VI. METHODOLOGY OF THE UNITED NATIONS POPULATION ESTIMATES AND PROJECTIONS

The preparation of each new *Revision* of the official population estimates and projections of the United Nations is a complex process involving several steps.

First, based on new empirical evidence and informed by recent findings in demography, sociology, political sciences and economics, specific guidelines about dominant future trends of fertility, mortality and migration are formulated. These guidelines ensure consistency between and comparability of the estimated and projected demographic trends for the world's countries. These guidelines are flexible and allow for the incorporation of country-specific trends that deviate from assumed general trends and they take into account a countries current position in the demographic transition from high to low fertility and mortality. Specific guidelines for fertility, mortality and migration are described in the main body of this chapter.

The population estimates and projections of this *Revision* cover 100 years, from mid-year 1950 to mid-year 2050. The 100 years covered in this *Revision* are subdivided into past estimates (1950-2005) and projections (2005-2050). Past estimates of demographic variables are either directly taken from national statistical sources, or estimated by the Population Division based on the best available national or international estimates, and if needed, adjusted for deficiencies such as age misreporting, underenumeration of populations or underreporting of vital events, such as birth and deaths. Adjustments are also made to international migration. The year 2005, separating the past estimates from the projections, is called the base year of the projections (sometimes the terms start year or jump-off year can be found). The projection period of this *Revision* covers 45 years and ends in 2050. Usually, population projections prepared by the United Nations Population Division are carried out for a number of variants and scenarios, all based on the medium variant.

The preparation of each new *Revision* of the official estimates and projections of the United Nations involves two distinct processes: (*a*) the incorporation of all new and relevant information regarding the past dynamics of the population of each country of the world, and (*b*) the formulation of detailed assumptions about the future paths of fertility, mortality and international migration. In order to ensure consistency and comparability, certain steps must be taken. New information is evaluated to determine recent changes in population dynamics that have an impact on the age and sex structure of the population at the base year (the year when the projection starts). The methods used to carry out this evaluation must be technically appropriate and consistently applied. With respect to the projection period, general guidelines are established regarding the paths that fertility, mortality and international migration are to follow in the future. However country-specific deviations from the general guidelines are common under two sets of circumstances: (*a*) when recent trends suggest that the population of a country is not yet ready to embark on the path determined by the general guidelines, a transition period between current dynamics and those formulated in the general guidelines has to be introduced, and (*b*) when the populations of certain countries are likely to experience a long-term deviation from the paths set by the guidelines, as is the case for countries where the prevalence of HIV/AIDS is high.

This chapter first describes the way past estimates were revised during the preparation of the *2004 Revision* of the United Nations population projections. It then examines the assumptions made and approaches used for projecting fertility, mortality and international migration up the year 2050. Finally, procedures and assumptions for population projections are described. This chapter also provides the technical documentation of models used for this *Revision* in an annex.

A. PAST ESTIMATES OF POPULATION DYNAMICS

One of the major tasks in revising the estimates and projections of the population of each country of the world is to obtain and evaluate the most recent information available on each of the three major components of population change: fertility, mortality and international migration. In addition, newly available census information or other data providing information on the age distribution of the population are also carefully evaluated. The process of updating and revising past population estimates usually entails not only the separate evaluation of the quality of the different estimates available but also and more importantly the search for consistency among them. The key task is therefore to ensure that for each country past trends of fertility, mortality and international migration are consistent with changes in the size of the population and its distribution by age and sex. In very general terms, for most countries in the more developed regions, the availability of detailed information on fertility and mortality trends over time and of periodic censuses of the population greatly facilitates the task of producing reliable estimates of past population dynamics. Yet, even for those countries, the data on international migration flows are generally inadequate. Consequently, consistency between population counts and the components of population change is often achieved by assigning to net international migration the residual estimate obtained by comparing actual intercensal population growth with independent estimates of natural increase.

