extended abstract

A Methodology for Projecting sub-National Populations Allowing for the Impact of HIV/AIDS and where Data are Limited and Defective

Rob Dorrington and Tom A Moultrie
Centre for Actuarial Research, University of Cape Town

Sub-national population projections are fraught with complexity at the best of times. It is axiomatic that as the population being modelled represents a smaller proportion of the national whole, the least understood of the three major demographic forces – migration (local and international) – plays an increasingly significant role, while sub-population specific estimates of the other demographic variables may not be available or reliable. This paper describes an approach to projecting sub-national population dynamics in 28 districts of a country (Botswana) whose aggregate population is less than one-fifth that of New York City. In doing so, the sum of the regional projections must remain consistent with that of the national population, while taking into account the careful incorporation of HIV epidemiological dynamics in each region and with the added difficulty of working with severely limited and defective census and survey data. In the process, interesting insights into the regional demographic dynamics of the country are developed.

1. Introduction

The projection of small-area (and sub-national) populations offers several distinct challenges to demographers. First among these is that as the area being projected gets smaller, closer attention needs to be paid to both regional (and international) migrations patterns levels and trends. The second is to parameterise accurately the other major demographic forces of fertility and mortality. These challenges become more complex when the sum of several small-area projections has to be constrained to reproduce that of a larger aggregate (e.g. national); when one needs take into account complex mortality dynamics associated with generalised (but nonetheless different in timing and level) AIDS epidemics in each of the constituent sub-populations; and when the quality of the data to hand to perform the projections is decidedly poor at a national level, and demonstrably worse for the constituent sub-populations.

In this paper, we describe a methodology applied to project the population of each of Botswana's 28 districts, from 1980 through to 2020, taking into account the impact of HIV/AIDS on the fertility and mortality of the population, as well as allowing for both internal and international migration.

2. Background

In June 2005, the Centre for Actuarial Research (CARe) at the University of Cape Town, South Africa, was awarded a tender by the United Nations Development Programme (UNDP, Botswana) and the Botswana National AIDS Co-ordinating Authority (NACA) to prepare (amongst other requirements) population projections of the Botswana population, by administrative and census district, through to at least 2020, making as much use of local data as possible (Dorrington, Moultrie and Tim 2006). The paper that will emerge from this extended abstract offers a case study of projecting sub-national population dynamics in 28 districts of a country (Botswana) whose aggregate population is less than one-fifth that of New York City. Four particular challenges arose in the process of meeting the objectives. First, the sum of the regional projections had to remain consistent with that of the national population (not only in terms of total population numbers and age-sex distributions, but also in terms of the country's epidemiological characteristics); second, the separate and distinct epidemiological dynamics of HIV in each region had to be taken into account; third, the projection had to take into account both current and future interventions (for example, roll-out of antiretroviral therapy); and – finally – we had to engage with the added difficulties of working with severely limited and defective data.

3. Structure of the paper

The overriding focus of the paper is on applied demography – describing the application of a host of demographic techniques, not least among which are those associated with small-area population projections, and the modifications applied to those methods to achieve the challenges described in the previous paragraph.

Consequently, the paper will first describe the factors (above) which make the production of sub-national population projections more challenging than is typically the case. Second, it will describe the method adopted for producing the constrained sub-national population projections. Third, we will present results from our projections showing the successful operation of the method; and finally, we will reflect on the process undergone and the extent to which lessons have been learned that may be of value to future practitioners.

4. Methodology

The complexity of the methodological approach adopted to project sub-national populations in an era of HIV/AIDS from limited and defective data is determined by the rigorousness of the method of allowing for the dynamics of the epidemic in the population(s) being studied; the availability and quality of the data to hand; and the procedure adopted to constrain the sub-national populations to maintain a multi-dimensional consistency with national projections.

