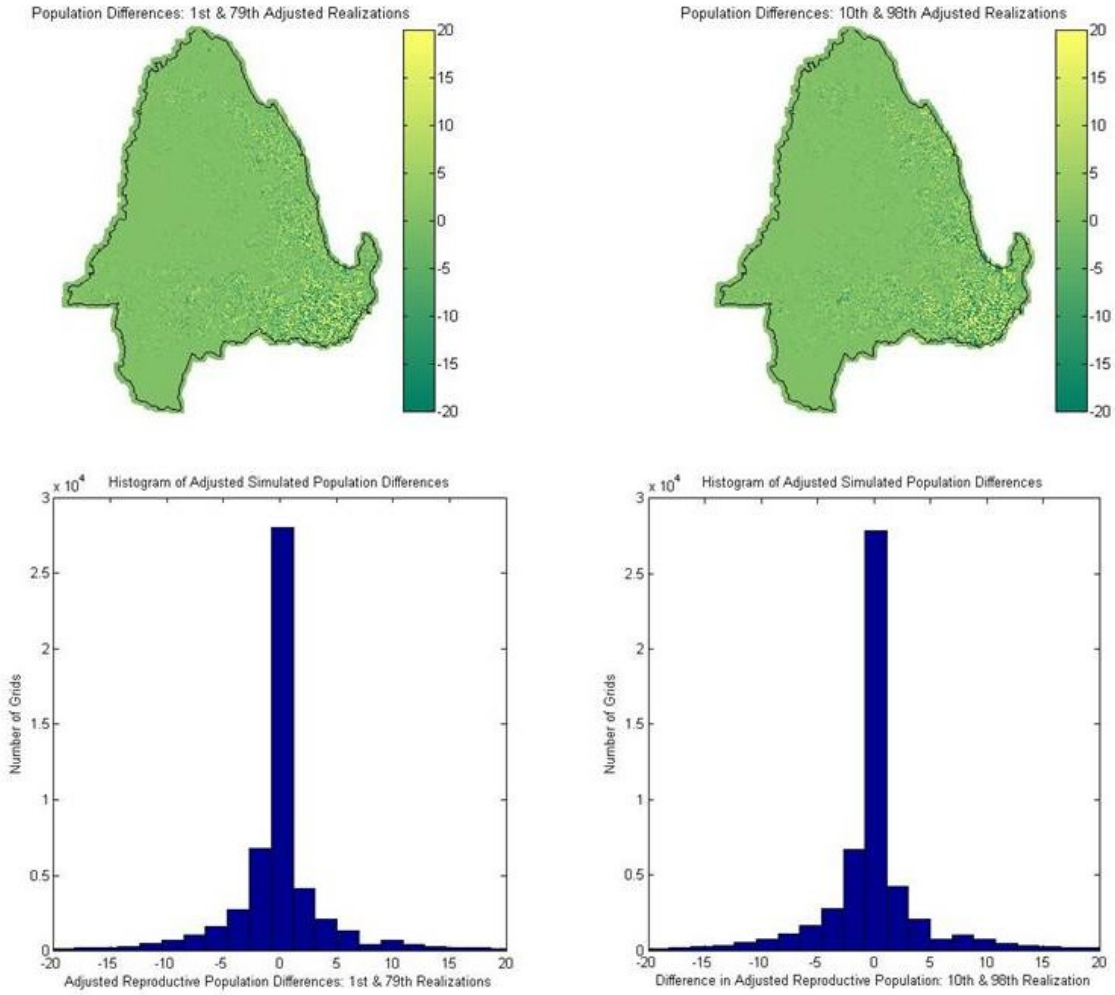

Figure 42. Kinondoni District: Maps and Histograms of Population Differences between some Adjusted Realizations of Reproductive Age

