

Consistent Subnational Population Projection

Griffith Feeney <feeney@gfeeney.com>

Version 1.04 26-Oct-2017

Paper presented to the XXVIII International Population Conference of the International Union for the Scientific Study of Population, Cape Town, South Africa, 29 October to 3 November 2017, Session 252, Projecting Populations and Demographic Components, organized by Thomas Spoorenburg, United Nations Population Division, chaired by Mathis Lerch, Max Planck Institute for Demographic Research, Friday 3 November 08:30–10:00 AM.

ABSTRACT

A new methodology for subnational cohort component projection is presented. It produces projected subnational numbers of births and deaths consistent with corresponding national numbers, subnational numbers of net internal migrants that sum to zero, and adjusted fertility, mortality and net migration input parameters consistent with final projected subnational numbers of births, deaths, net migrants, and projected age-sex distributions. It also provides a means of assessing the accuracy of subnational projection input parameters and a method for automatically generating future trends in subnational projection parameters from national parameter trends.

1. Introduction

Population projections are perhaps the most widely demanded product of national statistical systems throughout the world. National projections are important for many purposes, but subnational projections are equally important—often to a far larger number of users.

Programs to immunize infants and young children against vaccine-preventable diseases, for example, need estimates of annual births to know how many doses of vaccines to order each year. But they also need subnational estimates to know how best to distribute these doses throughout the country. Good subnational estimates help minimize wastage and maximize coverage.

Cohort component projection for a single population closed to migration is covered in standard demographic methods texts and taught in introductory courses. A base age-sex distribution is projected forward in time on the basis of anticipated future levels, trends and age patterns of fertility and mortality. Preparing these inputs may be demanding, but calculating a projection from the inputs is straightforward.

Subnational projections are more problematic. One commonly used approach involves three steps: calculate a national projection based on anticipated national fertility and mortality parameters; calculate preliminary subnational projections based on anticipated subnational fertility, mortality and internal migration parameters; calculate final projected subnational age-sex distributions by adjusting the preliminary distributions to ensure consistency with projected national distributions.

There are three problems with this procedure. First, projected subnational numbers of births and deaths do not sum to the projected national numbers. Second, projected subnational numbers of net migrants do not sum to zero, as logically they must. Third, the adjustment of the projected subnational age-sex distributions to make them consistent with projected national age-sex distributions makes them *inconsistent* with the subnational fertility, mortality and net migration parameters with which the projection began.

This results in presentation of projected subnational age-sex distributions that are inconsistent to an unknown degree with the subnational projection parameters on which they are purportedly based.

This paper present a new subnational projection methodology that solves all three of these problems. Preliminary projected subnational numbers of deaths and births are adjusted to be consistent with corresponding national numbers at each projection cycle. Preliminary projected subnational numbers of net internal migrants are adjusted so that they sum to zero. The consistency of projected numbers of births, deaths, and net migrants implies consistency of projected age-sex distributions. Adjusted subnational fertility, mortality and net internal migration projection input parameters consistent with final projected subnational numbers of births, deaths, and net internal migrants are calculated from these final projected numbers.

A software implementation in R ([R Core Team, 2015](#)) has been developed and is available on request. The paper also discusses incorporating international migration into national and subnational projections, the distinction between top-down and bottom-up subnational projection methods, and several practical advantages of top-down methods.

2. National Cohort Component Projection

This section describes component projection for a single population closed to migration. Projected components of population change, age-sex-specific numbers of births and deaths, as well as projected age-sex distributions, are regarded as projection outputs. Inputs and outputs are organized into a standard format used in the software implementation described in [Section 7](#).

Calculation of projected numbers is standard and is described in, for example, [Preston, Heuveline, and Guillot \(2001, section 6.3\)](#). Detailed formulas are presented here nonetheless because the presentation in following sections would be unintelligible without them.

[Table 1](#) organizes inputs and outputs for single projection cycle into a single table referred to as a *projection frame*. A frame is initialized by entering the four inputs to the calculation.

1. Numbers of persons in n year age groups with a concluding open-ended group at the beginning of the projection period (PopIN)
2. Life table ${}_nL_x$ values (nL_x) for the projection period with an open-ended group n years higher than the open-ended group for the age-sex distribution
3. The sex ratio at birth (SRB) for the projection period
4. Age-specific birth rates (ASBR) for the projection period

The “Births” rows are included in the table to show total female and male births during the projection period. These births may be thought of as persons in the negative n year age group ending in age zero at the beginning of the projection period. The female “Births” row also provides a place to enter the sex ratio at birth.

The dual open-ended age group rows accommodate the different open-ended age groups for the initial age-sex distribution and the life table ${}_nL_x$ values. The total row for females provides a place to enter total births during the projection period.

The total rows are also used to enter summary statistics that are useful for reference, but not necessary for calculating projection outputs. This includes the total fertility rate

calculated from the age-specific birth rates and the female and male expectations of life at birth calculated from the ${}_nL_x$. The latter calculation is one reason for taking ${}_nL_x$ values rather than survivorship ratios as mortality parameters. The projection frame is designed to display all projection inputs and all projection outputs for a single cycle of component projection.

