Subnational Projections Toolkit User's Guide

Version 2.0
November 2014

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Acknowledgments

The Subnational Projections Toolkit User's Guide was prepared at the International Programs Center for Demographic and Economic Studies, Population Division, U.S. Census Bureau, under an Inter-Agency Agreement (IAA) with the U.S. Agency for International Development (IAA GHT-A-00-08-00002).

The Guide was produced under the general direction of James D. Fitzsimmons, Assistant Division Chief for the International Programs Center for Demographic and Economic Studies. Peter D. Johnson, Lisa R. Lollock, and Thomas McDevitt were involved in its preparation. Peter Johnson also led the production of the Toolkit described in the Guide, with assistance from Amin Vafa on the software development and testing of the component workbooks and on the method descriptions.

We acknowledge and are grateful for the contributions of Eduardo Arriaga to the Toolkit. Several of the workbooks included in the Toolkit were originally developed by Arriaga as solutions to subnational projection problems encountered in statistically less developed countries while he was working at the Census Bureau.

We also wish to express our gratitude to colleagues at the U.S. Agency for International Development for their support to the project. We acknowledge with thanks the support of the technical staff working with the Census Bureau on the IAA – Rachel Marcus, Michelle Surdyk, and Yoonjoung Choi.

Table of Contents

Introduction 1

Methods and Software 8

Using the Toolkit 22

References 30

List of Figures

Figure 1. Overview of the cohort-component method for projecting a national population 3
Figure 2. Overview of the cohort-component method for projecting populations consistent across a country and its regions 4
Figure 3. Overview of projection using mathematical extrapolation with age-sex detail for districts 5
Figure 4. Overview of methods for projecting populations consistent across country, regions, and districts 6
Figure 5. Application of cohort-component method with RUP: bottom-up approach 11
Figure 6. Application of cohort-component method with RUP: top-down approach with a residual area 12
Figure 7. One-stage proportional adjustment with Toolkit 13
Figure 8. Two-stage proportional adjustment with Toolkit 14
Figure 9. Using Toolkit workbooks to incorporate migration 15
Figure 10. Illustration of a RUP input file 17
Figure 11. Input-output process of RUP and RUPAGG projections, bottom-up approach 18
Figure 12. RUPAGG input file for bottom-up projection (file Country.agg) 18
Figure 13. Input-output process and RUPAGG programming of RUP projections, residual approach 19
Figure 14. RUPAGG input file for top-down projection with a residual area (file Region3.agg) 19
Figure 15. Application of mathematical extrapolation with age-sex detail using CTBL32 and SALGST 21
Figure 16. Illustration of Graphical Analysis of Total Fertility Rates from India's Sample Vital Registration System (SRS) program and National Family Health Survey (NFHS) program 26
Figure 17. Flow of Toolkit calculations, from preliminary estimates to final subnational cohort-component projections: top-down approach 28
Figure 18. Input-output flow of mathematical extrapolation with age detail using Toolkit tools 29

Introduction

Population projections provide a glimpse into the future or, to be more precise, a set of implied future population values dictated by projection methodology, available data, and the necessary judgment of the analyst employing one or more analytical tools. Population projections provide a foundation for a wide range of program planning and evaluation tasks at national, regional, and local government authority levels. Projections help anchor subnational allocation of resources and evaluation of program performance to reasonable measures of population size and distribution. Projections provide the denominators for health programs coverage, measures of children in need of schooling, labor force-age population in need of jobs, older populations in need of health care, and households needing water, electrical power, and other services.

Methods for preparing population projections have been described elsewhere, and variants intended to improve on some of the methods at the subnational level have been considered in the literature for decades.[1] Subnational projection methods may be broadly grouped into three categories:

(1) Component methods, most commonly cohort-component projection of population, explicitly modeling population change as a function of initial population size and age and sex composition, fertility, mortality, and migration.

(2) Mathematical, or "trend," extrapolation of: (a) a single index of population (e.g., total population), (b) that index expressed as a ratio, or (c) totals or parts of the total (e.g., population disaggregated by age group and sex) expressed as proportions. The extrapolation is performed with one or more mathematical functions.

(3) Other methods. This group includes economic-demographic models (United Nations Population Division 1990; Smith, Tayman, and Swanson 2001), urban system models, and land use and housing unit methods (subsumed by Smith, Tayman, and Swanson (2001) under urban system models but treated separately in some earlier, less comprehensive works (e.g., Rives et al. 1995)). The distinguishing feature of these methods is that projected population depends not only on historical values of population variables but on other, non-demographic variables, as well.

The tools included in the U.S. Census Bureau's Subnational Projections Toolkit (referred to as the "Toolkit") and described in this User's Guide fall into the first two categories presented here: cohort-component projection and mathematical extrapolation of population totals. However, the Toolkit differs from earlier presentations of subnational and local projection methods in two ways. First, the Toolkit comprises software tools, associated procedural and methods documentation, and the Guide,

[1] See, for example, Economic Commission for Africa and Regional Institute for Population Studies (1985); Shryock, Siegel, and Associates (1971); George et al. (2004); Irwin (1977); Pittinger (1976); Smith, Tayman, and Swanson (2001); United Nations Department of International Economic and Social Affairs (1979, 1992); and United Nations Population Division (1990).

rather than projection technique descriptions only. Second, the Toolkit explicitly links cohort-component and non-cohort-component tools as parts of a coordinated approach to subnational projection.

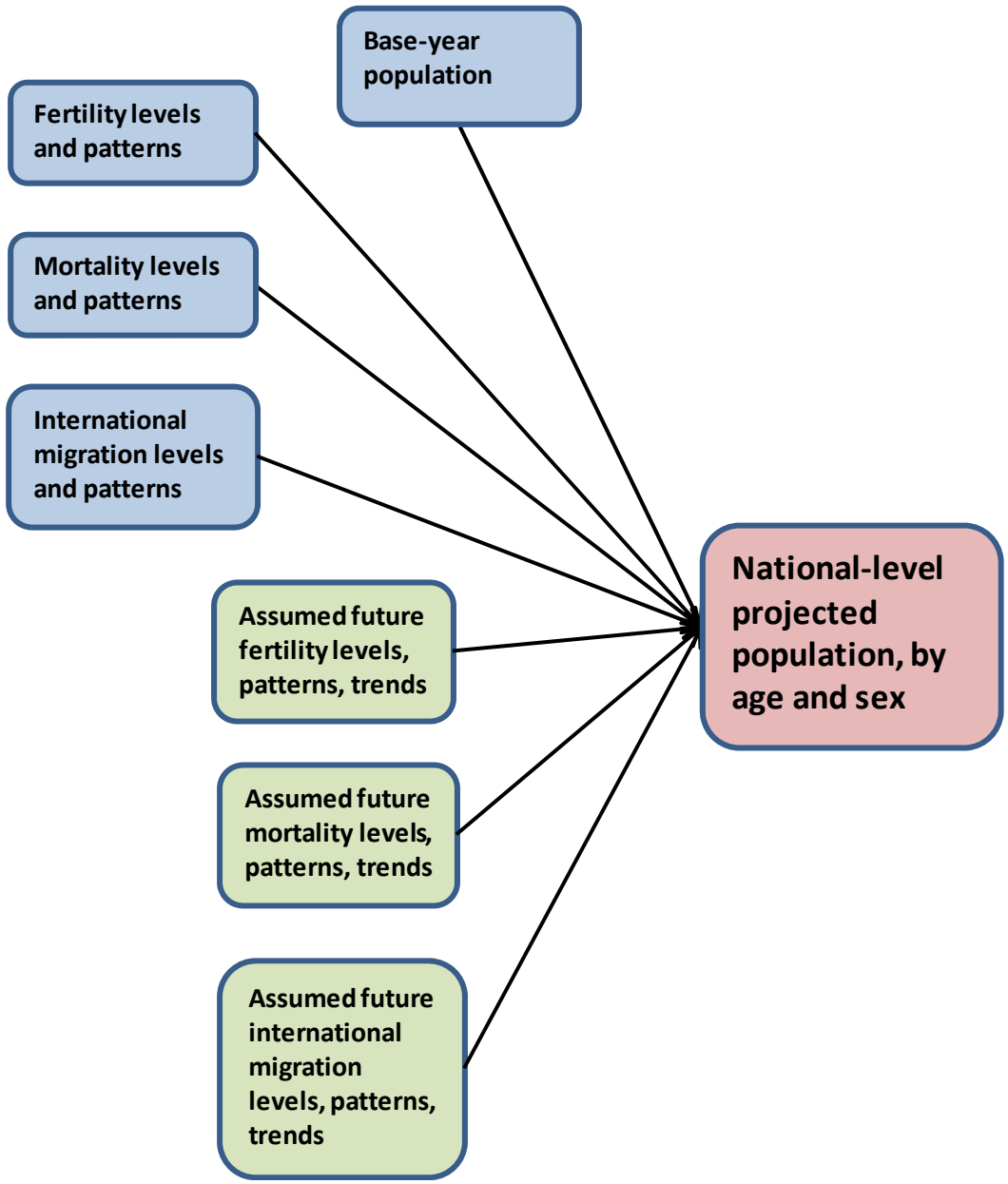
Subnational projections may be prepared for the populations of various administrative subareas of a country and for populations defined in terms of language, ethnicity or race, or other characteristics. Methodology employed in preparing projections varies with the purpose of the projection or projections, with data availability, and with the level of geographic detail.

In discussing subnational projections for geographic areas, demographers sometimes use the concept of hierarchical administrative divisions (ADMs), and the Guide follows this practice. The highest-level area, just below the national level, is referred to as the "first subnational administrative level," or ADM1. ADM1s may be called regions, provinces, states, or by some other name, depending on the country. The next level is the second subnational administrative level, or ADM2. ADM2s are called counties or districts in some countries, and by other names in other countries. The third subnational administrative level (or ADM3) may be termed sub-district, local government area level, or may carry other names. In this hierarchical system, the higher-level geography is referred to as a "parent" area. For example, in some countries, regions serve as parent areas for districts. Using this terminology to capture countries' hierarchical divisions, demographers would say that, due to data limitations and data management issues, cohort-component projection is the preferred method at the national and often at the first subnational level for population projections made by administrative level, but non-cohort-component methods are often preferred for projection at lower administrative (e.g., ADM2 or ADM3) levels.

The Toolkit provides projection tools appropriate for use at lower as well as higher administrative levels, and the Guide illustrates the use of outputs from national and ADM1 projections to control estimates prepared for lower administrative levels. In this Guide we will be using regions as examples of ADM1s and districts as ADM2s.

Schematically, the tasks involved in preparing cohort-component projections include assembling a consistent set of base-year population estimates disaggregated by age and sex; base-year fertility, mortality, and migration estimates; and assumed future fertility, mortality, and migration levels and age patterns (Figure 1).

