

©Copyright 2014

Yanjun He

# Probabilistic Population Projection for Countries with Generalized HIV/AIDS Epidemics

Yanjun He

A thesis  
submitted in partial fulfillment of the  
requirements for the degree of

Master of Science

University of Washington

2014

Committee:

Adrian E. Raftery, Chair

Samuel J. Clark

Program Authorized to Offer Degree:  
Statistics

University of Washington

**Abstract**

Probabilistic Population Projection for Countries with Generalized HIV/AIDS Epidemics

Yanjun He

Chair of the Supervisory Committee:  
Professor Adrian E. Raftery  
Statistics and Sociology

Population projection has long been an issue for researchers, governments and international organizations so that they can monitor and plan development and resources. The United Nation Population Division (UNPD) publishes the *World Population Prospects* (WPP) every 2 years, giving estimates of past and projections of future population by age and sex for all countries. The method used for population projection in the 2012 revision of WPP was developed by Raftery et al. (2012), which combines probabilistic projections of life expectancy at birth, fertility and mortality using a cohort component model. It was developed for countries without significant levels of HIV/AIDS prevalence. For countries with generalized HIV/AIDS epidemics, the mortality pattern is different for young and middle-aged adults, and therefore the effect of AIDS should be considered in estimation and projections. In this thesis, we incorporate HIV impact, including HIV prevalence and anti-retroviral therapy (ART) coverage in projecting life expectancy and age- and sex-specific mortality, and use them to probabilistically project population for countries with generalized HIV/AIDS epidemics. Through out-of-sample validation, we demonstrate that our method performs well in population projection for those countries.

**Keywords:** Probabilistic population projection, generalized HIV/AIDS epidemics, HIV prevalence, ART coverage

## TABLE OF CONTENTS

|  | Page |
|--|------|
| List of Figures . . . . .  | ii   |
| List of Tables . . . . .   | iii  |
| Chapter 1: Introduction . . . . .                                    | 1    |
| Chapter 2: Methodology and Results . . . . .                         | 5    |
| 2.1 HIV prevalence probabilistic estimation and projection . . . . . | 5    |
| 2.2 Life expectancy model and probabilistic projection . . . . .     | 10   |
| 2.3 Probabilistic population projection . . . . .                    | 13   |
| Chapter 3: Case Study . . . . .                                      | 17   |
| 3.1 Typical case: Kenya . . . . .                                    | 17   |
| 3.2 High HIV prevalence: Swaziland . . . . .                         | 20   |
| 3.3 Caribbean country: Jamaica . . . . .                             | 21   |
| 3.4 Negative discrepancy from WPP 2012: Botswana . . . . .           | 24   |
| 3.5 Positive discrepancy from WPP 2012: Sierra Leone . . . . .       | 26   |
| Chapter 4: Validation . . . . .                                      | 31   |
| Chapter 5: Discussion . . . . .                                      | 36   |
| Appendix A: Parameters $\phi$ and $r$ in EPP 2009 . . . . .          | 42   |
| Appendix B: Double logistic curve . . . . .                          | 43   |

## LIST OF FIGURES

| Figure Number  | Page |
|--|------|
| 1.1 Example of mortality rate comparison before and after entering generalized HIV epidemics for South Africa. The plot is drawn on the logarithmic scale of mortality for female in 1975-1980 (black) and 2000-2005 (red) respectively. | 3    |
| 2.1 The compartments in the HIV model with ART used in EPP 2009 (Brown et al., 2010).  | 6    |
| 2.2 Probabilistic HIV prevalence estimation (1980-2010) and projections (2011-2100) for Malawi and Uganda.   | 9    |
| 2.3 Lowess plot for observations with HIV prevalence less than 1% and greater than or equal to 1%.   | 12   |
| 2.4 Probabilistic life expectancy projections (2011-2100) for Malawi and Uganda.   | 14   |
| 2.5 Difference of population projection for the 40 countries in 2050 and 2100. Blue curves stand for positive difference and red is for negative.  | 16   |
| 3.1 Probabilistic HIV prevalence and life expectancy projections (2011-2100) for Kenya.  | 18   |
| 3.2 Summary of population projection for Kenya up to year 2100.  | 19   |
| 3.3 Probabilistic HIV prevalence and life expectancy projections (2011-2100) for Swaziland.  | 20   |
| 3.4 Summary of population projection for Swaziland up to year 2100.  | 22   |
| 3.5 Probabilistic HIV prevalence and life expectancy projections (2011-2100) for Jamaica.  | 23   |
| 3.6 Summary of population projection for Jamaica up to year 2100.  | 25   |
| 3.7 Probabilistic HIV prevalence and life expectancy projections (2011-2100) for Botswana.   | 26   |
| 3.8 Summary of population projection for Botswana up to year 2100.   | 27   |
| 3.9 Probabilistic HIV prevalence and life expectancy projections (2011-2100) for Sierra Leone.   | 28   |
| 3.10 Summary of population projection for Sierra Leone up to year 2100.  | 29   |
| B.1 The fitted curve for double logistic function under the “world” parameter $\theta^w = (12.2854, 40.9654, 5.5763, 17.5841, 3.7677, 0.6168)$ .   | 43   |

## LIST OF TABLES

| Table Number   | Page |
|--|------|
| 1.1 Countries with generalized HIV/AIDS epidemics, filtered by HIV prevalence higher than 1% in 1980-2010. They are focused in two different regions in the world, Africa and Caribbean/Latin America. . . . . | 2    |
| 2.1 Model fitted for the regression model of $\Delta e_{0,t}$ . . . . .  | 13   |
| 4.1 Model validation for life expectancy and population projection. MAE for population is in ( $\times 1,000$ ) unit. . . . .  | 33   |
| 4.2 Model validation for life expectancy based on different sets of HIV prevalence in fitting the model, EPP estimates and WPP 2012. . . . .   | 33   |
| 4.3 Validation for using data starting from different years and projection discrepancy from WPP 2012. . . . .  | 34   |

## ACKNOWLEDGMENTS

The author wishes to sincerely thank Professor Adrian E. Raftery and Professor Samuel J. Clark for their valuable advice and supervision. Also, the author thanks David J. Sharrow and Le Bao for discussion and providing help for this study. This thesis was partly supported by the Eunice Kennedy Shriver National Institute of Child Health and Development through grants R01 HD054511 and R01 HD070936.

## Chapter 1

## INTRODUCTION

Population projection has long been an issue for researchers, governments and international organizations. Getting a sense of how the population distribution changes and grows helps decision makers make good use of resources and plan development goals. The United Nations Population Division (UNPD) publishes the *World Population Prospects* (WPP) every 2 years, giving estimates of past and projections of future population by age and sex for all countries. Before 2010, WPP only offers deterministic estimation and projection on population and related variables, such as life expectancy at birth, total fertility rate (TFR) and mortality, etc. These estimates and projections are calculated using the standard cohort component method (Whelpton, 1936; Preston et al., 2001), which gives single valued projections. This method is based on the classical demography balancing equation, which calculates the population at time  $t+1$  summing up population projection in time  $t$ , the new births and immigrants during this time period, and subtracting the deaths and emigrants.

However, this method fails to account for uncertainty about future change on probabilistic basis. WPP 2010 (United Nations, Department of Economic and Social Affairs, Population Division, 2011a) and WPP 2012 (United Nations, Department of Economic and Social Affairs, Population Division, 2013a) uses projections of fertility and mortality rates from the Bayesian hierarchical models (BHMs) developed by Alkema et al. (2011) and Raftery et al. (2013). Though the models provide probabilistic projections, the two publications only use the predictive median as the deterministic projections. In July 2014, the UNPD has officially issued fully probabilistic projections using the same BHMs as the WPP 2012 projections.

Countries with generalized HIV/AIDS epidemics refer to those with HIV prevalence higher than 1% in the general population (i.e., not concentrated in a certain subpopulation such as females sex workers, male sex with male, etc.) (UNAIDS, 2012). We filter these

countries out by setting the threshold of HIV prevalence in 1980-2010 to be higher than 1% and checking the country progress reports published by UNAIDS for each country. The 40 AIDS countries we consider in this paper are shown in Table 1.1.

Table 1.1: Countries with generalized HIV/AIDS epidemics, filtered by HIV prevalence higher than 1% in 1980-2010. They are focused in two different regions in the world, Africa and Caribbean/Latin America.

| Area                        | Country                 |               |                             |
|-----------------------------|-------------------------|---------------|-----------------------------|
| Africa                      | Angola                  | Ethiopia      | Namibia                     |
|                             | Benin                   | Gabon         | Nigeria                     |
|                             | Botswana                | Gambia, The   | Rwanda                      |
|                             | Burkina Faso            | Ghana         | Sierra Leone                |
|                             | Burundi                 | Guinea        | South Africa                |
|                             | Cameroon                | Guinea-Bissau | Swaziland                   |
|                             | Central Africa Republic | Kenya         | Togo                        |
|                             | Chad                    | Lesotho       | Uganda                      |
|                             | Congo                   | Liberia       | United Republic of Tanzania |
|                             | Cote d'Ivoire           | Malawi        | Zambia                      |
|                             | Djibouti                | Mali          | Zimbabwe                    |
|                             | Equatorial Guinea       | Mozambique    |                             |
| Caribbean/<br>Latin America | Bahamas                 | Guyana        | Jamaica                     |
|                             | Belize                  | Haiti         |                             |

The method of probabilistic population projection for the other non-AIDS countries does not suit those countries due to the special characteristics of AIDS. The disease, different from other ones, causes death to mainly young and middle-aged adults.

Figure 1.1 shows the difference in mortality pattern in South Africa. The year 1981 marks the beginning of the generalized HIV epidemic in this country, so we compare the mortality rate in 1975-1980 (pre-epidemic) and that in 2000-2005 (in epidemic). The age

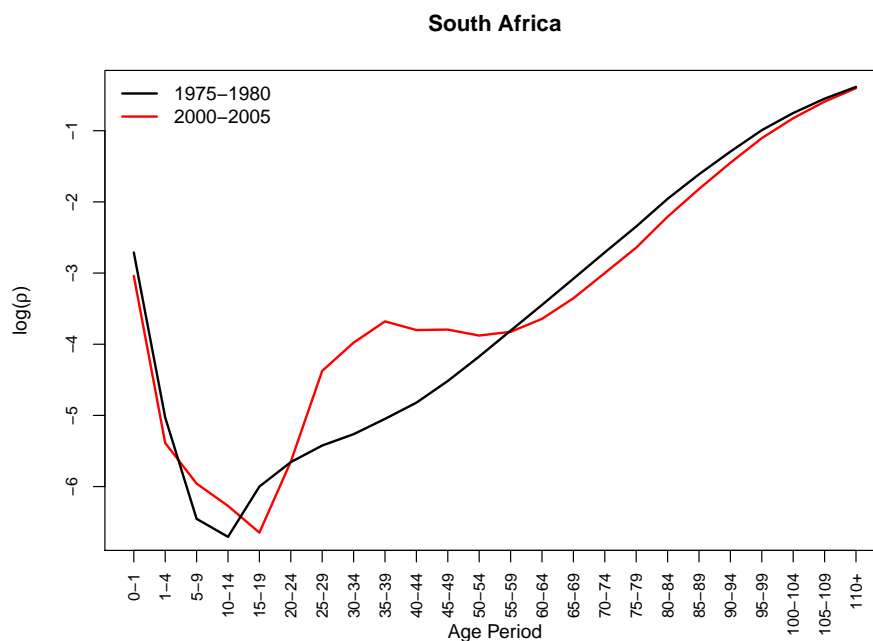


Figure 1.1: Example of mortality rate comparison before and after entering generalized HIV epidemics for South Africa. The plot is drawn on the logarithmic scale of mortality for female in 1975-1980 (black) and 2000-2005 (red) respectively.

pattern exhibits a sharp bump during the female reproductive period. Therefore, the HIV effect should be incorporated in mortality, and consequently, population projection of AIDS countries. Sharrow et al. (2014) developed a method to model the demographic impact of AIDS. The authors modeled the vector of age-specific mortality rates as a weighted sum of three independent age-varying components. They were derived from a Singular Value Decomposition of the matrix of age-specific mortality rate schedules. HIV prevalence is then considered in modeling the weights. It is different from how UNPD calculated the age- and sex-specific mortality for non-AIDS countries in WPP 2012, which used a multi-compartment model (United Nations, Department of Economic and Social Affairs, Population Division, 2013a). The new method of Sharrow et al. (2014), which can be viewed as extending the classic Lee-Carter model (Lee and Carter, 1992), converts the life

expectancy at birth ( $e_0$ ) projection to age- and sex-specific mortality. This method is able to catch the hump in young and middle-aged adults. It can also produce probabilistic projection given probabilistic ones in  $e_0$ .

For HIV prevalence, the UNPD currently uses the method developed by the UNAIDS Reference Group on Estimates, Modelling and Projections (Stover et al., 2012). The estimation and projection can be generated using the Estimates and Projection Package (EPP) software, which implements the UNAIDS model. The classic EPP software considers the change of HIV compartments and can produce projection up to year 2015 (Brown et al., 2010). Spectrum (Stover et al., 2012) is another software developed by UNAIDS which contains EPP in the latest version. The software itself stores the data on HIV prevalence and is easily accessible. In our application, we modified the classic EPP method to allow projections up to year 2100.

