Data Quality and Accuracy of United Nations Population Projections, 1950-95

Nico Keilman

Population Studies, Vol. 55, No. 2 (Jul., 2001), 149-164.

Stable URL:
http://links.jstor.org/sici?sici=0032-4728%28200107%2955%3A2%3C149%3ADQAAOU%3E2.0.CO%3B2-5

Population Studies is currently published by Population Investigation Committee.

Population Studies, 55 (2001), 149–164
Printed in Great Britain

Data quality and accuracy of United Nations population projections, 1950–95

NICO KEILMAN

Abstract. Between 1951 and 1998, the United Nations (UN) published 16 sets of population projections for the world, its major regions, and countries. This paper reports the accuracy of the projection results. I analyse the quality of the historical data used for the base populations of the projections, and for extrapolating fertility and mortality. I study also the impact this quality has had on the accuracy of the projection results. Results and assumptions for the sets of projections are compared with corresponding estimates from the UN 1998 Revision for total fertility and life expectancy at birth, total population, and the projected age structures. The report covers seven major regions (Africa, Asia, the former USSR, Europe, Northern America, Latin America, and Oceania) and the largest ten countries of the world as of 1998 (China, India, the USA, Indonesia, Brazil, Pakistan, Russia, Japan, Bangladesh, and Nigeria).

1. INTRODUCTION

Between 1951 and 1998, the United Nations (UN) published 16 sets of population projections for the world, its major regions, and nearly all countries. Details of these projections are given in Table 1, where they are labelled by base year. For some of these base years, the projections were revised a few years later: 1950I, 1950II, 1950III; 1975I and 1975II, etc.

Since the earliest projections were published, in the 1950s, they have expanded in regional and age detail, in time coverage, and in methodological sophistication (El-Badry and Kono 1986; Frejka 1994). Africa was absent from the 1950I series, because of data problems. Country detail was available only for Latin America and the Far East. The 1950II series attempted to derive country projections from the projected totals in each of the 25 regions. The 1960 and 1965 series used stable population theory and indirect estimation methods to estimate basic indicators from incomplete data. Available computer facilities made it possible to prepare the 1965 projections by age and sex for each country, and to compute a large number of other indicators. Various sets of model schedules for fertility, mortality, and migration were applied. Backward projections, starting from the base year 1965 and going back to 1950, were also prepared for each country. These developments continued into the 1970s. More detailed indicators were computed, base line data were improved, and the complex links between socioeconomic, political, and cultural factors in fertility and mortality change were taken into account. The World Fertility Survey (1974–82) greatly contributed to our understanding of those links. Finally, in the 1980s, the cycle of revisions was shortened from once every five years to once every two years.

Many of the historical estimates were adjusted in later years. At the time when projections are prepared, data on a country's base population are taken (or projected) from a recent census or a national population registration system. At that time, these data reflect the best estimates for the age pyramid in the projection's base year. Furthermore, assumptions on future levels of fertility, mortality, and migration are obtained on the basis of extrapolated historical values of key indicators, combined with substantive insights into the effects of socioeconomic, cultural, and policy variables on components of change. Data on those key indicators come from various sources, including surveys, vital registration, indirect estimation, and administrative sources.

New data (for example from a recent census or a post-enumeration survey), better estimation methods, and new analyses of fertility, mortality, and migration may all lead to improvements in older estimates. Historical demographic data are not fixed once and for all, because new findings and interpretations alter the picture we have of the past. For instance, the estimate of the world population in 1950 changed 17 times – most often in an upward direction – in UN Demographic Yearbooks published from 1951 to 1996. The largest revision in the estimate of the world population in 1950, by 44 million, occurred between 1953 and 1954. The main reason was a surprisingly high population estimate for China when the results of the census of 1953 became available.

Table 1. *UN population projections 1950–98*

No.	Label	Source (year of publication)	Base year	Fertility and mortality extrapolations	Projected age structure
1.	1950I	*Population Bulletin of the United Nations 1* (1951). Sales no. E52.XIII.2	1950		Total population size in 1980
2.	1950I	*Future population estimates by sex and age. Report I: The population of Central-America (including Mexico), 1950–1980* (1954) Sales no. 1954.XIII.3. *Report II: The population of South-America 1950–1980* (1955). Sales no. 1955.XIII.4. *Report III: The population of South-East Asia (Including Ceylon and China: Taiwan) 1950–1980* (1958) Sales no. 59.XIII.2. *Report IV: The population of Asia and the Far East 1950–1980* (1959) Sales no. 59.XIII.3	1950		Five-year age groups for the years 1955, 1960, ..., 1980
3.	1950II	*Proceedings of the World Population Conference 1954* Vol. III (1955) Sales no. E.55.XIII.8, pp. 265–328	1950		Update of 1950I; only total population for 1955, 1960, ..., 1980
4.	1950III	*The future growth of world population* (1958) Population Studies 28 Sales no. 58.XIII.2	1950		Update of 1950II; five-year age groups for 1960; broad age groups for 1960 and 1975; total population for 1960, 1965, ..., 2000
5.	1960	*World population prospects as assessed in 1963* (1966) Sales no. 66.XIII.2	1960	CBR and CBR for major regions	Broad age groups for 1960, 1965, 1970, ..., 1980, and 2000
6.	1965	*World population prospects as assessed in 1968* (1973) Population Studies 53 Sales no. E.72.XIII.4	1965	CBR, CDR, total fertility, and life expectancy for major regions and (up to 1980–85) for countries	Five-year age groups up to 70+ for 1970, 1975, ..., 2000 for major regions
7.	1970	*World population prospects as assessed in 1973* (1977) Sales no. E.76.XIII.4 and corrigenda	1970	CBR, CDR, total fertility, and life expectancy	Broad age groups for 1985 and 2000 only; total population for 1975, 1980, ..., 2000
8.	1975I	*Selected demographic indicators by country, 1950–2000: Demographic estimates and projections as assessed in 1978* (1980) ST/ESA/SER.R/38	1975	CBR, CDR, total fertility, and life expectancy	Five-year age groups for 1975,1980, 1990, 2000; broad age groups for 1975, 1980, ..., 2000
9.	1975II	*Demographic indicators of countries: Estimates and projections as assessed in 1980* (1982) ST/ESA/SER.A/82 and corrigendum	1975	CBR, CDR, total fertility, and life expectancy	Five-year age groups for 1975, 1980, ..., 2025
10.	1975II	*World Population Prospects as assessed in 1980* (1981). ST/ESA/SER.A/78	1975	CBR, CDR, total fertility, and life expectancy	Total population 1975, 1980, ..., 2025
11.	1980I	*World population prospects: Estimates and projections as assessed in 1982* (1985) ST/ESA/SER.A/86	1980	CBR, CDR, total fertility, life expectancy, and IMR	Five-year age groups for 1980, 1985, ..., 2025