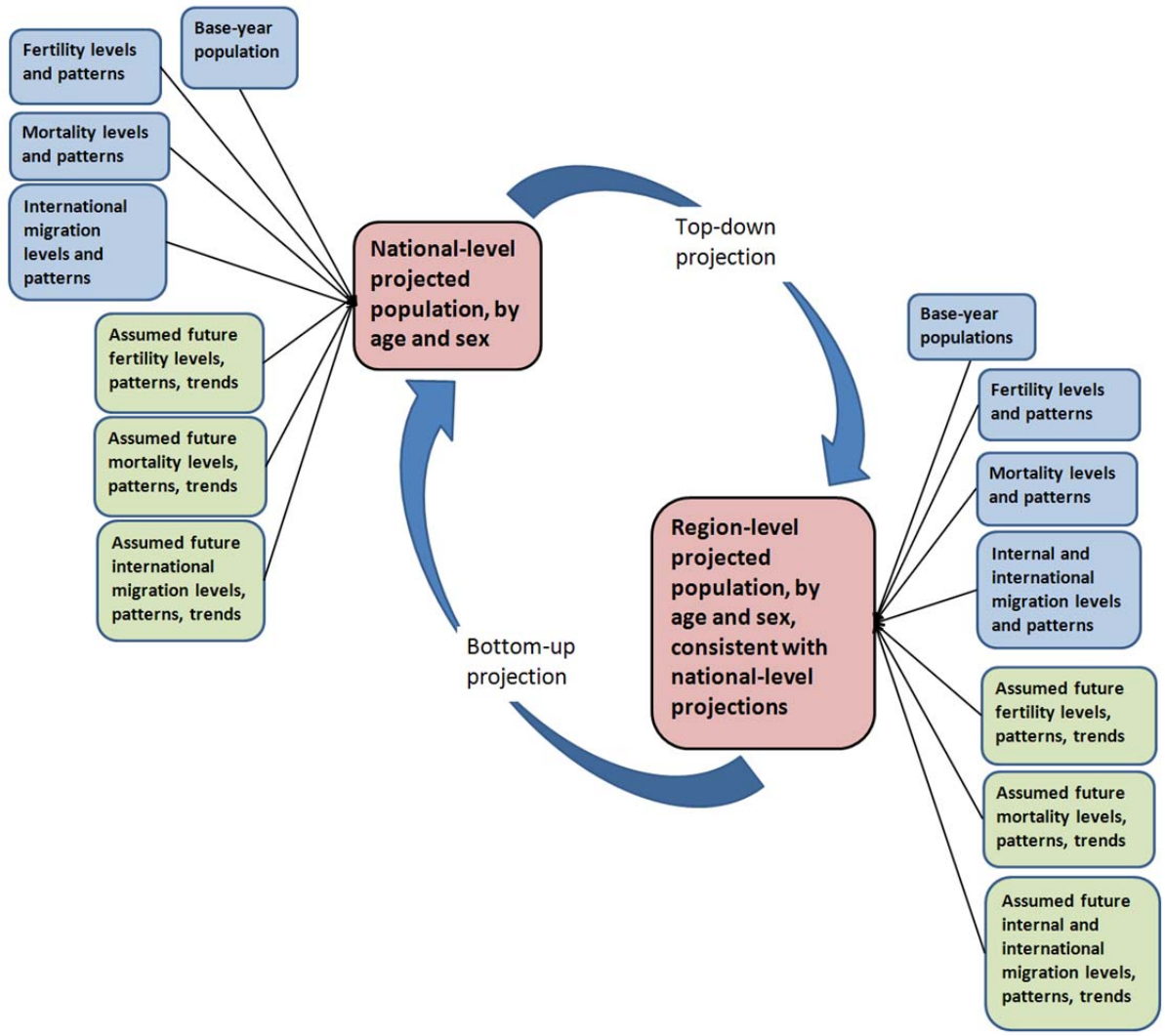
Figure 1. Overview of the cohort-component method for projecting a national population



The same set of tasks is used to prepare cohort-component projections at the national level and for larger subnational areas, such as regions or large states. However, two aspects of the subnational projections procedure distinguish it from the procedure used to prepare national projections. First, for subnational areas, we also need to develop estimates of internal migration (migration between subnational areas). Second, subnational projections can be prepared using procedures that force subnational area figures to agree with the projection values of the country. As will be further discussed in

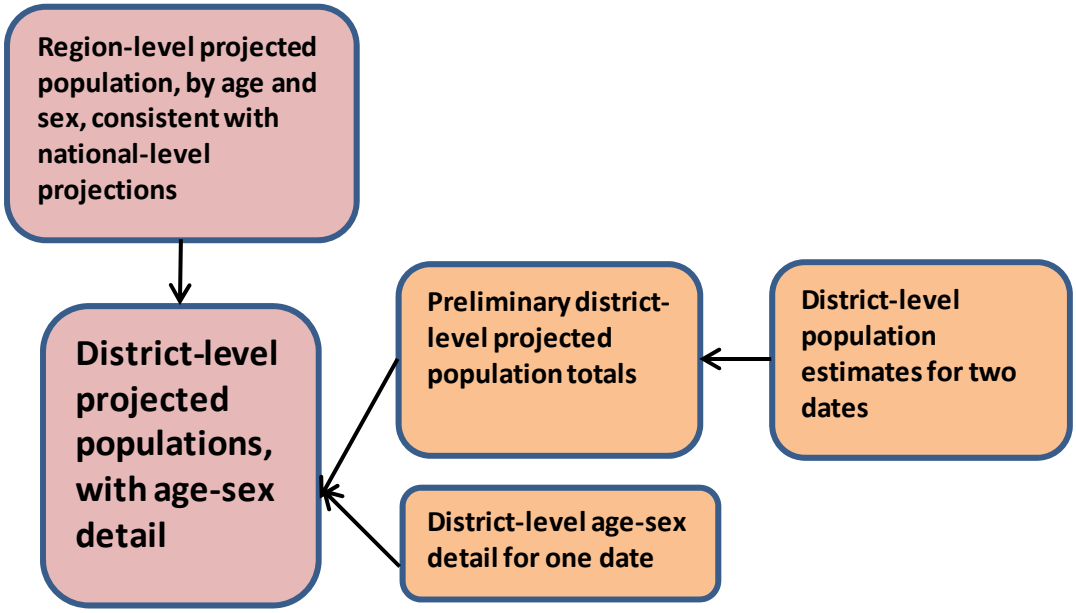
later sections of the Guide, there are different ways of establishing consistency between projections prepared for national and subnational areas; national projections may be calculated as the sum of subnational projections (Bottom-up, Figure 2), but other approaches force the summation of subnational area values to agree with national-level values (Top-down, Figure 2).

Figure 2. Overview of the cohort-component method for projecting populations consistent across a country and its regions



The tasks involved in preparing non-cohort-component projections of population for lower administrative levels also involve assembling a consistent set of base-year populations for the smaller, or local government administrative areas of the country, with (as shown in Figure 3) or without age-sex detail.

Figure 3. Overview of projection using mathematical extrapolation with age-sex detail for districts

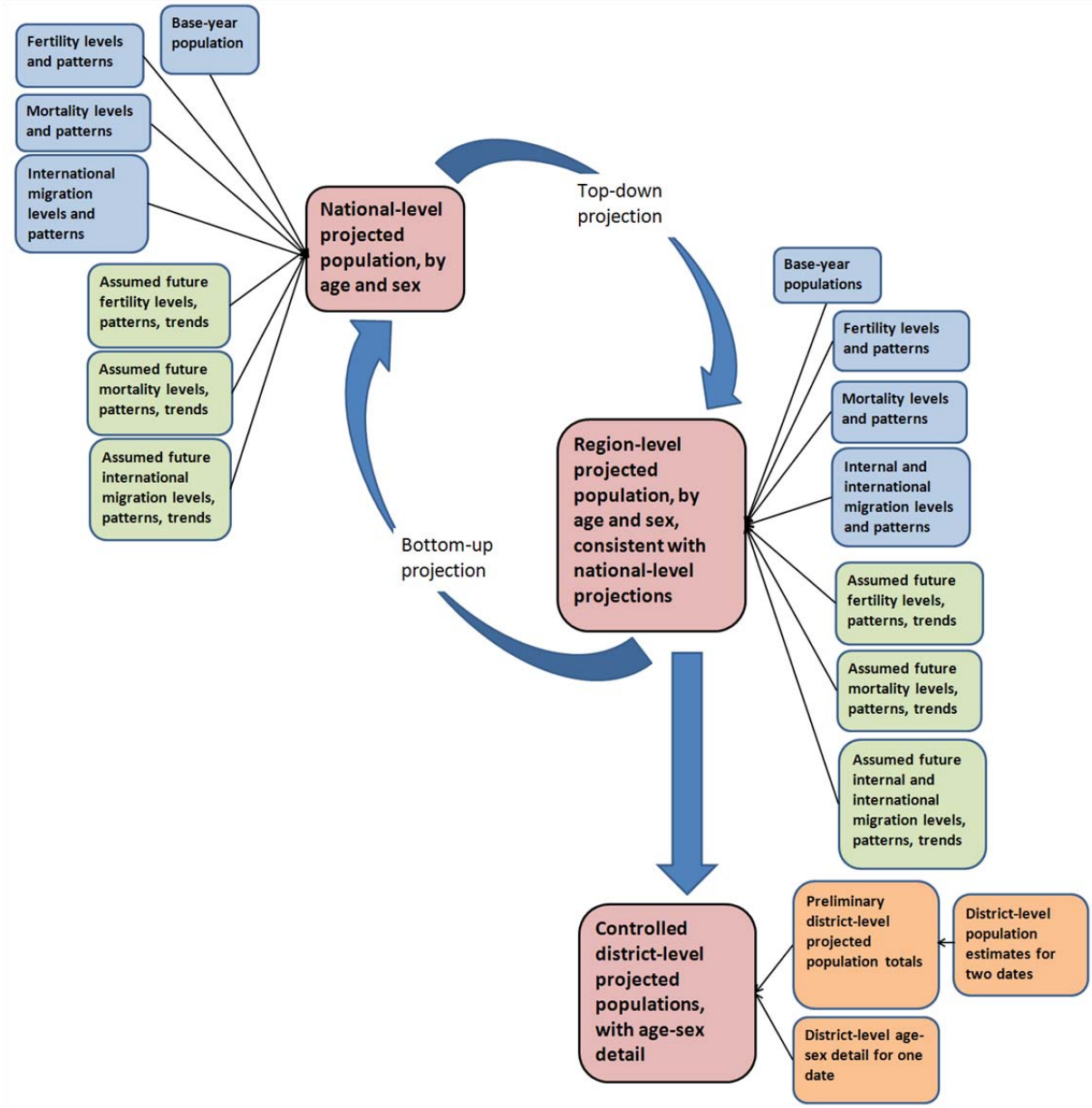


In either case, the extrapolations of smaller area populations are adjusted to add to a higher-level population total or age (or age-sex) distribution. Figure 4 shows the combination of higher-level cohort-component projection, lower-level cohort-component projection, and still lower-level mathematical extrapolation to generate consistent population projections at national and subnational levels.

This Guide describes the use of the Census Bureau's Rural-Urban Projection (RUP) program and Toolkit workbooks for preparing both cohort-component and non-cohort-component projections.[2] It also describes the use of cohort-component output to control either mathematical extrapolation of smaller area population totals or extrapolated smaller area totals with age-sex detail.

[2] RUP is described in detail in U.S. Census Bureau (2013c). For more information, see the Census Bureau Web site <http://www.census.gov/population/international/software/rup/>.