We follow the idea outlined in Raftery et al. (2012) of projecting population by starting from life expectancy at birth, total fertility rate and age- and sex-specific mortality rates. We adapt the existing method to account for the HIV effect so that we may apply it to AIDS countries. In Section 2 of this paper, we illustrate our method for estimating and projecting HIV prevalence, incorporating it into the model of life expectancy and, finally, projecting population. In Section 3, we apply our method to obtain probabilistic projections in selected AIDS countries including Kenya, Swaziland, Jamaica, Botswana and Sierra Leone. Finally, we discuss model validation and compare the method we developed with other current methods in Section 4.

## Chapter 2

**METHODOLOGY AND RESULTS**

To implement our population projection method, we need to draw projections of TFR, mortality rate and migration data, and combine them according to the standard cohort component model. The TFR projection can be calculated using the `bayesTFR`<sup>1</sup> package (Ševčíková et al., 2011) in R. This package implements the Bayesian hierarchical model discussed in Alkema et al. (2011). The median of the age-specific fertility rate projection turns out to be the same as WPP 2012. The projections of net international migration we use are also from WPP 2012. The impact of the AIDS epidemic motivates the need to modify age- and sex-specific mortality rates. We begin with estimation and projection of HIV prevalence and model life expectancy, taking into account the effects of HIV. Then we project life expectancy at birth ( $e_0$ ) based on the regression model we developed for life expectancy changes and the projection of HIV prevalence, and they are finally converted into age- and sex-specific mortality rates using the method of Sharrow et al. (2014). At the end, we make use of the same framework as Raftery et al. (2012) in dealing with non-AIDS countries: we combine the fertility, migration and modified mortality projection using the cohort component model. The details of the methodology will be discussed further in the next sections.

**2.1 HIV prevalence probabilistic estimation and projection**

As discussed above, we estimate and project HIV prevalence using the EPP software. The tool, developed by UNAIDS, is designed to fit the observed antenatal clinic (ANC) and

---

<sup>1</sup>The version we use is `bayesTFR` 4.0-3. The start year in simulating the MCMC is set to be 1750 as the default. In this thesis, we do not consider the historical data, the start year is then set to be 1950. In detail, the code for generating the chains is

```
> simulation.dir <- file.path(getwd(), "mylongrun")
> tfr.chains <- run.tfr.mcmc(nr.chains = 5, iter = "auto", start.year = 1950, present.year = 2012,
output.dir = simulation.dir, wpp.year = 2012, parallel = TRUE, nr.nodes = 5, cltype = "SOCK")
> tfr.predict <- tfr.predict(sim.dir = simulation.dir, end.year = 2100, burnin = 2000, nr.traj = 1000)
```

surveillance data to generate short-term projections of HIV/AIDS effect at a country level. The 2009 and subsequent versions of EPP take into account the effect of anti-retroviral therapy (ART). The therapy uses drug combinations which make the infectiousness of the disease lower by decreasing the viral load for a given HIV prevalence. Brown et al. (2010) introduce the dynamics of the compartments in the modified Reference Group model with ART used in EPP 2009. It is shown in Figure 2.1.

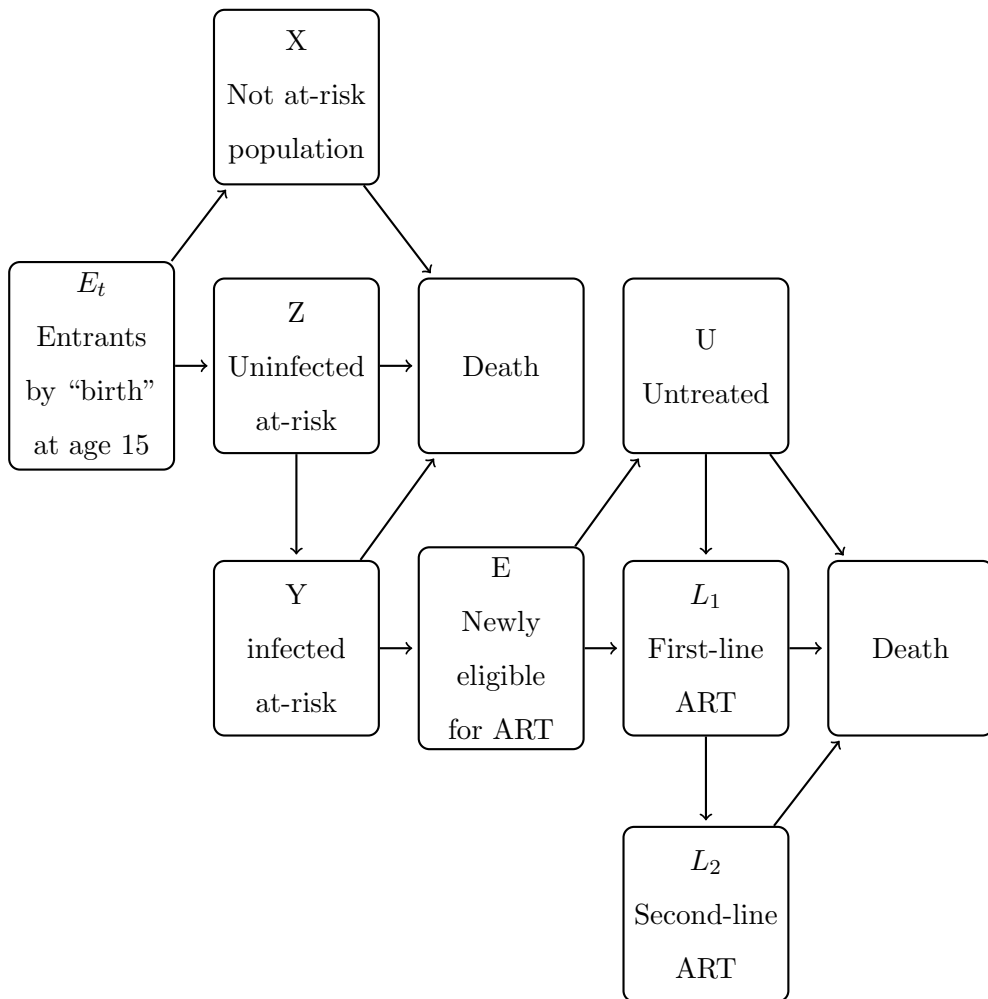


Figure 2.1: The compartments in the HIV model with ART used in EPP 2009 (Brown et al., 2010).

In this model, the population of interest is divided into three groups: those not at risk of HIV, those uninfected but at risk and those infected. Infected individuals have chances to be treated with ART through eligibility for two lines. People who fail to be treated in the first line of ART are moved to the second line. The newly eligible population, together with the eligible but not treated one at previous time, have equal chances to be treated.

However, EPP 2009 only provides short-term projections, up to the year 2015, whereas we want to be able to make projections up to the year 2100. The EPP software gives a flat projection after 2015. To solve this problem, we changed the R code version of the classic EPP model, developed by Le Bao, varying some previously fixed parameters in the infection dynamics. The changes are guided by the assumptions of UN about the trends in future HIV epidemics in AIDS countries (United Nations, Department of Economic and Social Affairs, Population Division, 2013a). The assumptions imposed on HIV prevalence changes for most countries work on two parameters which have remained constant in the past (See Appendix A). The first one is  $\phi$ , which reflects the rate of recruitment of new individuals into the high-risk or susceptible group. This is projected to decline by half every twenty years starting from 2013. The other one is  $r$ , which represents the force of infection. It is projected to decline by half every thirty years after 2013. The reduction in  $r$  reflects the assumption that changes in behavior among those subject to the risk of infection, along with increases in access to treatment for those infected, will reduce the chances of HIV transmission. As opposed to being flat after 2015, the projections of HIV prevalence decreases after 2013 and gets very close to 0 for most of the countries, i.e., the HIV epidemic is wiped out.

The data we used to generate the probabilistic HIV prevalence projection for the 40 AIDS countries are from UNAIDS [<http://apps.unaids.org/spectrum/>]. ANC data are available for all of the 40 countries, while surveillance data from national population based surveys (NPBS) are available for only 23 of them. The revised EPP software uses ANC data, together with the UNAIDS projections of ART coverage (also available from the UNAIDS website given above), to generate trajectories of HIV prevalence up to year 2100. For countries with NPBS data, the software calibrates the national survey result by shifting trajectories on the probit scale when calculating the likelihood for sampling. For countries

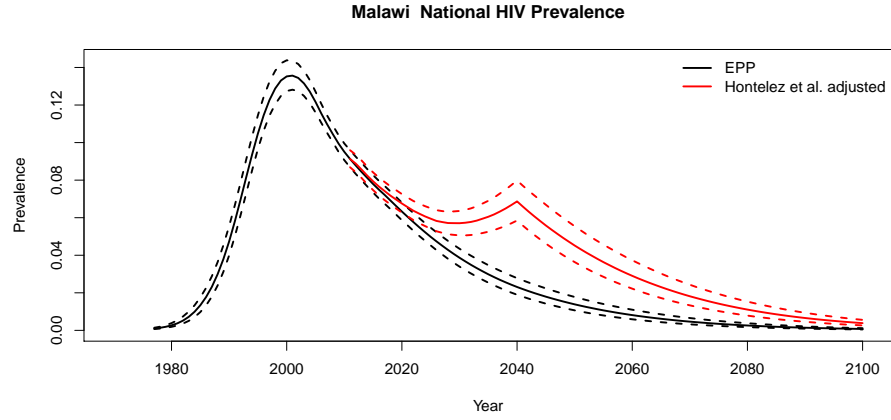
without such data, we manually shift the trajectories after fitting the model using ANC data. The method of calibration follows Alkema et al. (2008). The average bias between probit scaled population based survey prevalence and the ANC one was modeled as normal distributed. Urban and rural areas were treated separately. For each trajectory of HIV prevalence projection in countries without NPBS data, we first sample from the difference given in the literature and shift the trajectory in probit scale by the sampled constant. We then transform it back to the population scale. Formally,

$$\begin{aligned} c_{urban} &\sim N(0.11, 0.04^2), \\ V_{pop} &= V_{ANC} - c_{urban}, \\ \rho_{pop} &= \Phi(V_{pop}), \end{aligned}$$

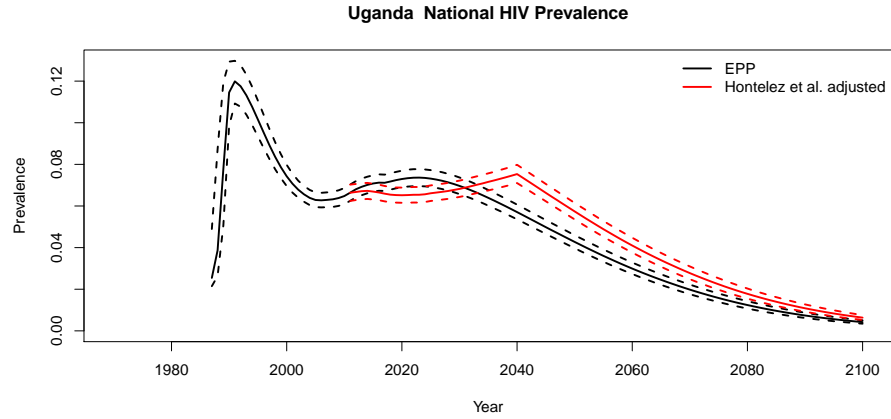
where  $c_{urban}$  is the constant we need to shift the probit-scaled trajectories  $V_{ANC}$  to the probit population prevalence  $V_{pop}$ .  $\rho_{pop}$  is then the prevalence we output after calibration. For rural areas, the constant is sampled from distribution  $c_{rural} \sim N(0.17, 0.05^2)$ .

The implementation of the revised R-version of EPP classic model above gives the HIV prevalence estimation up to year 2010 and projections starting from year 2011 for all the 40 countries. Some example are shown in Figure 2.2. The black curves represent the median estimate and projection for each country. The tail of the curve shows an exponentially decreasing slope towards 0 due to the decrease of at-risk population and of the force of infection. Some countries reach the only peak before 2010 (e.g., Malawi, Figure 2.2a) while others have a second peak during the projection period (e.g., Uganda, Figure 2.2b). This can be explained by characteristics of ART. The therapy not only reduces the infectiousness amongst people but also prolongs the survival of infected individuals. So in the near future, we can foresee people living longer with ART in some countries, which can increase HIV prevalence in the short term.