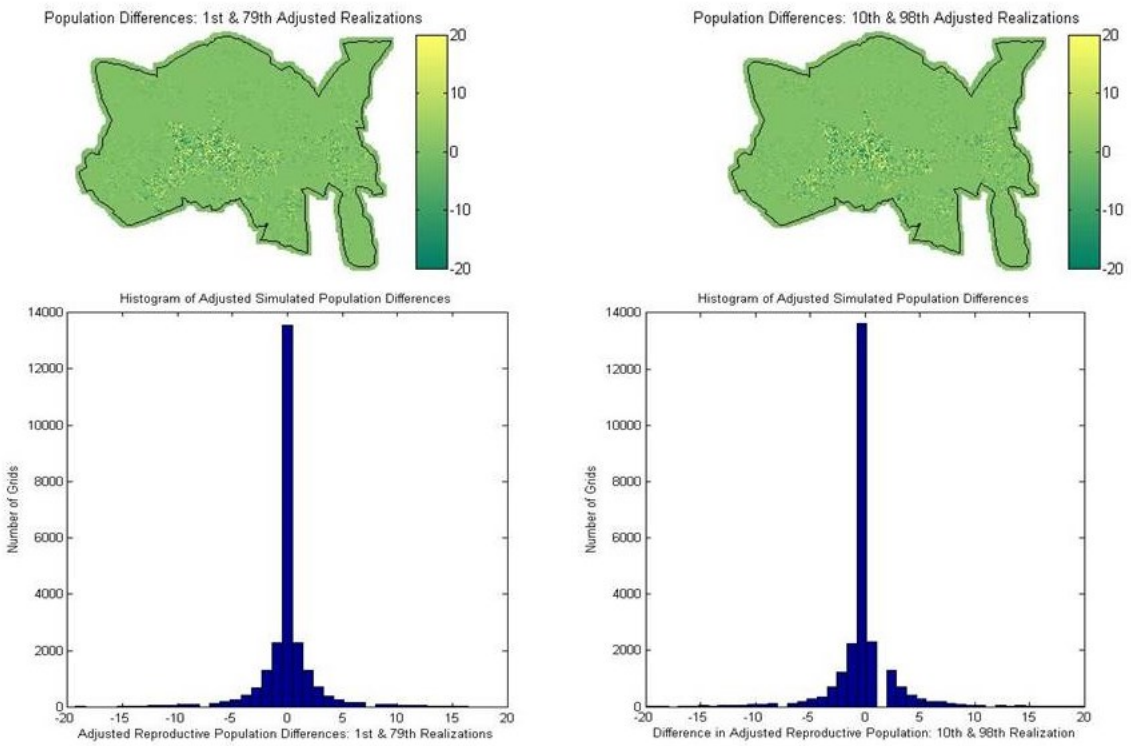

Figure 43. Mbeya Urban District: Maps and Histograms of Population Differences between some Adjusted Realizations of Reproductive Age