12.	1980II	*Global estimates and projections of population by sex and age: The 1984 assessment* (1987) ST/ESA/SER.R/70	1980	CBR, CDR, total fertility, life expectancy, and IMR	Five-year age groups for 1980, 1985, …, 2025
13.	1980II	*World Population Prospects. Estimates and Projections as assessed in 1984* (1986). ST/ESA/SER.A/98 and corrigendum	1980	CBR, CDR, total fertility, life expectancy, and IMR	Broad age groups
14.	1985I	*Global estimates and projections of population by sex and age: The 1988 revision* (1989) ST/ESA/SER/.R/93	1985	CBR, CDR, total fertility, life expectancy, and IMR	Five-year age groups 1985, 1990, …, 2025
15.	1985I	*World Population Prospects 1988* (1989). ST/ESA/SER.A/106	1985	CBR, CDR, total fertility, life expectancy, and IMR	Broad age groups
16.	1985II	*The sex and age distribution of population: The 1990 revision of the United Nations global population estimates and projections* (1991) ST/ESA/SER.A/122	1985	CBR, CDR, total fertility, life expectancy, and IMR	Five-year age groups for 1985, 1990, …, 2025
17.	1985II	*World Population Prospects 1990* (1991). ST/ESA/SER.A/120	1985	CBR, CDR, total fertility, life expectancy, and IMR	Broad age groups
18.	1990I	*The Sex and Age Distribution of the World Populations. The 1992 Revision* (1993). ST/ESA/SER.A/134	1990	CBR, CDR, total fertility, life expectancy, and IMR	Five-year age groups for 1990, 1995, …, 2025
19.	1990I	*World Population Prospects. The 1992 Revision* (1993). ST/ESA/SER.A/135	1990	CBR, CDR, total fertility, life expectancy, and IMR	Broad age groups
20.	1990II	*The Sex and Age Distribution of the World Populations. The 1994 Revision* (1994). ST/ESA/SER.A/144	1990	CBR, CDR, total fertility, life expectancy, and IMR	Five-year age groups for 1990, 1995, …, 2050
21.	1990II	*World Population Prospects. The 1994 Revision* (1995). ST/ESA/SER.A/145	1990	CBR, CDR, total fertility, life expectancy, and IMR	Broad age groups
22.	1995I	*The Sex and Age Distribution of the World Populations. The 1996 Revision* (1997). ST/ESA/SER.A/162	1995	CBR, CDR, total fertility, life expectancy, and IMR	Five-year age groups for 1995, 2000, …, 2050
23.	1995I	*World Population Prospects. The 1996 Revision* (1998). ST/ESA/SER.A/167	1995	CBR, CDR, total fertility, life expectancy, and IMR	Broad age groups
24.	1995II	*World Population Prospects. The 1998 Revision. Volume I: Comprehensive Tables* (1999) ST/ESA/SER.A/177. *Volume II: Sex and age* (1999) ST/ESA/SER.A/180. *Volume III: Analytical Report* (1999) ESA/P/WP.156	1995	CBR, CDR, total fertility, life expectancy, and IMR	Five-year age groups for 1995, 2000, …, 2050

The UN has repeatedly stressed the problems of inaccurate base data for age structures and vital components in certain regions and countries. Although data collection efforts have increased substantially over the past 20 years, 17 per cent of all countries or areas included in the UN 1998 Revision still had no post-1985 census data available on population size and age structure (Zlotnik 1999). In particular Africa has posed problems for several decades now. For instance, 14 of the 57 African countries in the 1995I-projections had their most recent census before 1985. Data on vital rates are often even more problematic. Countries accounting for no less than 40 per cent of world population lacked data more recent than 1990 on adult mortality. For fertility and child mortality, the data situation is generally much better (Zlotnik 1999).

This paper has two purposes. The first is to present general statistics on projection accuracy. The second is to analyse the extent to which data quality has influenced the accuracy of UN projections. From earlier empirical studies of the correlates of projection accuracy (Inoue and Yu 1979; Keyfitz 1981; Stoto 1983; Pflaumer 1988; Keilman 1998; National Research Council 2000) we know that the accuracy of UN projections is better for short than for long projection durations, and that it is better for large than for small countries and regions. We also know that projections of the old and the young tend to be less reliable than those of intermediate age groups, and that there are considerable differences in accuracy between regions. For instance, the crude death rate has been relatively difficult to project in Africa and Asia, while crude birth rate errors were above the world average for Asia, Northern and Latin America, and Oceania.

Relatively little attention has been given to the impact of data quality on projection accuracy. A recent study of nine past international projections (four by the UN, four by the World Bank, and one by the US Bureau of the Census) showed that absolute error in projected country populations averaged 5.5 per cent, taking the UN 1998 Revision as the benchmark (National Research Council 2000, Table B7). For countries that had a census in the 1990s, the error was 0.5 percentage points lower; for those without a census in the 1990s it was higher by 0.9 points. A recent census was also a strongly significant factor in the explanation of the error in projected total fertility and life expectancy. Whether or not countries have had a recent census is a useful proxy for data quality, but it masks important quality differences between countries and regions. One possible alternative is to measure the extent to which historical estimates have been revised in later years. This method has the advantage that it quantifies the quality of the old data. The approach taken in this paper is to inspect the error in the base population for each projection and the spread in the time series for birth and death rates.

The consequences of errors in the base population for age structure projections in later years can be assessed directly by a method described in Section 2.2.1. But for fertility and mortality extrapolations, it is impossible to determine explicitly to what extent these are too high or too low as a result of the quality of the historical data. That would require in-depth insight into the undocumented considerations of the projection makers, at the time when each of the projections was prepared. However, some indication is provided by bivariate correlations of projection error and data quality indicator. The data quality of a variable is operationalised here as the extent to which the estimate for that variable (for example, total fertility for 1970–74) had to be adjusted in subsequent revisions (for example in the 1980I round, the 1980II round, the 1985I round etc.). The range of adjustment for the 1970–74 total fertility estimates can be expressed in terms of the standard deviation, or some other measure of spread, of all 1970–74 total fertility estimates. A large spread indicates large problems with data quality, and, by hypothesis, low accuracy. See Section 2.2.2 for details.

For most projections, the UN computed more than one variant, typically a high, a medium, and a low one, as well as one based on an assumption of constant future fertility. In such cases I limited the analysis to the medium variant, as this is the one that users are most likely to select as the best guess. Since the focus in this paper is on comparative accuracy across regions and over time, it is unlikely that the choice of the medium variant has had any large impact on the conclusions.

The analysis will be restricted to seven major regions (Africa, Asia, the former USSR, Europe, Northern America, Latin America, and Oceania), and the largest ten countries of the world (China, India, the former USSR, USA, Indonesia, Brazil, Pakistan, Japan, Bangladesh, and Nigeria). Boundary changes created problems in the construction of time series in some instances. The USSR has been maintained as a separate regional entry into the 1990s. This means that the USSR is considered here both as one of the seven major regions, and as one of the largest ten countries. Corrections for Asian and European republics were

made accordingly. In other words, 'Europe' and 'Asia' should be interpreted in this paper as continents excluding republics of the USSR.[1]

The accuracy and the quality of the following demographic variables will be reported: total fertility and life expectancy at birth (both for the years before and after each projection's base year), total population size, and the population's age structure.

2. MEASURES OF PROJECTION ACCURACY AND DATA QUALITY

2.1 *Projection accuracy*

In this paper I assume that estimates published by the UN in the 1998 Revision for demographic variables up to 1995 are currently the best historical estimates one can obtain. In other words, the 1998 Revision is used as a benchmark. The projection error for a certain variable in a historical projection is simply taken as the difference between the projected value and the value as estimated in the 1998 Revision – 'estimated value' for short. This definition implies that a positive (negative) error indicates a projected value that was too high (low). To facilitate comparison between regions and countries of different size, many of the errors presented in this paper are in relative form, for instance the percentage error (PE, equal to (projected value – estimated value) × 100 / estimated value), or the absolute percentage error (APE, the absolute value of the PE). Percentage errors or absolute percentage errors are frequently averaged, for instance over a series of successive projection rounds. In this case we obtain the mean percentage error (MPE), or the mean absolute percentage error (MAPE). When an MPE is computed, positive errors (too-high projection results) are cancelled out by negative errors (too-low results). The MPE reflects the average *bias*: a positive value indicates that the projections tended to be too high on average, and a negative one that they tended to be too low. Where overestimates count as much as underestimates, the MAPE is a useful measure. It is a measure of *imprecision*, which tells us by how much the projections were in error, but not whether they were too high or too low.

2.2 *Data quality*

2.2.1 *Age structure projections*

Since I have adopted the 1998 Revision as the benchmark, there is already an error in the base population of many of the historical projections. Past projections often started off with an incorrect baseline population. Baseline errors are strongly age dependent, and these errors propagate through the age structure. As a consequence, the empirical error in the age structure after, say, ten years is the result of two sources. The first source is baseline error. The second one is an error caused by wrong assumptions for fertility, mortality, and migration for the ten-year projection period. I have simply assumed that, at any duration, the percentage error for each age group is the percentage baseline error for the corresponding birth cohort plus the percentage error caused by wrong assumptions. (This implies that the error in the base population is independent of that caused by wrong assumptions. Although this may seem a plausible assumption, it is not always a valid one. An example of dependence of base population errors on those in extrapolated mortality levels will be described below for the case of China.) Thus at each duration and for every age group, the share of the baseline error (for the birth cohort in question) in the total error indicates the degree to which data quality has contributed to the latter error. In general, the contribution of baseline error to the total error declines as projection length increases, and in the long run the dominant effect is that of wrong assumptions.