Hontelez et al. (2012) establishes a stochastic microsimulation model (STDSIM), which also considers the impact of rapidly expanding coverage of ART. By calibrating the projection offered by EPP through the result from STDSIM, it is believed to be more plausible. The authors offered us their deterministic projections for 43 sub-Saharan Africa countries for the years 2011 through 2040. We then follow the same calibration method of shifting



(a) Probabilistic HIV prevalence estimation and projection for Malawi



(b) Probabilistic HIV Prevalence estimation and projection for Uganda

Figure 2.2: Probabilistic HIV prevalence estimation (1980-2010) and projections (2011-2100) for Malawi and Uganda.

probit-transformed trajectories. Specifically, we first perform a probit transformation on the median of EPP projection from 2011 to 2100 ( $\rho_t^{MED}$ ) and take the difference between the transformed median from 2011-2040 and the probit-scaled STDSIM results ( $\rho_t^H$ ). Then we impute the probit level difference from 2041 to 2100 with the same value as the difference in 2040. Formally, the difference  $\lambda_t$  is given by

$$\lambda_t = \begin{cases} \Phi^{-1}(\rho_t^{MED}) - \Phi^{-1}(\rho_t^H) & t = 2011, \dots, 2040, \\ \lambda_{2040} & t = 2041, \dots, 2100. \end{cases}$$

Finally, we shift the probit-scaled EPP trajectories by the difference and transform them back to the original unit. That is, for each trajectory  $\rho_t$ , we adjust it to be

$$\rho_t^{adj} = \Phi(\Phi^{-1}(\rho_t) - \lambda_t).$$

Red curves in Figure 2.2 refer to the Hontelez et al. (2012) calibrated projections ( $\rho_t^{adj}$ ) in HIV prevalence. They do not differ much from the EPP outputs. Although the curves are less smooth, they are still plausible because people live long with ART treatment and possibly in the future, the death of those people and the gradually reducing HIV cases may result in a sharp decrease of HIV prevalence. We will use them as our probabilistic projection in the following paragraphs for life expectancy and mortality projection.

## 2.2 Life expectancy model and probabilistic projection

To prepare for the mortality projection, we need the life expectancy ( $e_0$ ) projection for the AIDS countries. For countries without significant level of HIV/AIDS epidemics, the UNPD uses the method developed by Raftery et al. (2013). The authors modeled the expected five-year change in female life expectancy in country  $c$  from time  $t - 5$  to  $t$  ( $\Delta e_{0,c,t} = e_{0,c,t} - e_{0,c,t-5}$ ) as a double logistic function of the life expectancy at time  $t - 5$  ( $e_{0,c,t-5}$ ) with country-specific parameter vector  $\theta^c$ , where  $c$  indexes country. They estimate the parameter for each country as well as the parameter for the rest of the world excluding AIDS countries using a Bayesian hierarchical model. The double logistic curve under the “world” parameters  $\theta^w$  is plotted in Figure B.1 (More details in Appendix B). If  $e_0$  for the current year is less than 60, the “world” double logistic curve predicts that the change of life expectancy in the next five years will increase and that the pace will decline as current  $e_0$  goes up. However, if  $e_0$  is larger than about 60, the change in the next five years will decrease and the change plateaus to around 0.6 years of life expectancy per five-year period. After fitting the model for female life expectancy in each country, Raftery et al. (2013) modeled the difference between male and female life expectancy for each country, using the method of Raftery et al. (2014). The life expectancy for males can thus be projected.

Here we consider the male and female life expectancy in the same country simultaneously and model the country-level life expectancy in a regression model. Due to the impact of AIDS and the use of ART, we consider HIV prevalence and ART coverage in addition to the effect of lagged life expectancy in the regression model. We build the regression model based on the data on life expectancy from WPP 2012, HIV prevalence from our median projection from EPP after adjusting according to Hontelez et al. (2012) results, and the ART coverage published by UNPD. All data are provided at a five-year frequency from 1990 to 2010. Here we use 1990 as a label for the mid-year of 1985-1990, and so forth.

We use observations pooled from all countries to estimate the model. The response variable in our regression model is the change in life expectancy from time  $t - 5$  to time  $t$ , i.e.,

$$\Delta e_{0,t} = e_{0,t} - e_{0,t-5} = \begin{pmatrix} e_{0,1,2010} \\ \vdots \\ e_{0,40,2010} \\ \vdots \\ e_{0,40,1995} \end{pmatrix} - \begin{pmatrix} e_{0,1,2005} \\ \vdots \\ e_{0,40,2005} \\ \vdots \\ e_{0,40,1990} \end{pmatrix},$$

where the second subscript in each element of the vector refers to the different countries. Our predictors include the five-year lagged life expectancy  $e_{0,t-5}$ , HIV prevalence  $\rho_t$  in percentage and the ART coverage  $A_t$  in percentage. Formally, they are

$$e_{0,t-5} = \begin{pmatrix} e_{0,1,2005} \\ \vdots \\ e_{0,40,2005} \\ \vdots \\ e_{0,40,1990} \end{pmatrix}, \quad \rho_t = \begin{pmatrix} \rho_{1,2010} \\ \vdots \\ \rho_{40,2010} \\ \vdots \\ \rho_{40,1995} \end{pmatrix}, \quad A_t = \begin{pmatrix} A_{1,2010} \\ \vdots \\ A_{40,2010} \\ \vdots \\ A_{40,1995} \end{pmatrix}.$$

Also, we organize the data by dividing the them into two parts, the first corresponding to  $\rho_t$  greater than or equal to 1%, the second to  $\rho_t$  less than 1%. This was inspired by the threshold of generalized HIV epidemics as the aim of the model is to capture the change of life expectancy over the period of the generalized HIV epidemics. The lowest plot for the two groups are shown in Figure 2.3. The shape of the lowest curve for the data observations

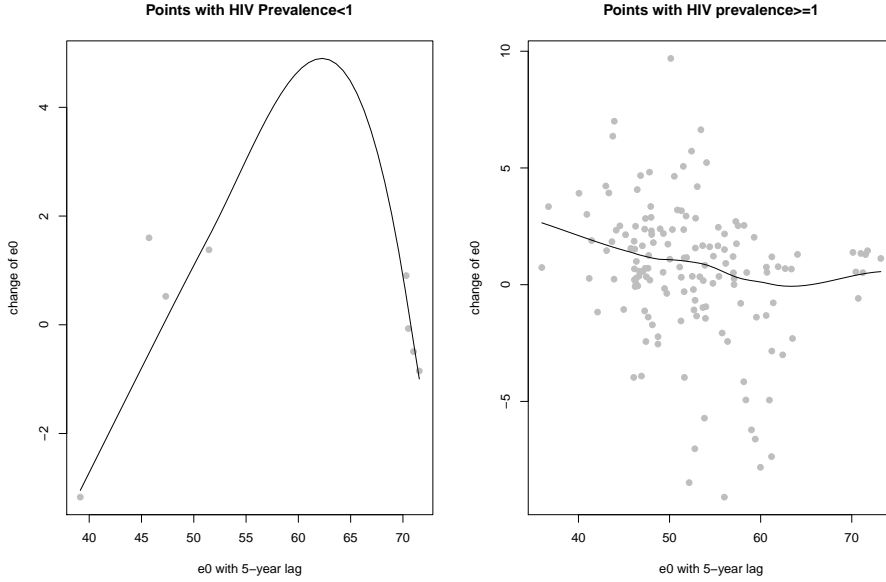


Figure 2.3: Lowess plot for observations with HIV prevalence less than 1% and greater than or equal to 1%.

pre-epidemic is similar to a double logistic curve, which matches the non-AIDS countries. For the set of observations with  $\rho_t \geq 0$ , this is not the case.

However, linear regression ends up with too sharp decreasing effect from HIV prevalence and may predict negative life expectancy in the future. Also we want the model to be close enough to Raftery et al. (2013) when those AIDS countries step out of epidemics. Therefore, in order to be compatible with the model for non-AIDS countries, we consider a double logistic transformation of  $e_{0,t-5}$  to  $g(e_{0,t-5})$ , where  $g(\cdot)$  refers to the double logistic curve with the “world” parameter  $\theta^w$  estimated by UNPD (See Appendix B for the definition of the function and the values of parameter). The model is

$$\Delta e_{0,t} = \beta_0 + \beta_1 g(e_{0,t-5}) + \beta_2 \gamma_t + \epsilon_t, \quad (2.1)$$

where  $\gamma_t = \rho_t \times (1 - A_t/100)$ , which approximates the proportion of the population that is HIV positive but not treated by ART. And the error term  $\epsilon_t$  is assumed to be i.i.d.  $N(0, \sigma^2)$ . In estimating the coefficients, we set  $\beta_1 = 1$  by default to match with the model

for non-AIDS countries. The fitted model is presented in Table 2.1.

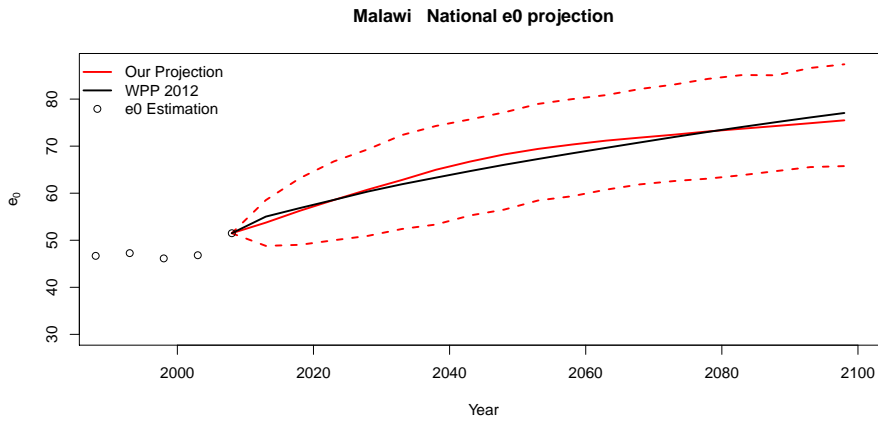
Table 2.1: Model fitted for the regression model of  $\Delta e_{0,t}$ .

| $R^2=0.32$ | Estimate | Standard Error |
|------------|----------|----------------|
| $\beta_0$  | -0.67    | 0.30           |
| $\beta_1$  | 1        | –              |
| $\beta_2$  | -0.28    | 0.03           |
| $\sigma^2$ | 6.38     |                |

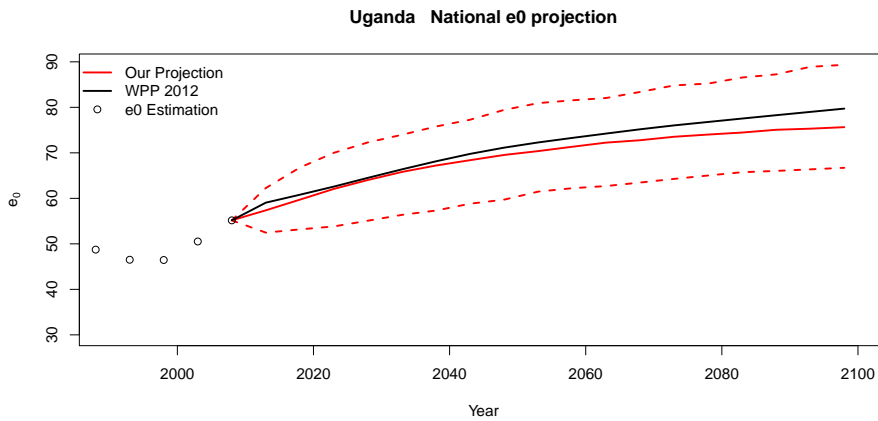
The model fitted shows that the impact of  $\gamma_t$  is negative, i.e., the smaller the number of HIV positive people not treated by ART, the longer the life expectancy at birth. Based on the fitted model above, we can project probabilistically the life expectancy five years by five years. Specifically, we take each trajectory of HIV prevalence in one country and its ART coverage up to year 2100. The life expectancy in 2015,  $e_{0,2015}$ , for the corresponding trajectory is predicted by adding  $e_{0,2010}$  to the change predicted from applying the model to  $e_{0,2010}$ ,  $\rho_{2015}$  and  $\gamma_{2015}$ . Similarly, we can get predictions for 2020 up to the year 2100. Figure 2.4 presents the life expectancy projections for Malawi and Uganda. The red curves stand for the projection from model (1) and the black one the deterministic projection from WPP 2012. The median projection matches the WPP 2012 deterministic projection well.

### **2.3 Probabilistic population projection**

After obtaining the probabilistic projection of HIV prevalence and life expectancy, we can incorporate them using the mortality model developed in Sharrow et al. (2014). The model takes life expectancy, HIV prevalence and ART coverage as inputs and returns a table for age- and sex-specific mortality. Total fertility rate is also needed for population projection. We use the fertility projection method proposed by Alkema et al. (2011); the method relies on a Bayesian hierarchical model which is implemented in the R package `bayesTFR`. This tool offers a probabilistic TFR projection for all countries. Our median projection of TFR is the same as the WPP medium projection, which was found by the same method.



(a) Probabilistic life expectancy projection for Malawi



(b) Probabilistic life expectancy projection for Uganda

Figure 2.4: Probabilistic life expectancy projections (2011-2100) for Malawi and Uganda.

For probabilistic population projection of countries without generalized HIV/AIDS epidemics, Raftery et al. (2012) proposed combining the projection on age-specific TFR and age- and sex-specific mortality rate in a standard cohort component model (Whelpton, 1936; Preston et al., 2001) to generate projection on population size. Since the effect of AIDS reflects the bump in mortality and we have incorporated it into the model and projection, we use the same method outlined in Raftery et al. (2012) to perform population projection.