For many countries in the less developed regions, the estimation of past trends is more complex. In these countries, information may be limited or lacking and the available data may be unreliable. In numerous cases, therefore, consistency can only be achieved by making use of models in conjunction with methods of indirect estimation (United Nations, 1983, 2002). In extreme cases, when countries have no data referring to the past decade or two, estimates are derived by inferring levels and trends from those experienced by countries in the same region that have a socio-economic profile similar to the country in question. However, since the 1970s the emphasis put on surveys and census-taking in the developing countries has considerably improved the availability of demographic information.

At the global level, data from censuses or reliable official estimates based on censuses, population registers and surveys referring to 1995 or later were available for 181 countries, 79 per cent of the 228 countries or areas for which projections are carried out. For 32 countries, data were available from the 1985-1994 global censuses round, for 13 countries, census information or estimates with comparable quality were available from the global 1975-1984 census round, and for 2 countries, data were available only for years before 1975.

Aside from relying on census information concerning the distribution of the population by age and sex, in order to estimate the population as of 1 July 2005, the base year, trends in fertility, mortality and international migration up to this date must be established. Ideally, complete time series of annual age-specific fertility rates, life tables and age-specific net international migration rates by sex would be needed. In practice, the information available is considerably less comprehensive, consisting often of no more than the average age-specific fertility rates experienced by women over one or two periods of various lengths, estimates of infant or child mortality at several points in time and, less commonly, one or two life tables for different periods. For developing countries, estimates of recent fertility and child mortality are often derived from surveys, especially when countries lack a civil registration system or have one that does not have sufficient coverage of all vital events. When countries have a reliable civil registration system, as is the case in most developed countries and in some developing countries, data on both fertility and mortality by age and sex are theoretically available on a continuous basis. However, owing to either delays in processing the data or the difficulty of estimating appropriate denominators to calculate age-specific fertility or mortality rates, fertility schedules and life tables may only be available for selected years. In preparing the revised estimates of the base-year population, such information has to be taken into account together with trends in other indicators, such as changes in the overall number of births.

To provide some assessment of the timeliness of the information on which the *2004 Revision* is based, tables VI.1a, b and c present the distribution of countries by region and time period, indicating the most recent information used for estimating fertility, child mortality and adult mortality. Such information pertains to a total of 192 countries, each of which was estimated to have a population of 100,000 inhabitants or more in 2000 and for which projections were prepared using the cohort-component method. For the other 36 countries considered, whose populations were below 100,000 in 2000, simplified projections were made. As a consequence, only total population figures and growth rates are available for these countries.

As table VI.1a indicates, this *Revision* incorporates relatively recent information on fertility for most countries: out of the 192 countries considered, 178 had information referring to 1995 or later. Only one country had no information whatsoever on fertility. All countries in Europe, Northern America and Oceania had information referring to 1995 or later, whereas in the developing regions the percentage of countries with recent information was lower, varying from a high of 96 per cent in Asia, to 94 per cent in Latin America and the Caribbean, and to a low of 81 per cent in Africa.

Child mortality, e.g. the probability to survive from birth to age five, is a lead indicator to asses

TABLE VI.1a. DISTRIBUTION OF COUNTRIES AND THE POPULATION ACCORDING TO THE MOST RECENT DATA USED FOR THE ESTIMATIONS OF FERTILITY

Topic and reference date	*Africa*	*Asia*	*Europe and Northern America*	*Latin America and the Caribbean*	*Oceania*	*Total*
			Number of countries			
No Information........	1	—	—	—	—	1
Before 1985	3	—	—	—	—	3
1985-1989...............	2	1	—	—	—	3
1990-1994...............	4	1	—	2	—	7
1995-1999...............	18	10	1	12	4	45
2000 or later............	26	38	40	21	8	133
TOTAL	54	50	41	35	12	192
			Population (millions)			
No Information........	—	—	—	—	—	—
Before 1985	62	—	—	—	—	62
1985-1989...............	11	29	—	—	—	40
1990-1994...............	40	22	—	1	—	63
1995-1999...............	243	1 293	—	126	2	1 664
2000 or later............	550	2 561	1 058	434	31	4 635
TOTAL	906	3 905	1 059	561	33	6 463
			Percentage of the population			
No Information........	0.0	—	—	—	—	0.0
Before 1985	6.9	—	—	—	—	1.0
1985-1989...............	1.2	0.7	—	—	—	0.6
1990-1994...............	4.4	0.6	—	0.2	—	1.0
1995-1999...............	26.9	33.1	0.0	22.4	5.0	25.7
2000 or later............	60.7	65.6	100.0	77.4	95.0	71.7
TOTAL	100.0	100.0	100.0	100.0	100.0	100.0