2.2.2 *Components of change*

Let $E_n(t)$ represent the estimate for a certain variable for the five-year period starting in year t, as assessed in the revision which the UN published together with the projection with base year n. This could be crude death rate, life expectancy at birth, infant mortality rate, total fertility etc. Thus, when analysing, say, life expectancy at birth, $E_{1980\text{I}}(1965)$ is the life expectancy estimate for the period 1965–69 according to the 1980I round of projections. The base year implied by label n is always later in time than the five-year period indicated by t; at most n represents the end of the period t, e.g. when n equals 1980I and t represents the period 1975–79. For a fixed period t, various base years n result in a series of estimates: $E_{1950\text{I}}(t)$, $E_{1950\text{II}}(t)$, $E_{1950\text{III}}(t)$, $E_{1960}(t)$, $E_{1965}(t)$,… The mean of this series is written as $m(t)$, and its standard deviation as $s(t)$. Such a series of estimates can be constructed for each period t, for each region or country, and for each demographic indicator. However, owing to missing data, the series for some indicators (crude birth rate, crude death rate) is longer than that for other indicators (total fertility,

life expectancy). The extent of missing data is also a little different between countries and regions.

3. THE ACCURACY OF TOTAL POPULATION SIZE

Figure 1 compares actual population size for the world as a whole with projected sizes according to a number of UN projections with base years starting in 1950. The first two projections with base year 1950 (i.e. 1950I, with published results for 1980 only, and 1950II) had both a population growth that was too low and a base population that was too small. The 1950III projections contained an upward revision of the world population compared to previous rounds, mainly as a consequence of the 1953 Census of China (El-Badry and Kono 1986, p. 37). The fact that the Chinese population was some 100 million larger than expected led in turn to substantial reductions in assumed mortality levels, and hence to stronger population growth starting with the 1950III projections. Averaging over all published UN projections for all calendar years, the resulting mean percentage error (MPE) shows that the world's projected population size was too low by 2.1 per cent. The overall mean absolute percentage error (MAPE) amounts to 2.8 per cent. The MAPE is one-third higher than the MPE, which suggests a fair amount of compensation of positive and negative errors in the MPE.

As is evident in Figure 2, projection accuracy diminishes for longer durations. The reason is that projections are based on smooth extrapolations of the past. The more years a projection covers, the greater the chance that unforeseen developments will produce unexpected changes in fertility, mortality, or migration. Note that even at duration zero, errors are not negligible because, as stated above, past projections often started off with an incorrect baseline population. These baseline errors reflect the quality of the data. Errors at longer durations are the combined result of data quality and the extent to which projection assumptions for the components of change (fertility, mortality and migration) were wrong. The relative contribution of data quality in the error declines with increasing duration (see below). Problems with the quality of the data for Africa are clearly visible in Figure 2. Only at durations of 10 years or more is the effect of data quality in African errors less than that of wrong assumptions in the demographic components.

Table 2 indicates that projections prepared in the 1950s had large errors, especially those for Africa, Asia, the former USSR, and Latin America. Asia

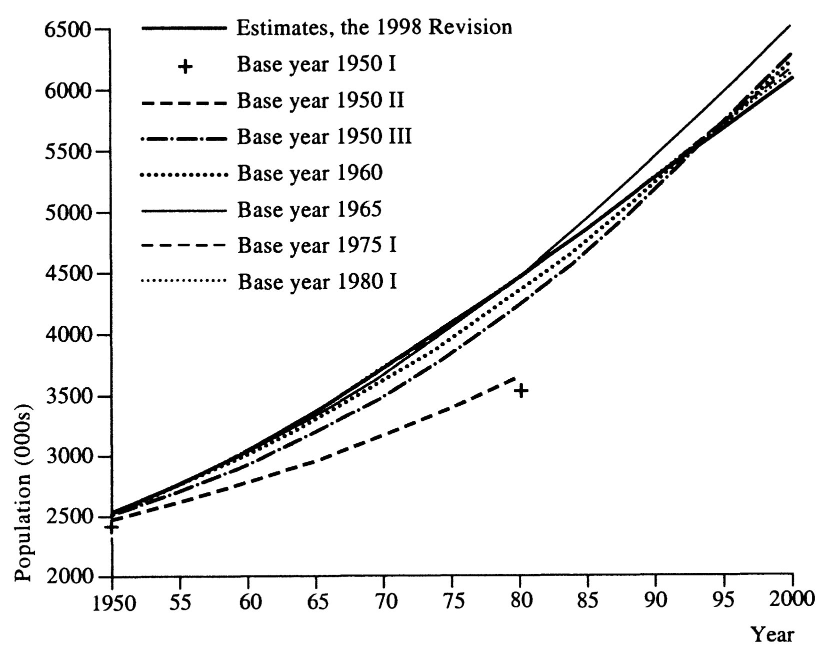


Figure 1. World population size 1950–2000, as projected and assessed since 1950.
Sources: See Table 1.

and Latin America are the only regions that had a more or less regular improvement in accuracy over the subsequent base years. In the other regions, errors in recent projections are not necessarily smaller than those in older ones. The reason is that, in addition to being based on demographic insights and data, any set of population projections is heavily based on current real trends in vital processes. Existing behavioural theories, even when extensively tested, may have limited predictive validity in time and space, may be strongly conditional, or may not be applicable without the difficult projection of non-demographic factors (Keyfitz 1982). That is why projection makers rely heavily on the extrapolation of historical trends in key demographic indicators, and thus why they face a problem when these indicators show unexpected changes in level or slope It will not be clear whether these are caused by random fluctuations, or whether there is a structural change in the underlying trends. A trend shift that is perceived as random will first lead to large projection errors. This effect is known in forecasting literature as assumption drag (Ascher 1978). Later, when the new trend is acknowledged, it will be included in the projection updates and the errors will diminish. On the other hand, random fluctuations that are perceived as a trend shift will cause projection errors, which will have a fluctuating effect on subsequent projections.

4. DATA QUALITY AND THE ACCURACY OF PROJECTED AGE STRUCTURES

UN projections for age structures since 1965 for seven large regions have been evaluated. Earlier UN projections had some age detail, but too little for a systematic analysis. For the largest ten countries in the world, age structures were evaluated starting with the projections with base year 1975.

The bias in the world's projected age structures is very modest: around two per cent on average for most five-year age groups, even when the projection period is 15 years. Exceptions are the young and the elderly segments of the world population. For ages 0–4, world projections since 1965 have been too high by up to five per cent after 15 years (and somewhat less for shorter durations). The elderly (aged 80 and over) have been underestimated by three, seven, and ten per cent for projections of five, ten, and fifteen years ahead, respectively.

The errors in projected age structures for post-transitional countries turn out to display similar patterns. The errors are large and positive for young age groups, and more or less equally large but *negative* for the elderly. Figure 3 illustrates the case of Europe. The over-projections among the young started around 1965, and were caused by unforeseen sharp declines in birth rates. The elderly were underestimated by up to 14 per cent after 15 years, because forecasters were too pessimistic about mortality, in particular for women. A similar shape in the error pattern for age structure projections was found for Northern America, Latin America, and Oceania. Earlier evaluations of projections prepared by statistical agencies in developed countries (Canada, Denmark, the Netherlands, Norway, and the United Kingdom) revealed the same error pattern as that in Figure 3 (Keilman 1997).