Table 1. Projection frame for single cycle of cohort-component projection

SexAge	PopIN	nLx	Deaths	SRB/ASBR	Births	PopOUT
BirthsF	2,143,154	5.0000	226,234	1.0500	-	-
f0-4	1,668,750	4.4722	71,003	-	-	1,916,919
f5-9	1,403,479	4.2819	21,291	-	-	1,597,747
f10-14	1,131,572	4.2170	14,597	-	-	1,382,187
f15-19	968,229	4.1626	24,259	0.1932	1,007,154	1,116,975
f20-24	812,542	4.0583	40,106	0.3048	1,338,462	943,970
f25-29	603,387	3.8579	42,633	0.2640	908,043	772,436
f30-34	605,647	3.5854	45,792	0.2040	594,865	560,754
f35-39	492,143	3.3143	34,609	0.1392	366,095	559,855
f40-44	352,303	3.0812	24,053	0.0744	150,630	457,533
f45-49	259,590	2.8708	17,812	0.0192	28,216	328,250
f50-54	193,245	2.6739	14,110	-	-	241,778
f55-59	151,872	2.4786	14,477	-	-	179,134
f60-64	119,802	2.2423	16,578	-	-	137,395
f65-69	105,697	1.9321	21,783	-	-	103,224
f70-74	66,812	1.5339	20,348	-	-	83,914
f75-79	45,158	1.0667	19,541	-	-	46,463
f80-84	23,489	.6051	15,942	-	-	25,617
f85-89	11,076	.1944	7,810	-	-	7,547
f90-94	3,224	.0573	2,872	-	-	3,266
f95+/95-99	947	.0063	861	-	-	438
f100+	-	.0006	-	-	-	-
TotalF	9,018,962	50.6927	696,712	5.9940	4,393,465	10,465,404
BirthsM	2,250,311	5.0000	242,541	-	-	-
m0-4	1,747,839	4.4611	80,006	-	-	2,007,770
m5-9	1,465,043	4.2569	25,849	-	-	1,667,833
m10-14	1,096,696	4.1818	15,232	-	-	1,439,194
m15-19	941,886	4.1237	20,208	-	-	1,081,464
m20-24	797,125	4.0352	31,629	-	-	921,677
m25-29	587,947	3.8751	36,020	-	-	765,497
m30-34	500,145	3.6377	39,227	-	-	551,928
m35-39	458,597	3.3524	42,701	-	-	460,918
m40-44	349,012	3.0403	37,188	-	-	415,896
m45-49	262,234	2.7163	29,975	-	-	311,824
m50-54	172,940	2.4058	19,172	-	-	232,259
m55-59	131,738	2.1391	17,078	-	-	153,768
m60-64	98,110	1.8618	18,222	-	-	114,660
m65-69	86,899	1.5160	23,059	-	-	79,889
m70-74	52,048	1.1137	19,458	-	-	63,841
m75-79	33,138	.6974	16,804	-	-	32,590
m80-84	14,345	.3437	10,314	-	-	16,333
m85-89	6,373	.0966	4,965	-	-	4,032
m90-94	1,362	.0213	1,289	-	-	1,408
m95+/95-99	386	.0011	351	-	-	108
m100+	-	.0001	-	-	-	-
TotalM	8,803,865	47.8773	731,288	0.0000	-	10,322,887

Note See text for explanation and formulas for calculation.

Calculation of the outputs for a single projection cycle requires four steps. In the formulas below ${}_nN_x$ denotes the number of persons age x to $x+n$ at the beginning of the projection

period, ${}_nN_x^p$ the projected number of persons age x to $x + n$ at the end of the period. The “SexAge” column in Table 1 shows conventional age-sex group labels, with ages in completed years, so that “0-4” refers to the interval beginning at exact age 0 and ending at exact age 5. Age group labels are prefixed by a letter to indicate female or male.

Step 1: Calculate deaths to the initial population during the projection period

$${}_n\text{Deaths}_x = {}_nN_x \times \left(1 - \frac{{}_nL_{x+n}}{{}_nL_x}\right), \quad x = 0, n, 2 \cdot n, \dots, z - n, \quad (1)$$

where z denotes the age at which the open-ended age group for the age-sex distributions begins, and

$${}_\infty\text{Deaths}_z = {}_\infty N_z \times \left(1 - \frac{{}_\infty L_{z+n}}{{}_\infty L_z}\right) = {}_\infty N_z \times \left(1 - \frac{{}_\infty L_{z+n}}{{}_n L_z + {}_\infty L_{z+n}}\right). \quad (2)$$

Both formulas apply to females and males. Life table ${}_nL_x$ values are for a life table with radix one (not a positive power of 10). The second formula shows why the life table open-ended age group must be n years higher than the open-ended group for the age-sex distributions. The values on the right hand side are available in the “nLx” column of the projection frame.

Deaths of persons age x to $x + n$ at the beginning of the projection period occur in this age group and in the following age group, but the subscripts indicate the age group at the beginning of the period. This convention results in the age group in formula (9) below being the same for the initial population and deaths.

