

DHS WORKING PAPERS

A Decomposition of Sources of Change in Population Size and Median Age, 1970–2020

Thomas W. Pullum Apoorva Jadhav

2021 No. 178

DEMOGRAPHIC AND HEALTH SURVEYS

August 2021

This document was produced for review by the United States Agency for International Development.

DHS Working Papers No. 178

A Decomposition of Sources of Change in Population Size and Median Age, 1970–2020

Thomas W. Pullum¹ Apoorva Jadhav²

> ICF Rockville, MD

August 2021

¹ Demographic and Health Surveys, ICF

² United States Agency for International Development

Corresponding author: Thomas W. Pullum, The DHS Program, ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA; phone: +1 301-407-6500; fax: +1 301-407-6501; email: Tom.Pullum@icf.com

Acknowledgments: The authors thank Richard Cincotta for a helpful review of an earlier draft.

Editor: Diane Stoy Document Production: Natalie Shattuck

This study was conducted with support from the United States Agency for International Development (USAID) through The DHS Program (#720-OAA-18C-00083). The views expressed are those of the authors and do not necessarily reflect the views of USAID or the United States Government.

The DHS Program assists countries worldwide in the collection and use of data to monitor and evaluate population, health, and nutrition programs. Additional information about The DHS Program can be obtained from ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850 USA; telephone: +1 301-572-0200, fax: +1 301-572-0999, email: info@DHSprogram.com, Internet: www.DHSprogram.com.

Recommended citation:

Pullum, Thomas W., and Apoorva Jadhav. 2021. *A Decomposition of Sources of Change in Population Size and Median Age*, 1970–2020. DHS Working Papers No. 178. Rockville, Maryland, USA: ICF.

CONTENTS

FIGURI ABSTR	ES RACT		v vii
1	INTRO	DUCTION	1
2	DATA .		3
3	METHO	ODS	5
4	RESUL	_TS	7
	4.1	Format for Presentation of Results	7
	4.2	Components of Change: The Global Pattern	8
	4.3	Components of Change: SDG Regions of the World	
	4.4	Components of Change: SDG Subregions	20
5	INTER	NATIONAL COMPARISONS	33
6	DISCU	SSION	
7	CONCI	LUSIONS	39
REFER	ENCES)	41
APPENDIX 1		DETAILED DESCRIPTION OF THE DECOMPOSITION	
	A1.1	A Model for Change in Population Size and Median Age	
	A1.2	Cohort Survival and Migration	
	A1.3	Components of Change between 1970 and 2020	
APPENDIX 2		COMPONENTS OF CHANGE FOR SELECTED COUNTRIES	

FIGURES

Change in population and median age: World 1970–2020	9
Change in population and median age: Sub-Saharan Africa 1970–2020	12
Change in population and median age: Northern Africa and Western Asia 1970–2020	13
Change in population and median age: Central and Southern Asia 1970–2020	14
Change in population and median age: Eastern and South-Eastern Asia 1970–2020	15
Change in population and median age: Latin America and the Caribbean 1970–2020	16
Change in population and median age: Australia and New Zealand 1970–2020	17
Change in population and median age: Oceania 1970–2020	18
Change in population and median age: Europe and Northern America 1970–2020	19
Change in population and median age: Eastern Africa 1970–2020	21
Change in population and median age: Middle Africa 1970–2020	22
Change in population and median age: Southern Africa 1970–2020	23
Change in population and median age: Western Africa 1970–2020	24
Change in population and median age: Northern Africa 1970–2020	25
Change in population and median age: Western Asia 1970–2020	26
Change in population and median age: Central Asia 1970–2020	27
Change in population and median age: Southern Asia 1970–2020	28
Change in population and median age: Eastern Asia 1970–2020	29
Change in population and median age: South-Eastern Asia 1970–2020	30
Change in population and median age: Europe 1970–2020	31
Change in population and median age: Northern America 1970–2020	32
Effect of each component on change in population, world regions 1970–2020	33
Effect of each component on change in median age, world regions 1970–2020	34
Change in population and median age, 15 world regions 1970–2020	34
Correspondence between median age and mean age, all countries and regions in 1970 and 2020	38
Change in population and median age: Afghanistan 1970–2020	50
Change in population and median age: Bangladesh 1970–2020	51
Change in population and median age: Benin 1970–2020	52
Change in population and median age: Burkina Faso 1970–2020	53
Change in population and median age: Cote d'Ivoire 1970–2020	54
Change in population and median age: Democratic Republic of the Congo 1970–2020	55
Change in population and median age: Ethiopia 1970–2020	56
Change in population and median age: Ghana 1970–2020	57

Change in population and median age: Guinea 1970–2020	
Change in population and median age: Haiti 1970–2020	59
Change in population and median age: India 1970–2020	60
Change in population and median age: Kenya 1970–2020	61
Change in population and median age: Liberia 1970–2020	62
Change in population and median age: Madagascar 1970–2020	63
Change in population and median age: Malawi 1970–2020	64
Change in population and median age: Mali 1970–2020	65
Change in population and median age: Mauritania 1970–2020	66
Change in population and median age: Mozambique 1970–2020	67
Change in population and median age: Nepal 1970–2020	68
Change in population and median age: Niger 1970–2020	69
Change in population and median age: Nigeria 1970–2020	70
Change in population and median age: Pakistan 1970–2020	71
Change in population and median age: Philippines 1970–2020	72
Change in population and median age: Rwanda 1970–2020	73
Change in population and median age: Senegal 1970–2020	74
Change in population and median age: South Sudan 1970–2020	75
Change in population and median age: Tanzania 1970–2020	76
Change in population and median age: Togo 1970–2020	77
Change in population and median age: Uganda 1970–2020	78
Change in population and median age: Yemen 1970–2020	79
Change in population and median age: Zambia 1970–2020	

ABSTRACT

The median age of a population is an easily interpreted summary of an age distribution that has been used as an analytical tool in research on the structure of the labor force, environmental changes, and political movements. Changes in the median age are determined by fertility, mortality, and migration, the same factors that are responsible for population growth. This paper develops a relatively simple unified model to account for changes in population size and median age. We apply this model to publicly available data files, produced by the UN's Population Division as part of World Population Prospects 2019, for the global population, and major SDG regions and subregions. The goal is to determine the relative importance of fertility, mortality, and migration for changes in population size and median age and median age during the half-century from 1970 to 2020. Change is defined as the difference between the observed values in 2020 and what would have been expected in 2020 if fertility and mortality rates in 1965-70 had not changed after 1970, and if there had been no international migration. We find that in most regions, population size was much less and the median age much older in 2020 than would have been expected. Major declines in fertility were by far the most important source of change, with relatively minor countervailing effects from improvements in mortality.