Figure 4. Overview of methods for projecting populations consistent across country, regions, and districts



The Guide provides an introduction to the subnational projection process and to the projection software constituting the Toolkit. The Guide offers directions on how to use the tools of the Toolkit, describes inputs needed for each tool, illustrates how output for some tools serve as input to others, and details how the output of several tools are linked to generate final projection products.

The second section of the Guide further explains the cohort-component and mathematical extrapolation approaches and provides the logic behind the selection and synthesis of specific techniques, given particular data types and the administrative level

to which projections apply. The second section also describes the subnational projection software that has been available on the Census Bureau's Web site (<http://www.census.gov/population/international/software/>), including the RUP software and the Excel-based CTBL32 workbook. Finally, this section introduces the Toolkit workbooks, consisting of: an upgraded version of CTBL32, and new workbooks –LTCSRMig, MIGSUB, PROJE032, PROJTFR32, RUPCompare, RUPSTCompare, RUPSubAdj, and SALGST.

The third section of the Guide provides a summary of the projection process, touching upon projection strategy, preparing projection assumptions, and conveying uncertainty in projections.

Methods and Software

This section presents two fundamental population projection approaches – the cohort-component method and mathematical extrapolation – and explains the selection criteria applied in using each. The application of each method, as well as their synthesis, is illustrated with examples, as implemented at the appropriate administrative level and with the proper data type. The subnational projection software available on the Census Bureau's Web site is also reviewed: the RUP projections software and the Toolkit Excel workbooks CTBL32, LTCSRMig, MIGSUB, PROJE032, PROJTFR32, RUPCompare, RUPSTCompare, RUPSubAdj, and SALGST.

Cohort-component versus mathematical extrapolation methods

In the cohort-component method, a base population is determined that agrees with known demographic characteristics of the nation or larger subnational area. Levels of fertility, mortality, and migration, referred to as "components of change," are determined for the base year and assumed for future years. Then the base population is projected into the future according to the projected components of change. Extrapolation, on the other hand, is undertaken by assuming that the base population (usually total population) grows according to a mathematical function (e.g., linear, exponential, or logistic). As a practical matter, a function is fitted to observed data, to summarize or smooth the data, and then extended beyond the periods for which the observed data are available to create projections for later years. The approach is simpler than the cohort-component method, and its data requirements are relatively minimal. In addition to serving as a means to project population, mathematical extrapolation can also be used to project change in fertility, mortality, and migration. These projected components of change can then serve as input for cohort-component projections.

The cohort-component method is generally recommended for projecting populations of nations or larger subnational areas – usually ADM1s, such as regions or large states, or groups of ADM1s – since the data available for these areas tend to meet the technique's requirements. Where data are available, the cohort-component method is preferable since it provides a means of projecting population more realistically and can often generate projections that turn out to be more accurate than those generated with the simpler extrapolation methods. However, for lower level administrative area populations, usually ADM2s, such as districts or counties, mathematical extrapolation is recommended if, as is often the case, there are data availability or data quality issues, and modeling internal migration involves substantial uncertainty in addition to computational burden.

Extrapolations for smaller area populations can be adjusted to agree with an independent projection of the larger area to which they belong. For example, the sum of ADM2 populations can be adjusted to agree with an independently projected

population of their corresponding, or "parent," ADM1. This adjustment process is referred to as *raking* or *controlling*. A common raking application involves using projected population totals for a parent area, generated from the cohort-component technique, to serve as controls for the summation of smaller area populations produced with mathematical extrapolation. In this way, the smaller area population extrapolations can partially benefit from the methodological advantages of the cohort-component method.

Although there are general recommendations for the selection of a population projection approach, the decision to apply the cohort-component method rather than mathematical extrapolation techniques needs to be made on a country-by-country basis, with careful consideration given to population size, data availability, and data quality. While data appropriate for the cohort-component method tend to be available for ADM1s of many countries, such data may be limited for others. In these cases, projection with mathematical extrapolation is preferred. For the purposes of this Guide, national-level and regional (ADM1) populations will be used to illustrate applications of the cohort-component technique. Mathematical extrapolation will be presented using examples involving controlling of district (ADM2) projections to regional populations.

It is generally recommended that the projection horizon of subnational areas – especially smaller areas – be limited to 25 years or less, given that the uncertainty surrounding projections for subnational areas tends to increase as we go down the administrative levels.

Cohort-component applications: bottom-up versus top-down

There are two broad categories of cohort-component subnational projection approaches, referred to as "bottom-up" and "top-down." These categories are based on what the underlying assumption is about the relative reliability of national versus subnational area estimates and projections.[3] The bottom-up approach can be used if the quality of data for subnational areas is considered good. In this approach, national-level estimates and projections are calculated as the sum of subnational estimates and projections. In some cases, the reason for using bottom-up is because it is thought that composition effects will cause national-level projections to have unusual trends (that is, difficult-to-explain trends attributable to the fact that the characteristics of faster-growing areas have more impact on national-level components of growth over time). The top-down approach, conversely, assumes that national-level figures are more reliable than

[3] The terms "estimate" and "projection" can vary across demographic literature and software manuals. They can refer to populations or components of change. Information about a present or past population (or component of change) not from a census or population register is often referred to as an estimate. Methods commonly applied to generate estimates include those that adjust recent census counts. However, in some cases, data suitable for estimation are not available, and techniques normally used for population projection are employed to calculate current and past populations. For the purposes of this Guide, estimates refer to a present or past population; projections refer to future populations.

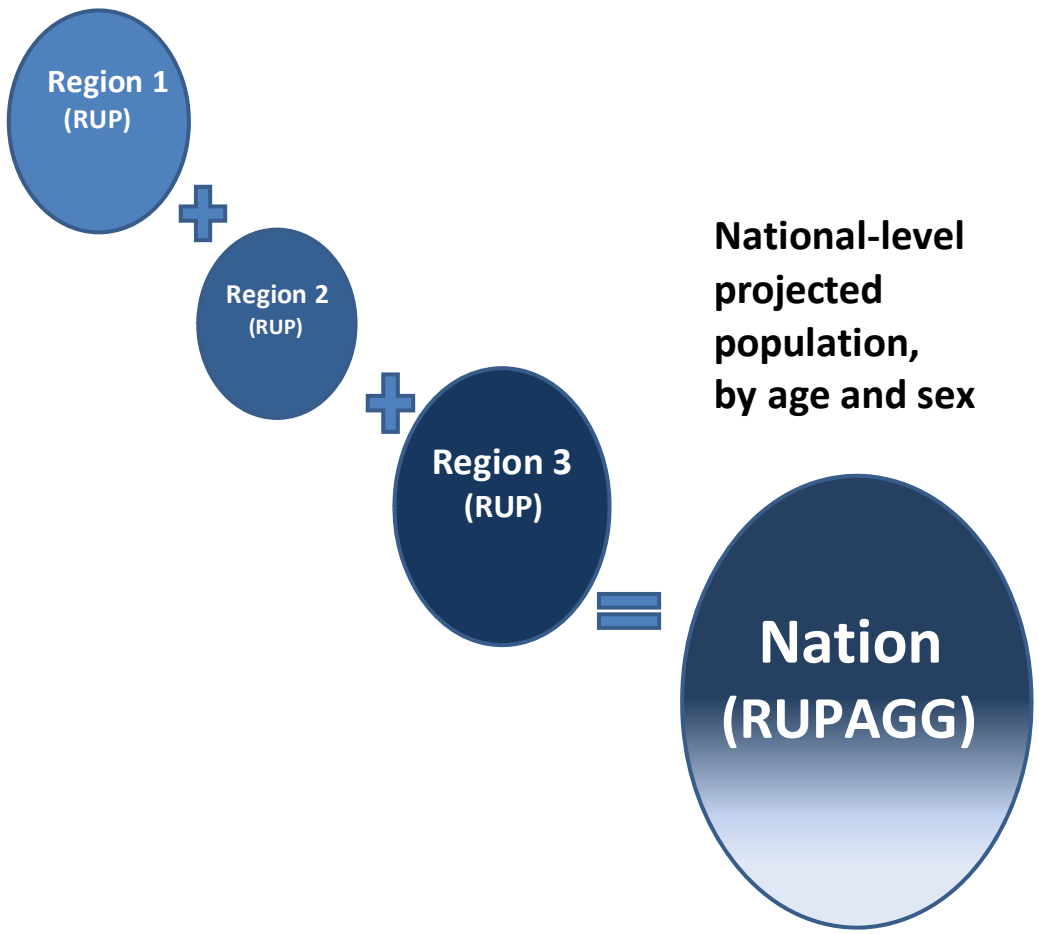
those at the subnational level and that, therefore, the subnational estimates and projections should be adjusted to be consistent with the independent national figures.

The concepts reflected in the top-down and bottom-up assumptions can be applied to lower geographic levels, as well, such as in the task of considering the relative accuracy of data for ADM1s compared to ADM2s. In this context, a bottom-up approach would assume that ADM2 estimates and projections are more reliable than those at the ADM1-level and that, therefore, the ADM1 projections should be calculated as the sum of the ADM2 figures. Conversely, the top-down approach would assume that the ADM1-level figures are more reliable than those at the ADM2 level and that, therefore, the ADM2 area projections should be controlled to the ADM1 figures. For the purposes of this Guide, examples of top-down and bottom-up applications will rely on the national-level and ADM1 populations to illustrate the assumptions and software applications involved in each.

The Census Bureau's Rural-Urban Projections (RUP) program and its Population Analysis System (PAS) workbooks support the application of both top-down and bottom-up approaches.[4] In the bottom-up approach, a RUP input file is developed to represent each subnational area, and the summation of the output projections for these areas (using program RUPAGG) is the national-level projection (Figure 5). The data used to create the input files – base population and measures of the components of change – can be generated using appropriate source data and PAS and Toolkit workbooks.

[4] The RUP and PAS software are available from the Census Bureau's Web site <http://www.census.gov/population/international/software/>.

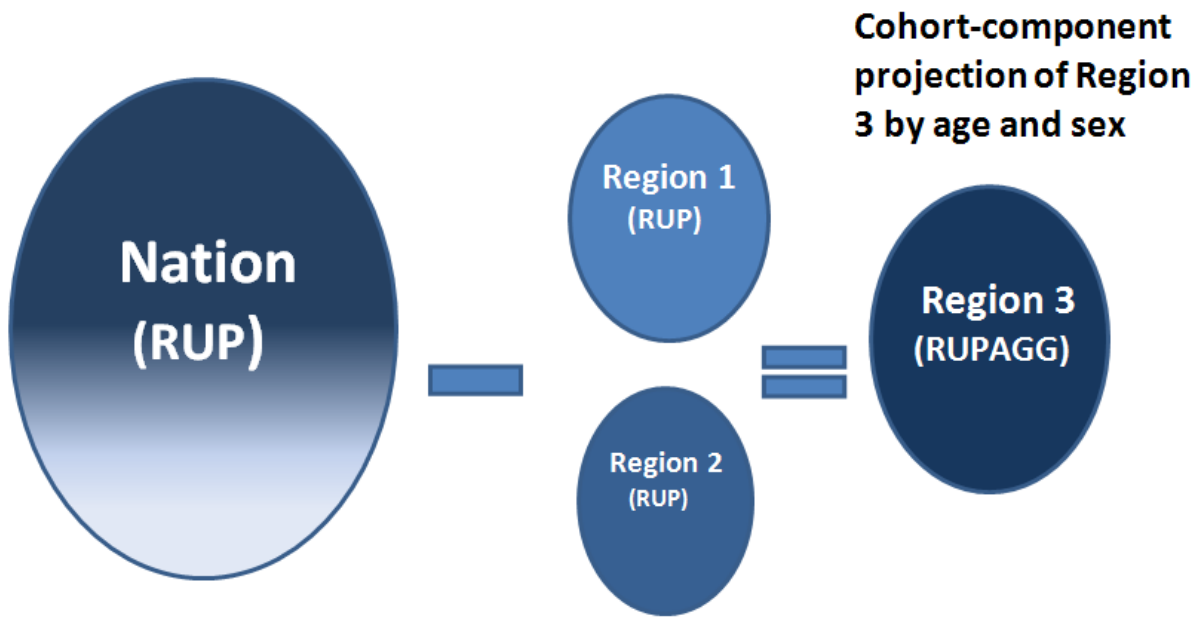
Figure 5. Application of cohort-component method with RUP: bottom-up approach



The top-down application is executed by developing one national-level RUP file and a series of RUP files for subnational areas. As a practical matter, the top-down approach may be implemented in one of two ways.