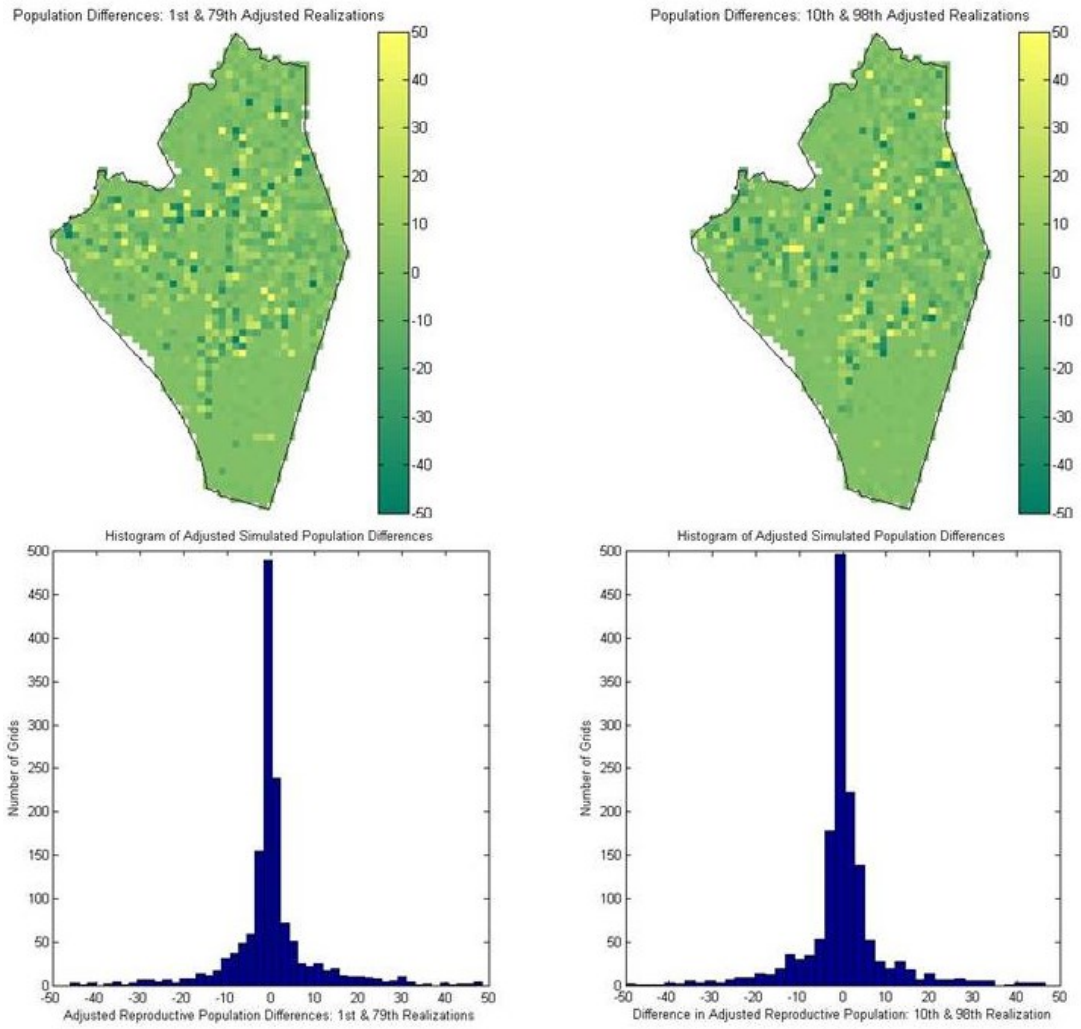

Figure 44. Mjini District: Maps and Histograms of Population Differences between some Adjusted Realizations of Reproductive Age

As shown in the quantile-quantile plots in figures 45 – 47, the adjusted realizations match the gridded population of reproductive age relatively well for populations of lower quantiles that comprise most of the grids in each district.

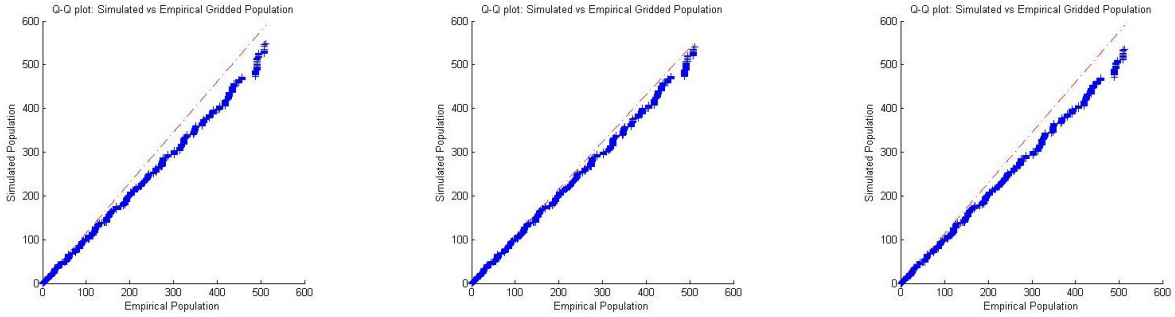

Figure 45. Kinondoni District: Q-Q Plots of 1st, 10th and 79th Adjusted Realizations

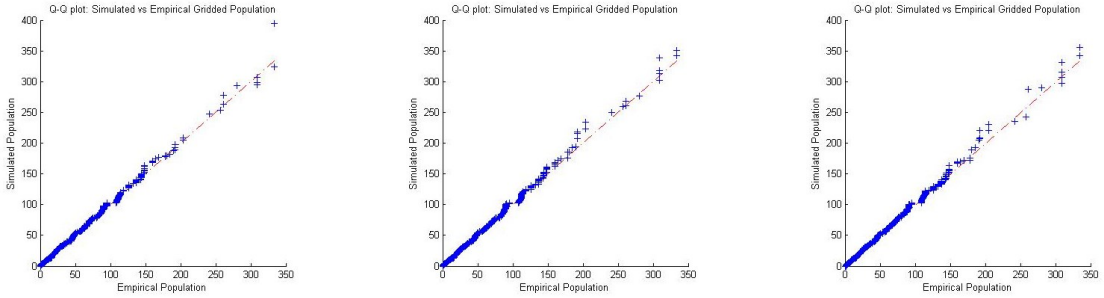

Figure 46. Mbeya Urban District: Q-Q Plots of 1st, 10th and 79th Adjusted Realizations