1 INTRODUCTION

It is well known that the changes in fertility, mortality, and migration rates, which drive population growth, are also sources of change in the shape of the age distribution (Preston and Stokes 2012). In particular, the median age¹ can increase or decrease, depending on the contributions of changes in fertility and mortality in different – and potentially opposite – ways. This eventual change in median age has implications for countries as they allocate resources for targeted investments in their populations as they age. Research has shown that myriad outcomes of a country, from economic growth (Bloom and Canning 2008) to political stability (Weber 2011), are linked to age structure, and are bolstered by expanding access to family planning (Starbird, Norton, and Marcus 2016). In this paper, we attempt to articulate these demographic dynamics in a systematic and interpretable manner. We describe population growth and change in the median age as joint consequences of changes in fertility, mortality, and migration. The approach is applied to world regions during the interval 1970-2020, as well as to specific countries (in Appendix 2). Findings from these analyses will help policymakers make the case for increased investment in health and human capital (Lutz 2019) and other development outcomes.

¹ An operational definition of the median age is provided below.

2 DATA

The analysis is applied to data at three levels: the entire world; the classification of countries into 8 major regions under the framework of the Sustainable Development Goals (SDG); and 12 selected subregions. Appendix 2 provides results for the priority countries of USAID's Office of Population and Reproductive Health. The data come from the United Nations (UN) Population Division's 2019 round of World Population Prospects (United Nations 2019). The UN data include all countries for the interval 1950-2020, in 5-year intervals of age and time. The data also include various alternative aggregations of countries or regions. This paper focuses on the 50-year interval 1970-2020. The 8 regions and 12 subregions are²:

Sub-Saharan Africa

- Eastern Africa
- Middle Africa
- Southern Africa
- Western Africa

Northern Africa and Western Asia

- Northern Africa
- Western Asia

Central and Southern Asia

- Central Asia
- Southern Asia

Eastern and South-Eastern Asia

- Eastern Asia
- South-Eastern Asia

Latin America and the Caribbean

Australia and New Zealand

Oceania

Europe and Northern America

- Europe
- Northern America

A synthesis of the observed patterns is based on the 12 subregions, plus the 3 regions (Latin America and the Caribbean, Australia and New Zealand, and Oceania) that are not subdivided into subregions. Together, this combination of 15 regions and subregions represents the entire world.

² SDG regions and subregions may differ from other similar classifications. For example, "Northern America" includes Canada and the United States and should not be confused with "North America," which conventionally includes Mexico.

Household surveys, including the Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS), were a major source of demographic data for the interval 1970-2020 for low- and middle-income countries (LMICs). The UN estimates also incorporate data from censuses and vital statistics systems. For each country, subregion, and region, the UN uses demographic models to ensure internal consistency.³

The following spreadsheets are used for each region:

- *Population by age.* The population (P) age, such as 0-4, 5-9, in a specific year such as 1950.
- *Median age*. The median age (m) in a specific year such as 1950.⁴
- Abridged life table. Survival ratios from one 5-year age interval to the next, 5 years later.
- Total fertility. The total fertility rate (TFR) in a specific time interval such as 1950-1955.
- *Births*. The number of births in the specified time interval such as 1950-1955.

The files describe males and females combined. The methods do not require disaggregation by sex. Estimates for a calendar year refer to July 1, the midpoint of the calendar year. The original Excel files were converted to Stata format and merged. All programming was done with Stata 16.

³ The numbers in the worksheets will be described as "observed" values, rather than estimates or projections, to avoid confusion with estimates from the model developed in this paper. However, some of the UN numbers, particularly for 2020, should be regarded as estimates and are subject to modification in future rounds of WPP. ⁴ We recalculate the median age within the procedure; the UN spreadsheet is only used for confirmation.

3 METHODS

The median age is sometimes represented by a whole number – for example, age 23. In this paper, the median is an "exact" age, estimated as the 50^{th} percentile of the continuous cumulative age distribution. The median is typically calculated by linear interpolation within the 5-year age interval that includes the 50^{th} percentile. For example, if the 50^{th} percentile is in the age interval 20-24, and is a fraction 0.69 into that interval, then the median is calculated as 20+0.69*5=23.45, which is interpreted to be 45% of the way between the 23^{rd} and 24^{th} birthdays. This number is equivalent to age 23 at last birthday but is more precise.

The technical approach used to develop the estimates is presented in detail in Appendix 1. Here we briefly introduce the analytical strategy with a sequence of statements about population dynamics, which provide a conceptual basis for the procedure.

- 1. Apart from migration (emigration and immigration), the size and age distribution of a population at any point in time are determined by the history of fertility rates and mortality rates.
- 2. Change in the size and age distribution of a population, between two time points such as 1970 and 2020, are determined by (a) the initial size and age distribution; (b) the initial or baseline fertility and mortality rates; and (c) changes in the rates during the interval. Even if the rates did not change from their baseline values, we would generally expect changes in size and the age distribution.⁵
- 3. The impact of any changes in fertility and mortality rates between two time points can be assessed in terms of how the *observed* population size and age distribution at the second time point deviate from *what would have been expected* with a continuation of the initial rates.
- 4. The determinants of population growth and the median age are identical—fertility, mortality, and migration—except for the level of detail. The median age will rise if there is a larger increase in the upper half of the age distribution than in the lower half. It is possible to construct a unified model for changes in population size and the median age.
- 5. International migration is also a determinant of population size and the age distribution, although for most countries it is secondary to fertility and mortality, and its pattern over time is less systematic. The effect of migration can be assessed separately from fertility and mortality.

Six types of rates are defined and calculated for each region or country and each 5-year interval from time t to time t+5:

TFR: The total fertility rate, a synthetic measure that can be interpreted as the mean number of children that would be born to a woman who survived from exact age 15 to exact age 50, and experienced the age-specific rates observed between time points t and t+5;

dch: Per 1,000 children born in the interval, the number who die before t+5;

dy: Per 1,000 people below the median age at time t, minus 5,⁶ the number who die before t+5;

⁵ With constant rates, growth would become exponential and the age distribution would eventually stabilize.

⁶ "The median age at time t, minus 5": m(t)-5, where m(t) is the median age at time t.

do: Per 1,000 people above the median age at time t, minus 5, the number who die before t+5;

ry: Per 1,000 people below the median age at time t, minus 5, the net number who emigrate before t+5;

ro: Per 1,000 people above the median age at time t, minus 5, the number who die before t+5.