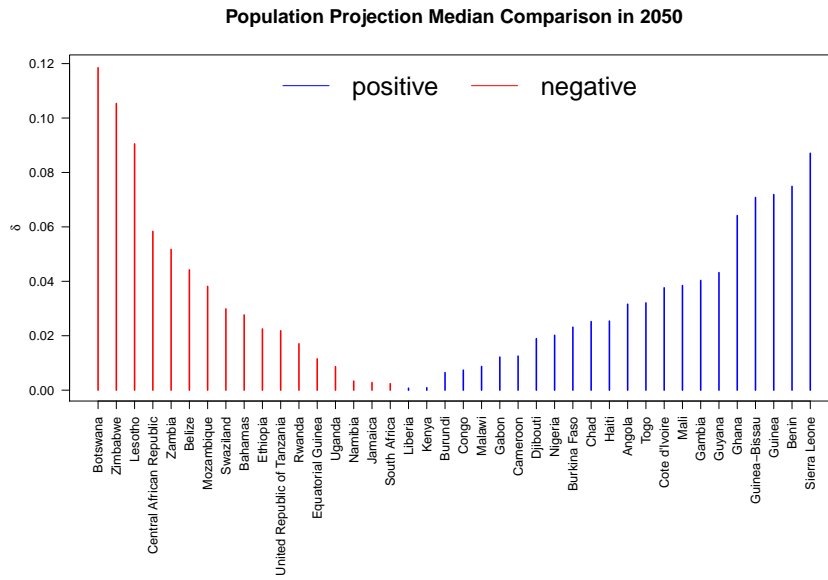
Technically, we can directly use the R package `bayesPop` (Ševčíková, 2013); however, we make slight adjustments. We want it to use the age- and sex-specific mortality rate directly instead of updating mortality with the life expectancy input inherent in the model. For each of the 40 countries, we have already obtained 1,000 trajectories for fertility and mortality projection every five years from 2015 to 2100. We then perform the adjusted `bayesPop` method to obtain 1,000 trajectories for each of the 40 countries.

Selected case results will be discussed in the next section. Figure 2.5 summarizes the discrepancy between the median of our projection ( $X_{MED}$ ) and projections offered in WPP 2012 ( $X_{WPP}$ ) for each of the 40 AIDS countries in 2050 and 2100. The difference is defined as

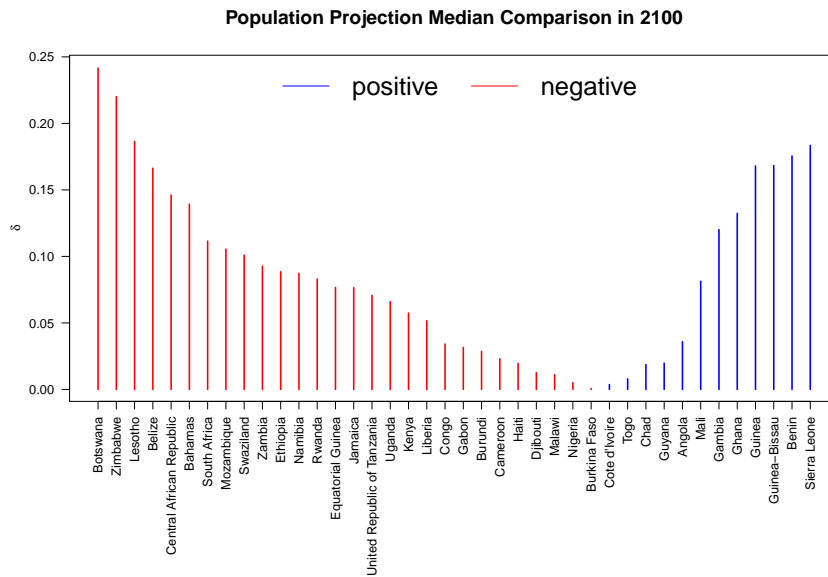
$$\delta = \frac{X_{MED} - X_{WPP}}{X_{WPP}}.$$

The y-axis of the plot is the absolute value of  $\delta$ . Red signifies negative differences, blue positive differences. Seventeen of the 40 countries have negative differences in 2050, showing that for those countries, we project lower population than UNPD does. Amongst them, Botswana and Zimbabwe have the largest difference, which is more than 10%. For the other 23 countries, we project a positive difference with Sierra Leone as the highest (around 9%).

In 2050, our projections for countries like South Africa and Liberia are similar to the ones from WPP 2012. The same cannot be said for 2100. Most of the countries (33 out of 40) are projected to have smaller population from our method compared to that projected in WPP 2012. Botswana and Zimbabwe have the largest negative difference in 2100, greater than 25%. On average, we project about 8.56 million more people in 2050 than WPP 2012 does, but around 82.51 million less in 2100. This can be explained by the effect of ART that keeps more HIV positive people alive longer in the near future. This is counterbalanced by the long term effect of the disease, despite its becoming rarer.



(a) Difference of population projection median with WPP 2012 in 2050



(b) Difference of population projection median with WPP 2012 in 2100

Figure 2.5: Difference of population projection for the 40 countries in 2050 and 2100. Blue curves stand for positive difference and red is for negative.

## Chapter 3

### CASE STUDY

In this section, we aim to explore the projections obtained using our method for selected countries: for example, countries with high HIV prevalence, countries with highest or lowest projection differences from WPP 2012 and countries outside of Africa.

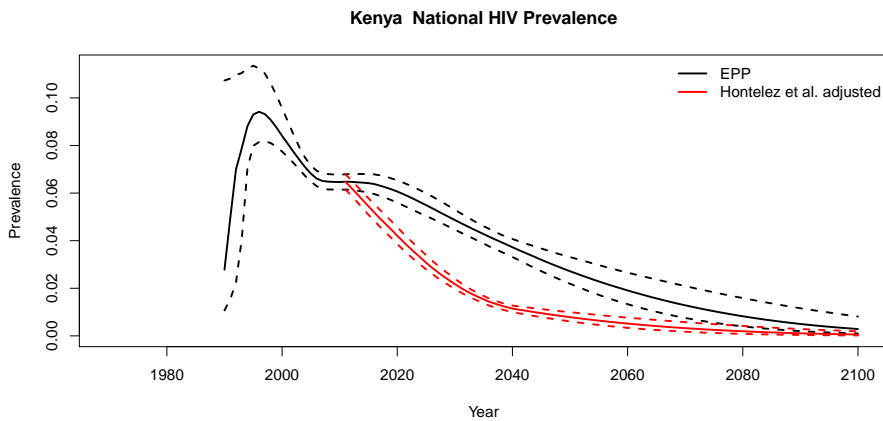
#### **3.1 Typical case: Kenya**

Kenya is a typical country with generalized HIV prevalence estimated at 6.5% over the period of 2005-2010 based on our EPP median output. UNAIDS reported the HIV prevalence in Kenya to be 6.2% as of December 2011 (UNAIDS, 2012). That matches well with our projection shown in Figure 3.1a. Kenya thus has a moderate generalized HIV/AIDS epidemics, so we can regard it as a typical case for study.

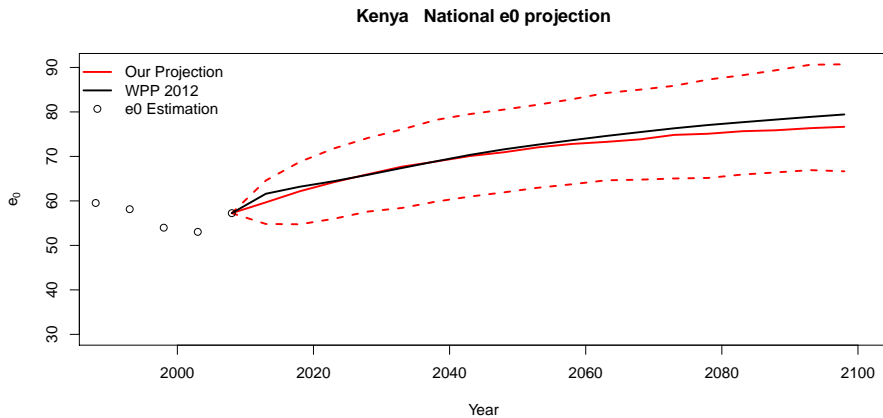
Life expectancy at birth is estimated to be 57.2 at mid-year of period 2005-2010, as reported by WPP 2012. As shown in Figure 3.1b, we project that, in the sense of median,  $e_0$  in Kenya will be around 70 in 2050 and about 75 at 2100. The increasing pace slows down as time goes by; as the life expectancy increases, its change in the next five years is influenced by the double logistic function (Figure B.1), which decreases when life expectancy is higher than about 60.

Kenya is one of the countries with the largest population among all the 40 countries. In 2005-2010, the UN estimated its population to be 41.0 million, ranking fifth among the 40 countries. Then for population projection, our median prediction there is 97.3 million people in Kenya in 2045-2050 (Figure 3.2a), close to the WPP 2012 prediction of 97.2 million. In 2095-2100, we project the population to be 151.2 million, compared with the projected 160.4 million from WPP 2012. The difference is not large, meaning that our projection is similar to that from WPP 2012. Figure 3.2 also plots the projected population of females of reproductive age (15-49, Figure 3.2b) as well as the population of new births (aged 0-4,

Figure 3.2c). We find that the difference from WPP 2012 is small. If we check the mortality change among women aged 10-44 (Figure 3.2d), the two projections align well, showing that the life pattern in Kenya projected from our method is similar to what UNPD projected in WPP 2012.

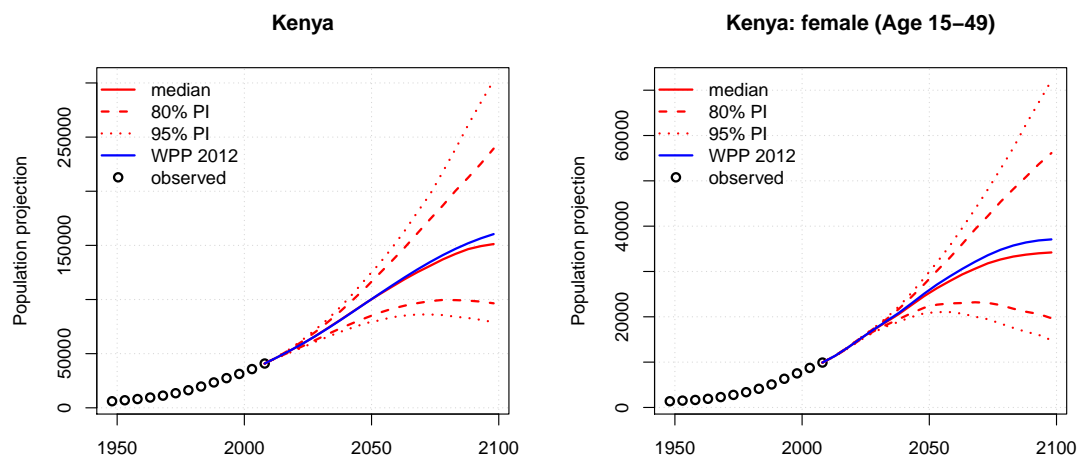


(a) Probabilistic HIV prevalence projection for Kenya

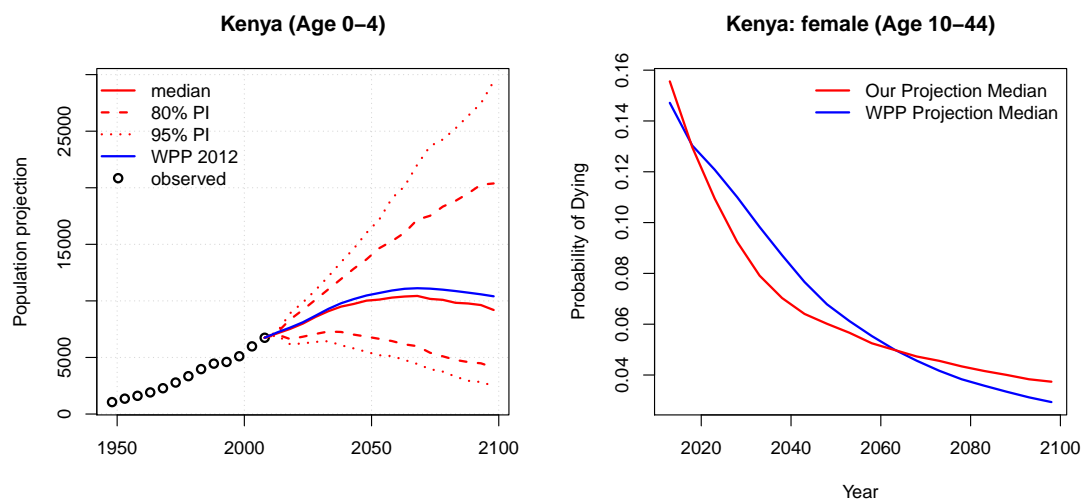


(b) Probabilistic life expectancy projection for Kenya

Figure 3.1: Probabilistic HIV prevalence and life expectancy projections (2011-2100) for Kenya.