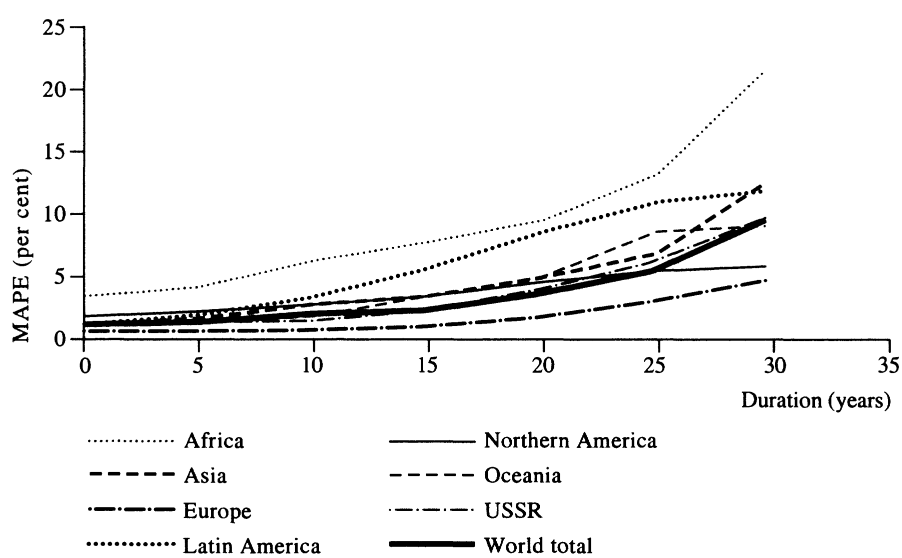


Figure 2. Mean absolute percentage error (MAPE) in projected total population size, by projection duration.
Sources: See Table 1.

Table 2. *Mean absolute percentage error (MAPE) in projected total population size in large regions, by base year*

	Africa	Asia	USSR[1]	Europe	Northern America	Latin America	Oceania	World total
Base year								
1950I	21.6	18.4	11.8	5.8	2.6	7.1	5.2	12.6
1950II	24.0	14.4	8.5	4.1	7.9	9.0	11.5	11.2
1950III	43.7	3.2	8.2	2.5	1.8	4.8	2.5	3.5
1960	3.3	3.9	4.9	0.9	4.4	6.4	2.2	1.8
1965	2.2	1.8	2.7	3.6	3.4	7.1	6.1	2.2
1970	1.1	2.3	1.4	1.6	2.6	5.9	3.8	1.5
1975I	1.8	1.1	1.1	0.4	3.9	5.7	0.9	0.6
1975II	2.4	1.0	0.8	0.1	3.0	2.7	0.9	0.2
1980I	4.5	1.1	1.2	0.1	2.2	2.2	1.0	0.2
1980II	4.7	1.1	1.3	0.2	2.1	1.8	0.6	0.2
1985I	5.7	0.5	0.9	0.4	2.3	1.8	0.5	0.9
1985II	4.9	0.5	0.9	0.3	2.3	1.8	0.4	0.9
1990I	5.6	0.7	0.9	0.4	1.8	0.3	1.1	1.1
1990II	3.7	0.4	0.3	0.1	1.4	0.3	0.1	0.6
1995I	3.2	0.0	0.5	0.1	0.0	0.7	0.6	0.4
Overall	11.7	3.6	3.8	1.7	3.3	4.8	3.2	2.8

Note 1: Russian Federation plus former USSR republics for 1995I and 1995II; including Baltic States for 1990I. Corrections accordingly were made for Europe and Asia, see endnote 1.
Sources: See Table 1.

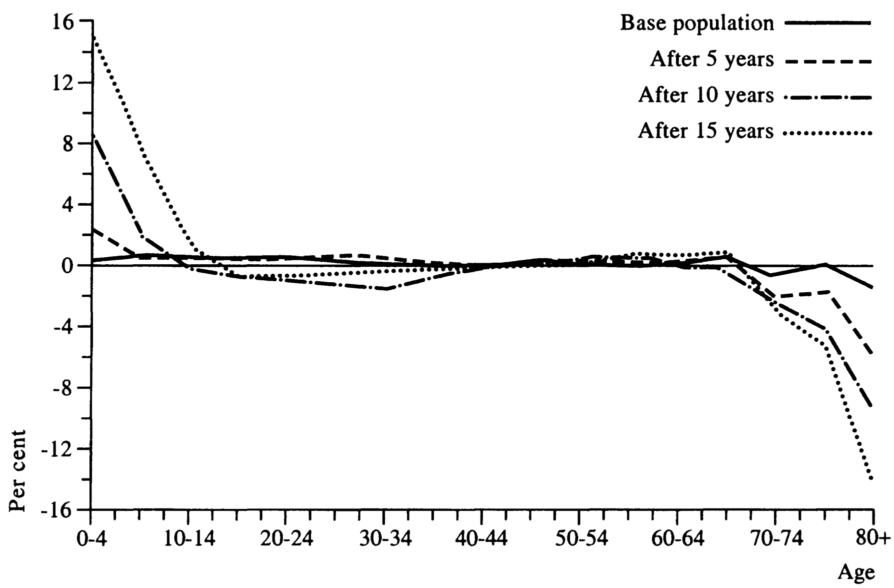


Figure 3. Mean percentage error in projected age structures, projections 1965 to 1990II. Europe.
Sources: Items 6–24 in Table 1.

Immigration flows into Europe observed in the 1970s and 1980s are hardly accounted for in the historical UN projections. This explains the moderate negative errors at ages between 20 and 45 in Figure 3.

Pre-transitional countries display much smaller relative errors in their age structure projections than do post-transitional countries. Figure 4 shows the case of Asia. With only a few exceptions, the errors are between −4 and +2 per cent – note however the age group 80+, where pessimistic mortality assumptions have had a stronger impact. But in contrast to their effects for Europe (Figure 3) and other post-transitional regions, wrong assumptions for fertility have played only a minor role in pre-transitional countries; cf. the small errors for children in Figure 4. Note also the pronounced error pattern in the base population. Age group 15–19 has been overestimated by nearly two per cent. This error propagates through the age

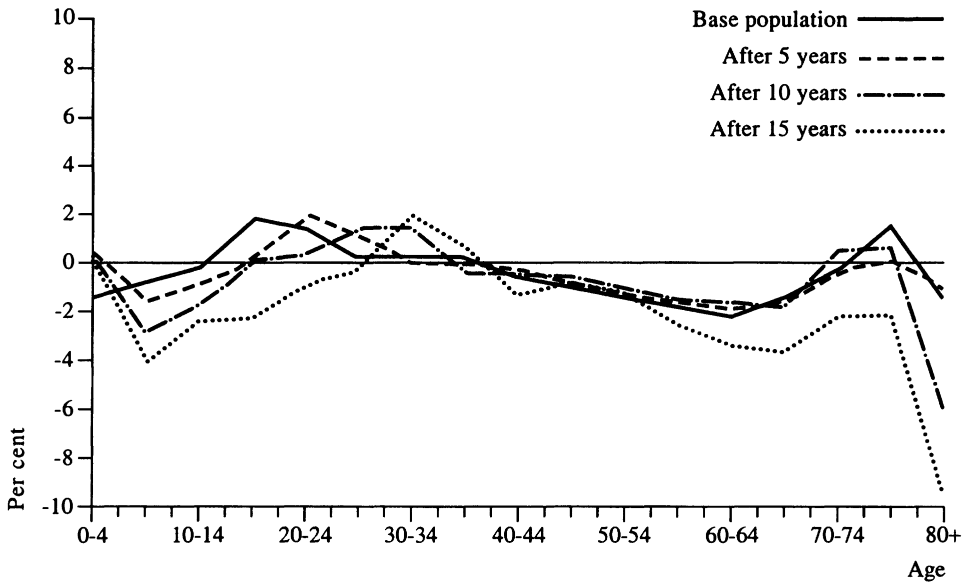


Figure 4. Mean percentage error in projected age structures, projections 1965 to 1990II. Asia.
Sources: Items 6–24 in Table 1.