Persons below the median age at time t, minus 5, are described as "the younger cohort." People above that age are "the older cohort."

More precise definitions of these six components of change are provided in Appendix 1. The labels for death rates begin with "d" and for migration rates begin with "r". The suffixes are "ch" for children under age 5, "y" for the younger cohort, and "o" for the older cohort. The labels for migration rates begin with "r" to emphasize that these two rates are based on residuals, which we interpret as measures of migration. The input data do not include direct measurements of migration.

The TFR and dch come from the WPP2019 spreadsheets. The other four rates are calculated from the observed data. All six rates are used to simulate hypothetical scenarios extending from 1970 to 2020. The principal scenarios can be described as:

S0: the observed trajectory of population size and median age (not a simulation);

S1: a simulation of what would have been observed if the 1965-70 levels of fertility and mortality had continued unchanged to 2020, and there had been no net migration.

The objective is to partition the difference S1-S0 into additive terms for the six components. A residual term is required because of potential internal inconsistencies and non-additivity. Empirically, the residual is small. The partitioning is tracked across the entire 50-year interval, with primary interest in the cumulative results in 2020.

4 RESULTS

4.1 Format for Presentation of Results

For each region, the results are presented in a standard set of six figures, numbered 1-6 and combined on a single page. Each figure contains, in turn, two to six subfigures. We first provide some general guidance for interpreting the figures.

Figure 1. Observed trajectories of population and median age and hypothetical alternatives. A subfigure for population is on the left and one for median age is on the right. Each subfigure shows the observed trajectory from 1970 to 2020 (scenario S0) with a line that is black and bolded. To facilitate comparisons, the population is given per 1,000 persons in 1970. Otherwise, the vertical axes for population and median age vary from one population to another. In general, although there are exceptions, both population and median age increased from 1970 to 2020.

Red lines in the subfigures show the hypothetical trajectories implied by a continuation of the fertility and mortality rates for 1965-70, and the absence of any net migration (scenario S1). Over time, as the implied age distribution becomes stable, the red line for population becomes exponential and the red line for median age flattens. In some populations, the red line for median age dips below the 1970 value before levelling out.

Blue lines show intermediate hypothetical trajectories implied if the observed fertility and mortality rates had been in effect and there was no migration. Thus, the difference between the blue and red lines is due entirely to change in fertility and mortality rates, and the difference between the blue and black lines is due entirely to net migration. The goal is to account for the difference between the black (observed) line and the red line, and to allocate the change to the separate components for fertility, mortality, and migration.

Figure 2. *Changes in the six components for 5-year time intervals between 1970 and 2020.* The upper row of subfigures tracks changes in TFR, dch, and dy, and the lower row tracks changes in do, ry, and ro. The vertical scales for the subfigures differ, depending on the range of values. Child mortality (dch) is expressed as deaths per 1,000 births in the interval. The dy, do, ry, and ro rates are expressed per 1,000 persons in the younger or older cohorts at the beginning of the interval. The age ranges for those rates thus change gradually if the median age changes. If the median age increases, for example, the "younger" age range expands and the "older" age range contracts.

Generally, the observed changes are declines in fertility and in the three mortality rates. Migration (inferred from residuals) is modeled as a decrement: a positive migration rate signifies net out-migration, with effects that are analogous to mortality. Changes in the rates tend to induce the following changes in population and median age:

If the TFR declines, the effect is to reduce population and increase the median age.

If child mortality (dch) declines, the effect is to increase population and reduce the median age.

If younger mortality (dy) declines, the effect is to increase population and reduce the median age.

If older mortality (do) declines, the effect is to increase population and increase the median age.

If younger net migration (ry) is predominantly negative (in-migration), the effect is to increase population and reduce the median age.

If older net migration (ro) is predominantly negative (in-migration), the effect is to increase population and increase the median age.

Fertility is the main engine of population growth. Therefore, TFR decline is usually the main restraint on population growth and the main source of increase in median age. Mortality decline, at any age, acts to increase population. Mortality decline and in-migration below the median age will reduce the median age, while mortality decline and in-migration above the median age will increase the median age. These relationships will reverse if fertility and mortality move in the opposite direction, or if net migration is in the opposite direction.

Figure 3. *Effects on population size of changes in the components*. This figure shows the effects on population size of the six components that were tracked in Figure 2, in the same sequence, and over the 50-year interval. For a specific population, all six subfigures are scaled to have the same vertical axis, determined by the observed range of positive and negative effects.

Figure 4. *Effects on median age of changes in the components*. This figure is similar to Figure 3 but shows the effects on the median age rather than population size.

Figure 5. *Cumulative effects on population size of changes in the components*. This figure shows the final cumulative effects, in 2020, of the six components. A horizontal bar for each component shows the extent to which that component, by itself, acted to change population size. For example, the bar for the TFR is generally negative, because a decline in the TFR caused the population to increase *by less than it would have increased if the TFR in 1965-70 had continued until 2020*. The amount of the effect is per 1,000 population in 1970. Mortality declines will result in positive bars, because lower mortality results in a larger population. The figure includes a bar labeled "sum of effects" that represents the net effect of the six components. Beneath that bar, another bar shows the difference that was to be explained, S1-S0 (the difference between the red and black bars in Figure 1, in 2020). Finally, the bottom bar in the figure shows the residual, which is the difference between the two preceding bars. The residual is generally small.

Figure 6. *Cumulative effects on median age of changes in the components*. The bars in this figure are analogous to those in Figure 5 but refer to the simulated cumulative effects on median age, in 2020, from the six components. As with population size, the main driver of change has been decline in fertility, which is partially offset by declines in mortality below the median age. The residual is generally small.

4.2 Components of Change: The Global Pattern

The global or world pattern is shown below.



The description of the global trajectories of population and median age begins with Figure 1, shown on the upper left. Within this figure, the subfigure on the left shows the observed population trajectory with a heavy black line. The vertical axis is keyed to a standard baseline population of 1,000 in 1970. The black line is very close to straight, with an approximate doubling of the initial population by 2020. The red line in the same subfigure shows that if the fertility and mortality rates observed in 1965-70 had continued throughout the interval, the trajectory would have become approximately exponential, and population size would have more than tripled. At the global level, and within most regions of the world, the observed trajectory is well below the projected trajectory. The objective is to understand the role of the components in bringing the red line down to the black line.

The right subfigure in Figure 1 describes the corresponding trajectories for median age. The black line shows a substantial and steady increase in observed median age. The red line shows that if the 1965-70 rates had continued, the median age would have declined. Both subfigures include a blue line, although it is only visible for the first few years in the subfigure for median age. In general, the blue line shows the trajectory that would have been followed if the observed changes in fertility and mortality had happened, but with no migration. Because there is no actual migration at the global level, the appearance of any migration at all should be interpreted as error.