Much work has gone into the development of a class of cohort-component projection models that incorporate HIV/AIDS explicitly in their projections. Broadly, there are two approaches to doing so. The first produces a projection ignoring the impact of HIV/AIDS and then estimates, usually as a derivation from a projection of the prevalence of HIV in the population, the number of people newly infected each year and the number of deaths from the infected survivors each year. These deaths are then removed from the non-HIV projection. Allowance can also be made, albeit somewhat crudely, for a reduction in the number of births due to HIV and a reduction in the number of AIDS deaths due to the provision of ART. Such is the approach adopted by DemProj and AIM components of the Spectrum suite of programmes developed by the Futures Group. Their fundamental limitation, of course, is that the estimates of HIV/AIDS incidence, prevalence, morbidity and mortality are effectively decoupled from the demographic projection and there is little to ensure consistency in assumptions between the two.

The second approach is to generate the number of new infections as part of the projection model (i.e. firmly embedding the epidemiological model within the demographic model) by making assumptions as to patterns of sexual behaviour and probabilities of transmission. Such models, although more demanding in terms of the number of assumptions that need to be made, are able to allow for a wider range of changes in behaviour and interventions, and do so more realistically.

It stands to reason that the second approach is better, even if more significantly more complex, provided one can reasonably estimate the required parameters. Hence, a better, albeit more difficult, approach is to apply cohort-component projection techniques to a multi-state population model, where the states reflect both demographic and epidemiological attributes (e.g. HIV-; asymptomatically HIV+, symptomatically HIV+; on antiretroviral therapy; dead from non-AIDS causes; AIDS-related deaths etc) and commencing the population projection at least a little time before the start of the epidemic. Such a model, the so-called ASSA model, has been developed for use in Southern Africa (Johnson and Dorrington 2006).

Addressing the challenges enumerated in the introduction

A brief outline of the solutions proposed to the first of the challenges is detailed below. The resolution of the other three challenges – the modelling of sub-national epidemics, the incorporation of interventions, and the appropriate handling of limited and defective data – affect the parameterisation of inputs, or the interpretation of outputs and are not addressed in this extended abstract in detail, but will be described in detail in the paper. Of particular interest here is the procedure adopted to maintain consistency between the sum of the regional projected populations, and the directly projected national population estimates.

1. The sum of all regions' populations (and their population weighted HIV-prevalences) must be constrained to emulate as closely as possible the national population (and HIV prevalence)

To address this constraint, a combination of a 'bottom up' and 'top down' approach is used to project the population and hence the demographic impact of HIV/AIDS. First, a model which projects the population of the country as a whole is constructed and calibrated. This is then used to help construct a prototype model for each of the census districts. The 'ratio method' is applied to the projected population by sex and age using the national model and the distribution of the population by sub-district, sex and age, at three censuses to provide a first estimate of the population by sex and age in each sub-district. These numbers in turn are compared for each sub-district to those projected using the prototype models without migration to determine a first estimate of numbers of migrants by sex and age on an annual basis. Each sub-district model is then adapted to include these numbers of migrants and recalibrated to sub-district specific data. The results of these projections are then aggregated to produce a second national projection. Of course, due to the non-linear nature of the models, these two projections cannot be expected to be exactly the same, however, if they differ significantly this indicates a need to improve one or more of the assumptions underlying the models.

There are several reasons for employing such a complicated and time-consuming approach. The first is that modelling the districts allows us to capture, to the extent that the data allow it, the different regional dynamics that there might be in the epidemic, in particular epidemics that started earlier or later that, grew faster or slower than, and/or have reached a plateau above or below that of the national average. The second is that it is more useful from a health (and other) management position to have a model of the local impact of the epidemic. On the other hand, many of the parameters are more reliably estimated at a national level and thus the constraint to have the district projections aggregate to the national projection places a limit on the errors in the district level projections.

In the case of Botswana, epidemiological evidence suggests that the epidemic began in the early 1980s. This requires, then, commencing a cohort-component population projection in 1980, which in turn requires suitable estimates of the initial populations, as well as trajectories of fertility, migration and the single decrement table associated with non-AIDS mortality, both nationally and for each district.

Estimation of the demographic input parameters

In summary, the model requires estimates (of rates or numbers) by individual age, by sex and district, for each year from 1980 to 2020, together with a base population by sex and single year of age for each district. The full process of demographic estimation will be described in detail in the final paper. Unit record data from the 1991 and 2001 census were provided, as well as the analytical reports from the 1981, 1991 and 2001 censuses. In addition, a range of different survey data sets were made available, although almost all of these had sample sizes too small, or produced results sufficiently at odds with those from the censuses, to be deemed incredible.