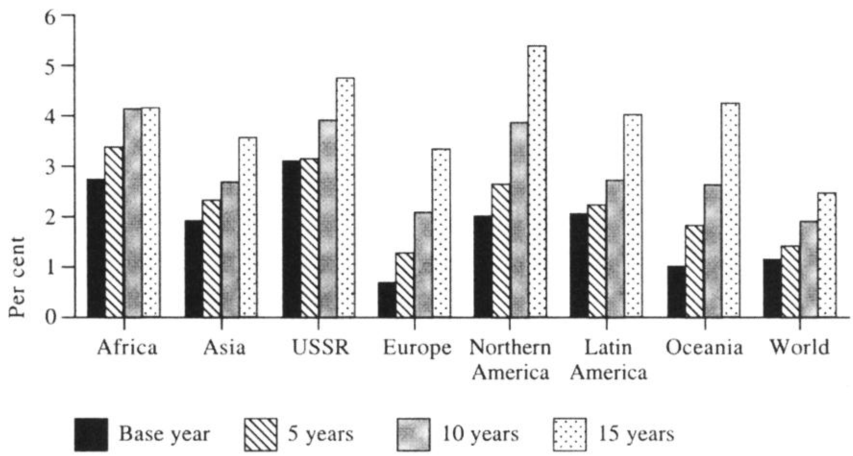
structure when projection duration increases; see age groups 20–24, 25–29, and 30–34 at durations 5, 10, and 15 years.

Figures 3 and 4 are based on the MPE, which reflects the bias in the projections. When imprecision is analysed and the MAPE is plotted, controlling for duration, it is clear that unexpected developments in fertility and mortality had a stronger impact on the accuracy of age structure projections in post-transitional countries than in pre-transitional countries, in spite of the very different levels of fertility and mortality. The shape of the MAPE-curves for Europe and Northern America is that of a rather symmetric bathtub, with large errors at a duration of 15 years both for the young (11–15 per cent for the age group 0–4) and the old (14–20 per cent, age group 80+). For Asia and Africa, I also find a bathtub type of curve, but the errors at extreme ages are less pronounced (8–10 per cent at ages 0–4, and 6–10 per cent for 80+) than those for post-transitional countries.

Graphs for age structure errors based on the MPE (similar to Figures 3 and 4) and the MAPE have been prepared for seven large regions, and for the largest ten countries in the world. Figures 5a and 5b show a summary of the MAPE-based age structure errors. They give the simple average (over all age groups) of the age-specific MAPEs, controlling for duration. These averages are obviously quite crude for regions and countries with a pronounced age pattern in the errors, but they allow us to make at least an approximate comparison across regions and selected countries with regard to the accuracy of the age structure projections.

For the world as a whole (Figure 5a), the age structures for the base populations in projections since 1965 were wrong by slightly more than one per cent. This is caused by revisions of the data for those base years up to 1998. The errors in projected age structures grow regularly with increasing projection duration, to about 2.5 per cent after 15 years. The steeper the increase, the larger the impact of wrong assumptions for fertility and mortality. Wrong assumptions have contributed relatively a lot to errors in Europe, Northern America, and Oceania. On the other hand, errors for the USSR, Africa, Asia, and Latin America are dominated more by base population errors and less by wrong assumptions, compared to the other regions. Note that after 15 years, the accuracy measured this way is lowest for age-structure projections in Northern America. To a large extent this is explained by a large error for the age group 80+ at a duration of 15 years: it turns out that projections since 1965 have been too low by 20 per cent for this age group. The mortality decline for the elderly in Northern America has obviously been difficult to project.

When we compare selected large countries (Figure 5b), the post-transition countries in the group (Japan and the US) show smaller errors than most of the pre-transition countries – yet Indonesia resembles the US and Japan. Note that fewer base years (1975 and later) than in Figure 5a (1965 and later) are included here, which means that the errors for the US in Figure 5b cannot be compared with



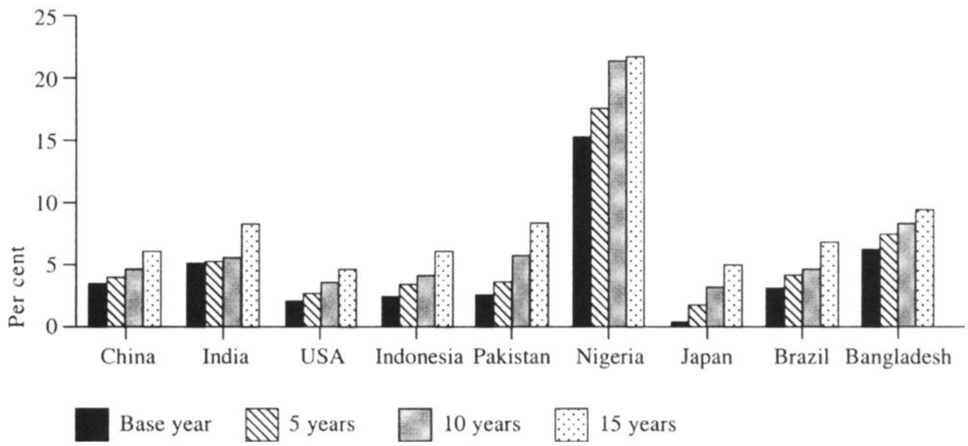


Figure 5a. Average error (MAPE) in age structure projections since 1965, by duration and region, world and large regions.
Sources: Items 6–24 in Table 1.
Figure 5b. Average error (MAPE) in age structure projections since 1975 for some selected countries, by duration.
Sources: Items 8–24 in Table 1.

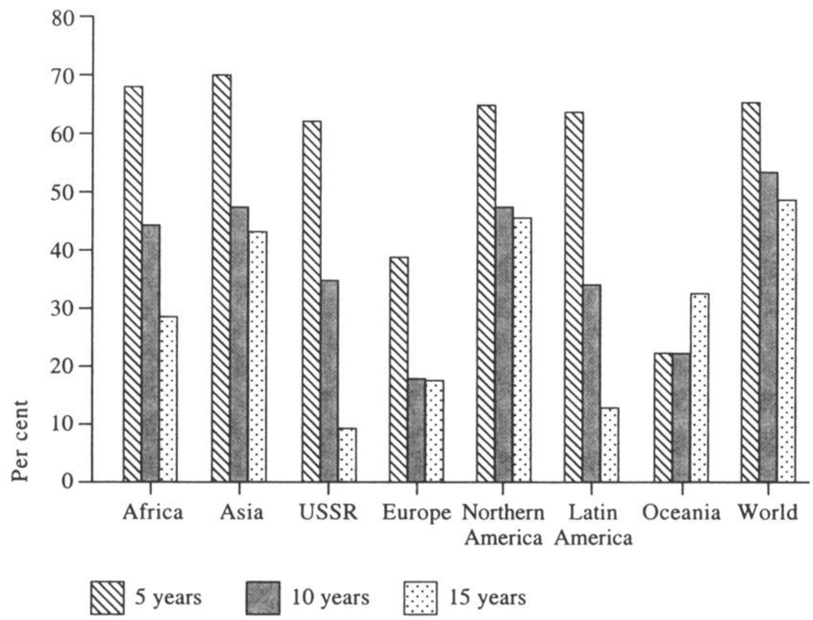
those for Northern America in Figure 5a. The case of Nigeria is remarkable. The November 1991 census resulted in a population size of 88 million, 35 million lower than the 123 million that had been expected. This in turn led to a major revision of historical data for Nigeria, causing errors of 20 to 40 per cent in projected population sizes since 1980.

Using the method described in Section 2.2.1, average errors in age structures presented in Figures 5a and 5b were broken down into two parts. One part had been caused by wrong assumptions about fertility, mortality, and migration for the regions concerned, and the other was a result of the error already present in the base population that had subsequently been propagated through the age structure. In the base year, i.e. at projection duration 0, the whole error is caused by the data quality factor. As projection duration increases, this factor loses importance, and a growing part of the error is caused by wrong assumptions. Figures 6a and 6b present the share of error caused by the data quality factor. The complement of each share is a product of wrong assumptions. After ten years, the data quality factor accounts for roughly one half of the average error in world age structure projections. In Europe and Oceania, the base population data are of good quality, and hence wrong assumptions already account for a large share of the errors after five years.