Figure 2 shows that the TFR and the three mortality rates in the model all declined from their 1965-70 levels. The vertical scales are different for each rate, but all the declines were substantial. Reductions in the mortality of the older cohort were concentrated in the first 20 years of the time interval. The migration rates, which fluctuate in a range of +/-1 person per 1,000, simply reflect measurement error. In the remaining discussion we omit any reference to migration at the global level.

Figures 3 and 4 show the effects of the changes in the six components. Fertility decline is very conspicuous as a force for reducing population growth and increasing the median age. For population, the three types of mortality decline had an opposing effect that was small, separately or in combination. For median age, the mortality decline below the median age had a small opposing effect, and mortality decline above the median age had a small supplementary effect, relative to fertility.

The horizontal bars in Figures 5 and 6 quantify the effects seen in Figures 3 and 4 at the end of the 50-year interval. Figure 5 shows that the decline in fertility acted to reduce the 2020 population by about 1,700, compared with a hypothetical global population in which the TFR remained at the 1965-70 level. The reduction of 1,700 can be interpreted as 1.7 times the global population in 1970. This effect was partly offset by declines in mortality, which added about 700, or 0.7 times the global population in 1970. A little more than half of the effect of mortality on population size was due to improvements in child survival. Only about a quarter was due to improvements in survival above the median age. The decomposition has a residual of 146, or 0.146 of the global population in 1970.

Figure 6 summarizes change in median age and partitions the difference between the global observed median age in 2020 (S0) and what would have been expected with no change in fertility and mortality rates between 1970 and 2020 (S1). The difference to be explained is 11.91 years. Two effects acted to increase the median age: the decline in fertility, which in itself would have induced an increase of 13.36 years, and improvements in survival above the median age, which would have induced an increase of 0.63 of a year. These effects were partially offset by reductions in child mortality (-1.74) and other mortality below the

median age (-0.53). The effects of changes in younger mortality (mortality below the median but not below 5) and older mortality (above the median) almost perfectly balance each other. The decline in fertility and the decline in under-5 mortality, which offset each other, are sufficient to explain the global movement in median age. The residual, due to the imperfect fit of the model, is 0.14 of a year, which is about 1% of the change that is to be explained.

For both population size and median age, the two main drivers are the decline in fertility and the decline in under-5 mortality. The model is additive, and the balance of these two effects is best described by their sum, although the ratio also gives a sense of their relative importance. For population, the fertility effect is approximately -4.5 times the under-5 mortality effect. For median age, the ratio is approximately -7.7.

There is considerable variation across regions of the world. We now describe the patterns in those regions, with the same set of figures.

4.3 Components of Change: SDG Regions of the World

The next eight pages provide similar graphical detail for the eight SDG regions: Sub-Saharan Africa; Northern Africa and Western Asia; Central and Southern Asia; Eastern and South-Eastern Asia; Latin America and the Caribbean; Australia and New Zealand; Oceania; Europe and Northern America.

















4.4 Components of Change: SDG Subregions

The next 12 pages show similar decompositions for 12 subregions: Eastern Africa, Middle Africa, Southern Africa, Western Africa (subregions of Sub-Saharan Africa); Northern Africa, Western Asia (subregions of Northern Africa and Western Asia); Central Asia, Southern Asia (subregions of Central and Southern Asia); Eastern Asia, South-Eastern Asia (subregions of Eastern and South-Eastern Asia); and Europe, Northern America (subregions of Europe and Northern America).
























5 INTERNATIONAL COMPARISONS

We now provide a synthesis of the regional and subregional patterns. The global pattern has been described. Here the interest is in variations within the global pattern. This synthesis includes 15 geographical units, which collectively represent the entire world. We include 3 SDG regions and 12 subregions. The 3 regions, which are not subdivided, are Latin America and the Caribbean; Australia and New Zealand; and Oceania. The 12 subregions are Eastern Africa; Middle Africa; Southern Africa; Western Africa; Northern Africa; Western Asia; Central Asia; Southern Asia; Eastern Asia; South-Eastern Asia; Europe; and Northern America. All are given equal weight to ensure that the results are not dominated by Southern Asia and Eastern Asia, which would be the case with weights proportional to population size. The results are presented with two boxplots, one for the effects of the components on population size and the other on median age. A scatterplot summarizes the net change in population and median age.

The first figure shows variation across regions in the effects on population size. There are six boxes, one for each type of effect. The horizontal bar within each box describes the median value of the effect. Decline in fertility has been the principal driver of population decline, which slows the amount of population increase. The largest of the offsetting effects of mortality decline has come from declines in under-5 mortality. Declines in mortality after age 5 have had little effect on population size. Migration effects are detectable in different regions, but in the aggregate, these effects have cancelled out. The median pattern is completely consistent with the global estimates reviewed in Section 4.2.



A second figure shows variation across regions in the effects on median age. Again, fertility decline was the main driver of increases in median age, and was supplemented to a very small extent by declines in mortality in the older cohort. The counterbalancing effect of declines in under-5 mortality and other mortality below the median age was very small. The imbalance between the fertility and mortality effects is even greater for change in median age than for the change in population. The effects of migration are small and effectively cancel out across the regions.



A scatterplot shows the changes in both population size and median age in the 15 regions / subregions. To calculate a regression line and correlation coefficient, the 15 points were given equal weight. The changes are strongly negatively correlated (r=-0.85). These so-called changes are differences between what was observed (or estimated by the UN Population Division) in 2020 and what would have been expected in 2020 with no change in the rates after 1970 and no migration.



The dots for Middle Africa and Eastern Africa are very close to a horizonal red line that represents no change in the median age. They are the only subregions with virtually no difference between the observed median age and what would have been expected. They experienced population increase, beyond what would have been expected, even as the median age remained steady. They are the only regions with larger populations in 2020 than would have been projected in 1970.

The dot for Australia and New Zealand and the dot for Northern America (Canada and the United States) are very close to a vertical red line that represents no change in population size. Their 2020 populations

were almost exactly what would have been expected or projected in 1970, assuming 1965-70 rates and no immigration. Fertility and mortality rates declined, but net in-migration almost completely offset the combined effects of changes in fertility and mortality. On average, immigrants were older than the median age (more so in Australia and New Zealand than in Northern America), and this led to increases in median age.