Step 2: Calculate births during the projection period

$${}_n\text{Births}_x = {}_n\text{ASBR}_x \times n \times {}_n\text{PYL}_x, \quad (3)$$

where the range of x ensures coverage of the reproductive age span and ${}_n\text{PYL}_x$ denotes person years lived during the projection period by women aged x to $x + n$. This is usually approximated by

$${}_n\text{PYL}_x = n \times \left(\frac{{}_nN_x + {}_nN_x^p}{2}\right) \quad (4)$$

Total births are calculated as the sum of age-specific births over all reproductive age groups.

$$\text{Female Births} = \text{Total Births} \times \left(\frac{1}{1 + \text{SRB}}\right), \quad (5)$$

where SRB denotes the sex ratio at birth defined as male births divided by female births (no multiplication by 100), and

$$\text{Male Births} = \text{Total Births} - \text{Female Births}. \quad (6)$$

Step 3: Calculate deaths during the projection period of female and male births during the period

$$\text{Deaths}_B = \text{Births} \times \left(1 - \frac{nL_0}{n}\right). \quad (7)$$

This formula applies to female and male births.

Step 4: Calculate the projected age-sex distribution The projected number of persons for the first age group is calculated, for females and males, by subtracting deaths of births from births,

$${}_nN_0^p = \text{Births} - \text{Deaths}_B \quad (8)$$

Projected numbers of persons for older age groups up to but not including the open-ended group are

$${}_nN_{x+n}^p = {}_nN_x - {}_n\text{Deaths}_x, \quad x = 0, n, 2n, \dots, z - n. \quad (9)$$

The projected number of persons in the open-ended age group is

$${}_\infty N_z = ({}_nN_{z-n} - {}_n\text{Deaths}_{z-n}) + ({}_nN_z - {}_n\text{Deaths}_z). \quad (10)$$

The projected number of persons in the open-ended age group consists of two components,

- survivors of persons in the oldest closed age group at the beginning of the projection period, who if they survive are in the next closed age group, z to $z + n$ at the end of the period (left term on the right hand side), and
- survivors of persons age $z+$ at the beginning of the period, who if they survive are in the open-ended group $(z + 5)+$ at the end of the period (right term on the right hand side).

3. Subnational projection: Deaths and births

Let a national projection be given as a series of national projection frames. These frames, together with base age-sex distributions and mortality, fertility, and internal migration parameters for each subnational area, are the inputs required for subnational projections.

Subnational projection frames (Table 2) include the eight columns of the national frame, an NIMR column for net internal migration ratios, an NIM column for numbers of net internal migrants, and a new PopOUT column for final projected age-sex distributions. The PopOUT column of the national frame becomes the Survivors column of the subnational frames. It shows the projected age-sex distributions for the subnational area before taking account of internal migration. The PopOUT column of the subnational frames shows projected age-sex distribution for subnational areas after taking account of internal migration.

Calculation of deaths, births, and projected population for subnational projection frames involves the same four steps as the national projection, but the first three of these steps include a constant factor adjustment to ensure that projected subnational deaths and births sum to the corresponding national numbers.

Table 2. Projection Frame single cycle of cohort component projection for subnational area

SexAge	PopIN	nLx	Deaths	SRB-ASBR	Births	Survivors	NIMR	NIM	PopOUT
BirthsF	213,813	5.0000	Deaths	-	1	-	-	-	-
f0-4	153,111	4.4722	22,570	-	-	191,243	0.0000	2	191,245
f5-9	131,298	4.2819	6,515	-	-	146,596	-0.0001	-14	146,583
f10-14	113,048	4.2170	1,992	-	-	129,306	-0.0016	-203	129,103
f15-19	90,621	4.1626	1,458	0.1932	97,668	111,590	-0.0005	-57	111,533
f20-24	87,252	4.0583	2,271	0.3048	133,809	88,350	-0.0020	-179	88,171
f25-29	70,229	3.8579	4,307	0.2640	101,095	82,945	-0.0016	-134	82,812
f30-34	52,504	3.5854	4,962	0.2040	60,063	65,267	-0.0010	-65	65,202
f35-39	38,834	3.3143	3,970	0.1392	30,404	48,534	-0.0017	-83	48,451
f40-44	31,940	3.0812	2,731	0.0744	12,656	36,103	0.0015	53	36,156
f45-49	24,863	2.8708	2,181	0.0192	2,622	29,759	0.0000	1	29,761
f50-54	19,860	2.6739	1,706	-	-	23,157	0.0005	11	23,168
f55-59	16,619	2.4786	1,450	-	-	18,410	0.0007	12	18,422
f60-64	13,782	2.2423	1,584	-	-	15,035	-0.0003	-5	15,030
f65-69	11,832	1.9321	1,907	-	-	11,875	-0.0008	-10	11,865
f70-74	8,744	1.5339	2,438	-	-	9,394	-0.0009	-8	9,385
f75-79	7,456	1.0667	2,663	-	-	6,081	-0.0009	-5	6,076
f80-84	3,797	0.6051	3,226	-	-	4,230	-0.0016	-9	4,221
f85-89	2,230	0.1944	2,577	-	-	1,220	-0.0029	-5	1,214
f90-94	786	0.0573	1,572	-	-	658	0.0022	2	659
f95+/95-99	336	0.0063	700	-	-	116	-0.0020	-	116
f100+	-	0.0006	305	-	-	-	-	-	-
TotalF	879,142	50.6921	-	5.9940	438,317	1,017,875	-0.0130	-694	1,019,175
BirthsM	224,504	5.0000	72,781	-	-	-	-	-	-
m0-4	150,892	4.4611	24,197	-	-	200,307	-0.0009	-183	200,123
m5-9	129,318	4.2569	6,907	-	-	143,985	0.0004	56	144,041
m10-14	110,798	4.1818	2,282	-	-	127,036	0.0009	115	127,152
m15-19	87,052	4.1237	1,539	-	-	109,259	0.0022	246	109,505
m20-24	73,832	4.0352	1,868	-	-	85,184	0.0010	85	85,270
m25-29	63,281	3.8751	2,930	-	-	70,902	0.0030	212	71,114
m30-34	50,478	3.6377	3,877	-	-	59,404	0.0013	78	59,482
m35-39	37,855	3.3524	3,959	-	-	46,519	-0.0006	-29	46,490
m40-44	29,628	3.0403	3,525	-	-	34,330	0.0030	104	34,435
m45-49	22,506	2.7163	3,157	-	-	26,471	0.0014	36	26,507
m50-54	17,853	2.4058	2,573	-	-	19,933	0.0012	23	19,956
m55-59	14,520	2.1391	1,979	-	-	15,874	0.0036	57	15,931
m60-64	11,758	1.8618	1,882	-	-	12,638	0.0029	37	12,675
m65-69	10,140	1.5160	2,184	-	-	9,574	0.0018	17	9,591
m70-74	7,136	1.1137	2,691	-	-	7,449	-0.0013	-10	7,439
m75-79	6,250	0.6974	2,668	-	-	4,468	0.0011	5	4,473
m80-84	3,308	0.3437	3,169	-	-	3,081	0.0000	-	3,081
m85-89	1,936	0.0966	2,378	-	-	930	-0.0012	-2	928
m90-94	735	0.0213	1,508	-	-	428	-0.0031	-2	426
m95+/95-99	848	0.0011	695	-	-	117	-0.0043	-1	116
m100+	-	0.0001	771	-	-	-	-	-	-
TotalM	830,124	47.8773	-	-	-	976,416	0.0123	846	978,735