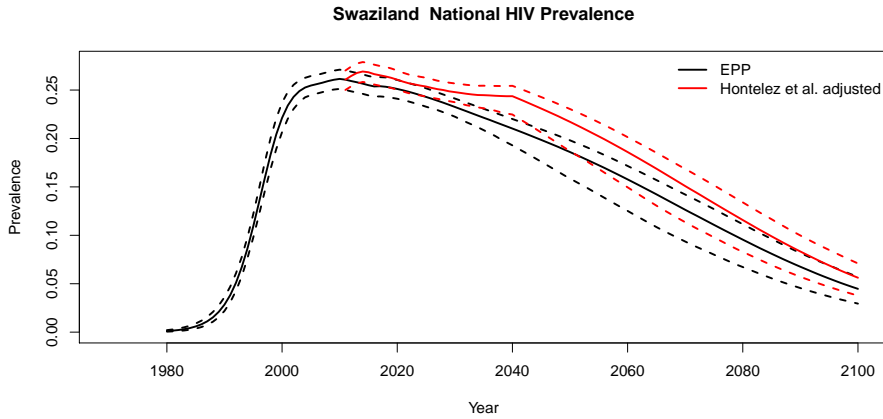
(a) Probabilistic population projection for Kenya  
 (b) Probabilistic population projection of female 15-49 for Kenya



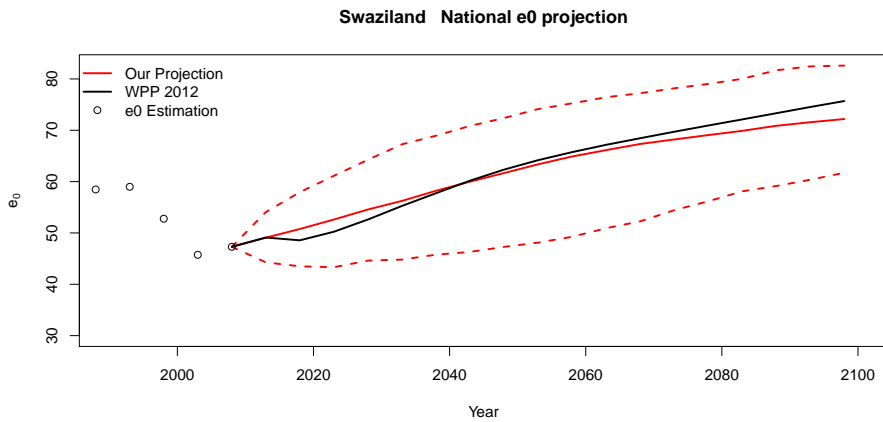
(c) Probabilistic population projection of age 0-4 for Kenya  
 (d) Mortality difference for female 10-44 from WPP 2012 for Kenya

Figure 3.2: Summary of population projection for Kenya up to year 2100.

### 3.2 High HIV prevalence: Swaziland



(a) Probabilistic HIV prevalence projection for Swaziland



(b) Probabilistic life expectancy projection for Swaziland

Figure 3.3: Probabilistic HIV prevalence and life expectancy projections (2011-2100) for Swaziland.

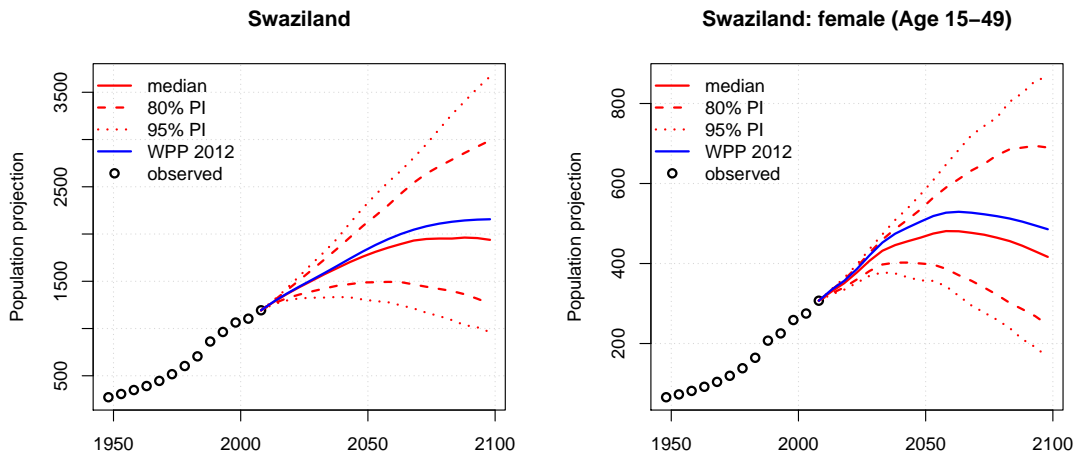
Swaziland claims the highest HIV prevalence in 2005-2010 at 16.5% as estimated from our model, compared to 26% in 2011 from UNAIDS (2012). It declines more slowly than Kenya does. If we compare Figure 3.3a with Figure 3.1a, in 2100, HIV prevalence is projected to

still be well above the threshold of generalized epidemics (only slightly less than 5%). From Figure 3.3b, we can see that the HIV prevalence in Swaziland reaches a peak at around 2011 and declines afterwards. In 2045-2050, we project the median HIV prevalence to be about 20%, higher than all the other countries. Compare with WPP 2012, our  $e_0$  projection covers the UN projection well.

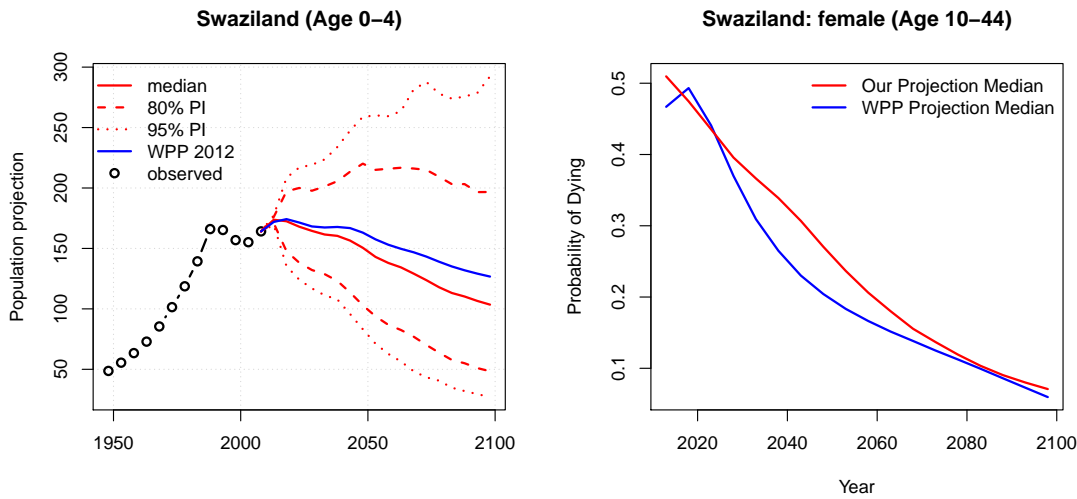
The population in Swaziland in 2005-2010 is estimated to be 1.12 million (Figure: 3.4a), which is relatively small in size (rank 35 out of 40) and the absolute discrepancy between our projection and the WPP 2012 projection is small—the former being 0.22 million less—but proportionally, our projection estimate is 10.1% less than that given in WPP 2012 for the year 2100. If we look at the components of population, the difference of population of female aged 14-49 and the new births are bigger proportionally (Figure 3.4b, 3.4c), compared with Kenya. Figure 3.4d on mortality comparison can explain some of the discrepancy. Since we project higher mortality rates for female aged 10-44 in the middle part of the century, more women are expected to die during or before entering the reproductive period. Therefore we predict a smaller female population and fewer births. The difference keeps non decreasing all the way to year 2100 since the mortality of WPP 2012 never exceeds the one we projected for females aged 10-44. The effect of AIDS on the population is thus reflected in changes in the mortality pattern.

### **3.3 Caribbean country: Jamaica**

Most countries with generalized HIV/AIDS epidemics are African countries, largely focused in the sub-Saharan region. Only five countries are from the Caribbean area or Latin America. Jamaica will be studied as an example here. HIV prevalence in Jamaica is estimated to be 2.4% in 2005-2010 from EPP, compared to 1.8% in 2010-2011 from UNAIDS (UNAIDS, 2012). Our HIV prevalence estimation and projection (Figure 3.5a) comes with large uncertainty bounds due to the small amount of data from ANC and lack of data from NPBS. This leads to a significant gap between our projected median for  $e_0$  and that from WPP 2012 (Figure 3.5b). For these Caribbean/Latin American countries, the WPP 2012 reports a linear increase of life expectancy. Our projection shows a flatter increase due to the current relatively high life expectancy and the effect of double logistic curve.

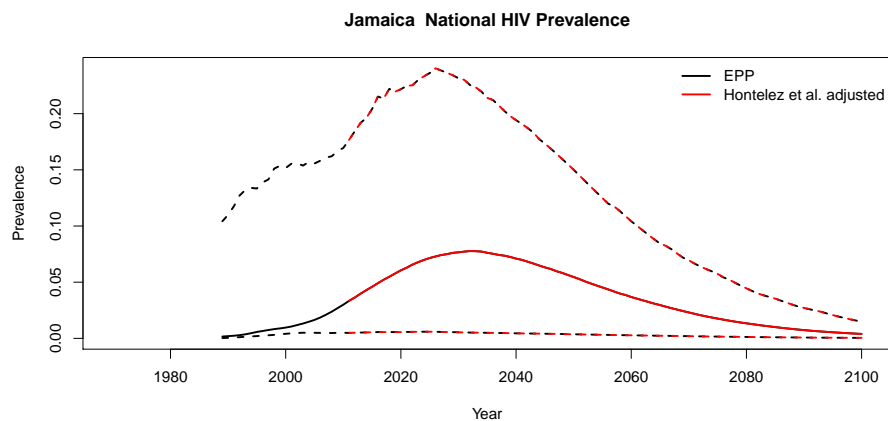


(a) Probabilistic population projection for Swaziland (b) Probabilistic population projection of female 15-49 for Swaziland

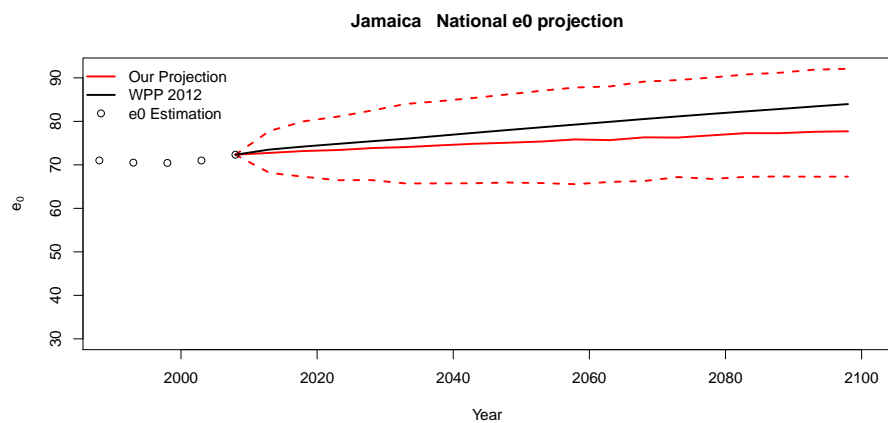


(c) Probabilistic population projection of age 0-4 for Swaziland (d) Mortality difference for female 10-44 from WPP 2012 for Swaziland

Figure 3.4: Summary of population projection for Swaziland up to year 2100.



(a) Probabilistic HIV prevalence projection for Jamaica



(b) Probabilistic life expectancy projection for Jamaica

Figure 3.5: Probabilistic HIV prevalence and life expectancy projections (2011-2100) for Jamaica.

Amongst the five Caribbean/Latin America countries, Jamaica has the second largest population, estimated to be 2.74 million in 2005-2010, in second place after Haiti. The population, as we can see from Figure 3.6a, will decrease after about 2030. This happens sooner than for most of the African countries. The discrepancy between our projection compared with that of WPP 2012 is not large, almost nonexistent in 2050 and only 0.16

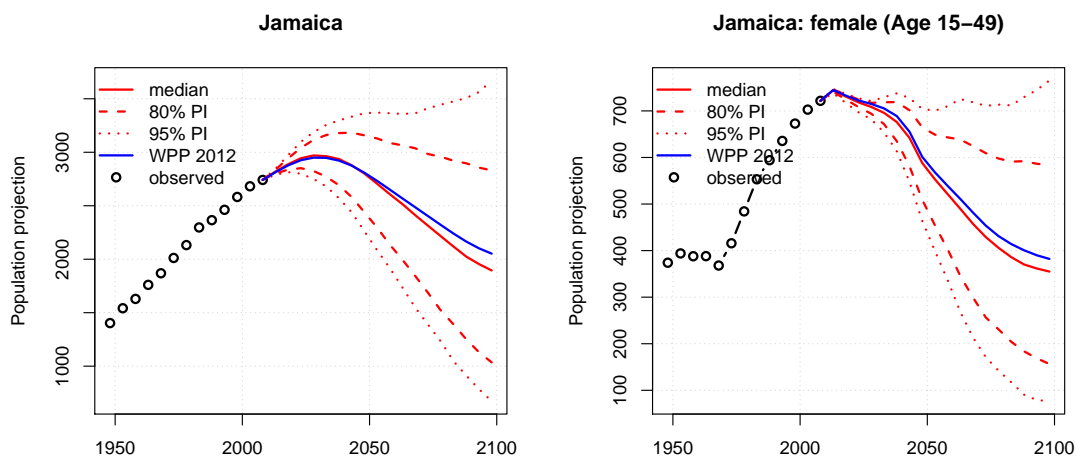
million in 2100. Since we project lower median life expectancy at birth, higher mortality is expected, which implies lower population projections. So the result is reasonable. The mortality comparison plot supports our previous explanation (3.6d). Overall, Jamaica, as a typical Caribbean AIDS country, has higher HIV prevalence, compared with other non-African countries. On the other hand, its life expectancy is higher than that of African countries. Both methods (for either AIDS countries or non-AIDS countries) for population projection can give reasonable results.