The largest and nearly identical population reductions, in a comparison of observed and expected population sizes (per 1,000 persons in 1970), occurred in three areas: Eastern Asia, South-Eastern Asia, and Latin America and the Caribbean. The magnitude of the difference was between two and three times the size of the 1970 population. These three areas also had the largest change or difference in median age, which was more than 12 years. Eastern Asia had by far the greatest increase in median age, which was the result of massive declines in fertility.

The best-fitting line through the scatterplot shows that the relationship between changes in population size and median age is very strong and nearly linear. We emphasize that the correspondence describes correlation rather than causation. Both changes have been shown to result from six components: fertility, mortality (in three age intervals, two of which have a shifting boundary), and net international migration (in two age intervals, with a shifting boundary). For population growth, the partitioning of mortality and net migration by age is not necessary, although making a distinction by age highlights the parallel between the sources of change in the two outcomes.

6 **DISCUSSION**

The model is conceptually simple but has some limitations. One is that the specification of the time interval is necessarily somewhat arbitrary. We have used the 50-year interval 1970-2020 because the WPP 2019 data extend to 2020. Within this interval, especially for specific countries, the trajectories of change were not always steady and consistent. A focus on the cumulative effects at the end has sometimes masked irregularities within the interval.

If comparable rates were available for the preceding half century, 1920-1970, or earlier, the results would have been much different. Some regions and countries would have shown major increases in population size because of dramatic declines in mortality, but not in fertility. The balance between fertility and mortality as sources of change would have been tilted much more toward mortality. The dominant role of fertility in our results is specific to the choice of reference period, and how it relates to the overall demographic transition.

There is also some sensitivity to the choice of dates for the initial (pre-1970) rates. We have used the rates for the 5-year interval 1965-70 in the simulations. Somewhat different results would be obtained with rates for 1960-70 or 1969-70.

The results using future versions of World Population Prospects would be somewhat different, because they will include updated, and more accurate, estimates of the quantities used in the model than were possible when WPP 2019 was prepared.

The model describes changes in median age, rather than mean age. For most purposes, the mean of a distribution is the best single summary statistic. However, for skewed distributions, the median is often more interpretable as the central value. The mean of any distribution is sensitive to the numerical values of the variable being analyzed. By contrast, a change in the median is affected only by changes in the number of cases below and above the previous median and not by changes in their distance from the median.

Over time, as the populations of most countries have aged and both the mean and median ages have increased, the age distributions have become less skewed and the mean and median have become closer. On average, giving equal weight to each population in the WPP 2019 data, from 1970 to 2020 the mean age increased from 25.8 to 32.7 and the median age increased from 21.2 to 30.7. In 1970, on average, the mean was 4.6 years higher than the median, with a difference ranging from 0.7 to 7.9 years. No country or region had an age distribution in which the median was above the mean. In 2020, on average, the mean was only 2.0 years above the median, with a difference ranging from -3.0 to 5.2.

The figure shown below compares the mean⁷ and median of the age distributions in 1970 (in blue) and 2020 (in red) using all countries and regions in the WPP 2019 database. The correlation between the mean and median was very high throughout the 50 years and increased slightly from 0.984 in 1970 to 0.996 in 2020.

⁷ The mean age is calculated from the spreadsheet for age distributions in 5-year age intervals, setting the mean age within each age interval at the midpoint of the interval. The median is taken directly from the spreadsheet for median age. Equal weight is given to each of 238 units (countries and aggregations of countries).



The model used in this paper, tracking the median rather than the mean over time, separates mortality and net international migration into components below and above the median. An alternative and more analytical approach, using the mean, was used by Preston and Stokes (2012). More recently, Canudas-Romo, Shen, and Payne (2021) developed a related approach. In both analyses the authors focused on aging, reductions in mortality at the oldest ages, and the increased relative size of the older population.

In the model for the median, no use was made of any specific age above the median. The only fixed age in the decomposition is age 5, with a child death rate (actually a probability) constructed for children born between times t and t+5 who do not survive beyond time t+5. Otherwise, the mortality components are based on the survival between t and t+5 of cohorts below and above the median, with a shift of the age boundary between the younger and older cohorts as the median changes.

The partitioning into additive components has required a residual term, primarily because of the nonadditive or multiplicative relationship of the fertility and child mortality components. We retain the residual term, and give it explicitly, although alternative decompositions could have allocated it among the one-way effects, which would have generated slightly different results. Within the procedure, a residual for just the combination of fertility and child mortality was allocated proportionately to those two components. As an alternative, that residual could have been combined with the residual for the other components.

The data do not include any direct estimates of net international migration, and the estimates given here are indirect. We repeat that the letter "r" in the labels for the migration components is intended to emphasize that net migration is being estimated from a residual. The migration estimates have face validity, although numerical values must not be over-interpreted. Migration is included as a rate, and in the simulations the rate may be applied to a changed denominator, which would produce a changed net number of migrants. It is expected that the numbers of births and deaths are proportional to cohort size. However, it is somewhat artificial to require that numbers of migrants are also proportional to cohort size. Some other, more technical limitations in the model are described in Appendix 1.

7 CONCLUSIONS

"How much will the world's population continue to grow?" is a recurrent research and policy question that resurfaces every time the UN Population Division releases the World Population Prospects (WPP) report that details population estimates and projections. The most recent revision of the WPP projects that from an estimated 7.7 billion people worldwide in 2019, the global population could grow to 8.5 billion in 2030, 9.7 billion in 2050, and 10.9 billion in 2100 (medium variant, United Nations 2019). According to the report, two-thirds of the projected growth of the global population through 2050 will be driven by the current age structure and would occur even if childbearing in high-fertility countries today were to fall immediately to replacement level.

Our results show that during the past half-century, changes in population size and median age – more specifically, the difference between the 2020 levels and what would have been expected in 2020 if the pre-1970 rates had continued – have been determined primarily by declines in fertility, with relatively small offsetting contributions from mortality, including child mortality.

Examining population growth has been key for demographers and population researchers who link it to development outcomes and economic growth. This is in part because change in population is typically accompanied by change in the shape of the age distribution. Recent work has found convincing evidence to invest in frameworks such as the Sustainable Development Goals (Abel, Barakat, Samir, and Lutz 2016), the demographic dividend (Bloom, Canning, and Sevilla 2003), and specific interventions such as family planning programs (Bongaarts 2016). Notably, an important focus of the research and policy debate has been the impact of population growth on climate change (O'Neill et al. 2010; Hardee and Mutunga 2010; Neumann et al. 2015) and directing resources that urge governments to take action to mitigate these effects. Identifying drivers of change in population age structure may guide policy on investing further in mortality reductions (child survival interventions, for example), fertility reductions, or migration policies.