Note See text for explanation and formulas for calculation.

Step 1: For each subnational area, calculate deaths during the projection period to persons in the population at the beginning of the period Initial projected numbers are calculated as for the national population. Final numbers are calculated by multiplying the initial numbers for each age-sex group by a constant factor chosen so that the sum of the final projected numbers over all subnational areas equals the corresponding national number.

More specifically, let the projected national number for a particular age-sex group be

denoted by T , and let the preliminary projected numbers for this age group for subnational areas be denoted x_1, x_2, \dots . The final projected subnational numbers are calculated as

$$\text{Factor} \cdot x_i, \text{ where Factor} = \frac{T}{\sum_i x_i}. \quad (11)$$

This ensures that the sum of the final projected numbers for the age-sex group over all subnational areas equals the projected national number. This procedure will be referred to as **Constant Factor Adjustment** (CFA). It is used repeatedly to ensure consistency of projected subnational numbers with the corresponding national numbers.

Step 2: For each subnational area, calculate births during the projection period

Initial numbers of births to women in each reproductive age group are calculated as for the national population. Final numbers are calculated by CFA. Final numbers of total births, female births, and male births for subnational areas are calculated as for the national population.

Step 3: For each subnational area, calculate deaths during the projection period of female births and of male births during the period

Initial numbers of deaths of female births and deaths of male births are calculated as for the national population. Final numbers are calculated from the initial numbers by CFA.

Step 4 For each subnational area, calculate projected numbers of survivors of persons in the area at the beginning of the period, or born in the area during the period, by sex and age at the end of the period

The calculation is the same as the national calculation of projected numbers of persons at the end of the period. Adjustment is not necessary because the consistency of the subnational components of growth implies consistency of the projected age-sex distributions.

The adjustment factors obtained in Steps 1-3 are important subsidiary outputs of the subnational projection calculations. If mortality and fertility parameters for subnational areas accord well with the corresponding national parameters, the adjustment factors will be close to one. Factors departing substantially from one indicate a problem with national and/or subnational projection inputs that should be addressed before proceeding.

Steps 1-4 are the first four of six steps required to complete subnational projection frames for a single projection period. They calculate numbers of births, deaths, and survivors only. The last two steps deal numbers of net internal migrants and final projected age-sex distributions. These steps are described in [Section 5](#).

4. Internal migrants: Definitions

The Survivors column of [Table 2](#) shows what projected numbers of persons would be if there were no internal migration. Internal migration will lower the projected numbers in the PopOut column of the table by the number of out-migrants and increase the projected number by the number of in-migrants. The following definitions elaborate and clarify these terms ([Feeney, 1973](#)). They apply to each subnational area (SNA), projection period, and age-sex group at the *end* of the period.

- Def. 1. **Survivors** are persons who were in the SNA at the beginning the period, or were born in the SNA during the period, who survived to the end of the period

- Def. 2. *Net internal out-migrants* are persons who were in the SNA at the beginning of the period, or who were born in the SNA during the period, who survived to the end of the period and were at this time in some other SNA
- Def. 3. *Net internal in-migrants* are persons who were in some other SNA at the beginning of the period, or were born in some other SNA during the period, who survived to the end of the period and were at this time in the SNA
- Def. 4. *Net internal migrants* is the number of net internal in-migrants *minus* the number of net internal out-migrants

The terms defined by the first three definitions may refer to persons, sets of persons, or to numbers of persons in these sets. The fourth definition refers to numbers only. In Definitions 2 and 3, “net” refers to the net effect of mortality and migration during the projection period, *i.e.* to net-over-time. Persons who migrate during the period but die before the end of the period are not counted as migrants, nor is any account taken of multiple migrations during the projection period. Whether a person who survives to the end of the projection period is a net migrant depends only on where they were at the beginning and end of a projection period.