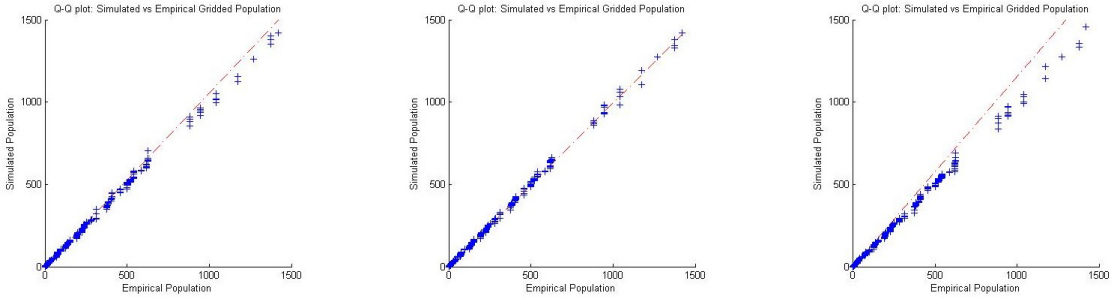

Figure 47. Mjini District: Q-Q Plots of 1st, 10th and 79th Adjusted Realizations

D. Utility Functions for Antiretroviral Costs

In *Kinondoni* district, the minimization of the linear loss function under the three decision-making attitudes reveals best estimates of HIV prevalence values that are all approximately 15% higher than the aggregate prevalence values ($\bar{v}_k$) and almost four times higher than the sum of spatially varying prevalence ($\bar{\tilde{v}}_k$). The minimax regret attitude yields a best HIV prevalence estimate of 87,774 people, a value that is equivalent to the median value in the distribution of varying HIV prevalence values ($\tilde{v}_k$) (see table 6 and 8). As expected, the maximax attitude caters to the highest HIV prevalence estimate (87,880 people) but at a greater impact on ARV cost ($1167.10). That said, the best estimate of HIV prevalence under the maximax attitude is 208 people more than the best estimate in the budget-conscious maximin attitude.

In *Mbeya Urban* district, the minimization of the linear loss function under the three decision-making attitudes reveals best estimates of HIV prevalence values that are all approximately 46% lower than the aggregate prevalence values ($\bar{\mathrm{v}}_{\mathrm{k}}$) and approximately 10% higher than the sum of spatially varying prevalence ($\bar{\tilde{\mathrm{v}}}_{\mathrm{k}}$). The minimax regret attitude yields a best HIV prevalence estimate of 5,158 people, a value that is equivalent to the median value in the distribution of varying HIV prevalence values ($\tilde{v}_k$) (see table 6 and 9). The maximax attitude caters to the highest optimized HIV prevalence estimate (~5181 people) but at a greater impact on ARV cost ($196). That said, the best HIV prevalence estimate under the maximax attitude is 45 people more than the best estimate in the budget-conscious maximin attitude.

In *Mjini* district, the minimization of the linear loss function under the three decision-making attitudes reveals HIV prevalence values that are two times higher than the aggregate prevalence values ($\bar{v}_k$) and approximately 26% higher than the sum of spatially varying prevalence ($\bar{\tilde{v}}_k$). The minimax regret attitude yields a best HIV prevalence estimate of 2,703 people, a value that is equivalent to the median value in the distribution of varying HIV prevalence values ($\tilde{v}_k$) (see table 6 and 10). The maximax attitude yields the highest best HIV prevalence estimate (~2707 people) but at a greater impact on ARV cost ($29.40). That said, the best HIV prevalence estimate under the maximax attitude is only 3 people more than that of the more budget-conscious maximin attitude.