5. DATA QUALITY AND THE ACCURACY OF FERTILITY AND MORTALITY PROJECTIONS

The accuracy of extrapolated values of the total fertility and the life expectancy at birth has been evaluated for ten projection rounds, starting with the 1965 round. Errors were computed as the simple difference between extrapolated and estimated

values. Since total fertility levels vary strongly between regions, countries, and over time, I also calculated percentage errors for this indicator. For the world as a whole, the extrapolated total fertility was wrong by 7.5 per cent, or 0.2 children per woman on average during the years 1965–95; see Table 3. Total fertility-extrapolations for the world that are too low are very rare (five out of 31 projection errors, not shown here), and therefore the mean absolute error is only slightly higher than the mean error. When successive projection rounds are considered, the projections prepared at the end of the 1970s and the beginning of the 1980s show smaller errors than the most recent ones. Thus there is no consistent trend towards more accurate projections. There is a strong duration effect in the set of total fertility projections, with absolute errors increasing from 0.12 children per woman for total fertility projections 0–5 year ahead, to 0.25 for durations of 10–15 years, and 0.62 children per woman for durations of 20–25 years. When successive projection periods since the early 1980s are



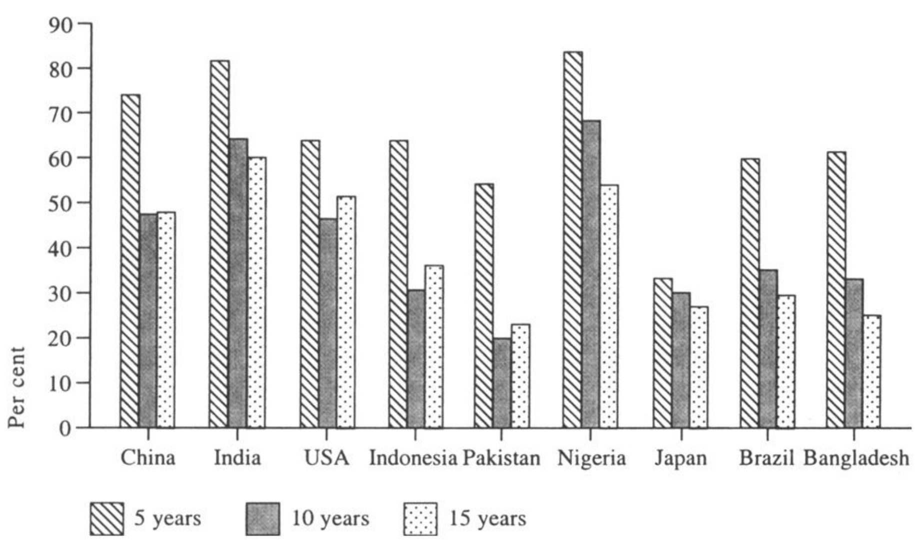


Figure 6a. Average error in age structure projections. Proportion caused by base population error, large regions.
Sources: Items 6–24 in Table 1.
Figure 6b. Average error in age structure projections. Proportion caused by base population error, large regions.
Sources: Items 8–24 in Table 1.

considered, it turns out that projecting world fertility for the period 1965–75 was relatively error-free (errors below 0.16 children per woman, or 3 per cent), compared to later years (between 0.2 and 0.3, or 5–12 per cent). A plausible explanation is the fact that the fall in birth rates for the world as a whole was much steeper after the mid-1970s, than in the years before 1975.

Mean absolute errors in Table 3 show that it has been difficult to make accurate projections of absolute total fertility levels for Asia (for instance, for India, China, Indonesia, Pakistan, and Bangladesh), Latin America (Brazil), and Nigeria. When percentage errors (MAPE) are analysed, Europe, Northern America, and Japan have to be added to this list. Apparently the MAEs in fertility extrapolations tend to be higher in countries with initially higher fertility and initially lower life expectancy. But there are exceptions, cf. the moderate error for Africa as a whole. African fertility was more or less stable at a high level of around 6.5 children per woman until the mid-1980s. Total fertility then fell rapidly, to reach a level of 5.5 children per woman in 1990–95. The average errors in total fertility for the first half of the 1990s are much higher than those for earlier periods. The example of Africa suggests that stability in base series tends to increase accuracy. Indeed, in many cases a lack of stability is caused by a trend shift, which may easily lead to assumption drag and hence inaccurate forecasts, as discussed earlier.

Life expectancy has been difficult to extrapolate in the USSR, Oceania, China, Indonesia, Pakistan, Bangladesh, and Nigeria. Except for the USSR and Nigeria, projection makers have been too pessimistic about future mortality in these countries, and have predicted life expectancy levels that were too low on average – *much* too low in many cases. For the USSR the mean error equals the mean absolute error – the life expectancy extrapolations have consistently been too high since 1965. There has been a clear tendency towards greater accuracy in extrapolated life expectancy at birth for the world as a whole in projections prepared since the early 1980s. The projections with base years 1970–801 had an average absolute error in life expectancy of between 2 and 2.4 years, whereas the later rounds were in error by 1.4 years or less. Accuracy also improves when we consider subsequent five-year periods for which the extrapolations have been made: the MAE dropped from 2.9 years for the years 1965–70 to 1.3 years thirty years later.

Figures 7 and 8 show that the fall in total fertility and the increase in life expectancy for the world as a whole have been predicted consistently, but that

Table 3. *Mean absolute error (MAE) and mean absolute percentage error (MAPE) in extrapolated total fertility and life expectancy at birth by major region and for selected large countries, period 1965–95*[1]

	Total fertility		Life expectancy
	MAE (children/ woman)	MAPE (%)	MAE (years)
World total	0.24	7.5	1.79
Regions:			
Africa	0.24	4.1	1.62
Asia	0.34	10.3	1.14
Europe	0.26	14.7	0.80
USSR[3]	0.18	8.6	2.89
Northern America	0.22	11.3	1.25
Latin America	0.57	16.9	0.81
Oceania	0.24	9.1	2.14
Large countries:			
China	0.31	10.2	2.74
India	0.34	8.0	1.52
USA	0.18	9.5	1.03
Indonesia	0.45	11.5	3.53
Brazil	0.69	22.7	1.59
Pakistan[2]	0.45	7.6	3.30
Japan	0.21	12.8	1.35
Bangladesh[2]	0.79	20.8	2.81
Nigeria	0.47	8.0	2.56

Note 1: Base years 1965–1990II, unless indicated otherwise.
Note 2: Base years 1970–1990II.
Note 3: See Table 2, note 1.
Sources: Items 6–24 in Table 1.

slopes and levels have varied a great deal. Yet the errors in historical values seem to be somewhat smaller than those in extrapolated values. The thick solid unmarked lines represent values taken from the UN-1998 Revision. Time series from a number of projections with base years between 1965 and 1990 are given as marked lines (solid lines for values before the base date, dashed lines for extrapolated values). Figure 7 demonstrates very clearly the assumption drag referred to earlier: the new trend exhibited by total fertility after 1965 was not picked up by projection makers until some ten years later: compare the extrapolations in the 1965 and 1970 projections with those in the 1975II-rounds and later. Assumption drag is weaker in Figure 8, cf. the low life expectancy extrapolations in the projections of 1965 and 1975II.