The rationale for these definitions is twofold. First, migration that does not affect projected age-sex distributions for subnational areas is irrelevant to projection. Definitions suited to projections may therefore differ from definitions for other purposes. Second, these definitions conform reasonably closely to what census data on place of residence five years ago (or place of previous residence and duration of residence) will provide.

- Def. 5. The *net internal out-migration proportion* is the number of net out-migrants from the SNA divided by survivors for the SNA
- Def. 6. The *net internal in-migration ratio* is the number of net in-migrants for the SNA divided by the survivors for the SNA
- Def. 7. The *net internal migration ratio* is the number of net internal migrants for the SNA *divided by* survivors for the SNA

These statistics may be calculated from tabulations of census data on place of residence five years ago or place of previous residence and duration of residence.

5. Subnational projection: Net internal migrants

Initial numbers of net internal migrants for each subnational area and age-sex group are calculated by multiplying the numbers of survivors in the Survivors column of the subnational projection frames by the net internal migration ratios in the NIMR column of the frames.

Final numbers are obtained by adjusting the initial numbers to ensure that summing final numbers of net internal migrants for each age-sex group over all subnational areas gives zero, as logically it must. Constant factor adjustment (CFA) does not work here—if applied, it unhelpfully forces numbers of net internal migrants for every subnational area and age-sex group to zero.

Consider a particular age-sex group, let $x_i, i = 1, 2, \dots$ denote the number of net internal migrants for this age-sex group for subnational area i , and define

$$P = \sum_{x_i > 0} x_i \quad \text{and} \quad N = - \sum_{x_i < 0} x_i. \quad (12)$$

If the sum of x_i over all subnational areas is not zero, then $P \neq N$. Suppose for specificity that $P > N$. Equality may be obtained by reducing P and increasing N . Let the values of x_i that are greater than zero be multiplied by $1 - z$, z a small number greater than zero, and the values of x_i that are less than zero be multiplied by $1 + z$. The resulting adjusted values will sum to zero if and only if

$$P(1 - z) = N(1 + z), \quad (13)$$

which implies

$$z = \frac{P - N}{P + N}. \quad (14)$$

Adjusted numbers of net internal migrants are thus calculated as

$$y_i = \begin{cases} x_i(1 - z) & \text{if } z_i > 0; \\ x_i(1 + z) & \text{if } z_i < 0. \end{cases} \quad (15)$$

The following numbered steps continue from Steps 1-4 in [Section 3](#).

Step 5: For each subnational area, calculate the number of net migrants during the projection period Calculate initial numbers by multiplying numbers of survivors by the net internal migration ratios and final numbers as

$${}_n\text{NIM}_x^i = \begin{cases} \text{Initial } {}_n\text{NIM}_x^i \cdot (1 - z) & \text{if } \text{Initial } {}_n\text{NIM}_x^i > 0; \\ \text{Initial } {}_n\text{NIM}_x^i \cdot (1 + z) & \text{if } \text{Initial } {}_n\text{NIM}_x^i < 0. \end{cases} \quad (16)$$

Step 6: For each subnational area, calculate the projected number of persons at the end of the projection period

$${}_nN_x^i = {}_n\text{Survivors}_x^i + {}_n\text{NIM}_x^i. \quad (17)$$

These numbers are shown in the PopOUT column of the subnational projection frame illustrated in [Table 2](#).

6. Adjusting input parameters

The preceding section concludes the calculation of projected numbers of deaths, births, net internal migrants and projected age-sex distribution for each subnational area. Because the preliminary numbers of deaths, births, and net migrants have been adjusted, however, the input parameters with which the projection began no longer correspond to projected numbers. This section shows how to calculate input parameters consistent with the final projected numbers of deaths, births, and net internal migrants.

From formula (7),

$${}_nL_0 = n \times \left(1 - \frac{\text{Deaths}_B}{\text{Births}} \right) = n \times P_B, \quad (18)$$

where P_B is introduced to denote the expression in parentheses in the middle term, which is the proportion of births during the projection period who survive to the end of the period.

From formula (1),

$$\frac{{}_nL_{x+n}}{{}_nL_x} = 1 - \frac{{}_n\text{Deaths}_x}{{}_nN_x} = {}_nP_x, \quad x = 0, n, 2n, \dots, z-n, \quad (19)$$

where ${}_nP_x$ is introduced to denote the middle term, which is the proportion of persons aged x to $x+n$ at the beginning of the period who survive to the end of the period, and z denotes the age at which the open ended age group for age-sex distributions begins. From this formula it follows that

$${}_nL_{x+n} = {}_nL_x \times {}_nP_x, \quad x = 0, n, 2n, \dots, z-n. \quad (20)$$

From formula (2),

$$1 - \frac{\infty\text{Deaths}_z}{\infty N_z} = \infty P_z = \frac{\infty L_z}{{}_nL_{z-n} + \infty L_z}, \quad (21)$$