VII. Model Limitations

While our study advocates for the incorporation of spatially varying datasets if available, disaggregating health and socio-economic indicators remains a matter of an informed guess, at best. In this study, the disaggregation of the district level HIV prevalence rates was necessitated by the lack of discretized HIV prevalence data to begin with and few degrees of freedom existed for exploring spatial heterogeneity in our population variables. Essentially, the population value at each pixel was simulated independently of other pixels because of lack of extraneous information dictating population dynamics within each district. While the AfriPop dataset ensured that the spatial pattern in our model remained consistent with a pre-established gridded dataset, our non-homogenous Poisson model failed to model spatial dependence that one would expect in Tobler's first law of geography (Tobler, 1970). The assumption here is that the local (pixel-specific) AfriPop estimates, once adjusted to account

for the temporal difference, provides all the information necessary for determining the uncertainty in the unknown population value at each pixel. In essence, all the complexity of the spatial pattern is encapsulated via this model into the local adjusted AfriPop population estimates. Furthermore, since the gridded population was modeled within each district at a time, it was difficult to account for spatio-temporal demographic changes, for example migration, between census years and geographical areas. We tried to mitigate this spatial uncertainty by applying national net migration rates and urbanization rates to decrease and increase population, respectively, yet what happens on the ground between each pixel, district and census years remains uncertain and difficult to model. Lastly, our spatial adjustment algorithm did not quite distinguish population dynamics between urban and rural areas. HIV Prevalence is known to be generally higher in urban areas than in rural areas across gender (TACAIDS, 2014), an observation that our model did not consistently capture across all districts, mostly due to the compounded uncertainty of areas defined as urban and rural.

VIII. Future Research

The next step in this research that is arguably the most crucial would be to validate the population distribution in the original and derived gridded population datasets. One approach would be to use recent MODIS satellite imagery to distinguish urban and rural extents within any global region of interest (e.g. Schneider et al., 2009). With the consistently scarce data on population in remote rural communities, it is likely that rural population and HIV prevalence rates therein are underestimated. As such, research

emphasis on prediction of population dynamics is necessary in order to estimate population at the pixel level and temporal scales between and beyond the census years. For instance, the role of fertility rates and migration in rural and urban areas could be investigated (e.g. Carr & Burgdorfer, 2013). Prediction of urban population dynamics would benefit from simulation of urban growth and the resulting land use change, as performed for other urban areas using Clarke's SLEUTH model (Clarke et al, 1997, 2008).

In terms of methodology, further work into how the choice of pseudo-random number generator may introduce potential bias during the simulation process is required. A study by Van Niel & Laffan (2003) demonstrated that the choice of random number generator employed in the Monte Carlo simulation of multiple realizations of a GIS layer could lead to significantly different results in the output. Future research could also entail the exploration of other models of simulating errors resulting from population downscaling. For instance, instead of a non-homogenous simulation process, correlated simulation could be used to explore dynamics at the pixel level. For areas with incomplete and/or irregular observation units, geostatistical simulation methods such as that developed by Kyriakidis and Yoo (2005) could be adopted to generate realizations at the pixel level.

Finally, from a decision-making standpoint, further work is needed on the development on concepts of loss functions and estimation theory which can be applied to HIV studies that utilize empirical data. In the meantime, other decision-making attitudes such as the Bayesian approach coupled with other types of loss functions could be investigated and possibly adopted for larger-scale regions. Additionally, further research would also entail a study of the potential role of ambiguity as far as spatial uncertainty is concerned. A consistent finding across several fields reveals that many decision-makers tend to be averse to ambiguity

(Highhouse & Hause, 1995; Mukerji & Tallon 2004). When this ambiguity is likened to spatial uncertainty, potential aversion by decision-makers could present a bias in which they avoid regions in which uncertainty is included (Van Dijk & Zeelenberg, 2003).

IX. Conclusion

Our findings show that the utilization of coarse aggregated values such as census data at the district level instead of spatially varying values may hinder effective decision analysis in situations that call for the use of a spatial population dataset. As such, we show that ARV cost assessment is sensitive to the level of aggregation of the population datasets and as such, may result in potentially significant discrepancies in cost that might affect policy.

However, we recognize that up-to-date fine resolution datasets for some regions, especially in developing countries, are not always readily available. Furthermore, the few that exist typically contain source data and methods that are inherently uncertain and difficult to explicate. In such cases, decision-makers may benefit from the application of multiple simulated spatial distributions of fine scale population along with the associated cost estimates. From this distribution of cost estimates, decision-makers should feel relatively confident selecting optimized impact values based on an explication of risk attitude, thus avoiding unforeseeable consequences of underestimating or overestimating impact assessment outcomes.