Median age analysis and change has implications for countries as they plan resources for targeted investments in their populations as they age. Research has shown that myriad outcomes of a country, from economic growth (Bloom and Canning 2008) to political stability (Weber 2011), are linked to age structure. This paper seeks to contribute to the discussion in four ways. First, it provides a robust empirical analysis of population dynamics over a period of time for all regions and major subregions (and some countries) of the world. Second, the dominant role of fertility decline should give global family planning programs a stimulus to justify the importance of voluntarism and informed choice in family planning in the face of many competing health interests. Third, the findings underscore the importance of investing in all drivers of fertility decline, and not just family planning. More effort could be directed to each proximate determinant of fertility (Bongaarts 2015) and to identifying the direct biological and behavioral influences. Important cultural, economic, social, and environmental distal factors work in conjunction with the proximate determinants in affecting total fertility rates. Finally, analyses to contextualize discussions and future research on the implications of population growth and change in median age are important, particularly with respect to population aging and its implications at the global, regional, and country levels.

REFERENCES

Abel, G. J., B. Barakat, K. C. Samir, and W. Lutz. 2016. Meeting the Sustainable Development Goals Leads to Lower World Population Growth. *Proceedings of the National Academy of Sciences of the United States of America* 113 (50) 14294-14299. https://doi.org/10.1073/pnas.1611386113.

Bloom, D. E. and D. Canning. 2008. Global Demographic Change: Dimensions and Economic Significance. *Population and Development Review* 34: 17-51. https://www.jstor.org/stable/25434758?seq=1.

Bloom, D. E., D. Canning, and J. Sevilla. 2003. *The Demographic Dividend: A New Perspective on the Economic Consequences of Population Change*. Santa Monica, CA, USA: The Rand Corporation. https://www.rand.org/content/dam/rand/pubs/monograph_reports/2007/MR1274.pdf.

Bongaarts, J. 2015. Modeling the Fertility Impact of the Proximate Determinants: Time for a Tune-up. *Demographic Research* 33: 535-560. https://www.demographic-research.org/volumes/vol33/19/33-19.pdf.

Bongaarts, J. 2016. Development: Slow Down Population Growth. *Nature* 530: 409-412. https://doi.org/10.1038/530409a.

Canudas-Romo, V., T. Shen, and C. Payne. 2021. The Role of Reductions in Old-Age Mortality in Old-Age Population Growth. *Demographic Research* 44: 1073-1084. https://www.demographic-research.org/volumes/vol44/44/44-44.pdf.

Hardee, K., and C. Mutunga. 2010. Strengthening the Link between Climate Change Adaptation and National Development Plans: Lessons from the Case of Population in National Adaptation Programmes of Action (NAPAs). *Migration and Adaptation Strategies for Global Change* 15 (2): 113-126. http://dx.doi.org/10.1007/s11027-009-9208-3.

Lutz, W, J. C. Cuaresma, E. Kebede, A. Prskawetz, W. C. Sanderson, and E. Striessnig. 2019. Education Rather Than Age Structure Brings Demographic Dividend. *Proceedings of the National Academy of Sciences of the United States of America* 116 (26): 12798-12803. https://doi.org/10.1073/pnas.1820362116.

Neumann, B., A. T. Vafeidis, J. Zimmermann, and R. J. Nicholls. 2015. Future Coastal Population Growth and Exposure to Sea-Level Rise and Coastal Flooding - A Global Assessment. PLoS ONE 10 (3): e0118571.

https://doi.org/10.1371/journal.pone.0118571.

O'Neill, B. C., M. Dalton, R. Fuchs, L. Jiang, S. Pachauri, and K. Zigova 2010. Global Demographic Trends and Future Carbon Emissions. *Proceedings of the National Academy of Sciences of the United States of America* 107 (41): 17521-7526. https://doi.org/10.1073/pnas.1004581107.

Preston, S. H., and A. Stokes. 2012. Sources of Population Aging in More and Less Developed Countries. *Population and Development Review* 38 (2): 221-236. https://doi.org/10.1111/j.1728-4457.2012.00490.x.

Starbird, E., M. Norton, and R. Marcus. 2016. Investing in Family Planning: Key to Achieving the Sustainable Development Goals. *Global Health Science and Practice* 4 (2): 191-210. https://doi.org/10.9745/GHSP-D-15-00374.

United Nations, Department of Economic and Social Affairs, Population Division. 2019. World Population Prospects 2019: Highlights (ST/ESA/SER.A/423). https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf.

Weber, H. 2012. Demography and Democracy: The Impact of Youth Cohort Size on Democratic Stability in the World. *Democratization* 20 (2): 335-357. https://doi.org/10.1080/13510347.2011.650916.

APPENDIX 1 DETAILED DESCRIPTION OF THE DECOMPOSITION

A1.1 A Model for Change in Population Size and Median Age

In the absence of external migration, in a closed population, the basic equation for population change is P(t+5) = P(t) + B(t) - D(t). Here P(t) is the population at time t and P(t+5) is the population at time t+5. B(t) and D(t) refer to the numbers of births and deaths, respectively, between times t and t+5.

Define m(t) to be the median age at time t. For example, m(1970)=23.45 means that on July 1, 1970, half of the population had not yet reached exact age 23.45 and the other half was older.⁸ In general, at time t, the population above the median is P(t)/2 and the population below the median is P(t)/2. The difference between the numbers above and below the median is 0.

Typically, the total population will increase between t and t+5. If half of the *increase* in population is *above* 23.45 (for example) and half is *below* 23.45, the median will not change. Otherwise, the amount of change in the median in the next 5 years is determined mainly by the *difference* between (a) the increase in the population above the median and (b) the increase in the population below the median. If (a) is larger than (b), there is a general aging of the population and m(t+5) will be greater than m(t). If the reverse occurs, the median age will decline.

It will help to define more terms. Define PO(t+5) to be the number of people at time t+5 (such as July 1, 1975) who are older than *the previous* median, 23.45. This number consists of *survivors* of people who were alive and above exact age 18.45 (m(t) – 5 = 23.45 - 5) in 1970. The letter "O" in "PO" refers to "the older cohort." Similarly, define PY(t+5) to be the number of people at time t+5 who are *survivors* of "the younger cohort," people who were younger than exact age m(t)-5 at time t. In addition to the younger cohort, the population at time t+5 includes children age 0-4 who are *survivors of the births* between t and t+5. That number will be labelled PCH(t+5). The difference between the number at t+5 who are above the previous median and the number below the previous median is Δ =PO(t+5) – [PY(t+5) + PCH(t+5)]. If Δ >0, the median will increase and if Δ <0 it will decrease.