∞P_z denoting the term at left, which is the proportion of persons in the open-ended age group at the beginning of the projection period who survive to the end of the period. The right equality may be written

$$a = \frac{c}{b+c} \quad \text{which implies} \quad c = b \times \frac{a}{1-a}. \quad (22)$$

Reverting to the notation of (21) gives

$$\infty L_{z+n} = {}_nL_{z-n} \times \frac{\infty P_z}{1 - \infty P_z} \quad (23)$$

Formulas (18), (20), and (23) may be used to calculate the life table ${}_nL_x$ from adjusted age-sex-specific deaths, sex-specific births during the projection period, and the age-sex distribution of the population at the beginning of the period. The calculation may be thought of as consisting of two steps: first calculate the survival proportions P_B , ${}_nP_x$, and ∞P_z , then calculate final ${}_nL_x$ from these survival proportions.

Net internal migration rates consistent with the adjusted numbers of net internal migrants are calculated as

$${}_n\text{NIMR}_x = \frac{{}_n\text{NIM}_x}{{}_n\text{Survivors}_x}, \quad (24)$$

where ${}_n\text{NIM}_x$ and ${}_n\text{Survivors}_x$ are adjusted numbers of net internal migrants.

Age-specific birth rates consistent with adjusted numbers of births by age of mother are calculated as

$${}_n\text{ASBR}_x = \frac{{}_n\text{Births}_x}{{}_n\text{PYL}_x}, \quad (25)$$

where ${}_n\text{Births}_x$ and ${}_n\text{PYL}_x$ denote adjusted values of births and person years lived.

7. Software implementation

The projection methodology described in preceding sections has been implemented in R (R Core Team, 2015). Implementation files are available from the author on request. Input parameters are read from spreadsheet files and projection results may be written to spreadsheet files, making it possible to produce projections with minimal knowledge of R. Inputs may be given by single years or five year age groups with any open-ended age group consistent with the age grouping.

The national projection frames defined in Section 2 and the subnational frames defined Section 3 are implemented as matrices with named rows and columns. A national projection is an R list of projection frames, one for each projection cycle. Producing a national projection consists of two main steps, initialization and calculation.

Initialization creates the list, enters the base age-sex distribution into the projection frame for the first cycle, and enters the projection parameters ${}_nL_x$, SRB, ASBR, and NIMR into the appropriate columns of the projection frame for each cycle.

Calculation for every cycle except the first begins by assigning the age-sex distribution in the PopOUT column of the frame for the preceding cycle to the PopIN column of the frame for the current cycle. Age-sex-specific deaths and births and the projected age-sex distribution are then calculated as described in Section 2.

A subnational projection may be thought of as a generalized matrix of projection frames with a row for each subnational area and a column for each projection cycle. It is implemented as a double list: a list with one component for each projection cycle, with each component of this list a list of projection frames for subnational areas.

A national projection is one input to subnational projections, so the first step in calculating a subnational projection is to calculate a national projection. The second step is initialization of the subnational projections. This creates the double list of projection frames, assigns base age-sex distributions for subnational areas to the frames for the first projection cycle, and assigns mortality, fertility and net migration parameters for each subnational area and each projection cycle to the appropriate column in the appropriate frame.

The third step is calculation of subnational projections as described in Section 3 and Section 5. Preliminary numbers of deaths, births, and net migrants in projection frames are overwritten by the final numbers as these become available, but the adjustment factors are written to a list of matrices, one for each projection cycle, with columns showing adjustment factors for deaths, births, and the z of Formula (14) of Section 4.

The fourth and final step is calculation of adjusted input parameters consistent with the final projected numbers of deaths, births, and net internal migrants, as described in Section 6. The original input parameters are over written, but are available in the projection input files.

Base age-sex distributions, one of the four inputs for the national and subnational projections, are implemented as a matrix with named rows and columns, one row for each

age-sex group, one column for the national population, and one subsequent column for each subnational area.

Input ${}_nL_x$ values are represented as a list of matrices with rows for age-sex groups and columns for projection cycles. The first matrix shows values for the national population, subsequent matrices values for subnational populations. The same list-of-matrices structure is used for ASBR-SRB (sex ratio at birth) inputs and for NIMR inputs.

The projection methodology applied to the specific case shown in Tables 1 and 2—five year age groups with open-ended age group 95+, three subnational regions, and three projection cycles—is implemented in a spreadsheet for comparison with the R implementation.

The R implementation may be used to produce projections for any number of subnational areas and projection cycles for single year age groups, five year age groups, or n -year age groups with any open-ended age group consistent with the age grouping. It does not at this writing provide supplementary facilities necessary for routine application with minimal knowledge of R, such as utility functions to facilitate calculation of projection inputs and summary tables of projection results.

8. International migration

Given the problematic nature of international migration data, incorporating international migration into population projections often resolves into three problems: how to determine the total number of international migrants during each projection period; how to distribute these totals by age and sex for the national population; and how to distribute the age-sex-specific national numbers to subnational areas. The methods of preceding sections may be adapted to deal with the second and third problems. The same ideas apply if total international migrants are given separately for females and males.