(1) National projection and projections for some subareas. The difference between the national projection and the summation of the subnational projections– or "residual"– is then generated as the projection of the missing area or areas (Figure 6). The residual may be one or more administrative areas (i.e., the sum of areas not modeled individually). As discussed below, this application is implemented using the Census Bureau's RUPAGG program.

Figure 6. Application of cohort-component method with RUP: top-down approach with a residual area

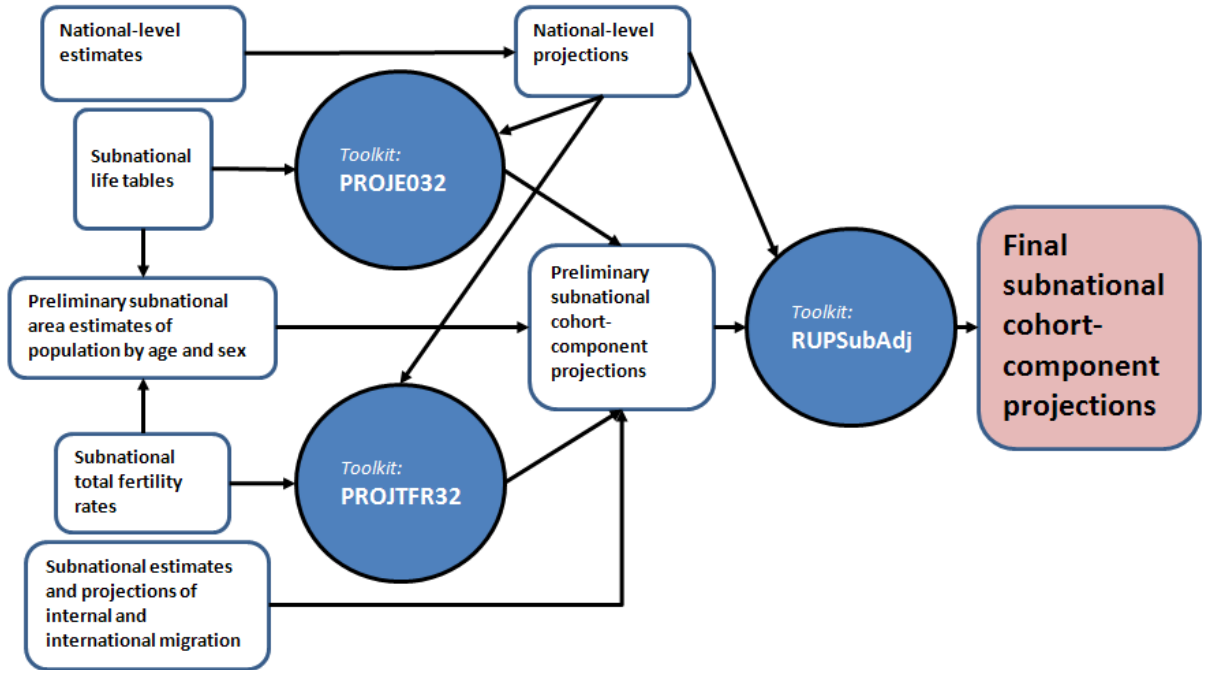


(2) The national projection and a complete set of regional projections. In this case, the subarea projected populations, births, deaths, and international migration are all forced to sum to the national control projection and net internal migration across regions is forced to equal zero in every projection year. The calculations are made using one national-level RUP input file, a series of preliminary estimates and projections for each subnational area, and proportional adjustment with one or both of the Toolkit workbooks CTBL32 and RUPSubAdj (see Figure 7 and Figure 8). As with the bottom-up approach, input files for RUPs involved in the top-down approach can be constructed using PAS-based applications of demographic techniques and appropriate source data.

As a practical matter, bottom-up, top-down with a residual area, and top-down with proportional adjustment to a control projection all assume that the analyst begins by creating a series of national and subarea base-year populations adjusted for age misreporting and age-sex-area-specific coverage errors, which are also consistent (particularly for the under-10 age group) with estimated fertility, mortality, and migration. In the top-down with residual area approach (Figure 6), any discrepancy between the national projection and the sum of regional projections for all but one (or more) area(s) is subsumed into the residual region. However, if the regional projections are carefully constructed, the residual area's projection will represent projected population, births, deaths, and migrants rather than error. In the top-down with proportional adjustment framework, having calculated adjusted, but still preliminary, base-year populations, the analyst may choose one of two approaches.

(1) One-stage procedure. In the base year and in each year of the projection, population, births, deaths, and international and internal migration are adjusted to match a national control projection's population, births, deaths, and migrants. Subnational area rates are then re-calculated. This procedure, implemented using the Toolkit workbook RUPSubAdj, is illustrated in Figure 7.

Figure 7. One-stage proportional adjustment with Toolkit



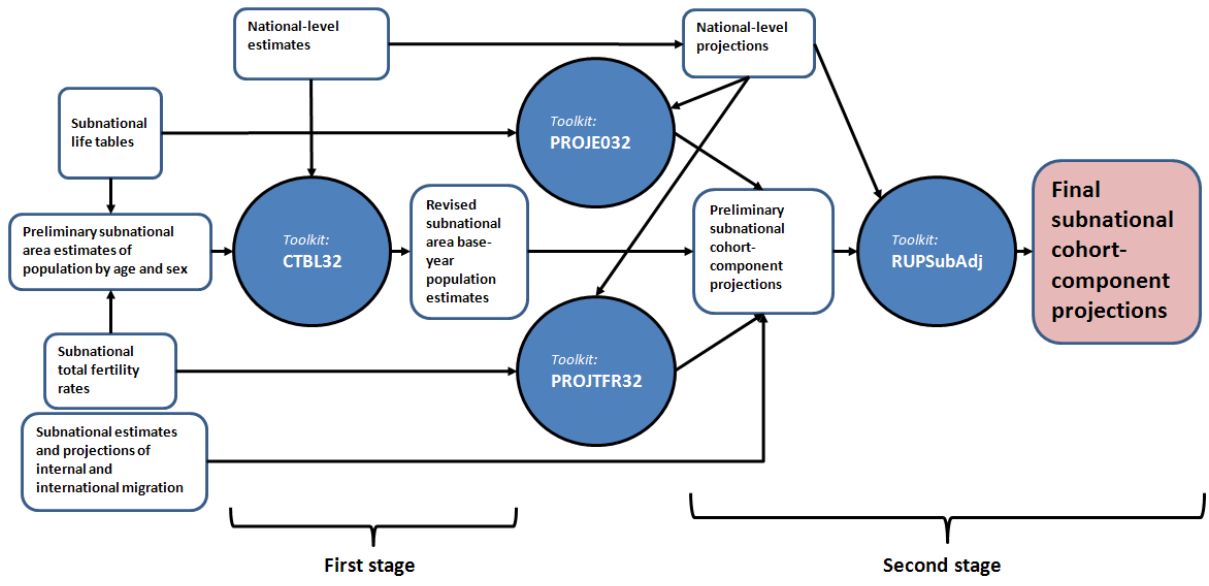
(2) Two-stage procedure. This approach refers to the situation where estimated base populations do not add up to the national total. Base-year populations are adjusted to match the national population by age and sex using Toolkit workbook CTBL32 (Figure 8).

Regardless of whether a one-stage or two-stage procedure is used, the preliminary regional projection files may be prepared using one of the PAS workbooks -- BASEPOP or BPSTRNG (for consistent base population construction) -- and the Toolkit workbooks PROJTFR32 and PROJE032 (for fertility and mortality projection). In addition, internal migration by age and sex may be calculated to ensure zero net internal migration across regions, and international migrants must be allocated across regions for the base year.[5] The final adjustment, implemented with the Toolkit workbook RUPSubAdj,

[5] The Toolkit does not currently provide software for adjusting (raking) regional net international migration to match national net international migration apart from RUPSubAdj. If all age-sex-specific net international migration were of the same sign (that is, either all emigration or immigration), CTBL32 could be used to ensure consistency in assumed numbers of international migrants in the base year as well as in other years of the projection. Otherwise, the plus-minus approach described in Shryock, Siegel, and

guarantees that in each year of the subarea projections, projected populations, births, deaths, and international migrants are adjusted to match the national control projection values, and net internal migration is forced to zero (Figure 7 or Figure 8).

Figure 8. Two-stage proportional adjustment with Toolkit



Incorporating migration

Preparing migration assumptions for subnational cohort-component projection, whether for internal, subarea-to-subarea migration, or for international migration to and from each subarea, presents special estimation and modeling challenges. However, projections at the subnational level depend critically on having reasonable migration assumptions, and this is increasingly true the lower into the administrative hierarchy of subnational areas one projects with a component model. The task of making migration assumptions for subnational projections, when using the cohort-component method, is especially important because migration often explains much of the variation in population growth at lower administrative levels.

Methods for measuring and modeling international and internal migration have been described in many publications, including the Census Bureau's ***Population Analysis with Microcomputers*** (Arriaga, 1994). Other discussions dating back to the early 1950s include Bogue, Hinze, and White, 1982; Edmonston and Michalowski, 2004;

Associates (1971:705-706); Smith, Tayman, and Swanson (2001); and elsewhere would be needed. Similarly, forcing net internal migration to be zero in every projection year could be accomplished by estimating numbers of internal migrants for each region for every fifth year or so and ensuring the zero using formulas published by Leete (1992). RUPSubAdj applies a modified version of the calculations described in Leete (1992). For more information see the Methods Description document for RUPSubAdj (U.S. Census Bureau 2013b).