### ***3.4 Negative discrepancy from WPP 2012: Botswana***

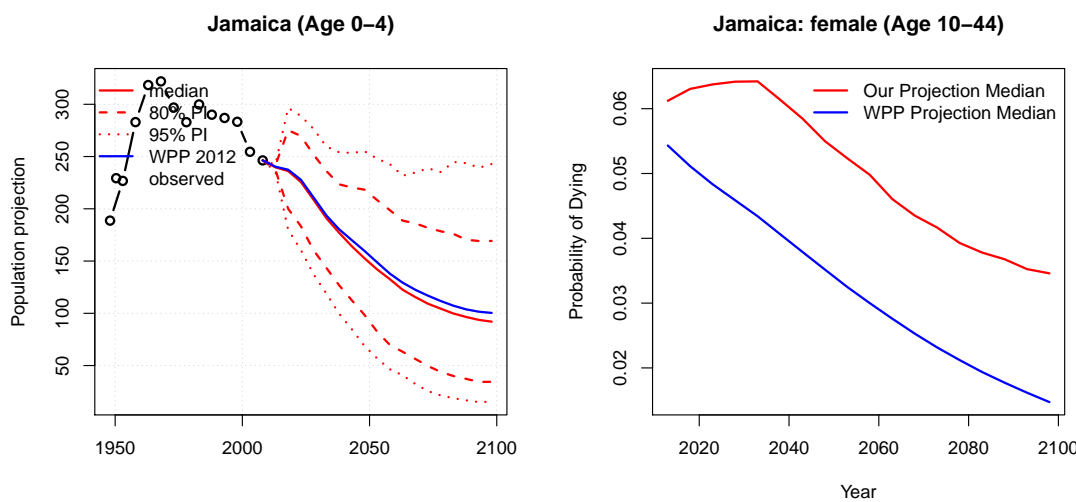
As we summarized at the end of Section 2, Botswana presents the highest negative population projection discrepancy from WPP 2012 in both 2050 and 2100, as high as 25% less in 2100 (median). Botswana is experiencing one of the largest HIV/AIDS epidemics in the world with prevalence estimated at 24.7% in 2005-2010 (23.4% from UNAIDS (UNAIDS, 2012)). Though declining over time, it is still projected to be as high as 15% in 2050 and roughly 2% in 2100 (Figure 3.7a). This result in a sustained lower projected life expectancy than what is given in WPP 2012 (Figure 3.7b).

Based on our calculation, our median projection for 2050 is 2.45 million for Botswana, compared with the 2.78 million projected in WPP 2012. The difference increases from 0.33 million in 2050 to 0.73 million in 2100. Due to the lower life expectancy we get, shorter life are expected under our settings and thus fewer reproductive female aged 15-49 and newborns are predicted, as seen from Figure 3.8b, 3.8c.

The large discrepancy is due to the effect of mortality difference. Since we consider HIV impact in mortality model, we project much higher mortality after 2020. The discrepancy is maintained up to year 2100 (Figure 3.8d). So accumulatively, we expect to see fewer and fewer people in Botswana than predicted by WPP 2012. We also note that both our method and WPP 2012 predict the beginning of population decline at around 2075, and this can be explained by the high mortality rate arising from high HIV prevalence and declining fertility.

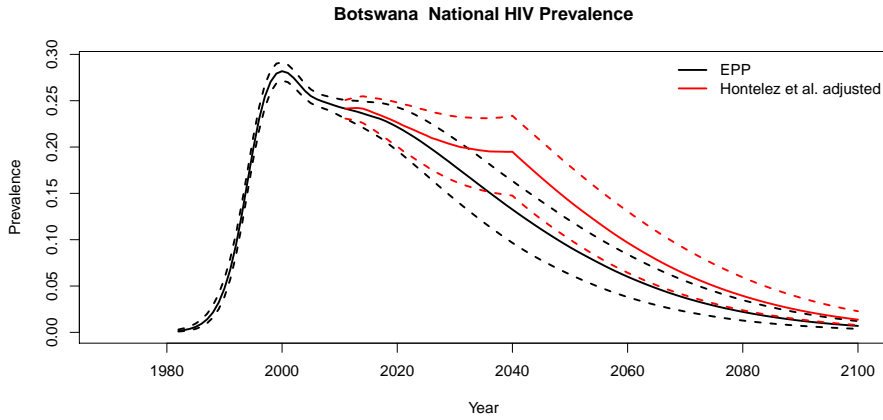


(a) Probabilistic population projection for Jamaica  
 (b) Probabilistic population projection of female 15-49 for Jamaica

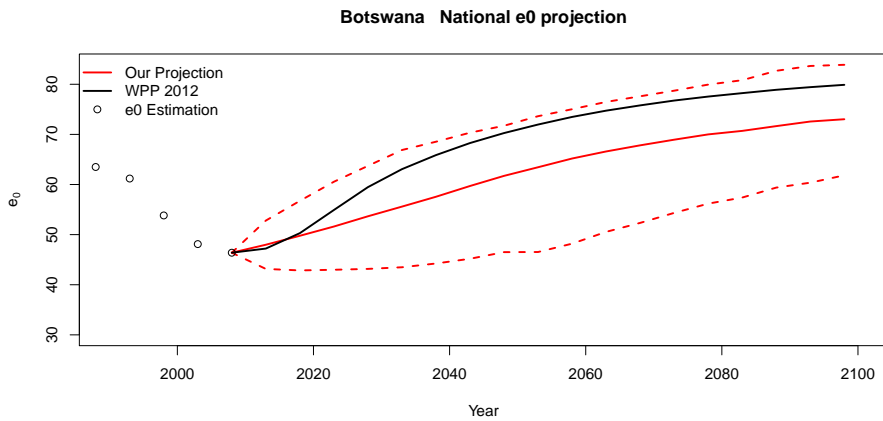


(c) Probabilistic population projection of age 0-4 for Jamaica  
 (d) Mortality difference for female 10-44 from WPP 2012 for Jamaica

Figure 3.6: Summary of population projection for Jamaica up to year 2100.



(a) Probabilistic HIV prevalence projection for Botswana

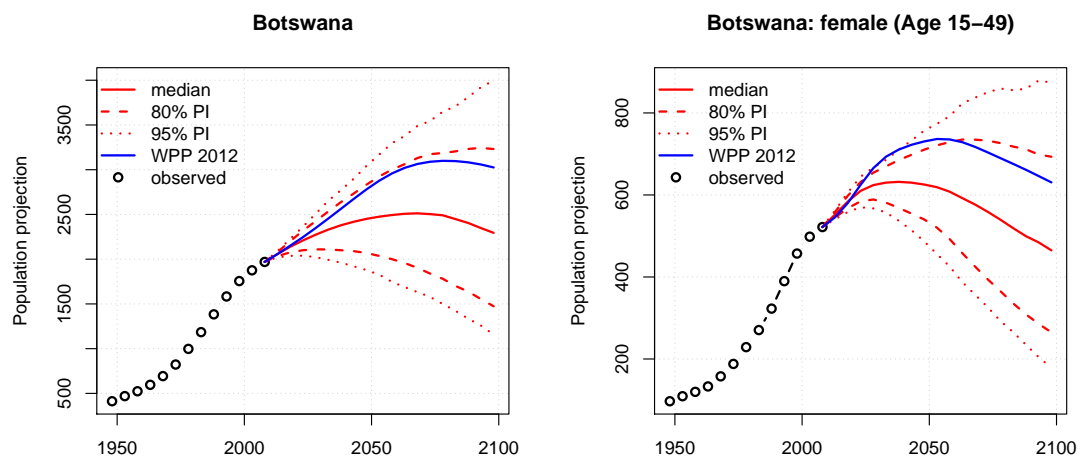


(b) Probabilistic life expectancy projection for Botswana

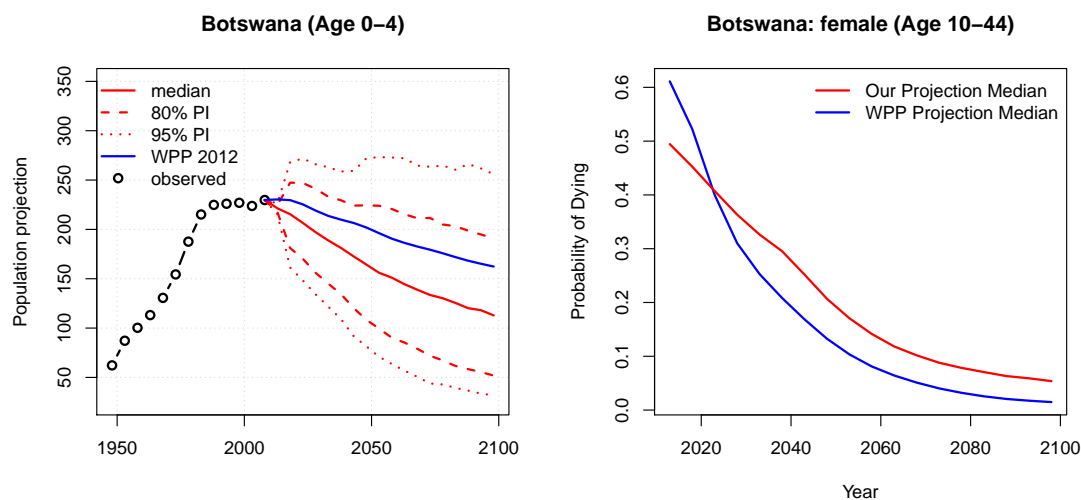
Figure 3.7: Probabilistic HIV prevalence and life expectancy projections (2011-2100) for Botswana.

### 3.5 Positive discrepancy from WPP 2012: Sierra Leone

The last case we consider in the paper is Sierra Leone, for which we project a high positive difference in projected population size from WPP 2012 in both 2005 and 2100. First of all, its HIV prevalence is estimated to be 1.7% in 2005-2010, compared to 1.6% from UNAIDS

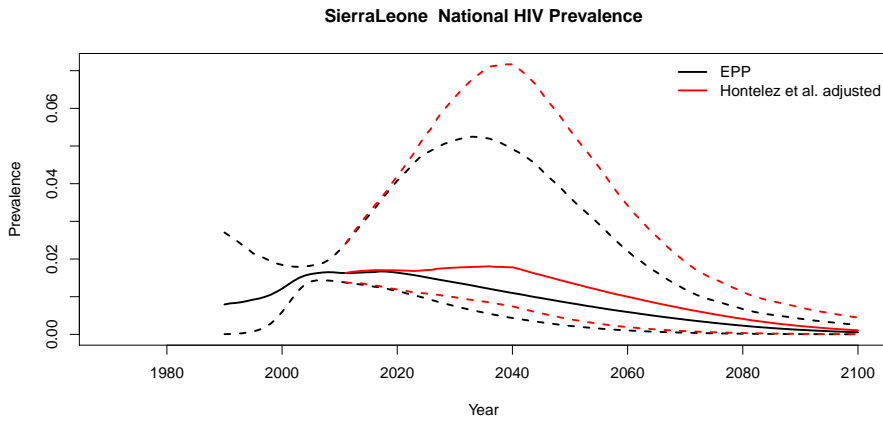


(a) Probabilistic population projection for Botswana (b) Probabilistic population projection of female 15-49 for Botswana

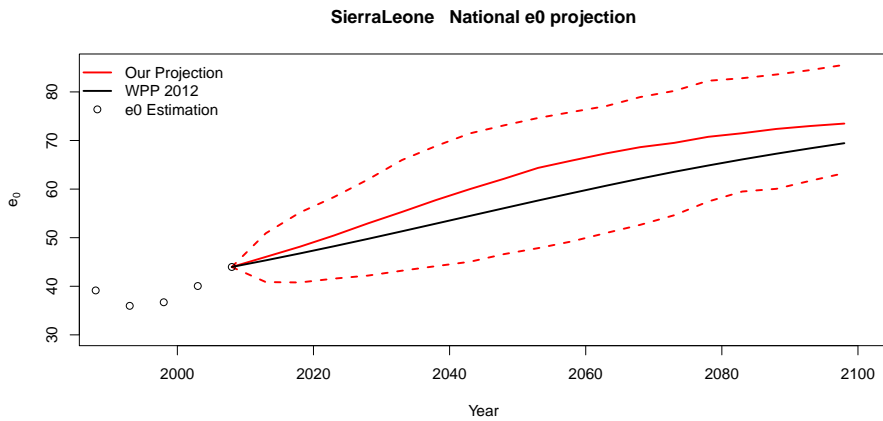


(c) Probabilistic population projection of age 0-4 for Botswana (d) Mortality difference for female 10-44 from WPP 2012 for Botswana

Figure 3.8: Summary of population projection for Botswana up to year 2100.