Note the connection with population increase. The total population at time t+5 is the sum P(t+5) = PO(t+5) + [PY(t+5) + PCH(t+5)]. Projecting the population from t to t+5 can be expressed as an allocation into the same three subgroups, and then changing a sign from "-" to "+".

We next describe the relationship between Δ and the change in the median, describing it in a context where the median is rising. The increase in the median, m(t+5)-m(t), must be sufficient to "capture" half of the difference Δ , which is $\Delta/2$ people at time t+5. The increase in the median will also depend on the number of people per year of age in the vicinity of the preceding median. In effect, $\Delta/2$ is the area of a rectangle with base m(t+5)-m(t), in units of years, and height h, where h is approximately the number of people in the single year of age between age m(t)-0.5 and age m(t)+0.5.⁹ Thus, because the area of a rectangle is the

⁸The median is calculated by interpolation with many decimal places, and no one is "at" the median.

⁹This number is calculated from the age distribution at time t+5, using the median age from time t.

product of the base and the height, we have $\Delta/2 = [m(t+5)-m(t)]^{*}h$, leading to $m(t+5) = m(t) + \Delta/(2^{*}h) = {PO(t+5) - [PY(t+5) + PCH(t+5)]}/(2^{*}h)$.

This formula is not exact because both h and the median are calculated with interpolation, using the discrete age intervals in the data.

A1.2 Cohort Survival and Migration

To track the older cohort, younger cohort, and births from one survey to the next, we define four age intervals in the age distribution at time t and also at time t+5, using the median m(t), as follows:

- 1. below exact age 5 (a 5-year age interval)
- 2. exact age 5 to exact age m(t)-5, where m(t) is the median at time t
- 3. exact age m(t)-5 to exact age m(t) (a 5-year age interval)
- 4. exact age m(t) and above.

The populations in these four intervals are labeled P1(t), P2(t), P3(t), and P4(t) at time t; and P1(t+5), P2(t+5), P3(t+5), and P4(t+5) at time t+5. At time t, P4(t) = P1(t)+P2(t)+P3(t) = P(t)/2. This is a shifting notation for a pair of successive time points, using the median at time t, m(t), as a reference point at time t+5 as well as time t. That is, the median at time t is carried over to the distribution at time t+5. The only age interval that is fixed over time is the first one, for children below exactly age 5.

The link between the distribution at time t+5 and that at time t is as follows:

P1(t+5): survivors of the births between t and t+5 P2(t+5)+P3(t+5): survivors of P1(t)+P2(t) (the younger cohort) P4(t+5): survivors of P3(t)+P4(t) (the older cohort)

At this point in the procedure, we introduce the possibility of international migration. When P4(t+5) is compared with P3(t)+P4(t), for example, the difference reflects emigration or immigration, as well as survival. It is possible, and in some countries is observed, that a cohort is larger at time t+5 than at time t because the net number of immigrants exceeded the number of deaths.

The UN data do not include direct information about migration. The number of migrants is inferred as a residual after the number of survivors has been estimated with the 5-year survival ratios that are included in the spreadsheet with abridged life tables.

We are thus able to calculate five rates:

dch, the death rate for children age 0-4;

dy, the death rate for the younger cohort between t and t+5;

do, the death rate for the older cohort between t and t+5;

ry, the net immigration rate for the older cohort between t and t+5;

ro, the net immigration rate for the older cohort between t and t+5.

The "death rate" (dy or do) described here is a ratio, the number of deaths to a cohort, divided by the initial size of the cohort. This is not an occurrence/exposure rate. For children age 0-4, the death rate (dch) is the life table estimate of the proportion of children born between t and t+5 who do not survive to time t+5.¹⁰ These ratios are labeled "rates" for simplicity.

The age ranges for the younger and older cohorts shift gradually as the median age changes. This is not considered to be a serious issue because the mortality risk is usually relatively low and relatively flat in the vicinity of the median age. However, it should be considered when interpreting the balance, between the younger and older cohorts, of the effects of changes in mortality and migration.

For the younger and older cohorts, the difference between the observed size of the cohort at time t+5 and the initial observed size of the cohort at time t, minus the expected number of deaths, is interpreted as a measure of migration. That number, when divided by the initial observed size of the cohort, is a net emigration rate. Migration is treated as a decrement, like mortality, so that a positive value describes net emigration; a negative value describes net immigration. We do not include a migration component for children born between t and t+5.

The residual described above as migration may arise in part from reconciliations among the components of the complex model used by the UN. The Population Division has produced trajectories for all countries of the world that have a remarkably high level of internal consistency. Nevertheless, much of the input data available to the UN for those calculations is incomplete or inconsistent. For many countries, direct estimates of international migration are unavailable. Also, although the deaths between t and t+5 are calculated by applying the life table survival ratios to the population at time t, in 5-year age intervals, deaths to migrants between t and t+5 are not included as deaths in the model, and persons who immigrated and died between t and t+5 are not included.

The number of deaths to the births between t and t+5 is calculated with the life table under-5 survival information. Some of what would normally be understood as under-5 mortality is included in the mortality of the younger cohort (dy) and is not included with the deaths to children born between t and t+5 (dch). Some of the under-5 deaths between t and t+5, typically about 30%, occur in children who were born during the 5 years *before* t, rather than in the 5 years *after* t. It could be desirable to shift those under-5 deaths from dy to dch, but that shift would disrupt the cohort perspective and the "accounting" structure of the model. The number of deaths that are included in dch are proportional to the number of births.

The number of births between t and t+5 is available in the UN data, as are the crude birth rate (CBR) and the total fertility rate (TFR). Of these three measures of fertility, the TFR is preferred for this model. The TFR is readily interpreted as the expected number of children to a woman who survives the childbearing ages, with current age-specific fertility rates, and is widely used in policies.

We describe reproduction in terms of the ratio of the number of children under age 5 at time t+5 to the total population at time t: P1(t+5)=K*TFR*(1-dch)*P(t). Here dch is calculated from the life table for the time interval, as described. TFR(t) comes directly from the UN data for the interval and is used as an alternative to CBR(t). The only term that is unknown is K, which is also specific to the time interval. The term K is a reconciliation term required because the measure of fertility is the TFR rather than the CBR. It is calculated

¹⁰ The child death rate, dch, is calculated as $1-5L_0/500,000$, using a UN life table with radix 100,000.

by rearranging the equation algebraically and substituting the known values of the other terms. It is found that K varies empirically within a narrow range, usually centered on approximately 0.4.