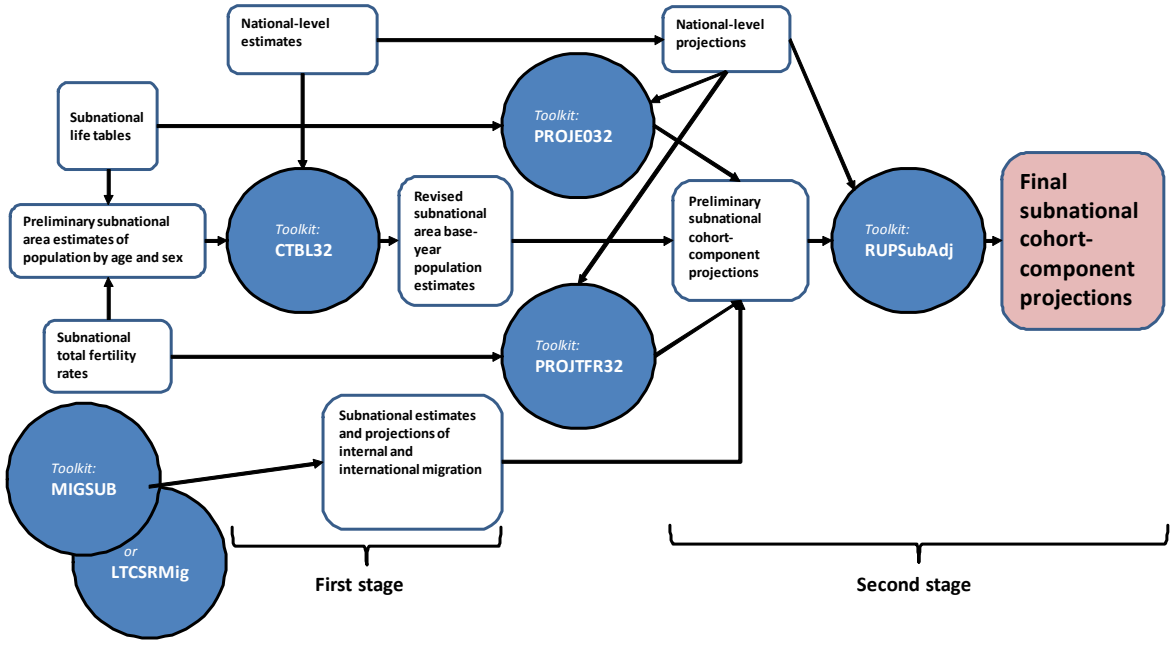
Morrison, Bryan and Swanson, 2004; Rogers and Castro, 1981; Rogers, Little and Raymer, 2010; Shryock and Siegel, 1971; Siegel and Hamilton, 1952; Smith, Tayman and Swanson, 2001; United Nations, Department of International Economic and Social Affairs, 1970 and1992; United Nations Statistical Office, 1978; United Nations Statistics Division, 1998.

The Toolkit offers two workbooks designed to facilitate the preparation of migration assumptions for subnational cohort-component projection (see Figure 9):

- MIGSUB supports direct estimation of migration for component projection. MIGSUB estimates area-specific in-migration, out-migration, and net migration for subnational areas (subareas) based on tabulated responses to a question about residence a fixed number of years prior to a census or from a published table of flows of in-migrants and out-migrants for that period. The pattern of internal migration from the reference census is used, along with user-provided assumptions about trend in migration levels, to project numbers of internal migrants for a user-specified time horizon. The same pattern is used to model international migration into and out of subareas over the projection horizon. In addition, total net migration (internal net migration plus international net migration) for each subarea is generated for use in projection models that do not distinguish migration streams.

- LTCSRMig supports indirect estimation of migration. LTCSRMig calculates residual estimates of net migration for subareas using intercensal survival ratios. Estimates are generated using forward and backward life table survival, an average of the

Figure 9. Using Toolkit workbooks to incorporate migration



forward and backward estimates, census survival ratios, and a composite of census survival and life table survival ratios. User-selected child-woman ratios are used to complete the age pattern of net migration.

Cohort-component projection approaches with RUP and Toolkit software

Up to this point, we have presented overviews of the input-output flow of software applications used to generate subnational projections. We are now going to illustrate the detailed mechanics and programming requirements necessary to implement the projection process using the Census Bureau's RUP and Toolkit software.

The RUP input file, mentioned on previous pages, is illustrated in Figure 10. This is a RUP input file for the hypothetical country, Republic of Demographica. The RUP input file can be developed with the help of the Census Bureau's PAS software, as applied to appropriate source data.[6] For descriptions of how PAS software may be used to help develop the elements of a RUP input file, see Arriaga (1994) and the description of cohort-component methodology used by the Census Bureau to generate national-level estimates and projections for its International Data Base (U.S. Census Bureau 2013a).

In preparing bottom-up subnational projections, one RUP input file (an ".in" file) is usually developed for each subnational area, and the output of all subnational RUP projections, called intermediate output files (or ".io1" files)," are aggregated to generate national-level projections. Figure 11 illustrates the aggregation of a series of three RUP projections (*.io1 files) created by executing RUP input files for subnational areas and the generation of a combined projection for the country. The RUPAGG input file is shown in Figure 12.

[6] For more information, see <http://www.census.gov/population/international/software/pas/>.

Figure 10. Illustration of a RUP input file

```
TITL                1
POPULATION PROJECTIONS FOR DEMOGRAPHICA: 2000-2050
*******************************************************************************
N5                 17
PROJ  2050
SXRB             1.05
REG                 1
TOT
CODE                1
*******************************************************************************
POP M52000               502652
   2488460   2448469   2266525   2033047   1688670   1314334   1005099    875681
    802584    699060    561363    409943    310059    223745    154793     88353
     64818
POP F52000               484046
   2414789   2404821   2185023   2044376   1719848   1365436   1029479    892138
    831229    729187    579612    428034    333684    251292    192309    118832
    105671
*******************************************************************************
MX  M42000
   0.05220   0.00494   0.00208   0.00150   0.00266   0.00391   0.00486   0.00548
   0.00671   0.00870   0.01189   0.01595   0.02235   0.03170   0.04705   0.06907
   0.10219   0.13369
MX  F42000
   0.03771   0.00491   0.00197   0.00131   0.00180   0.00221   0.00260   0.00310
   0.00411   0.00555   0.00782   0.01079   0.01556   0.02233   0.03414   0.05231
   0.08277   0.10471
MXM M 2020     66.17
MXM F 2020     72.93
...
*******************************************************************************
ASFR 52000
    0.0650    0.2000    0.1990    0.1360    0.0810    0.0490    0.0060
ASFR 52005
    0.0683    0.1664    0.1416    0.0860    0.0561    0.0215    0.0061
ASFR 52008
    0.0620    0.1512    0.1259    0.0803    0.0423    0.0158    0.0047
TFR   2020      2.05
...
*******************************************************************************
MIGNM52000
         0         0         0     20032     36772     39540     18221     16948
     14902         0         0         0         0         0         0         0
         0
MIGNF52000
         0         0         0         0         0         0         0         0
         0         0         0         0         0         0         0         0
         0
...
*******************************************************************************
OUTP                1
OMX                 1
*******************************************************************************
END
```

General information about the projection

Population data

Mortality data

[more data]*

Fertility data

[more data]*

Migration data

[more data]*

Output control

* A typical RUP input file will have mortality inputs (deaths, age-specific death rates, or life expectancies) for additional years; may have additional fertility inputs (births, age-specific fertility rates, or total fertility rates); and will show assumed migration numbers or rates for more than one year.

Figure 11. Input-output process of RUP and RUPAGG projections, bottom-up approach

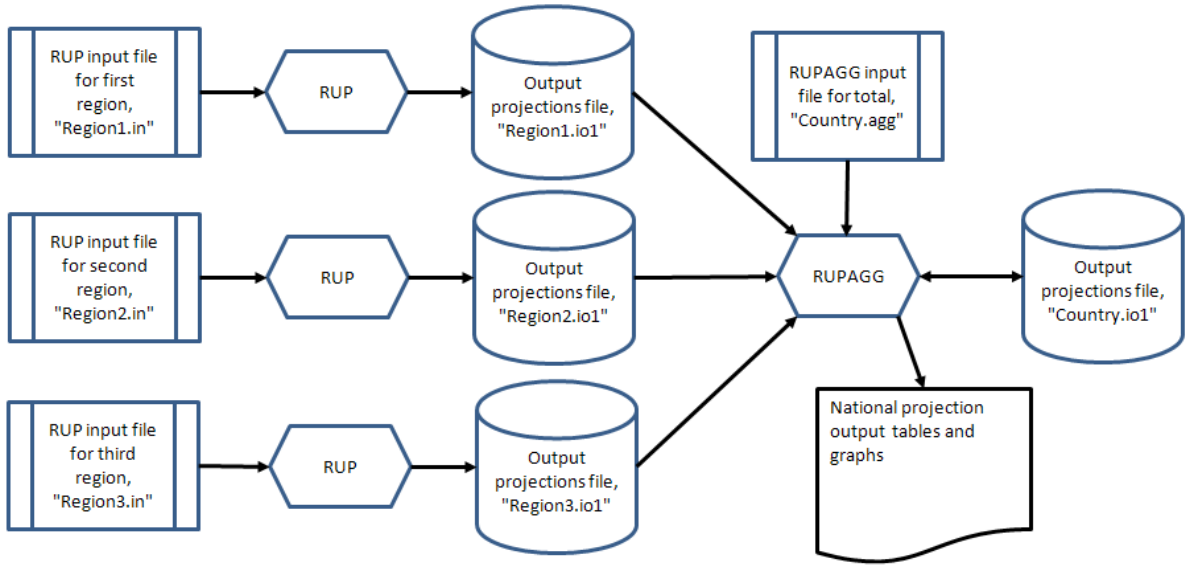


Figure 12. RUPAGG input file for bottom-up projection (file Country.agg)

```
TITL                2
Aggregation of RUP files for Regions 1, 2, and 3 of
the Republic of Demographica
BASE  2000
PROJ  2025
N                  17
TOT
CODE                3
*********|*********|*********|
INPF                1
Region1.IO1
*********|*********|*********|
INPF                1
Region2.IO1
*********|*********|*********|
INPF                1
Region3.IO1
*********|*********|*********|
OUTP                1      2025
OMX                        2025
END
```

The alternative of preparing projections for a subset of country regions and computing the residual as the difference between the national and the sum of the projected regions is illustrated in Figure 13 and the RUPAGG input file is shown in Figure 14 . The Republic of Demographica's third regional projection is calculated here as the national projection (population, births, deaths, migrants) minus the Region 1 and Region 2 projections. The projected national and regional populations and events (births, deaths, and numbers of international and internal migrants) will be fully consistent.