Imagine for a moment that the projection makers in the past had the data from the 1998 Revision at their disposal. Would the errors in total fertility and life expectancy have been smaller in that case? Only a partial answer can be given. Given the

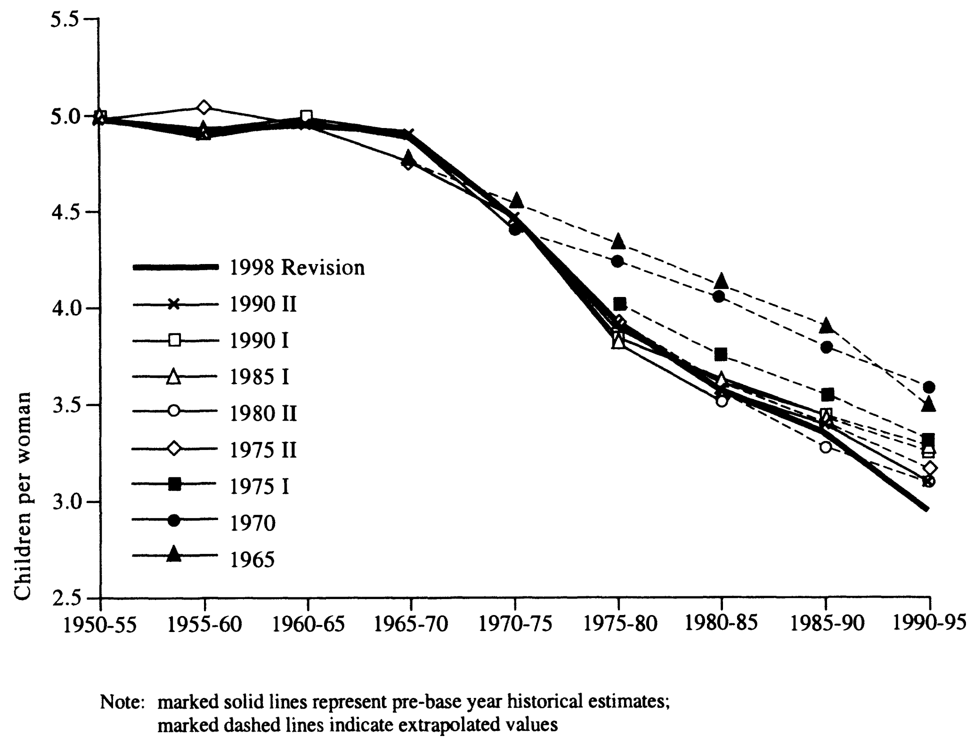


Figure 7. Total fertility, world.
Sources: Items 6–24 in Table 1.

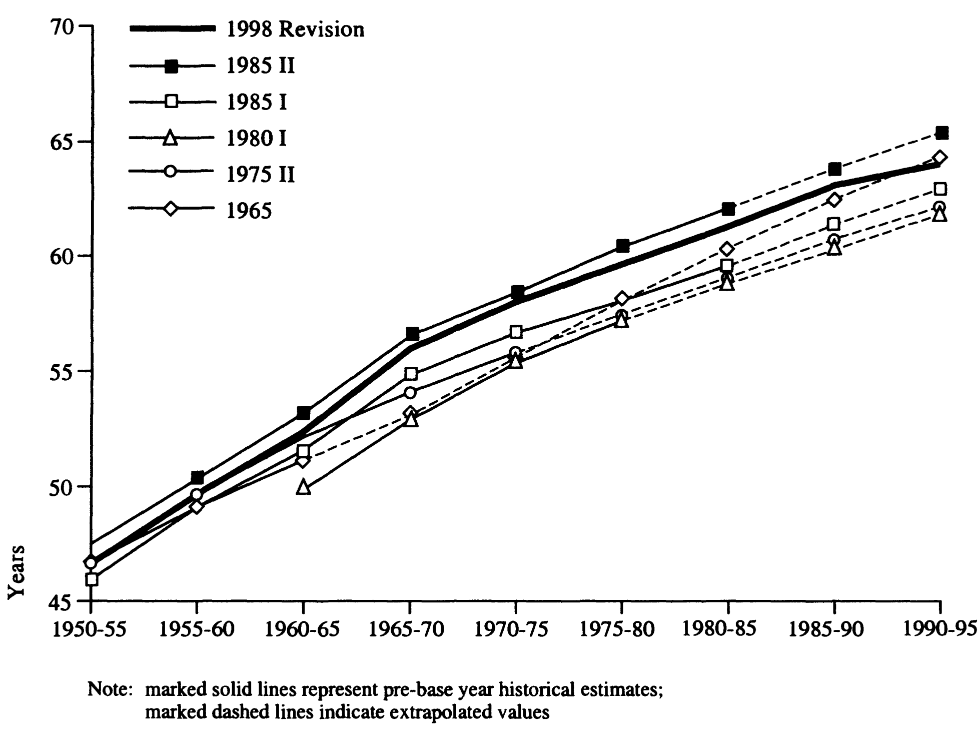


Figure 8. Life expectancy at birth, world.
Sources: Items 6–24 in Table 1.

benchmark, the errors in extrapolated values are determined by the starting point and the slope (or change in slopes, if non-linear trends are assumed) of the extrapolations. By a parallel shift of the extrapolated curves so that each curve matches the benchmark values in the five-year period before the base year, one may correct for the error caused by the starting point, but not for that caused by the slope. When I carried out such a shift for the total fertility – and the life expectancy – curves of Figures 7 and 8, the errors diminished considerably. The mean error in total fertility fell from 0.25 to 0.14 children per woman, and the mean absolute error from 0.27 to 0.14 children per woman. For life expectancy, the reduction was from –1.56 to –0.05 years for the ME, and from 1.79 to 0.86 years for the MAE. This simple descriptive statistical finding suggests that a considerable part of projection errors in total fertility and life expectancy may be explained by the quality of the data. A more formal analysis follows below.

Table 4 shows summary statistics on data quality by region and country. Using the method of Section 2.2.2, standard deviations were computed for historical estimates of total fertility and life expectancy for the following periods: 1950–55, 1955–60, 1960–65 (all three for historical estimates worked out in connection with the projections of 1965 and later), 1965–70 (projection 1970 and later), 1970–75 (projection 1975I and later), 1975–80 (projection 1980II and later), 1980–85 (projection 1985I and later), 1985–90 (projection 1990I and later), and 1990–95 (projections 1995I and 1995II). The columns show the value of the simple average of the nine period-specific standard deviations for each country/region. A large value of this average indicates large adjustments over subsequent projection rounds, controlling for duration, and thus poor data quality. In other words, the indicator reflects the average spread in the bundle of time series similar to the one in Figure 7 or 8 (excluding extrapolated values). I experimented also with alternative specifications of the data quality indicator. The *overall* standard deviation was computed for all historical estimates together. It combines data quality and time trend in the historical estimates: the steeper the trend or the stronger the adjustments or both in combination, the larger the overall standard deviation. The *weighted* average standard deviation took the number of estimates in each period into account: twelve for each of the periods 1950–55 and 1960–65; ten for 1955–60 and 1970–75; eleven for 1965–70; eight for 1975–80; six for 1980–85; four for 1985–90; and two for 1990–95. The *relative* standard deviation (the simple, the weighted, and the overall one) was obtained by dividing the standard deviation $s(t)$ by the corresponding mean value $m(t)$. These three alternative specifications resulted in conclusions about the link between data quality and projection accuracy similar to those reported here, although the link between data quality and overall standard deviation was generally much weaker than that with the other two standard deviations.

Table 4. *Average standard deviation (s.d.) for historical total fertility and life expectancy at birth, major regions and selected large countries, period 1965–95*

	s.d. for total fertility (children/woman)	s.d. for life expectancy (years)
World total	0.03	0.70
Regions		
Africa	0.09	0.37
Asia	0.05	0.71
Europe	0.03	0.31
USSR[1]	0.05	0.66
Northern America	0.03	0.06
Latin America	0.04	0.33
Oceania	0.03	0.57
Large countries		
China	0.14	1.58
India	0.19	0.51
USA	0.03	0.04
Indonesia	0.12	0.79
Brazil	0.06	0.59
Pakistan	0.19	1.18
Japan	0.02	0.09
Bangladesh	0.12	0.97
Nigeria	0.29	1.24
Correlation with MAE	0.36	0.74*
Correlation with MAE 0–4 yrs	0.74*	0.92*

*: Significantly larger than zero at the five-per cent level; Fisher's z-test (one-sided).
Note 1: See Table 1, Note 1.
Sources: Items 6–24 in Table 1.

Table 4 reflects the well-known fact that birth data are of poor quality in Africa and in Asia. See also the results for Nigeria, China, Pakistan, and Bangladesh. The high values for China and Bangladesh are caused by poor data quality in the period before 1975: recent values are much lower than reported in Table 4. Pakistan had problems with its data during the years 1970–85, whereas the high value for Nigeria is to a large extent explained by data adjustments in the period 1990–95. The three Asian countries just mentioned also have problems with the quality of their mortality data. For the regions, the African mortality data are of average quality, comparable to those for Latin

America and the world as a whole. The USSR data are of poor quality.