To summarize, for a specific country, by using the UN data, we can link the population and median age at time t+5, with the population and median age at time t, with the following equations, in which K, TFR, dch, dy, ry, do, and ro refer to the interval between t and t+5:

- Children under age 5: P1(t+5)=K*TFR*(1-dch)*P(t)
- The younger cohort: P2(t+5)+P3(t+5)=(1-dy-ry)*[P1(t)+P2(t)]
- The older cohort: P4(t+5)=(1-do-ro)*[P3(t)+P4(t)]

If K and the UN median age are recalculated at each step, we have essentially an exact accounting framework. If the reconciliation term, K, is set at its median calculated value for the series, we can obtain an excellent tracking of the series of age distributions. When the recursive formulas are applied, the change in population size and median age are thus driven by six rates: TFR, dch, dy, ry, do, and ro.

A1.3 Components of Change between 1970 and 2020

The trajectories of the six rates across 5-year intervals of time between 1970 and 2020 describe changes in fertility, survivorship, and net migration during the 50-year time span. The next step in the procedure is the articulation of change in population size and median age into components for the six rates. This decomposition or partitioning is done under the following scenarios, which are applied for the successive 5-year intervals. Each scenario refers to a trajectory of values of either population size or the median age. First define the following:

- S0: The observed trajectory of population size or median age.
- S1: A hypothetical trajectory that would have been observed if TFR, dch, dy, and do were all fixed at 1965-1970 levels, and ry and ro were zero.

The procedure allocates the difference S1-S0 into additive components that can be interpreted as the cumulative effects of changes in TFR, dch, dy, do, ry, and ro. These effects can be assessed throughout the interval, but the greatest interest lies in their final values in 2020. To accomplish this, we construct a set of scenarios of "partial" changes in the six components, collectively referred to as S2. The label "all" refers to TFR, dch, dy, and do as a set, not including ry and ro.

- S2(TFR): TFR changes as observed; dch, dy, and do are fixed at 1965-1970 levels; ry=ro=0;
- S2(dch): dch changes as observed; TFR, dy and do are fixed at 1965-1970 levels; ry=ro=0;
- S2(TFR, dch): both TFR and dch change as observed; dy and do are fixed at 1965-1970 levels; ry=ro=0;
- S2(dy): dy changes as observed; TFR, dch, and do are fixed at 1965-1970 levels; ry=ro=0;
- S2(do): do changes as observed; TFR, dch, and dy are fixed at 1965-1970 levels; ry=ro=0;
- S2(all). TFR, dch, dy, and do change as observed; ry=ro=0;

S2(all, ry): TFR, dch, dy, do, and ry change as observed, ro=0;

S2(all, ro): TFR, dch, dy, do, and ro change as observed, ry=0.

The final scenario, S3, is an application of the model in which all the components change as observed:

S3: TFR, dch, dy, do, ry, and ro all change as observed.

Because the model is basically an accounting framework, we expect S3 and S0 to be very close. There will necessarily be some discrepancies, at least in part because the K factor has been replaced by its median value in all the scenarios. The difference S3-S0 is therefore calculated as a residual.

Effects, which are differences between scenarios, are constructed in a hierarchical manner, as follows.

D(TFR)=S2(TFR)-S1 D(dch)=S2(dch)-S1 D(TFR, dch)=S2(TFR, dch)-S1 D(dy)=S2(dy)-S1 D(do)=S2(do)-S1 D(ry)=S2(all,ry)-S2(all)D(ro)=S2(all,ro)-S2(all)

The effects of dy, do, ry, and ro are additive, while the effects of TFR and dch are not, because the number of child deaths is proportional to the number of births. As a result of the non-additivity of TFR and dch, D(TFR,dch) is not equal to D(TFR)+D(dch). The difference is a type of two-way interaction common in decompositions. It could be reported in the results as a residual, but instead we allocate it to D(TFR) and D(dch). Typically, in decompositions, a two-way interaction term is divided by two and allocated equally to the one-way effects. Here, however, the one-way effects have different signs. Equal allocation could change the sign of a one-way effect, and potentially lead to a misinterpretation of the counterbalancing effects of fertility and child mortality. Therefore, the interaction term is allocated in proportion to the one-way effects. D(TFR) and D(dch) are replaced by D(TFR)' and D(dch)', respectively, defined with D(TFR)'=F*D(TFR) and D(dch)'=F*D(dch), where F=D(TFR, dch) / [D(TFR)+D(dch)]. F is a time-varying factor that forces additivity and maintains the signs of the TFR and dch effects.

The migration rates are not systematic and can be positive or negative. Their effects are estimated after the other effects have been taken into account, as shown with the definitions of D(ry) and D(ro).

The difference S3-S1 can be partitioned into a sum of differences plus a residual:

S3-S1=D(TFR,dch)+D(dy)+D(do)+D(ry)+D(ro)+R31

Here R31 is calculated as R31=(S2-S1)-[D(TFR,dch)+D(dy)+D(do)+D(ry)+D(ro)].

To summarize, the terms in the overall decomposition are given by:

S1-S0=D(TFR)'+D(dch)'+D(dy)+D(do)+D(ry)+D(ro)+R,

where R=R31+(S3-S0) is a final residual term. R31 and (S3-S0) are usually small and are aggregated, rather than reported separately. The final residual term for median age is usually less than one year in magnitude. When it appears to be large relative to the difference to be explained, it is usually because the difference to be explained is small.

A disaggregation or separation step within the simulations needs to be clarified. The survival of the younger cohort produces the sum P2(t+5)+P3(t+5), and for the next step in the simulation it is necessary to specify an allocation into age intervals P2 and P3. This is done with a statistical model based on the observed balance between these two numbers as a function of median age. For the observed data between 1970 and 2020, the observed ratio of population counts P3/(P2+P3) is regressed on the observed ratio of the widths of the age intervals, 5/(median age - 5). Coefficients b0 and b1 are produced. Then, when we have a new median age and the sum (P2+P3), P3 is estimated as [b0+b1*5/(median age - 5)]*(P2+P3). P2 is the complement, obtained by subtracting the estimate of P3 from (P2+P3). Thus, the balance between P2 and P3 relies on statistical, rather than demographic modeling, based on consistency with the balance in the observed distributions for each specific region or country. Alternative models to separate P2+P3 into P2 and P3 are certainly possible, but we believe the results would be similar.

APPENDIX 2 COMPONENTS OF CHANGE FOR SELECTED COUNTRIES

This appendix presents decompositions and figures for the countries classified as high priority by the Office of Population and Reproductive Health at USAID. The countries are listed in alphabetical order.




























