In the case of the estimation of fertility and adult and child mortality, after correction of evident errors (for example reference period errors and undercount of children) in the data from the 1991 and 2001 censuses, a variety of indirect techniques were applied to the data to provide point estimates of vital rates. These needed to be back-projected to 1980, the starting year of the projection, while remaining as consistent as possible with estimates published from the 1981 census. Throughout this process, care had to be taken to isolate the possible confounding effects of HIV/AIDS as well as violations of the underlying assumptions on the methods adopted on the results produced. (One such example would be the assumed independence of the mortality of mothers and their children in the Children Ever Born-Children Surviving method, under conditions of a generalised HIV/AIDS epidemic.) In summary, the principal methods used to derive required fertility and mortality rates were:

Fertility	Levels	Relational Gompertz models
	Trends	Interpolation between point estimates
		Exponential decay outside of point estimates
Child mortality	Levels	Brass Children Born- Children Surviving method, after removal of impact of HIV
	Trends	Exponential decay of non-AIDS mortality
Adult mortality	Levels	Data on deaths reported by households, combined with combined application of the Generalised Growth Balance method and Synthetic Extinct Generations method (to adjust for differential undercount between censuses, and differential completeness of death registration)
	Trend	Exponential decay of non-AIDS mortality

Migration

For most countries international migration is a relatively small component of the national demographic balancing equation. This is not the case in Botswana which has experienced non-trivial flows over the 20 years from 1981 to 2001. In addition to this it is likely that, in common with South Africa, the country will have experienced its share of hidden migration from Zimbabwe, particularly in recent years. Unfortunately, as is the case with most countries, particularly developing countries, it is extremely difficult to document accurately these flows of migrants, and particularly little can be done to estimate the numbers of migrants of which there is nothing other than anecdotal evidence.

We adopted a two-stage approach to deriving estimates of the numbers of migrants. The first stage was to estimate the flow of immigrants net of emigrants by sex and age over each of the intercensal periods by making use of the change in the 'stock' of each of the foreign-born population resident in Botswana, and of the Batswana population resident outside the country between two censuses. The second stage was, as part of the reconciliation of the census populations, to check if there was an excess in the 1991 and 2001 censuses that could be reasonably explained by hidden migration – if so then the numbers of immigrants or emigrants could be adjusted accordingly.

The number of surviving immigrants less the number emigrants in five-year age groups up to the open interval 75+ as at the end of the intercensal period were derived as the difference between the number foreign-born less the number of absentee Batswana at that age in the second census, less the same figure ten years younger in the first census. To these numbers were added back the number that might have been expected to have died before the second census on the assumption that migration took place uniformly over the intercensal period. Numbers of migrants at individual ages were derived from these numbers using Beers formula and then converted to numbers for each of the years.

5. Results and conclusions

Examples of the output which will be presented appear on the following pages. The essential point in terms of this extended abstract is to convey the extent of regional variation in essential demographic indicators arising from the projections.