the welfare of children. The availability of information on it was also high (table VI.1b), with 186 countries having information referring to 1995 or later, including all of the developed countries. At the regional level, only one country in Africa and one in Latin America and the Caribbean had no information on childhood mortality while three African countries and one Asian country had data referring to dates prior to 1995. This implies that for 99.9 per cent of the world population information from 1995 and later was used in the estimation of child mortality.

In contrast with the availability of information on fertility and on mortality in childhood, information on adult mortality was sparse and often outdated (table VI.1c). Data on adult mortality referring to 1995 or later was used in little more than a third of all countries considered (37 per cent) and no empirical data on age specific mortality was available for the estimation of adult mortality in 41 per cent of the countries. Information was especially lacking or of insufficient quality among the countries of Africa (91 per cent) and Asia (36 per cent). However, it is important to under-

TABLE VI.1b. DISTRIBUTION OF COUNTRIES AND THE POPULATION ACCORDING TO THE MOST RECENT DATA USED FOR THE ESTIMATION OF CHILD MORTALITY

Topic and reference date	*Africa*	*Asia*	*Europe and Northern America*	*Latin America and the Caribbean*	*Oceania*	*Total*
			Number of countries			
No Information	1	—	—	—	—	1
Before 1985	2	—	—	—	—	2
1985-1989	1	—	—	—	—	1
1990-1994	0	1	—	1	—	2
1995-1999	14	14	2	12	5	47
2000 or later	36	35	39	22	7	139
TOTAL	54	50	41	35	12	192
			Population (millions)			
No Information	—	—	—	—	—	—
Before 1985	5	—	—	—	—	5
1985-1989	3	—	—	—	—	3
1990-1994	0	1	—	—	—	1
1995-1999	224	1 461	43	98	2	1 828
2000 or later	674	2 443	1 015	463	31	4 626
TOTAL	906	3 905	1 059	561	33	6 463
			Percentage of the population			
No Information	0.0	—	—	—	—	0.0
Before 1985	0.5	—	—	—	—	0.1
1985-1989	0.4	—	—	—	—	0.1
1990-1994	0.0	0.0	—	0.0	—	0.0
1995-1999	24.7	37.4	4.1	17.5	5.6	28.3
2000 or later	74.4	62.6	95.9	82.5	94.4	71.6
TOTAL	100.0	100.0	100.0	100.0	100.0	100.0

line that life expectancy estimates are often derived from more recent information than what is documented in table VI.1c with respect to adult mortality. In particular, life expectancy at birth was often arrived at by using recent information of infant and child mortality and appropriate model life tables. In addition, official estimates of adult mortality were sometimes considered, but not necessarily used because their quality was not adequate or due to methodological constraints (e.g. implementation of the AIDS methodology and actual base year of the projection). Thus, table VI.1c reflects the data that was used in this *Revision*, not the availability of the data.