Figure 13. Input-output process and RUPAGG programming of RUP projections, residual approach

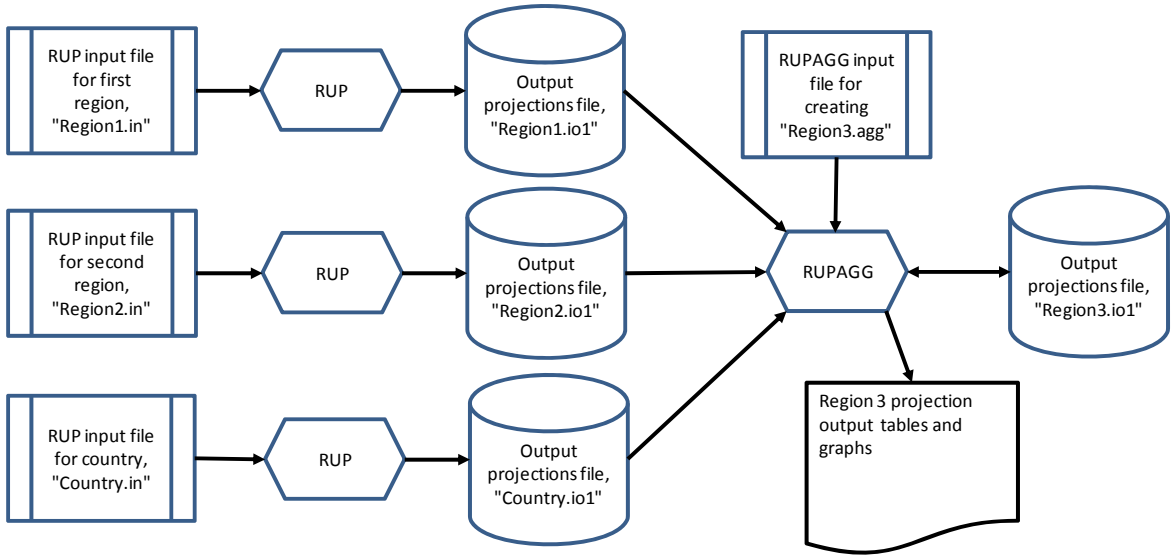


Figure 14. RUPAGG input file for top-down projection with a residual area (file Region3.agg)

```
TITL                1
Residual projection of Demographica's Region 3
BASE  2000
PROJ  2025
N                  17
TOT
CODE                3
*********|*********|*********|
INPF                1
Country.IO1
*********|*********|*********|
INPF               -1
Region1.IO1
*********|*********|*********|
INPF               -1
Region2.IO1
*********|*********|*********|
OUTP                1      2025
OMX                        2025
END
```

Note the **-1** for Region1 and Region2 indicating these should be subtracted from the national totals in Country.io1.

One step in the process of preparing subnational projections with cohort-component methodology and RUP, whether using RUPAGG to create a residual region or RUPSubAdj to proportionally adjust lower-level projections to match a higher-level projection, has not been mentioned up to now. This is the evaluation step.

The analyst will want to check that subarea projection outputs (population, births, deaths, and international migrants) sum to projected national figures. The analyst will also want to check that there is no discontinuity in the age distribution generated by the process of proportional adjustment. This check can be made by generating a series of population pyramids for the projection launch year and the following two to five years from the RUPEX Control sheet.[7] A disconnect, evident in a shelf or a notch in the age distribution of one or more subareas, may indicate a need to recalculate affected base populations using adjusted fertility and mortality from RUPSubAdj. In addition, the analyst should check that the resulting mortality, fertility, and net migration values by age and their trends for each subnational area look reasonable. These checks can also be made using the output option buttons on the RUPEX workbook.

Some additional evaluation is warranted when projecting subarea populations. Not only must subarea populations add to the larger projected population to which they belong, and not only must internal net migration sum to zero, but trends in fertility, mortality, migration, and population across areas should be both as expected by the analyst and plausible. Toolkit workbooks CTBL32, PROJE032 and PROJTFR32, and RUPSubAdj help ensure internal consistency of national and subnational projections. In addition, however, workbooks RUPCompare and RUPSTCompare facilitate comparison of projected subarea populations and component indicators to further improve projection quality at the subnational level.

Mathematical extrapolation

In developing population projections with mathematical extrapolation, CTBL32 may be used to generate smaller area age distributions, or age-sex distributions, that are consistent with the age-sex distribution of the population of the higher-level geography to which the areas belong. For example, CTBL32 can generate ADM2 age distributions that are consistent with the age distribution of their “parent” ADM1. The method employed in this process is referred to as “two-way raking.”[8]

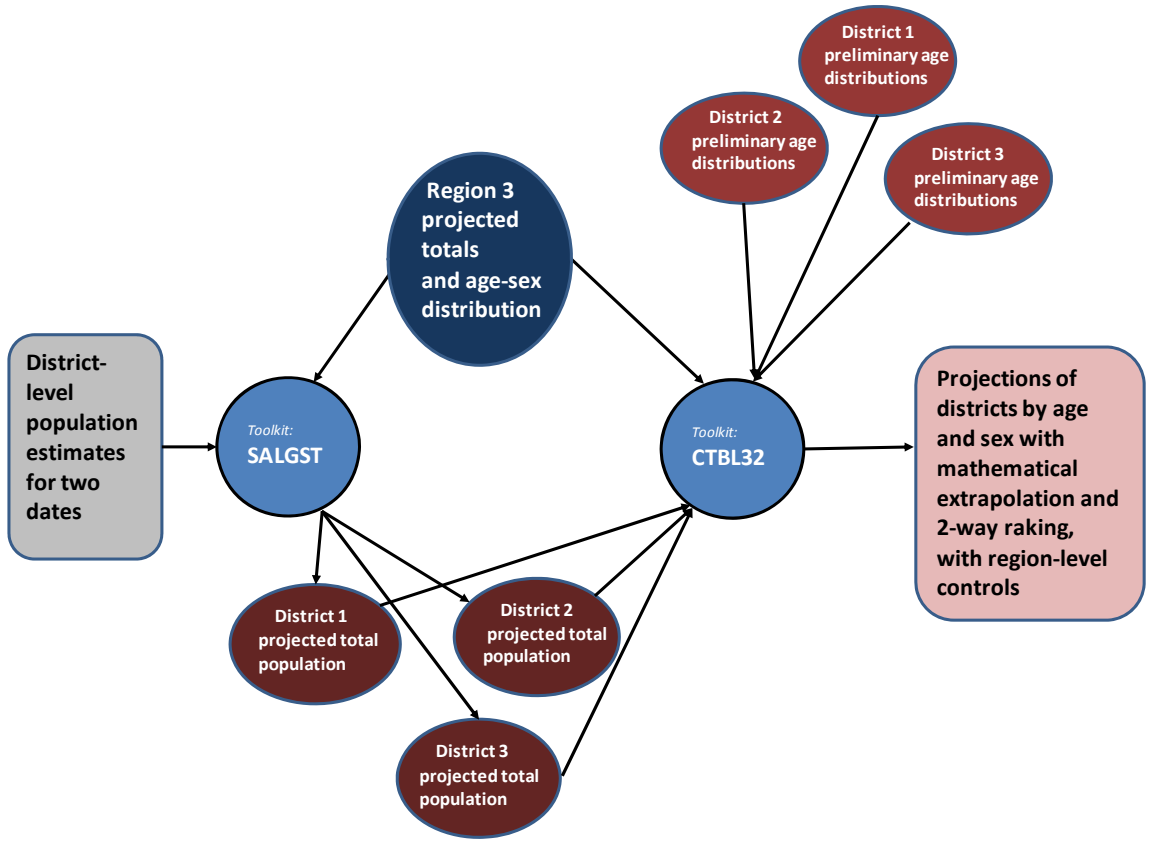
In applying two-way raking to develop ADM2 age structures, CTBL32 controls the populations within the ADM2 preliminary age groups to two variables: (1) the ADM1 projected population by 5-year age group and (2) the ADM2 projected population totals.

[7] The launch year referred to in the Guide is the beginning, or base, year of the projection.
[8] Shryock, Siegel, and Associates (1971: 707, 795). The two-way raking procedure is also referred to as N-dimensional controlling (Smith, Tayman, and Swanson 2001: 260), contingency table adjustment, the method of iterative proportions or iterative proportional fitting, and the Deming method (Shryock, Siegel, and Associates (op cit., 707). In addition to age groups, two-way raking can be applied to other variables such as educational attainment, labor force, income, languages, and ethnicity.

This is illustrated in Figure 15 where SAGLST is used to project district population totals.

If, as would normally be the case, age and sex consistency is the goal, two SALGST extrapolations and two CTBL32 tables should be used, one for males and one for females. Note that the Region 3 projected populations by age and sex in Figure 15 can come from RUP projections or RUPAGG adjusted projections produced by RUPSubAdj.

Figure 15. Application of mathematical extrapolation with age-sex detail using CTBL32 and SALGST



Using the Toolkit

This third and last section of the User's Guide offers a summary of the population projection process, with emphasis at each step on the use of RUP software and the Toolkit workbooks. By now, the reader has realized that the Census Bureau's Subnational Projections Toolkit represents a sort of "capstone," bringing together the Census Bureau's Rural-Urban Projection software, its Population Analysis System spreadsheets, and the nine workbooks constituting the Toolkit, listed in the table below, as an integrated set of tools for preparing subnational projections.

Toolkit Workbooks and Methods Used for Developing Subnational Projections		
Workbook	**Function**	**Data Required**
Cohort-component projection		
CTBL32	Two-way raking of base populations	National base population by sex and age and ADM1 preliminary estimates of population by sex and age
PROJE032	Mathematical extrapolation of life expectancy at birth (e(0))	National estimates and projections of e(0) for several dates and subnational estimates of e(0) for one date
PROJTFR32	Mathematical extrapolation of total fertility rate (TFR)	National estimates and projections of TFR for several dates and subnational estimates of TFR for one date
LTCSRMig	Indirect, survival ratio-based estimation of net migration for an intercensal period	National and subarea age-sex distributions 5 or 10 years apart plus, for life table survival, national and subarea life table survival ratios
MIGSUB	Projection of internal and international migrants by sex and subnational area	Population by subarea at the time of a census and at a prior time point, user-specified growth of total population, internal migration rate, and international migration
RUPSubAdj	Proportional adjustment of population/events	National and subnational RUPs
RUPCompare	Comparison of RUP output for two populations	Two RUP projections
RUPSTCompare	Comparison of RUP output for a user-defined number of populations	RUP projections for each of a user-defined number of populations

Non-cohort-component methods / mathematical extrapolation		
SALGST	Mathematical extrapolation of population using a simple logistic curve	ADM1 estimates and projections of population for several dates, and ADM2 estimates for two dates
CTBL32	Two-way raking of projected populations, one year at a time	An ADM1 projection of population by sex and age, ADM2 projections of total population, and ADM2 preliminary estimates of population by sex and age

We begin this section with guidance on projection methodology (cohort-component versus mathematical extrapolation) and conclude with extrapolation of smaller areas with age detail.

Choosing a projection strategy involving a combination of cohort-component and mathematical extrapolation methods

The first section of the Guide touched upon the fundamental decision of whether to prepare cohort-component projections or non-cohort-component (mathematical) projections at the national, ADM1, and lower ADM levels. No advice was given about how to choose. And, while the cohort-component method is technically superior because it takes changing fertility, mortality, and migration into account, it is also more data- and time-intensive, and therefore more expensive. The competing considerations of technical superiority and cost mean that neither general approach is inherently "better" for a specific country or subarea level of that country. What we can say is this:

- Population projection at the national level should be undertaken using cohort-component methodology.

- Projection at the first subnational administrative level may be made using the cohort-component method if available fertility, mortality, and internal migration data for the subareas are of good quality. However, the cost of working with these data must be considered.

- Projection at the ADM2 and lower levels may be made using non-component methods whether projections are to be made with or without age detail, providing that administrative boundaries have not changed from census to census, or populations can be estimated for a consistent set of defined bounded areas.

- Regardless of projection method selected for subnational areas, national and lower ADM projections must be made consistent. Lower-level ADM projections, if

prepared by age and sex, should be adjusted to add to higher-level population control projections by age and sex (top-down approach) or higher-level projections may be calculated as the sum of lower-level projections (bottom-up approach).