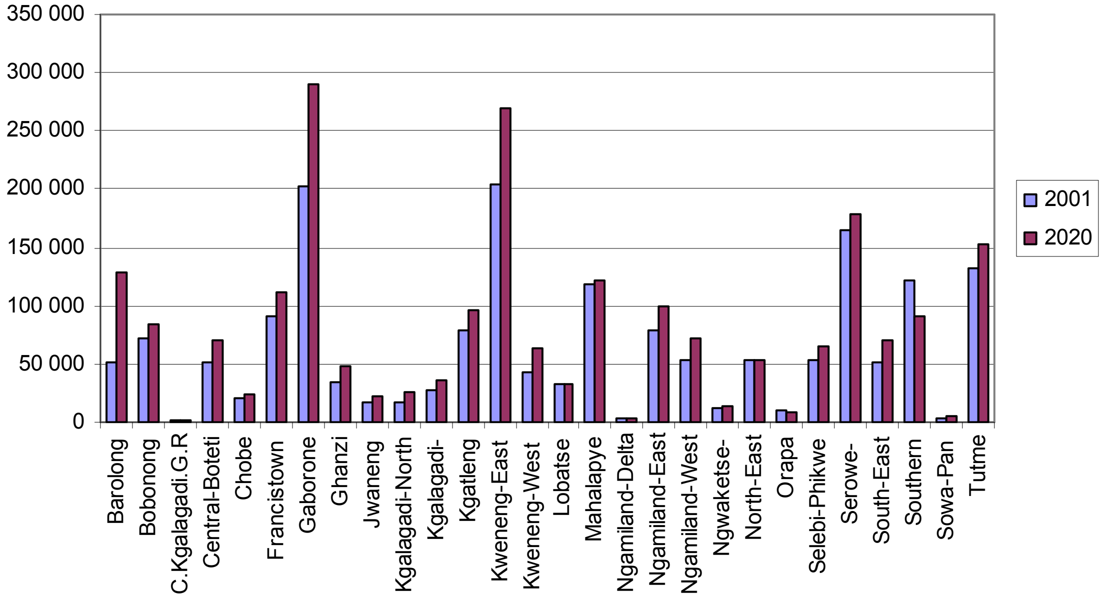
Several important conclusions arise from this exercise in applied sub-national demography. First, that – despite the obvious difficulties and inherent errors involved in an exercise of this magnitude under such conditions as were experienced – the exercise can provide information, which is not only valuable for policy purposes, but also more nuanced than that which might have been produced by adopting a simpler approach. Second, the use of an iterative top-down-bottom-up approach has been shown to add value to situations where one has to constrain the sum of many sub-national projections to a whole, and has been shown to work when the constraints are more significant than mere aggregations by age and sex, There is no reason why the approach presented here could not be adapted to any situation requiring constrained projection on any number of dimensions. Third, the approach presented allows the practitioner to establish some parameters iteratively at a national level first, where small-area data may be too sparse or unknown.

Dorrington, R. E., Moultrie, T. A. and Tim, D. 2006. *Modelling the impact of HIV/AIDS in Botswana*. Gaborone: UNDP and NACA, Botswana.

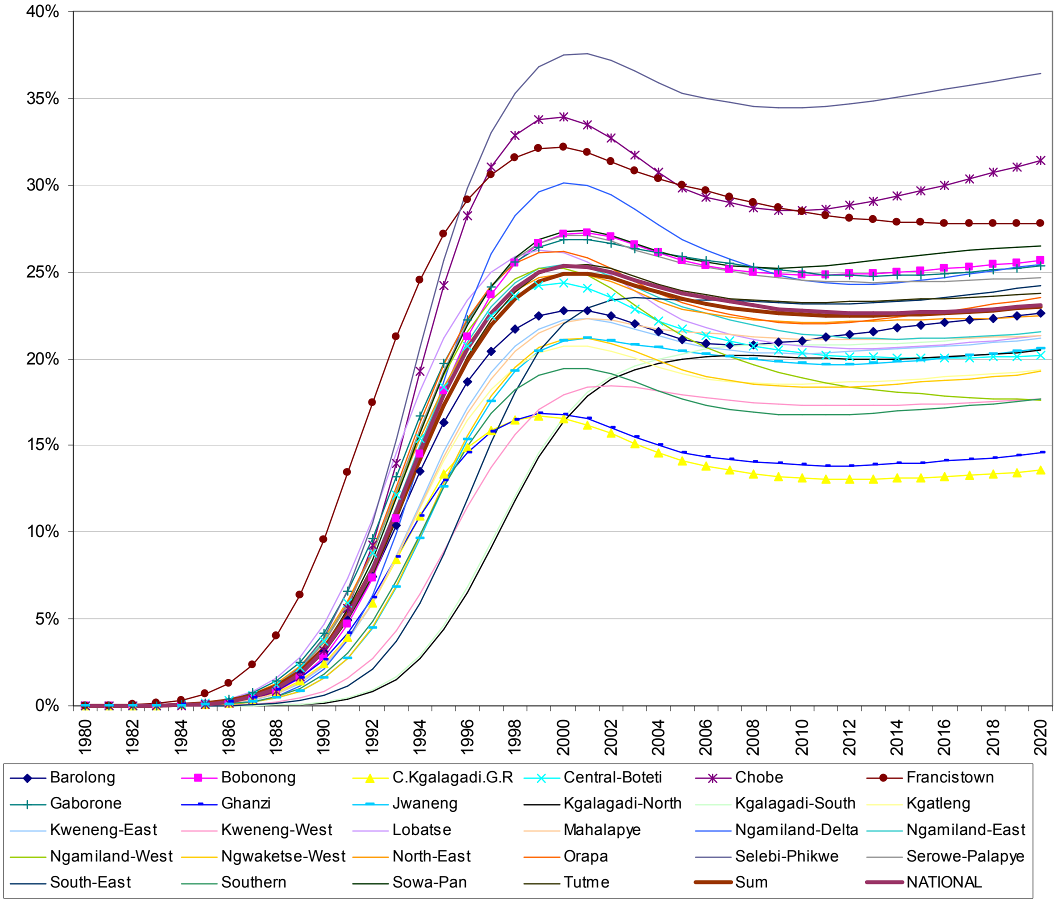
Johnson, L. and Dorrington, R. E. 2006. "Modelling the demographic impact of HIV/AIDS in South Africa and the likely impact of interventions", *Demographic Research* **14**:541-571.