It is important to consider the implications of data availability for the quality of the population estimates and projections made. One way of assessing the probable overall impact of the uncertainty involved in making estimates on the basis of non-existent or outdated information is to calculate the proportion of the population to which the less reliable or outdated estimates refer. With regard to information on child mortality, the total population of countries lacking recent information was, with fewer than 10 million people, very small. As for fertility, the population of countries that either lacked data entirely or whose most recently used estimates referred to periods before

TABLE VI.1c. DISTRIBUTION OF COUNTRIES AND THE POPULATION ACCORDING TO THE MOST RECENT DATA USED FOR THE ESTIMATION OF ADULT MORTALITY

Topic and reference date	*Africa*	*Asia*	*Europe and Northern America*	*Latin America and the Caribbean*	*Oceania*	*Total*
Number of countries						
No Information	49	18	—	7	4	78
Before 1985	1	1	—	3	1	6
1985-1989	1	3	2	2	0	8
1990-1994	1	12	4	11	1	29
1995-1999	1	6	24	2	5	38
2000 or later	1	10	11	10	1	33
TOTAL	54	50	41	35	12	192
Population (millions)						
No Information	854	537	—	40	1	1 432
Before 1985	1	30	—	10	—	41
1985-1989	33	88	7	7	—	135
1990-1994	9	1 495	322	212	—	2 039
1995-1999	8	1 301	558	9	27	1 903
2000 or later	1	455	171	282	4	914
TOTAL	906	3 905	1 059	561	33	6 463
Percentage of the population						
No Information	94.3	13.7	—	7.2	1.9	22.2
Before 1985	0.1	0.8	—	1.7	1.5	0.6
1985-1989	3.6	2.3	0.7	1.3	—	2.1
1990-1994	1.0	38.3	30.4	37.8	0.5	31.5
1995-1999	0.8	33.3	52.7	1.6	83.8	29.4
2000 or later	0.1	11.7	16.2	50.3	12.3	14.1
TOTAL	100.0	100.0	100.0	100.0	100.0	100.0

1995 amounted to 3 per cent of the world population. However, the proportion of the population lacking equally recent estimates of adult mortality amounted to 56 per cent. Therefore, the most serious weakness faced in producing the 2005 base year estimates of the population of each country was the lack of recent information on mortality and especially on adult mortality. Owing to the wide availability of data from recent census enumerations, the lack of recent mortality data would be less critical if those enumerations were accurate. However, considerable evidence exists that coverage errors in census enumerations are not necessarily small. Furthermore, as an increasing number of countries face sharp rises in mortality brought about by war, civil strife, major social and economic changes, or the HIV/AIDS epidemic, direct information on those trends is necessary to derive appropriate estimates of their impact on the population's age and sex structure. In the absence of direct and reliable information on the timing and magnitudes of such 'shocks' to mortality, estimations and approximations are necessarily less accurate.

A final consideration in the *Revision* of past estimates of population dynamics concerns the sources of information regarding international migration. In preparing this *Revision* particular attention was given to official estimates of net international migration or its components (immigration and emigration), to information on labour migration or on international migration flows recorded by receiving countries, to estimates of undocumented or irregular migration by origin, and to data about refugee flows and stocks prepared by the Office of the United Nations High Commissioner for Refugees. Even by combining these numerous data sources, it is difficult to produce comprehensive and consistent data of net migration over time. In those cases net international migration was estimated as the residual not accounted for by natural increase between two successive enumerations of the population. Clearly, therefore, the paucity of reliable and comprehensive data on international migration should also be singled out as one of the limitations in producing more accurate estimates of the population for the base year.

B. THE PROJECTION OF FERTILITY

This section provides a detailed account of how future levels and age patterns of fertility were projected for countries in different groups of current fertility levels. In the discussion that follows, assumptions and methodology will be described in terms of the following groups of countries:

1. High-and medium fertility countries. High fertility countries are countries that until 2005 had had no fertility reduction or only an incipient decline, and medium-fertility countries are those where fertility has been declining but whose level was still above 2.1 children per woman in 2000-2005;
2. Low-fertility countries: Countries with total fertility at or below 2.1 children per woman in 2000-2005.

The projection of fertility in this *Revision* follows the methodology that was introduced in the *2002 Revision*, with only minor changes. In this *Revision*, as in the previous one, it is assumed that countries in the transition from high to low fertility will ultimately approach a fertility floor of 1.85 children per woman, regardless of their current position in the fertility transition. The transition from the current level of fertility to the fertility floor is expressed by models of fertility change over time. These models have been formalized for this *Revision*; they are described below and documented in the annex of this chapter. The assumption for countries currently below replacement level have been slightly altered, namely it is now assumed that the fertility recovery will follow a uniform pace. As a consequence, individual countries will reach the fertility floor at different years in the future and not, as in the previous *Revision*, between 2045 and 2050.