- Within a top-down framework or bottom-up framework, a national cohort-component projection made using RUP can be made fully consistent with a series of ADM1 projections by using RUPAGG. In the bottom-up case, the analyst may use RUPAGG to add the subarea-specific projections; in the top down case, the analyst may use projections for a subset of subareas and a higher administrative level RUP projection to calculate a projection for a residual area by subtraction. Within the top-down case, a series of ADM1 RUP projections may be forced to be consistent with a national RUP projection by using RUPSubAdj. Mathematically extrapolated populations (e.g., at ADM1 or ADM2 level) may be prepared and forced to match a higher-level administrative area cohort-component projection using SALGST. Age detail for these non-cohort-component projections may be added using CTBL32.

Choosing between top-down and bottom-up projection for subarea populations

The above discussion has bypassed the question of whether the top-down or the bottom-up approach is generally preferable. Earlier it was noted that the choice between the two approaches depends on the analyst's assessment of the relative reliability of national versus subnational area estimates and projections.

However, an equally likely case is the one in which data for one or more subarea populations are questionable. In this case, top-down forced consistency either distributes error proportionately across subareas, whether more or less affected by error (the RUPSubAdj approach), or it forces the error into a residual subarea, which may be a subarea with otherwise relatively better data. It might be preferable to accept population, fertility, mortality, or migration for subareas that appear to have better quality data and then to adjust population and components of change for subareas with less convincing data, combining projections for subareas with better data with projections for weaker subareas after adjustment.

If, on the other hand, data are problematic or generally missing at lower levels, or if internal migration data are especially problematic, the residual approach may be recommended (the RUPAGG approach with a residual area).

Moreover, some national statistical agencies publish national population projections prior to subnational projections and are reluctant to revise those national projections in order to take advantage of better quality subarea projections made later on. In that situation, the agency is forced to use a top-down approach, controlling subnational projections to the previously released national projection.

Preparing the launch year projection inputs

The tasks of preparing a base population and launch year estimates of fertility, mortality, and migration for cohort-component projection were introduced earlier. The Census Bureau's 1994 publication, ***Population Analysis with Microcomputers***, lays out the full process, including the preparation of the crucial launch year estimates of population and the components of change, and that overview is worth recalling here. From ***Population Analysis with Microcomputers***, volume I, chapter VIII:

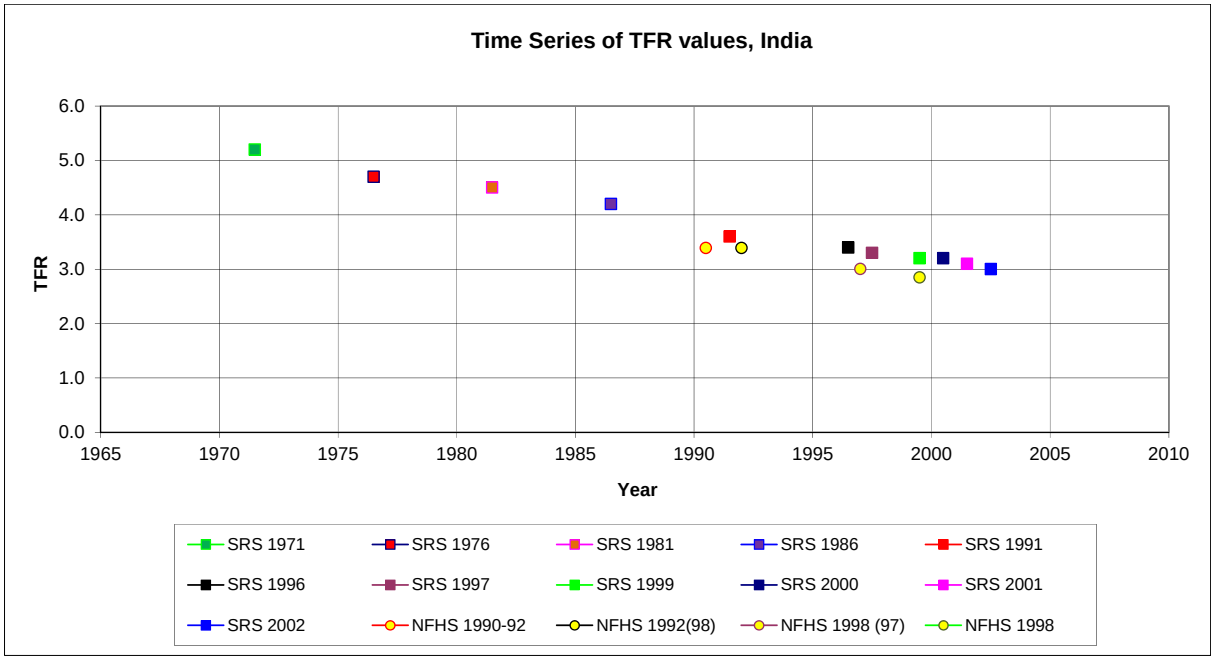
> A base population is determined that agrees with known demographic characteristics of the country. Levels of mortality, fertility, and migration are determined for the base year and projected to future years. Then the base population is projected into the future according to the projected components of change.
>
> - A base population is obtained after analyzing the available age and sex structure …The Census Bureau's PAS spreadsheets GRPOP-YB, PYRAMID, SINGAGE, AGESEX, AGESMTH, BASEPOP, and BPSTRNG … are strongly recommended.
>
> - Two aspects of mortality have to be projected: the level and the pattern by age and sex. The Census Bureau's PAS spreadsheets E0LGST, FITLGSTC, LOGISTIC, INTPLTM, and INTPLTF [as well as] the United Nations MATCH program in the MORTPAK package [are especially useful for this task].[9] Those underlined are strongly recommended.
>
> - Two aspects of fertility have to be projected: the level and the age pattern. The Census Bureau's PAS spreadsheets TFRLGST, TFRSINE, FITLGSTC, LOGISTIC [are especially useful for this task] … Those underlined are strongly recommended.
>
> - If more than one area within a country is to be projected, two aspects of internal migration have to be estimated for the past and projected into the future: the total number of migrants and their age and sex composition.
>
> - Two aspects of international migration have to be estimated: the total number of migrants and their age and sex composition.

Twenty years later, these five steps still constitute the backbone of the cohort-component projection process. At the same time, new tools have been developed and increased attention has been given to the visual representation of data as a guide in preparing projection inputs. Graphical cohort analysis (GRPOP-YB) is no less useful

[9] The United Nations' MORTPAK package provides twenty software tools for demographic analysis, with special emphasis on mortality estimation. MORTPAK is available at http://www.un.org/en/development/desa/population/publications/mortality/mortpak.shtml.

today than it was two decades ago in identifying age-misreporting and age-sex-selective coverage problems in censuses. Graphing indicators such as the total fertility rate, under-5 mortality rate, or life expectancy at birth from multiple sources for multiple time points continue to be recognized as an important way of comparing and assessing the plausibility of divergent estimates from censuses, vital statistics systems, and surveys, and for determining trends in those indicators. Figure 16, for example, shows a series of total fertility rate estimates from India's sample vital registration and a national household survey program. Some difference in the level of fertility is apparent between the two series though the trend is consistent across series.

Figure 16. Illustration of graphical analysis of total fertility rates from India's Sample Vital Registration System (SRS) program and National Family Health Survey (NFHS) program



MIGSUB and LTCSRMig are new tools based on established methodologies for estimating levels of internal migration from a census with a question about residence X years prior to the census and from the age distributions from two censuses using survival ratios, respectively.

Preparing projection assumptions

The task of making assumptions about fertility, mortality, internal migration, and international migration (including refugees) for subnational cohort-component projections involves ensuring consistency between national and subnational projections. However, it also requires the analyst to consider whether subarea fertility and mortality, at least, should follow similar trends, and whether internal and international migration for

various subareas may be assumed to follow common trends. Toolkit workbooks PROJTFR32 and PROJE032 allow the analyst to consider the possibility that subarea TFRs and life expectancies generally move together toward some limit level. MIGSUB allows the analyst to model the volume of internal and international migration in future years as a function of the pattern of internal migration for a fixed reference period prior to a census. These three tools, taken together, provide useful alternatives to the assumption, sometimes used in subnational projections, that fertility, mortality, and internal migration will remain constant in future, or will increase or decrease independently of what happens in other subareas.[10]

Conveying uncertainty in population projections

Some users of subnational population projections, including those interested in using projections as the basis for programmatic and business decisions, may prefer to have some indication of the uncertainty associated with projected population. Smith, Tayman, and Swanson (2001:339-340, 354-355) list several approaches to quantifying the uncertainty associated with projections, including (a) preparing several sets of cohort-component projections, each with a different assumption or set of assumptions, to illustrate the range of likely outcomes, (b) constructing model-based or empirically based statistical prediction intervals, and (c) providing the user with a summary of past forecast errors. Smith et al. argue that projections are no more likely to be too high than too low, but they and George et al. (2004:593-594) suggest that relative error increases with length of projection horizon and inversely with geographic, or administrative, level. That is, mean absolute percent error is likely to be several times larger for a 30-year projection period than for a 5-year projection period, and larger to small administrative areas, such as local government areas, than for larger administrative regions. The discussion in this Guide assumes that the analyst is focusing on preparing one set of subnational population projections. The analyst is not prohibited from preparing multiple sets of projections for each subarea if time and budget permit.

Combining cohort-component projections for higher ADM level populations with mathematical extrapolation for lower ADM-level populations

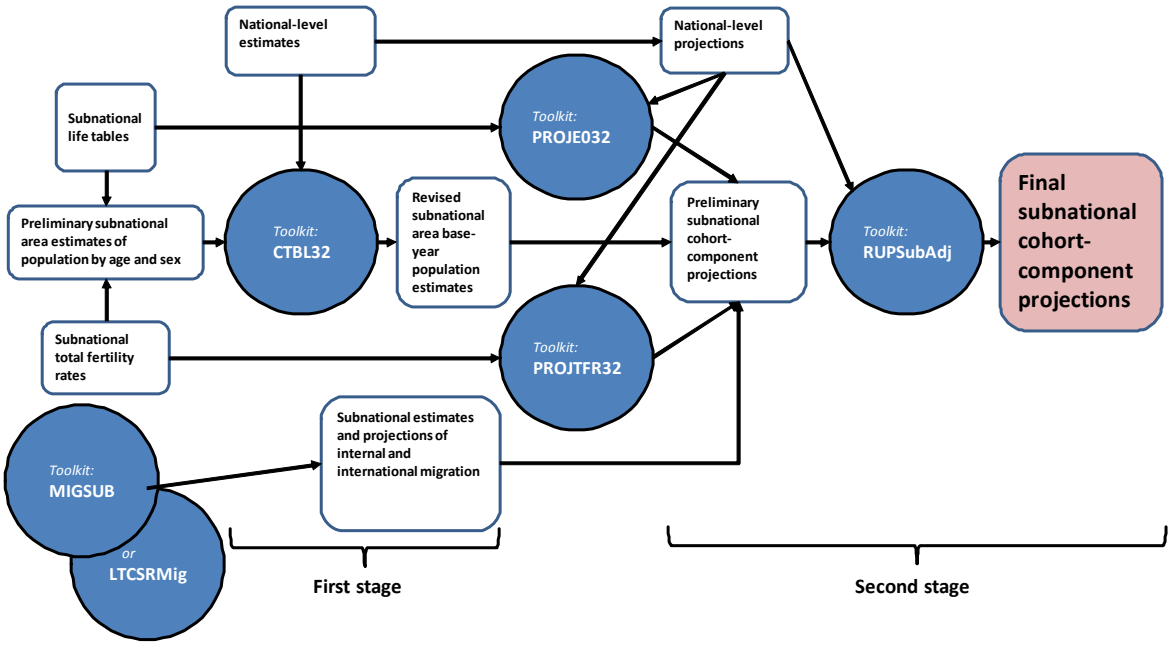
A typical set of subnational projections will involve cohort-component projection at the national and first ADM levels with mathematical extrapolation for lower ADM levels. That is, base populations and launch year fertility, mortality, and migration estimates are generated and, if a top-down approach is being used, population, births, deaths, and

[10] McDevitt (2008) showed that, with Demographic and Health survey data, the procedures used in PROJTFR32 reduce subarea forecast error for fertility. McDevitt (2009) showed that extrapolation algorithms for subnational areas that use the national trend in mortality as a guide, including the algorithm used in PROJE032, tend to yield more accurate forecasts of future mortality level than linear and simple logistic models, which are not guided by national trend.

numbers of migrants for lower-level ADMs are adjusted, if necessary, to ensure that they are fully consistent. Assumptions are made about future fertility, mortality, and migration. A preliminary set of projections based on these estimates and assumptions – previously referred to as the first stage of a two-stage projection process and depicted as such in Figure 17- is followed in a second stage ensuring that population, births, deaths, and migration are consistent over the full projection horizon.