International migration may be incorporated into a national projection in the same way that internal migration is incorporated into subnational projection. The definitions of [Section 4](#) generalize from internal to international migrants. Age-sex-specific net international migration ratios for the national population and each subnational area are defined in the same way as net internal migration ratios. National and subnational projection frames are extended by adding columns for net international migration ratios and numbers of net international migrants.

Population census questions on household members abroad may provide information on numbers and age-sex distribution of international out-migrants in recent years. Place of residence five years ago or place of previous residence and duration of residence may provide similar information on international in-migrants. Data of this kind are likely to be problematic, but may nonetheless provide information useful for distribution total international migrants. Both points are illustrated by the 2014 census of Myanmar ([Myanmar Department of Population, 2017](#), Appendix B, Section B.6).

9. Top-down and bottom-up methods

The subnational projection methodology described in this paper is “top-down”. It begins with a national projection and requires that subnational projections, however specifically calculated, add up to the national projection. Subnational projections may also be made “bottom-up”. Subnational projections are calculated first, then the national projected is calculated by adding up subnational projected numbers of births, deaths, and projected age-sex distributions. National projection parameters consistent with the national projections may be calculated as described in [Section 6](#).

Bottom-up methods are easier to understand and easier to calculate than top-down methods, and they ensure by construction that projected subnational numbers are consistent with projected national numbers, but top-down methods have several important practical advantages.

- They are expected to be more robust against errors in subnational projection parameters
- The adjustment factors provide a valuable tool for analyzing the accuracy of national and subnational projection inputs
- The adjustment procedure may be used to automatically generate subnational fertility, mortality and migration trends from national trends
- Top-down methods make it possible to produce subnational progressions in stages for progressively smaller subnational areas

9.1 Robustness When data quality is problematic, the accuracy of subnational projection parameters is likely to be substantially lower than the accuracy of national parameters. Forcing subnational projections to be consistent with a national projection is a way of “disciplining” the subnational inputs (Brass, 1971).

There are several reasons for expecting the accuracy of subnational inputs to be lower than the accuracy of national inputs. In the absence of a well developed civil registration and vital statistics system, fertility and mortality parameters are usually obtained from population surveys or indirectly estimated from population census data (Moultrie et al., 2013). Survey estimates are problematic for subnational areas because sample size tends to high standard errors. Indirect estimation techniques typically assume a population closed to migration, sometimes only implicitly, which tends to make subnational estimates less accurate than national estimates.

Internal migration inputs are another matter. Population censuses are the main source, usually tables derived from a question on residence five years prior to the census or questions on previous residence and duration of residence. There seems to have been little systematic study of the accuracy of this information, but the example of the 1960 United States census question on place of 1955 residence is instructive.

A published table of persons age five and over by region of residence at the census and region of 1955 residence (US Census Bureau, 1963, Table 2, pages 4-7) suggests numbers of persons not responding to the place of 1955 residence question that are modest—2-3 percent or less—in relation to total numbers of persons, but very large in relation to numbers of persons reporting a different region of residence in 1955.

The level of non-response is unfortunately clouded by the grouping persons who did not respond to the 1955 residence question with persons who reported themselves abroad in 1955, but the numbers in this category are so large relative to the numbers in the interregional migration flows that they cast doubt on the accuracy of these flows. For every one of the nine regions, the numbers abroad in 1955 or not reporting 1955 residence is larger than the largest out-migration stream from the region.

9.2 Analyzing the accuracy of projection inputs The importance of the adjustment factors that ensure consistency between subnational and national projections was mentioned in passing at the end of Section 3. If the accuracy of the national and subnational base age-sex distributions and fertility, mortality and migration parameters is high, the projected subnational age-sex distributions and numbers of births, deaths and net mi-

grants should be nearly consistent with the projected national numbers. If this is true, the adjustment factors will be close to 1.

The adjustment factors are therefore a tool for assessing the accuracy of the national and subnational projection inputs. Adjustment factors close to 1 provide useful if not definitive evidence accurate inputs. Adjustment factors departing substantially from 1 provide definitive evidence of errors in projection inputs. Larger departures from 1 evidently indicate lower accuracy, smaller departures higher accuracy. Going beyond these generalities requires knowledge and experience that can come only from application of the new method to producing subnational projections for different countries.

9.3 Autogeneration of subnational parameter trends Working out anticipated future trends in national and subnational projection parameters necessarily begins with estimates of fertility, mortality and migration parameters for the recent past, which will be extrapolated by some means to the first projection period. Assume that these subnational estimates are available and that a national projection has been produced.

Subnational fertility and mortality trends may be produced by inputting the estimated base projection period parameters for all future projection periods. The adjustment procedure will produce projected subnational numbers of births, deaths and migrants consistent with the national numbers. The procedure described in Section 6 may then be used to calculate parameter trends for each subnational area that are consistent with the national projection.

The resulting adjustment factors will not necessarily be close to 1 except for the first projection period. Adjustment factors for future periods will reflect changes in projected national numbers of births, deaths and net migrants determined by the national base age-sex distribution and projection parameters.

9.4 Producing subnational projections in stages The main labor of producing population projections is the preparatory work of adjusting the base age-sex distributions, estimating current fertility, mortality and migration parameters, and projecting these parameters into the future. Preparatory work for subnational areas tends to be more difficult than for the country as a whole, so work for n subnational areas is likely to require more than n times as much time and effort as doing it for the national population. Top-down methods make it possible to provide users with a national projection relatively quickly, followed by subnational projections when subnational projection inputs have been prepared.