The figure below compares the projected population by census sub-district in 2020 with that of the populations in 2001, at the time of the census.

Population by census sub-district in 2001 and 2020

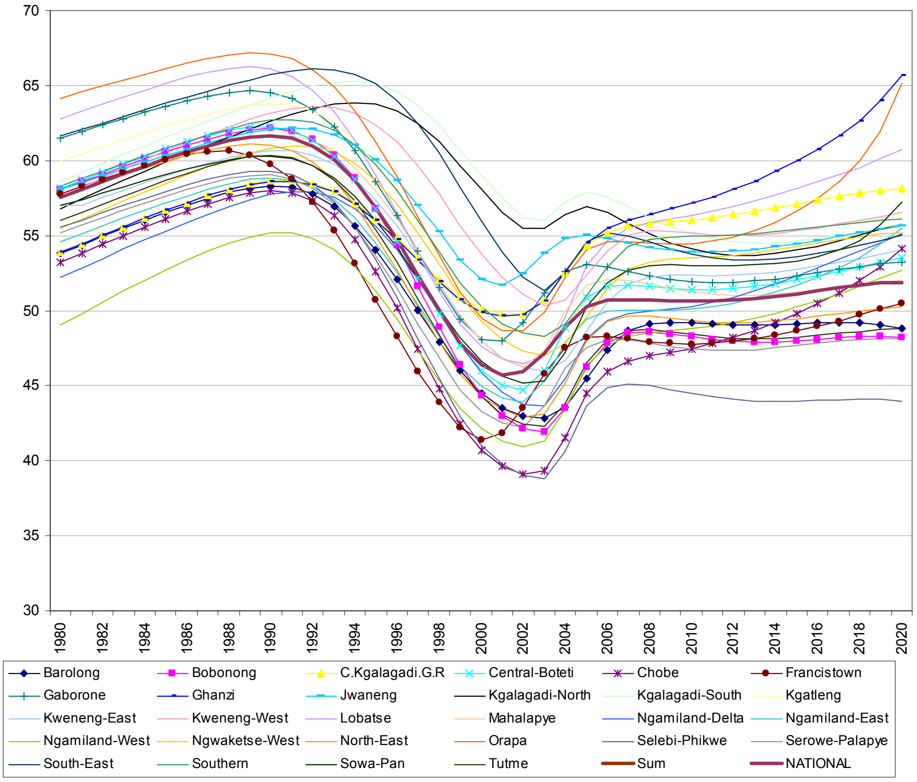


Prevalence of adults 15-49 by census sub-district



Although the different levels of infection of the various sub-districts result in different impacts on the life expectancy and under five mortality by sub-district (both shown below), comparison of the estimates in the 1980s shows that some of the difference is due to differing non-AIDS mortality by sub-district.

Life expectancy by census sub-district



Under five mortality rate by census sub-district