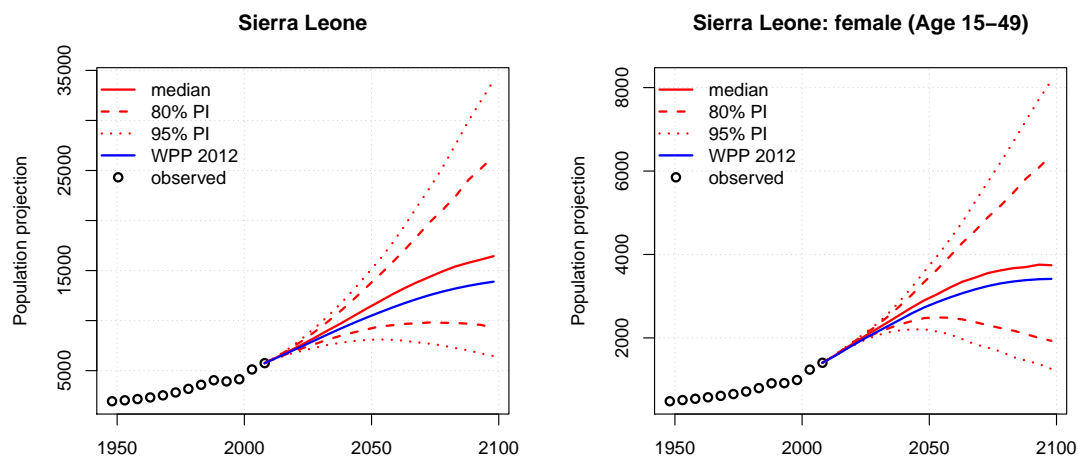
(a) Probabilistic HIV prevalence projection for Sierra Leone



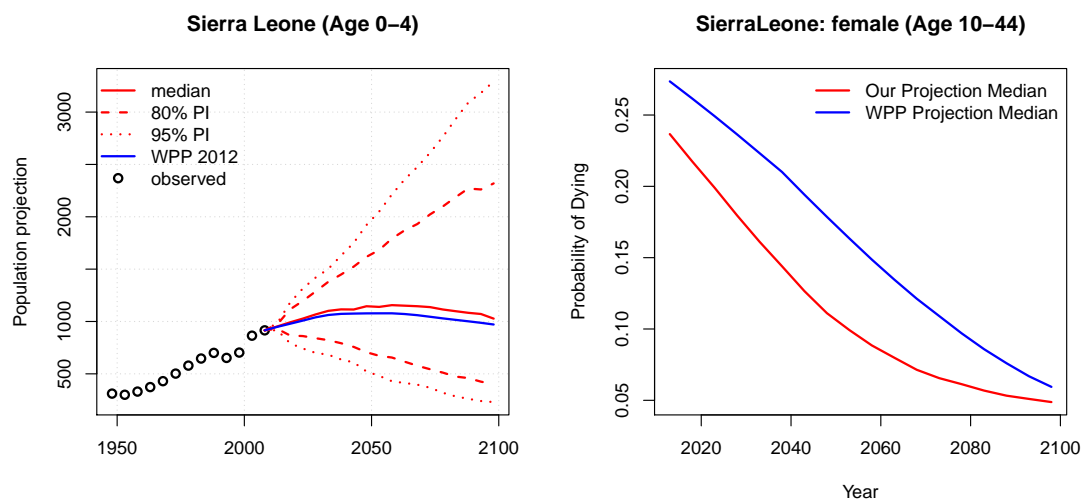
(b) Probabilistic life expectancy projection for Sierra Leone

Figure 3.9: Probabilistic HIV prevalence and life expectancy projections (2011-2100) for Sierra Leone.

(UNAIDS, 2012). Since only a small amount of ANC data is available for Sierra Leone, the uncertainty of the HIV prevalence projection is high, but overall, the HIV prevalence in the country is low. The epidemic is projected to die out soon (Figure 3.9a). Therefore, the decrease in life expectancy resulting from HIV effect is small and the “world” parameter for double logistic provides a higher projection for Sierra Leone than WPP 2012 projects.



(a) Probabilistic population projection for Sierra Leone (b) Probabilistic population projection of female 15-49 for Sierra Leone



(c) Probabilistic population projection of age 0-4 for Sierra Leone (d) Mortality difference for female 10-44 from WPP 2012 for Sierra Leone

Figure 3.10: Summary of population projection for Sierra Leone up to year 2100.

Hence, some countries like Sierra Leone may encounter higher life expectancy projection under our method, compared to that given by WPP 2012 (Figure 3.9b). The subsequent influence is thus higher population projection (Figure 3.10a).

Sierra Leone possesses 5.75 million people in 2005-2010. From the method discussed in this paper, we project a median population size of 1.12 million in 2050, compared to 1.03 million projected by WPP 2012. This difference is 0.09 million in 2050 and it increases to 2.55 million in 2100. The discrepancy in the projections is due to the difference in the projected number of females aged 15-49 and newborns (Figure 3.10b, 3.10c). This is because we project a lower number of deaths, as shown in Figure 3.10d. We calculate far fewer deaths than WPP 2012 for females aged 10-44. So we expect there to be more females of reproductive age over the next five years. This results in more newborns.

Chapter 4  
**VALIDATION**

To assess the performance of our method, we perform model validation for our method, with emphasis on life expectancy and population projections. We compare the results obtained using our method with the ones obtained by ignoring the effect of HIV as is done in **bayesLife** and **bayesPop**.

Since the HIV data collection began not too long ago, we only perform one-out-of-sample validation, which means that we remove all the data and observations in 2005-2010 and use only data before 2005 to do fit models and do projections. Note that the both life expectancy and population projection is based on the estimation of HIV prevalence. However, there are no ANC data or NPBS data for Sierra Leone and Liberia, so we can only work on 38 countries (excluding Sierra Leone and Liberia from the 40 countries) for validation. After getting the median estimation of probabilistic HIV prevalence before 2005, we fit the same regression model as equation (1), but with response

$$\Delta e_{0,t} = e_{0,t} - e_{0,t-5} = \begin{pmatrix} e_{0,1,2005} \\ \vdots \\ e_{0,38,2005} \\ \vdots \\ e_{0,38,1995} \end{pmatrix} - \begin{pmatrix} e_{0,1,2005} \\ \vdots \\ e_{0,38,2005} \\ \vdots \\ e_{0,38,1990} \end{pmatrix},$$

and predictors

$$e_{0,t-5} = \begin{pmatrix} e_{0,1,2000} \\ \vdots \\ e_{0,38,2000} \\ \vdots \\ e_{0,38,1990} \end{pmatrix},$$

$$\gamma_t = \rho_t \times (1 - A_t/100) = \begin{pmatrix} \rho_{1,2005} \\ \vdots \\ \rho_{38,2005} \\ \vdots \\ \rho_{38,1995} \end{pmatrix} \times \left( \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix} - \frac{1}{100} \begin{pmatrix} A_{1,2005} \\ \vdots \\ A_{38,2005} \\ \vdots \\ A_{38,1995} \end{pmatrix} \right).$$

They are the same as the variables used in fitting equation (1), except that all the data points after the year 2005 are removed in this case.

Following the same method discussed in Sharrow et al. (2014), the authors helped build a mortality model for validation with the data points after 2005 removed. Based on the validation version of TFR projection, we can obtain population projections for all the 38 countries. Recall that for both life expectancy and population, we will obtain 1,000 trajectories. So for 2005-2010, we can get 1,000 projected values for each country.

Then we consider several criteria for validation. The first one is MAE, the mean absolute error, taking the mean absolute difference between our median projection and the true values. Secondly, we consider SAPE, standardized absolute predictive error, which assesses the calibration of the probabilistic projection. It is calculated as:

$$SAPE = \sqrt{\frac{\pi}{2}} \frac{|l_c - \hat{l}_c|}{\hat{\sigma}_{pred,c}},$$

where  $l_c$  refers to the true value and  $\hat{l}_c$  is for our median projection. Then we standardize the absolute difference by the standard deviation of prediction. For a good probabilistic projection, the mean SAPE value should be close to 1. Besides the two measures above, we also assess the model by the coverage of our predicted distribution over the true values among all the 38 countries.

Four summaries of the results are shown in Table 4.1. It is shown that we have MAE for life expectancy validation close to each other for the two methods and both are almost 2, which means that both methods, on average, are 2 years off in life expectancy projection. We have better MSAPE and coverages, which shows that our projection provides wider confidence intervals in projecting life expectancy, covering more uncertainties. For population, we have much lower MAE than `bayesPop` without HIV effect does. The MSAPE from our method is 0.97—very close to the theoretical value of 1 for a calibrated probabilistic forecast,

Table 4.1: Model validation for life expectancy and population projection. MAE for population is in ( $\times 1,000$ ) unit.

|              | Life Expectancy |           | Population   |          |
|--------------|-----------------|-----------|--------------|----------|
|              | HIV Adjusted    | bayesLife | HIV Adjusted | bayesPop |
| MAE          | 2.09            | 1.92      | 159.46       | 218.46   |
| MSAPE        | 1.44            | 2.82      | 0.97         | 1.52     |
| 95% Coverage | 84.2%           | 57.9%     | 89.5%        | 78.9%    |
| 90% Coverage | 71.1%           | 55.3%     | 84.2%        | 73.7%    |
| 80% Coverage | 65.8%           | 50.0%     | 84.2%        | 65.8%    |

and far better than 1.52 from using `bayesPop`. Also, the coverages of population projection from our method are closer to the true percentages than `bayesPop`'s are. To summarize, the method we have developed in this paper covers the true values better, thus can be regarded as a useful method for probabilistic population projection for AIDS countries.

Table 4.2: Model validation for life expectancy based on different sets of HIV prevalence in fitting the model, EPP estimates and WPP 2012.

| Data         | EPP Median | WPP 2012 |
|--------------|------------|----------|
| MAE          | 2.09       | 2.12     |
| MSAPE        | 1.44       | 1.47     |
| 95% Coverage | 84.2%      | 78.9%    |
| 90% Coverage | 71.1%      | 71.1%    |
| 80% Coverage | 65.8%      | 65.8%    |

In addition to compare the performance of the above methods, we use validation to explain the selection of data as well. In other words, we can use validation to explain why we chose EPP median estimates as inputs for regression model of life expectancy as opposed

to the the estimates from UNAIDS and why we used data starting from 1990 but not 1980 (the earliest record available). To answer the first question, we compare the same criteria in one-out-of-sample validation under two different versions of the data for life expectancy. Table 4.2 shows the comparison that if EPP estimation is used in building the model, the MAE is 2.09 versus 2.12 for using UNAIDS estimations. Also MSAPE is closer to 1 if we use EPP estimates. The coverage under two conditions are similar with 95% coverage better for using EPP estimation. Overall, the summary in the table indicates that using EPP estimation in building life expectancy model is better, since we consistently use the same source.

Table 4.3: Validation for using data starting from different years and projection discrepancy from WPP 2012.

|                 | Validation, one-out-of-sample |       |       |       |       |
|-----------------|-------------------------------|-------|-------|-------|-------|
| Starting Year   | 1980                          | 1985  | 1990  | 1995  | 2000  |
| MAE             | 2.09                          | 1.97  | 1.88  | 1.81  | 1.61  |
| MSAPE           | 1.44                          | 1.38  | 1.30  | 1.13  | 0.91  |
| 95% Coverage    | 84.2%                         | 81.6% | 84.2% | 84.2% | 92.1% |
| 90% Coverage    | 71.1%                         | 76.3% | 73.7% | 84.2% | 86.8% |
| 80% Coverage    | 65.8%                         | 65.8% | 68.4% | 71.1% | 78.9% |
|                 | Projection, no points removed |       |       |       |       |
| Starting Year   | 1980                          | 1985  | 1990  | 1995  | 2000  |
| Mean Abs. Diff. | 3.02                          | 3.01  | 2.88  | 2.90  | 3.91  |

To answer the question why we used data starting from 1990 rather than 1980, we also ran the one-out-of-sample validation for life expectancy prediction based on different subsets of the data, varying the starting year from 1980 to 2000. The end year is fixed at 2005. Table 4.3 gives the comparison in MAE, MSAPE and coverages under these cases. Although the performance is best when starting from 2000, we choose 1990 as the start year. This is

because that we lose too much information if we only use data from 2000 onwards and the data before 1990 are of poor quality. To support our reasoning, we also consider the mean prediction discrepancy from WPP 2012 projections. We put back the removed data points in 2005-2010 and use data starting from different years to project life expectancy up to year 2100. After that, we calculate the mean absolute difference between median projection and the WPP 2012 projections across all the years and countries. The mean absolute difference is summarized in Table 4.3 as well. The 1990 column contains the smallest value, i.e., it is reasonable to use data starting from 1990.

## Chapter 5

**DISCUSSION**

We have developed a probabilistic method of population projection for countries with generalized HIV/AIDS epidemics. The basic idea follows Raftery et al. (2012) by combining fertility, mortality and immigration projection in a cohort component model for projection. The method of Raftery et al. (2012) is designed for countries without substantial levels of HIV/AIDS epidemics. To adjust the method for AIDS countries, we incorporate HIV effect into life expectancy and mortality projection in the form of HIV prevalence and ART coverage. For the 40 countries with a high level of the HIV epidemic, we provide probabilistic projection of population using the method developed above. This method is capable of providing an intuition on HIV trends and future population development for researchers and policy makers.