Meanwhile, population totals may be mathematically extrapolated for ADM2 and smaller subareas based on at least two censuses, preferably after adjustment for age-and-sex-specific undercount. SALGST may be used for this purpose, projecting males and females separately. Finally, CTBL32 may be used to prepare ADM2 or lower-level projections consistent with cohort-component projections for ADM1 and national projections.

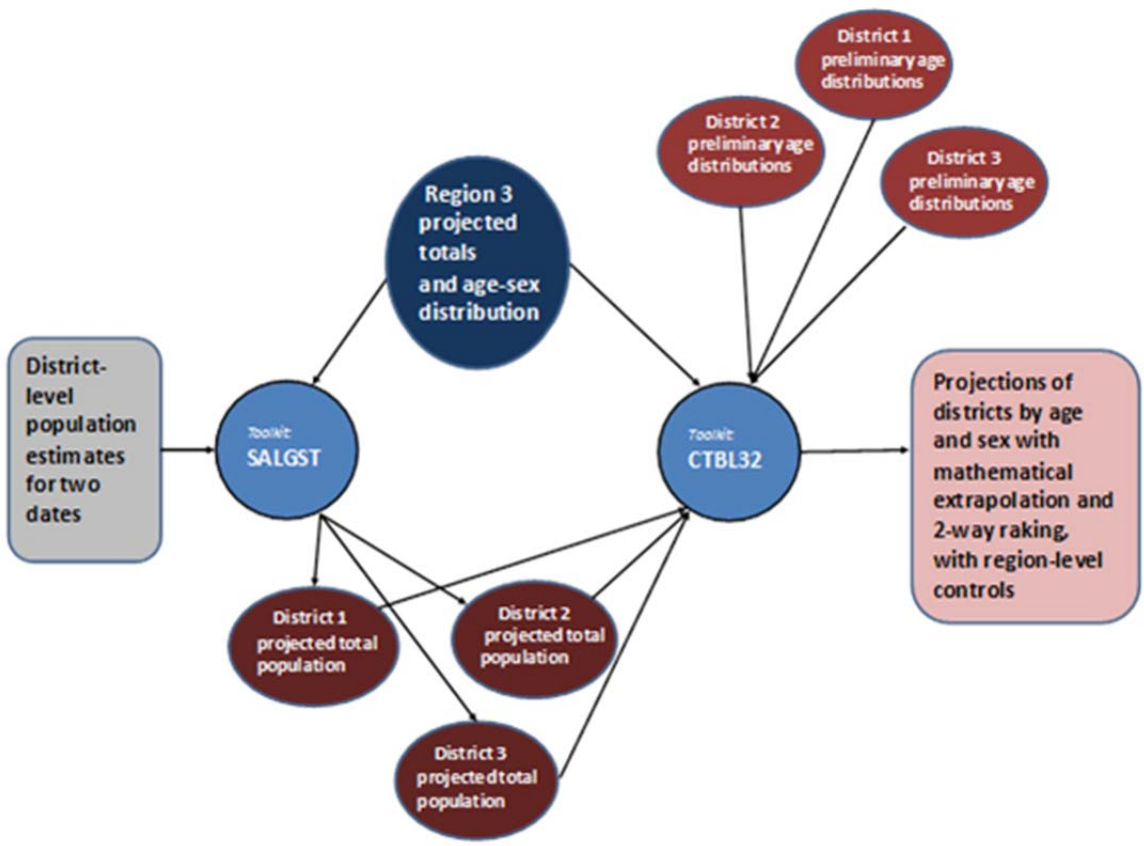
Figure 17. Flow of Toolkit calculations, from preliminary estimates to final subnational cohort-component projections: top-down approach



The combination of SALGST and CTBL32 are illustrated in Figure 18. SALGST is used to project district-level total populations by fitting a logistic curve to the proportions of the ADM1 population resident in three districts. The extrapolated district populations are controlled to match a higher-level control projection, in this case for Region 3. SALGST's use of the logistic curve in extrapolating population ensures that individual subarea population proportions projected are always greater than zero and less than one. SALGST's use of a higher-order control projection ensures that the sum of projected subarea populations always matches the projected population of the higher-level administrative region. On the right half of Figure 18, the district totals are combined with (1) a projected sex-specific regional age distribution and (2) preliminary

age distributions for the districts (perhaps from the most recent census or survey) to generate a new set of projected district age distributions controlled to the Region 3 age distribution.

Figure 18. Input-output flow of mathematical extrapolation with age detail using Toolkit tools



References

Arriaga, Eduardo E. 1994. ***Population Analysis with Microcomputers***. With Peter D. Johnson and Ellen Jamison. Washington, D.C.: U.S. Bureau of the Census, U.S. Agency for International Development, and United Nations Population Fund.

Bogue, Donald J., Kenneth Hinze, and Michael White. 1982. ***Techniques of Estimating Net Migration***. Chicago: University of Chicago, Community and Family Study Center.

Economic Commission for Africa and Regional Institute for Population Studies. 1985. ***Report of the Regional Training Workshop on Demographic Estimates and Projections in Africa***, Vol. 2, Appendix: Workbook on National and Subnational Projections. Accra, Ghana.

Edmonston, Barry and Margaret Michalowski. 2004. "International Migration," in Jacob S. Siegel and David A. Swanson, eds. 2004. ***The Methods and Materials of Demography***, Second Edition. San Diego, CA: Elsevier, pp.455-492.

George, M.V., Stanley K. Smith, David A. Swanson, and Jeffrey Tayman. 2004. "Population Projections," in Jacob S. Siegel and David A. Swanson, eds. 2004. ***The Methods and Materials of Demography***, Second Edition. San Diego, CA: Elsevier, pp.561-601.

Irwin, Richard. 1977. "Guide for Local Area Population Projections." Technical Paper No. 39. Washington, D.C.: Bureau of the Census.

Leete, Richard, 1992. ***PEOPLE***. With Fong Cheng Hong, Cheng Fan Soon, and Geoff Simmons. Software and documentation distributed by the United Nations Economic and Social Commission for Asia and the Pacific (ESCAP). Bangkok.

McDevitt, Thomas. 2008. "Consistent Projection of Fertility for Subnational Areas Using a Logistic Curve." Poster presented at the annual meetings of the Population Association of America, New Orleans, April 17-19, 2008.

McDevitt, Thomas. 2009. "Consistent Projection of Mortality for Subnational Areas Using a Logistic Curve." Paper presented at the annual meetings of the Population Association of America, Detroit, April 29 – May 2, 2009.

Morrison, Peter A., Thomas M. Bryan, and David A. Swanson. 2004. "Internal Migration and Short-Distance Mobility," in Jacob S. Siegel and David A. Swanson, eds. 2004. ***The Methods and Materials of Demography***, Second Edition. San Diego, CA: Elsevier, pp.493-521.

Pittinger, Donald B. 1976. ***Projecting State and Local Populations***. Cambridge, Massachusetts: Ballinger.

Rives, Norfleet W., Jr., William J. Serow, Anne S. Lee, and Harold F. Goldsmith, and Paul R. Voss, eds. 1995. ***Basic Methods for Preparing Small-Area Population Estimates***. Madison, Wisconsin: University of Wisconsin-Madison/Extension, Applied Population Laboratory.

Rogers, Andrei and Luis J. Castro. 1981, ***Model Migration Schedules***. Laxenburg, Austria: International Institute for Applied Systems Analysis.

_____, Jani Little, and James Raymer. 2010. ***The Indirect Estimation of Migration: Methods for Dealing with Irregular, Inadequate, and Missing Data***. New York: Springer.

Shryock, Henry S., Jacob S. Siegel, and Associates. 1971. ***The Methods and Materials of Demography.*** Washington, D.C.: U.S. Government Printing Office.

Siegel, Jacob S. and C. Horace Hamilton. 1952. "Some Considerations in the Use of the Residual Method of Estimating Net Migration," ***Journal of the American Statistical Association***, 47:480-83.

Smith, Stanley, K., Jeff Tayman, and David A. Swanson. 2001. ***State and Local Population Projections and Analysis***. New York: Kluwer Academic/Plenum.

United Nations, Department of International Economic and Social Affairs. 1970. ***Methods of Measuring Internal Migration***. Manuals on methods of estimating population, Manual VI. Population Studies No. 47. ST/SOA/Series A/47. New York.

_____. 1979. ***Prospects of Population: Methodology and Assumptions***. Population Studies, No. 67. New York.

_____.1992. ***Preparing Migration Data for Subnational Population Projections***. By Alden Speare. New York.

United Nations Population Division. 1990. ***Projection Methods for Integrating Population Variables into Development Planning***. Volume I (Methods for Comprehensive Planning), Module Two (Methods for Preparing School Enrolment, Labour Force and Employment Projections). New York.

United Nations, Department of International Economic and Social Affairs, Statistical Office. 1978. ***Statistics of Internal Migration: A Technical Report***. Series F, No. 23. New York.

United Nations, Department of International Economic and Social Affairs, Statistics Division. 1998. ***Recommendations on Statistics of International Migration***, Revision 1. Statistical Papers Series M, No. 58, Rev. 1. New York.

U.S. Census Bureau. 2013a. "International Data Base, Population Estimates and Projections Methodology."
http://www.census.gov/population/international/data/idb/estandproj.pdf

_____. 2013b. "RUPSubAdj." Subnational Projections Toolkit. Methods Documentation.
http://www.census.gov/population/international/files/sptoolkit/RUPSubAdj.pdf

_____. 2013c. ***The Rural-Urban Projection (RUP) Program, A User's Guide***,
<http://www.census.gov/population/international/files/rup/RUPUsersGuide2013.pdf>