References

1. AfriPop.,2012, AfriPop Project, Retrieved June 13, 2013, from http://www.afripop.org/

2. AllAfrica., 2012, Tanzania: Local Firm to Produce ARVs Soon, Retrieved Sept 21, 2013, from http://allafrica.com/stories/201205211078.html

3. Ang, A.H-S., &Tang, W.H., 1990, Probability Concepts in Engineering Planning and Design. Volume II: Decision, Risk, and Reliability, *John Wiley & Sons*, New York.

4. Clarke, K.C., Hoppen, S., & Gaydos, L., 1997, A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay Area, *Environmental and Planning B* 24: 247–61

5. Clarke, K.C., 2008, Mapping and modeling land use change: An application of the SLEUTH Model. In Pettit C, Cartwright W, Bishop I, Lowell K, Pullar D, and Duncan D (*eds*) *Landscape Analysis and Visualization: Spatial Models for Natural Resource Management and Planning*. Berlin, Springer: 353–66

6. López-Carr, D., & Burgdorfer, J., 2013, Deforestation Drivers: Population, Migration, and Tropical Land Use. *Environment: Science and Policy for Sustainable Development,* (55)1: 3-11.

7. CIA., 2012a, Central Intelligence Agency The World Factbook, 2012, Retrieved May 2, 2014, from https://www.cia.gov/library/publications/the-world-factbook/rankorder/2155rank.html

8. CIA., 2012b, Central Intelligence Agency The World Factbook for Tanzania, 2012, Retrieved May 2, 2014, from http://cia-world-fact-book.findthedata.org/l/1001/Tanzania

9. CIA., 2012c, Central Intelligence Agency The World Factbook, 2012, Retrieved May 2, 2014, from https://www.cia.gov/library/publications/the-world-factbook/geos/tz.html

10. DHS., 2013, The DHS Project, Retrieved Oct 12, 2013, from www.measuredhs.com

11. DHS., 2013b, The DHS Project, Retrieved Oct 12, 2013, from http://dhsprogram.com/what-we-do/survey/survey-display-393.cfm

12. Eicher, C.L., & Brewer, C.A., 2001, Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science,* 28:125 – 138.

13. Fisher, P.F., & Langford, M., 1995, Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A*, 27: 211-224.

14. Highhouse, S. & Hause, E., 1995, Missing information in selection: An application of the Einhorn-Hogarth Ambiguity Model. *Journal of Applied Psychology*, 80: 86-93.

15. Hunter, G. J., & Beard, K., 1992, Understanding error in spatial databases. *The Australian Surveyor,* 37(2):108-119

16. Hunter, G. J., & Goodchild, M. F., 1997, Modelling the uncertainty of slope and aspect estimates derived from spatial databases. *Geographical Analysis*, 29(1): 35-49.

17. IOM., 2014, International Organization for Migration, Retrieved May 4, 2014, from https://www.iom.int/cms/tanzania

18. Kyriakidis, P.C., & Yoo, E.H., 2005, Geostatistical prediction and simulation of point values from areal data. *Geographic Analysis,* 37(2): 124 – 151.

19. Langford, M., 2006, Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems*, 30: 161-180.

20. Linard, C., Gilbert, M., Snow, R.W., Noor, A.M., Tatem, A.J., 2012, Population Distribution, Settlement Patterns and Accessibility across Africa in 2010. *PLoS ONE* 7:e31743.

21. Linard, C., & Tatem, A.J., 2012, Large-scale spatial population databases in infectious disease research. *International Journal of Health Geographics,* 11:7.

22. McCleary, G.F., Jr., 1984, Cartography, geography and the dasymetric method. In *Proceedings 12th conference of international cartographic association, Perth Australia*, 1:599 – 610.

23. Mennis, J., 2003, Generating surface models of population using dasymetric mapping. *The Professional Geographer* 45: 317-322.

24. Mondal, P., & Tatem, A.J., 2012, Uncertainties in Measuring Populations Potentially Impacted by Sea Level Rise and Coastal Flooding. *PLoS ONE* 7(10): e-48191.

25. Mukerji, S., & Tallon, J. M., 2004, Ambiguity aversion and the absence of wage indexation. *Journal of Monetary Economics*, 51(3):653-670.