A simple correlation between mean absolute errors in Table 3 and the quality indicators in Table 4 shows no relationship for total fertility, and a rather strong one for life expectancy. However, when the MAEs are restricted to the first five years of the projection period, the relationship with data quality indicators in Table 4 becomes much stronger, both for the fertility data (r=0.74) and the mortality data (r=0.92). (For total fertility, correlations between the standard deviation and the *MAPE* (rather than the *MAE*) were non-significant (even negative, –0.23 for all durations) or rather weak (0.43 for errors in projections 0–4 years ahead).) This again demonstrates that the data quality effect in the error declines for later durations, when the impact of wrong fertility and mortality extrapolations becomes more important.

6. CONCLUSIONS

This paper presents descriptive findings for the accuracy of historical UN projections, and the quality of the UN data in large countries and major regions. The paper has addressed three broad questions:

- Do data quality and projection accuracy differ between regions and large countries?
- Did data quality and accuracy improve over time?
- Does poor quality go together with poor accuracy?

There is considerable variation among large countries and major regions in historical projection accuracy and data quality. Not surprisingly, problems are largest in pre-transition countries, and especially in Asia. The quality of UN data for total fertility and life expectancy has been problematic in the past for China, Pakistan and Bangladesh. The poor data quality for these countries went together with large errors in projected total fertility and life expectancy. For Africa as a whole, data on total population and age structure have been revised substantially in the past, and this is a plausible reason for the poor performance of the projected population. Nigeria, which is the only African country in this analysis, underwent major revisions in its data in connection with the Census of 1991. In turn, historical estimates of fertility and mortality indicators had to be adjusted, and this explains large projection errors in age structure, total fertility and the life expectancy for this country. The problematic data situation for the former USSR is well known, in particular that for mortality data. Mortality trends in the former Soviet Union indicated significant health problems before its dissolution. Between 1970 and 1985, life expectancy stagnated, and, for males, even declined. These trends came unexpectedly, and the earlier levels were not well enough reflected in the available data. The result was that, on average, life expectancy projections were too high by 2.9 years, which in turn caused large errors in projected age structures for the elderly. For Europe and Northern America, data quality is generally good. Yet the two regions have large errors in long-range projections of their age structures, caused by unforeseen trend shifts in fertility and mortality in the 1960s and 1970s. The fertility decline in Latin America has also been underestimated, and this has led to moderate errors in projected age structures.

With few exceptions, there is a clear improvement over time in the quality of total fertility and life expectancy data for the regions and countries that I considered. A temporary worsening of the data quality, presumably in connection with population censuses held in the 1970s and 1980s, is clearly visible for Bangladesh, Indonesia and the United States. But regarding projection accuracy, there is no general trend towards smaller errors for recent projections. In many cases, the projections made at the end of the 1970s or the early 1980s are more accurate than the other ones, or subsequent projections show a more or less stable mean error. In other cases, the trend is slightly downwards, possibly because the errors for the more recent projections were caused by short projection durations.

Finally, and not surprisingly, I found that poor data quality for total fertility and life expectancy tends to go together with poor projection performance. The relationship is stronger for mortality than for fertility, and stronger for short-term than for long-term extrapolations. Between 50 and 80 per cent of the variation (across regions and countries) in the mean error five years ahead for the two indicators could be explained by the instability in the time series of historical estimates. For life expectancy, the relationship between data quality and the error for the first five projection years is close to a straight line. This means, for example, that if the quality of mortality data for the former Soviet Union could become twice as good as it was in the past (as measured by the quality indicator used in this paper), one could expect a reduction in the short-term error in life expectancy projections from 1.9 to 0.7 years. This level is close to the

current errors for Europe (0.8) and Latin America (0.6).

It will be clear that data quality is not the only factor that accounts for forecast accuracy. Projection length, population size, unexpected developments, and assumption drag contribute too. For instance, the HIV/AIDS epidemic has led to a stagnation of African life expectancy at a level around 51 years. Past projections extrapolated the increases in life expectancy observed until 1985–89, and this has caused large errors for the period 1990–94. Future projections take HIV/AIDS into account, and this will probably reduce life expectancy errors.

NOTES

Nico Keilman is at the Department of Economics, University of Oslo, Norway. The author gratefully acknowledges financial support from the US National Research Council, expert research assistance by Svenn-Erik Mamelund, and comments by Øystein Kravdal, Randy Bulatao, and participants in the 34th meeting of the German Association for Population Science (Bochum 5–7 April 2000).

[1] Changes of country borders since 1950 were dealt with in the following way. *India*: The 1960 and 1970 projections include Sikkim. *Indonesia*: Since 1963, the population of Indonesia includes that of West Irian (former West New Guinea). The population of West Irian has been added to the Indonesian population for the 1950III projection. *Bangladesh and Pakistan*: The first projection with figures for Bangladesh is the 1970 projection. The 1950II, the 1950II, and the 1960 and 1965 projections for Pakistan include Bangladesh. Compared to estimated population numbers for the period 1950–65 of the current Pakistan, the projections until 1965 give population numbers that are too high for Pakistan. Therefore, Bangladesh and Pakistan were only considered from the 1970 projection and onwards. *Japan*: Starting with the 1970 projection, the population of Japan includes that of Ryukyu Islands. The population of Ryukyu Islands has been added to the Japanese population for the projections 1950III, 1960 and 1965. *USSR*: The USSR has been maintained as a regional entry. For the 1990I projections, the populations of the Baltic States have been subtracted from that of Europe, and were added to that of the former USSR. For the projections since 1990II, the 15 states of the former USSR (Asiatic and European USSR) have been aggregated and added to the Russian Federation. At the same time, the population of the European USSR was subtracted from that of Europe, and Asiatic USSR was subtracted from Asia.

REFERENCES

Ascher, Willam. 1978. *Forecasting: An Appraisal for Policy-Makers and Planners*. Baltimore and London: The Johns Hopkins University Press.

El-Badry, M.A., and S. Kono. 1986. "Demographic estimates and projections", *Population Bulletin of the United Nations* **19–20**: pp. 35–44.

Frejka, Tomas. 1994. "Long-range global population projections: Lessons learned", in W. Lutz (ed.), *The Future Population of the World: What Can we Assume Today?* London: Earthscan Publications Ltd. Pp. 3–15.

Inoue, Shunichi, and Y.C. Yu. 1979. United Nations New Population Projections and Analysis of Ex Post Facto Errors. Paper presented at the Annual Meeting of the Population Association of America, Philadelphia, April 1979.

Keilman, Nico. 1997. "Ex-post errors in official population forecasts in industrialized countries", *Journal of Official Statistics* **13**: pp. 245–277.

Keilman, Nico. 1998. "How accurate are the United Nations' world population projections?", *Population and Development Review* Supplement **24**: pp. 15–41.

Keyfitz, Nathan. 1981. "The limits of population forecasting", *Population and Development Review* **7**: pp. 579–593.

Keyfitz, Nathan. 1982. "Can knowledge improve forecasts?", *Population and Development Review* **8**: pp. 729–751.

National Research Council. 2000. John Bongaarts and Rodolfo Bulatao (eds.), *Beyond Six Billion: Forecasting the World's Population.* Panel on Population Projections. Committee on Population, Commission on Behavioral and Social Sciences and Education. Washington DC: National Academy Press.

Pflaumer, Peter. 1988. "The accuracy of U.N. population projections", in *Proceedings Annual Meeting American Statistical Association,* New Orleans, August 1988, Social Statistics Section, pp. 299–304.

Stoto, Michael. 1983. "The accuracy of population projections", *Journal of the American Statistical Association* **78**: pp. 13–20.

Zlotnik, Hania. 1999. World Population Prospects: The 1998 Revision. Paper presented at the joint ECE-Eurostat Work Session on Demographic Projections, Perugia, Italy, 3–7 May.