At the end of Section 2, we stated that the population projected for 2045-2050 is slightly higher than what WPP 2012 projects for the same period. This is explained by the fact that we account for the effect of ART, which prolongs life for treated individuals. So in the short term, it is reasonable to expect higher population levels than that projected without accounting for ART. In the long run, the population projection is affected by the projection of mortality patterns. The existence of high HIV epidemics results in lower life expectancy projected on average. Their incorporation results in higher mortality for young adults, which influence the population of reproductive female and newborns. This explains why we have far lower population projected in 2100 than WPP 2012.

Following Raftery et al. (2013), we modeled the expected change in life expectancy using a double logistic curve. The parameter we used is a “world ” posterior mean estimate. It can be a decent approximation, but country-specific parameters may be more appropriate for capturing country-specific characteristic. Instead of projecting life expectancy for two sexes as a pool, Raftery et al. (2014) projected female and male life expectancy jointly

by projecting female age-specific life expectancy first and then projecting the gap between male and female life expectancy. So for future work, we may want to model life expectancy separately for different sexes. Moreover, we can try to model the AIDS and non-AIDS countries together—since we use the double logistic function to model  $e_0$  in both cases—using a dummy indicator for generalized HIV status.

The cohort component model requires projection of net international migration as well. However, we only prepared probabilistic projection on fertility and mortality. Research is currently under way on the development of probabilistic forecasting methods for net international migration (Azose and Raftery, 2013). The probabilistic population projection will be more complete if we can incorporate uncertainty about net migration in the future.

Finally, Bhavan et al. (2008) and Önen et al. (2010) mentioned the aging of HIV epidemics, which we have not yet considered. As more and more research and treatment has been carried out for AIDS in recent years, people with HIV positive live longer, while fewer people get infected over time. Then the composition of age groups with HIV undergoes many changes. Thus it is possible that we could improve our projections by considering the effect of the aging of HIV epidemics.

## BIBLIOGRAPHY

- Alkema, Leontine, Adrian E Raftery, and Tim Brown. 2008. "Bayesian melding for estimating uncertainty in national HIV prevalence estimates." *Sexually Transmitted Infections* 84:i11–i16.
- Alkema, Leontine, Adrian E Raftery, Patrick Gerland, Samuel J Clark, François Pelletier, Thomas Buettner, and Gerhard K Heilig. 2011. "Probabilistic projections of the total fertility rate for all countries." *Demography* 48:815–839.
- Azose, Jonathan J and Adrian E Raftery. 2013. "Bayesian Probabilistic Projection of International Migration Rates." *arXiv preprint arXiv:1310.7148* .
- Bhavan, Kavita P, Vinay N Kampalath, and Edgar Turner Overton. 2008. "The aging of the HIV epidemic." *Current HIV/AIDS Reports* 5:150–158.
- Boily, Marie-Claude, Rebecca F Baggaley, Lei Wang, Benoit Masse, Richard G White, Richard J Hayes, and Michel Alary. 2009. "Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies." *The Lancet Infectious Diseases* 9:118–129.
- Brown, Tim, Le Bao, Adrian E Raftery, Joshua A Salomon, Rebecca F Baggaley, John Stover, and Patrick Gerland. 2010. "Modelling HIV epidemics in the antiretroviral era: the UNAIDS Estimation and Projection package 2009." *Sexually Transmitted Infections* 86:ii3–ii10.
- Burns, David N, Carl W Dieffenbach, and Sten H Vermund. 2010. "Rethinking prevention of HIV type 1 infection." *Clinical Infectious Diseases* 51:725–731.
- Cohen, Myron S, George M Shaw, Andrew J McMichael, and Barton F Haynes. 2011. "Acute HIV-1 infection." *New England Journal of Medicine* 364:1943–1954.

- Hollingsworth, T Déirdre, Roy M Anderson, and Christophe Fraser. 2008. "HIV-1 transmission, by stage of infection." *Journal of Infectious Diseases* 198:687–693.
- Hontelez, Jan AC, Sake J de Vlas, Rob Baltussen, Marie-Louise Newell, Roel Bakker, Frank Tanser, Mark Lurie, and Till Bärnighausen. 2012. "The impact of antiretroviral treatment on the age composition of the HIV epidemic in sub-Saharan Africa." *AIDS (London, England)* 26.
- Lalic, Nevena. 2011. *Joint probabilistic projection of female and male life expectancy*. Master's thesis, University of Washington.
- Lee, Ronald D and Lawrence R Carter. 1992. "Modeling and forecasting US mortality." *Journal of the American Statistical Association* 87:659–671.
- Lingappa, Jairam R, James P Hughes, Richard S Wang, Jared M Baeten, Connie Celum, Glenda E Gray, Wendy S Stevens, Deborah Donnell, Mary S Campbell, Carey Farquhar, et al. 2010. "Estimating the impact of plasma HIV-1 RNA reductions on heterosexual HIV-1 transmission risk." *PLoS One* 5:e12598.
- Mayer, Kenneth H and Kartik K Venkatesh. 2010. "Antiretroviral therapy as HIV prevention: status and prospects." *American Journal of Public Health* 100:1867–1876.
- Miller, William C, Nora E Rosenberg, Sarah E Rutstein, and Kimberly A Powers. 2010. "The role of acute and early HIV infection in the sexual transmission of HIV." *Current Opinion in HIV and AIDS* 5:277.
- Önen, Nur F, E Turner Overton, Warren Seyfried, Emily R Stumm, Mariea Snell, Kristin Mondy, and Pablo Tebas. 2010. "Aging and HIV infection: a comparison between older HIV-infected persons and the general population." *HIV Clinical Trials* 11:100–109.
- Powers, Kimberly A, Azra C Ghani, William C Miller, Irving F Hoffman, Audrey E Pettifor, Gift Kamanga, Francis EA Martinson, and Myron S Cohen. 2011. "The role of acute and early HIV infection in the spread of HIV and implications for transmission prevention strategies in Lilongwe, Malawi: a modelling study." *The Lancet* 378:256–268.

- Preston, S, Patrick Heuveline, and Michel Guillot. 2001. "Demography: measuring and modeling population processes." *Pop. Dev. Rev* 27:365.
- Raftery, Adrian E, Leontine Alkema, Patrick Gerland, et al. 2014. "Bayesian population projections for the United Nations." *Statistical Science* 29:58–68.
- Raftery, Adrian E and Le Bao. 2010. "Estimating and projecting trends in HIV/AIDS generalized epidemics using incremental mixture importance sampling." *Biometrics* 66:1162–1173.
- Raftery, Adrian E, Jennifer L Chunn, Patrick Gerland, and Hana Ševčíková. 2013. "Bayesian probabilistic projections of life expectancy for all countries." *Demography* 50:777–801.
- Raftery, Adrian E, Nan Li, Hana Ševčíková, Patrick Gerland, and Gerhard K Heilig. 2012. "Bayesian probabilistic population projections for all countries." *Proceedings of the National Academy of Sciences* 109:13915–13921.
- Ševčíková, Hana. 2013. "Bayesian Probabilistic Population Projections: Do It Yourself." .
- Ševčíková, Hana, Leontine Alkema, and Adrian E Raftery. 2011. "bayesTFR: An R package for probabilistic projections of the total fertility rate." *Journal of Statistical Software* 43:1–29.
- Sharrow, David J, Samuel J Clark, and Adrian E Raftery. 2014. "Modeling Age-Specific Mortality for Countries with Generalized HIV Epidemics." *PloS one* 9:e96447.
- Stover, John, Tim Brown, and Milly Marston. 2012. "Updates to the Spectrum/Estimation and Projection Package (EPP) model to estimate HIV trends for adults and children." *Sexually Transmitted Infections* 88:i11–i16.
- UNAIDS. 2012. *Global Report: UNAIDS Report on the Global AIDS Epidemic: 2012*. UNAIDS.
- United Nations, Department of Economic and Social Affairs, Population Division. 2011a. *World Population Prospects: the 2010 Revision*. New York: United Nations.

- United Nations, Department of Economic and Social Affairs, Population Division. 2013a. *World Population Prospects: the 2012 Revision*. New York: United Nations.
- Wandel, S, M Egger, R Rangsiri, KE Nelson, C Costello, C Lewden, T Lutalo, A Ndyanaabo, J Todd, L Van der Paal, et al. 2008. “Duration from seroconversion to eligibility for antiretroviral therapy and from ART eligibility to death in adult HIV-infected patients from low and middle-income countries: collaborative analysis of prospective studies.” *Sexually Transmitted Infections* 84:i31–i36.
- Whelpton, Pascal K. 1936. “An empirical method of calculating future population.” *Journal of the American Statistical Association* 31:457–473.

Appendix A  
**PARAMETERS  $\phi$  AND  $R$  IN EPP 2009**

According to Brown et al. (2010), the model given by Figure 2.1 is given by

$$N_t = X_t + Y_t + Z_t + U_t + L_{1t} + L_{2t},$$

where  $t$  refers to the time point. At time  $t$ ,  $N_t$  is the total population,  $X_t$  is the not-at-risk compartment,  $Y_t$  is the infected population but not eligible for ART and  $Z_t$  is the at-risk but uninfected group.  $U_t$  refers to the infected and eligible for ART but not treated,  $L_{1t}$  is the group going on first-line ART, and  $L_{2t}$  is for those on second-line ART.

The two parameters  $\phi$  and  $r$  are used in the following way. For  $\phi$ , the behavior-change parameter, it shows the rate of how many new individuals are recruited into the high-risk group. Denote  $f_t$  as the fraction of population entering the at-risk compartment at time  $t$ , i.e., the new population entering group  $Z$  is  $Z_t^{new} = f_t E_t$ , where  $E_t$  is the new entrants at time  $t$ .

$$f_t = \frac{\exp(\phi_t)}{\exp(\phi_t) + 1/f_0 - 1},$$

It is a function of  $\phi$  and the initial fraction  $f_0$ .  $\phi$  is assumed to decrease by half every twenty years after 2013. Therefore, for  $t$  greater than 2013, we formulate  $\phi_t$  as

$$\phi_t = \phi_{2013} \cdot \exp\left[-\frac{\log(2)}{20}(t - 2013)\right].$$

The other parameter  $r$  refers to the force of infection.

$$Y_t^{new} = [r_t(Y_t + U_t + R_{inf}(L_{1t} + L_{2t}))/N_t + \eta_t]Z_t,$$

where  $Y_t$  refers to the infected population and  $Y_t^{new}$  is for the newly infected at time  $t$ . Here  $R_{inf}$  adjusted the infection rate of population on ART treatment, which is given by UNAIDS. Similar as  $\phi$ ,  $r_t$  is assumed to decline by half every thirty years, i.e.,

$$r_t = r_{2013} \cdot \exp\left[-\frac{\log(2)}{30}(t - 2013)\right].$$

## Appendix B

## DOUBLE LOGISTIC CURVE

The double logistic function with parameter  $\theta = (\Delta_1, \Delta_2, \Delta_3, \Delta_4, k, z)$  is given by

$$g(x|\theta) = \frac{k}{1+\exp(-\frac{A_1}{\Delta_2}(x-\Delta_1-A_2\Delta_2))} + \frac{z-k}{1+\exp(-\frac{A_1}{\Delta_4}(x-\sum_{i=1}^3 \Delta_i-A_2\Delta_4))}.$$

The constants  $A_1$  and  $A_2$  is arbitrary, but we take  $A_1 = 4.4$  and  $A_2 = 0.5$  so that the first four parameters can be interpreted. The results do not change if the product  $A_1A_2$  is unchanged. As shown in Figure B.1, the first four parameters  $\Delta_1, \dots, \Delta_4$  identify intervals of life expectancy when the rate of the life expectancy gains in the next five years is changing. Parameter  $k$  is an approximation of the maximum gain in life expectancy and  $z$  is the asymptotic rate of five-year gains as life expectancy increases.

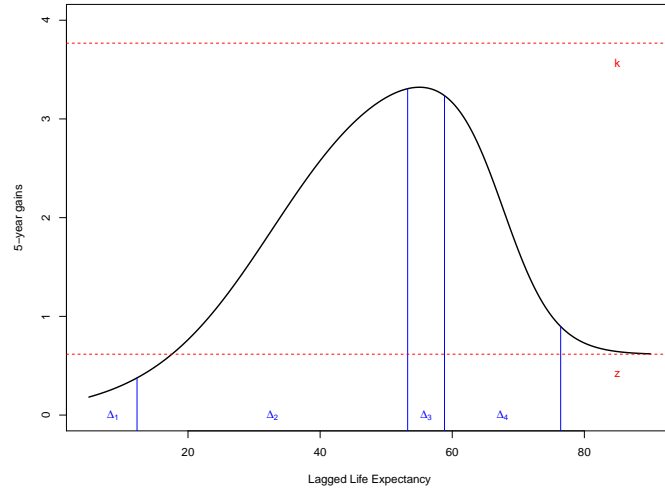


Figure B.1: The fitted curve for double logistic function under the “world” parameter  $\theta^w = (12.2854, 40.9654, 5.5763, 17.5841, 3.7677, 0.6168)$ .