26. NBS., 2012, National Bureau of Statistics, Retrieved Sep 8, 2013, from http://www.nbs.go.tz/.

27. NBS., 2012b, Tanzania in Figures 2012, Retrieved Dec 2, 2013, from http://www.nbs.go.tz/nbs/takwimu/references/Tanzania_in_figures2012.pdf

28. Oyelami, B. O., & Ogidi, J. A., 2013, Most Effective Utility Function for HIV Drug using Cost and Efficacy Analysis, *Research Desk*, 4: 350-361

29. Rubinstein, R.Y., & Kroese, P.K., 2007, Simulation and the Monte Carlo Method, 2nd Ed., *John Wiley & Sons,* Hoboken, NJ.

30. Savage, L.J., 1954, The Foundations of Statistics, *Wiley*, New York.

31. Savage, S.L., 2009, The Flaw of Averages, *John Wiley & Sons,* Hoboken, NJ.

32. Schneider, A., Friedl, M.A., Potere D, 2009, A new map of global urban extent from MODIS satellite data. *Environmental Research Letters* 4(4).

33. Simon, H.A., 1955, A Behavioral Model of Rational Choice. *Quarterly Journal of Economics* 69(1):99-118.

34. TACAIDS., 2014, Tanzania Commission For AIDS, Retrieved May 10, 2014, from http://www.tacaids.go.tz/index.php?option=com_content&view=article&id=138:about-hiv-and-aids&catid=31:hiv-and-aids-information-&Itemid=158

35. Tatem, A.J., Noor, A.M., Hay, S.I., 2004, Defining approaches to settlement mapping for public health management in Kenya using medium spatial resolution satellite imagery. *Remote Sensing of Environment* 93:42-52.

36. Tatem, A.J., Noor, A.M., Hay, S.I., 2005, Assessing the accuracy of satellite derived global and national urban maps in Kenya. *Remote Sensing of Environment* 96:87–97.

37. Tatem, A.J., Noor, A.M., von Hagen, C., Di Gregorio, A., Hay, S.I., 2007, High Resolution Population Maps for Low Income Nations: Combining Land Cover and Census in East Africa. *PLoS ONE* 2(12).

38. Tatem, A.J., Campiz, N., Gething, P.W., Snow, R.W., Linard, C., 2011, The effects of spatial population dataset choice on estimated of population at risk of disease. *Population Health Metrics* 9:4.

39. Tatem, A.J., Adamo, S., Bharti, N., Burgert, C.R., Castro, M., Dorelien, A., Fink, G., Linard, C., John, M., Montana, L., Montgomery, M.R., Nelson, A., Noor, A.M., Pindolia, D., Yetman, G., Balk, D., 2012, Mapping population at risk: improving spatial demographic data for infectious disease modelling and metric derivation, *Population Health Metrics* 10(8).

40. Thompson, K.M., & Graham, J.D., 1996, Going beyond the single number: Using probabilistic risk assessment to improve risk management, *Human and Ecological Risk assessment*, 2(4): 1008-1034.

41. Tobler, W.R., 1979, Smooth Pycnophylactic Interpolation for Geographic Regions. *Journal of the American Statistical Association,* 74:519-536.

42. Tobler, W.R., 1970, A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46: 234–240

43. UN Population Division., 2013, United Nations Department of Economic and Social Affairs, Population Division, Retrieved May 10, 2014, from http://www.un.org/en/development/desa/population/

44. Van Niel, K., & Laffan, S.W., 2003, Gambling with randomness: the use of pseudorandom number generators in GIS. *International Journal of Geographic Information Science*, 17(1):49-68

45. Van Dijk, E., & Zeelenberg, M., 2003, The discounting of ambiguous information in economic decision making. *Journal of Behavioral Decision Making*, 16(5): 341-352.

46. Wald, A., 1950, Statistical Decision Functions, *Wiley*, New York.

47. Wright, J. K., 1936, A method of mapping densities of population: With Cape Cod as an example. *Geographical Review* 26: 103 – 110.

48. Yuan, Y., Smith, R.M., & Limp, W.F., 1997, Remodeling census population with spatial information from Landsat TM imagery. *Computers, Environment and Urban Systems*, 21: 245 – 258.