More generally, the method described in [Sections 2-7](#) may be used to produce projections for 2nd level administrative areas consistent with previously produced projections for 1st level administrative areas, and similarly for lower level areas. The only modification required is to choose a number other than zero to divide subnational numbers of net migrants for each age-sex group in [Formula 12](#) of [Section 5](#). Zero net migration is a logical necessity only for the national population.

10. Conclusion

Subnational population projections are among the most widely demanded products of national statistical systems throughout the world, but methodology for producing them has been neglected. It is not covered in standard demographic methods texts and receives scant attention in the research literature. Multiregional projection ([Rogers, 2015](#)) is an obvious and important partial exception to the general rule, but its applicability is limited.

Why this methodological inattention to so important a subject? One explanation may be the notion that cohort component projection is an established method that was developed long ago, has been thoroughly standardized, and does not require methodological attention. But this is true only of its application to national populations closed to migration. It is emphatically *not* true of cohort component methods for subnational projections or for national populations open to international migration.

It is not uncommon, for example, for subnational projections to be made by adjusting preliminary projected subnational age-sex distributions directly to ensure that they sum to given projected national age-sex distribution. This results in projected subnational numbers of births and deaths that do not sum to projected national numbers, and in projected subnational numbers of net migrants do not sum to zero, as logically they must. It also results in subnational projection results do not satisfy the demographic equation, invalidating subnational analyses of the components of population change.

Adjusting projected subnational age-sex distributions directly also means that they are no longer consistent with the subnational fertility, mortality and internal migration parameters with which the projection began. This results in presentation of projected subnational age-sex distributions that are inconsistent to an unknown degree with the subnational projection parameters on which they are purportedly based.

The subnational cohort component projection method presented in this paper solves these three problems. Projected subnational age-sex-specific numbers births and deaths are adjusted to be consistent with national numbers at each projection cycle. Projected subnational numbers of net internal migrants are adjusted at each projection cycle to ensure that they sum to zero. The consistency of the subnational components of growth ensures the consistency of the projected subnational age-sex distributions. Finally, the method includes formulas for calculating subnational mortality, fertility and net internal migration projection input parameters consistent with final projected subnational numbers of deaths, births, net internal migrants, and age-sex distributions.

The new method goes beyond producing subnational projections fully consistent with a given national projection. The adjustment factors that ensure the consistency of projected subnational and national numbers provide a tool for analyzing the accuracy of national and subnational projection inputs. The adjustment procedure provides a method for automatically generating subnational projection parameters trends from national trends.

Given the computation-intensive nature of subnational projection, the practical utility of any method depends on the availability of implementation software. The R implementation described in [Section 7](#) is available on request from the author.

Acknowledgements

Thanks to Ricardo Neupert for continuing email discussions and penetrating comments on several drafts of this paper, and to Samuel Feeney for indispensable suggestions on the R implementation. Development of the methodology presented in this paper began with work on a population projection report produced by the Myanmar Department of Population, Ministry of Labour, Immigration and Population, supported by UNFPA, following the 2014 census of Myanmar. It was facilitated by training workshop interaction with members of the projection report team, Khaing Khaing Soe, Yin Yin Kyaing, Khin Myo Khine, Hlaing Phwe Thu, May Myint Bo, Mar Lar Htun, Tin Tin Lay, and Wai Wai Hlaing Zin, and by comments by Thomas Buettner, Andreas Demmke, Werner Haug, Fred Okwayo, and Ian White on a draft report.

References

- Brass, William. 1971. Disciplining Demographic Data. *International Population Conference, London, 1969*, Volume 1, pages 183-204. Liège: International Union for the Scientific Study of Population.
- Feeney, Griffith. 1973. Two models for multiregional population dynamics. *Environment and Planning* **5**:31-43.
- Moultrie, Tom, Rob Dorrington, Allan Hill, Kenneth Hill, Ian Timus and Basia Zaba. 2013. *Tools for Demographic Estimation*. Paris: International Union for the Scientific Study of Population. Available at demographicestimation.iussp.org/content/get-pdf-book-website, accessed 6 September 2015.
- Myanmar Department of Population. 2017. *Thematic Report on Population Projections for the Union of Myanmar, States/Regions, Rural and Urban Areas, 2014 - 2050*. Census Report Volume 4-F. Nay Pyi Taw: Ministry of Labour, Immigration and Population. Available at myanmar.unfpa.org/en/node/15104, accessed 18 September 2017.
- Preston, Samuel H., Patrick Heuveline, and Michel Guillot. 2001. *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers Inc.
- R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at www.r-project.org/, accessed 7 February 2016.
- Rogers, Andrei. 2015. *Applied Multiregional Demography: Migration and Population Redistribution*. London: Springer International Publishing.
- United States Census Bureau. 1963. *Lifetime and Recent Migration: State of Residence in 1960 by Geographic Division of Birth and Residence in 1955*. United States Census of Population: 1960. Final Report PC(2)-2D. Subject Reports. Washington, DC: US Government Printing Office. Available at www.census.gov/prod/www/decennial.html (Census of Population and Housing, 1960 → 1960 Census of Population → Vol. II. Subject Reports → 2D - 2E), accessed 18 September 2017.