1. The high- and medium-fertility countries

The projection of fertility for the high-fertility and medium-fertility countries is carried out through a unified model. First introduced in the *2002 Revision* (United Nations, 2004a, pp. 183-184), it is formulated in terms of the pace of fertility decline. An important feature of the fertility model is its assumption that fertility in all countries will eventually fall below replacement level, but not necessarily by the end of the projection period in 2050. In light of evidence that fertility in a growing number of countries in less-developed regions has already dropped below replacement level, or is rapidly approaching it, this *Revision* keeps the assumption of a fertility "floor" of 1.85 children per woman.

a. The pace of fertility decline

For all countries where fertility was above the replacement level in 2000-2005, fertility was projected in this *Revision* using a model based upon the combined experience of all countries that underwent fertility decline between 1950 and 2000.

The model relates changes in fertility over a specified period to the level of fertility at the beginning of that period. When examining the empirical evidence, a particular pattern emerges: decline in fertility is relatively slow when the fertility transition is starting but accelerates until fertility is between 4 and 5 children per woman. The pace of decline then decreases near the end of the transition. Further analysis of empirical pathways of fertility decline suggested some variations of the general pattern, associated with the pace or speed of fertility decline at the beginning and at the end of the fertility transition.

Three models have been identified to capture best the variety of pathways from high to low fertility. One model represents pathways with a slower decline at both high and low fertility, labelled Slow/Slow. A second model exhibits faster decline at both high and low fertility levels and is labelled Fast/Fast. Empirical evidence suggests a third model that combines a fast decline at high levels of fertility and slow decline as lower fertility levels are approached. The latter model is labelled Fast/Slow. The three models of fertility decline have been implemented as logistic functions and are documented in the annex to this chapter.

The models are expressed as fertility decline in the current year given a certain level of fertility in the previous year. For most countries, the new model was used to project fertility beginning in 2005, based on the estimated levels of total fertility in 2000-2005. However, in the high-fertility countries where there has been no evidence of fertility decline to date, it was assumed that fertility would remain constant until 2010 and begin to fall according to the model after that year.

Figures VI.1a and 1b illustrate the different trajectories of the three models for different base levels of fertility, here in 2000. A high-fertility country with a total fertility of 8 children per woman in 2000 (figure VI.1a) would reach the fertility floor of 1.85 children per woman with the Fast/Fast model not before the year 2083. Assuming the Fast/Slow and the Slow/Slow model, the fertility floor would be reached after 2100, e.g. after a period of more than 100 years. Figure VI.1a also shows that the Fast/Fast model and the Fast/Slow model show different trends only after fertility declined to about 3 children per woman.

Figure VI.1b shows trajectories of fertility decline for a country with a total fertility of 4 children per woman in 2000. It will reach the fertility floor after 39 year (Fast/Fast model), 61 years (Fast/Slow model) and 62 years (Slow/Slow model). The figure also shows that for countries with medium and low levels of fertility, the Fast/Slow and the Slow/Slow model converge.

b. Age pattern of fertility

For both the high-fertility and medium fertility countries the age pattern of fertility was projected by interpolating linearly between a starting proportionate age pattern of fertility and a target model pattern. The target pattern is usually attained in either 2045-2050 or in the period when the country reaches its lowest fertility level. Several model patterns of fertility, shown in the annex to this chapter, are available (table VI.3, VI.4, VI.5). In certain cases, the proportionate age pattern of fertility was held constant for the projection period.

2. Low-fertility countries

a. The pace of fertility recovery

Low-fertility countries are those where the total fertility was 2.1 or below in 2000-2005. For those

Figure VI.1a. High fertility: Models of fertility decline

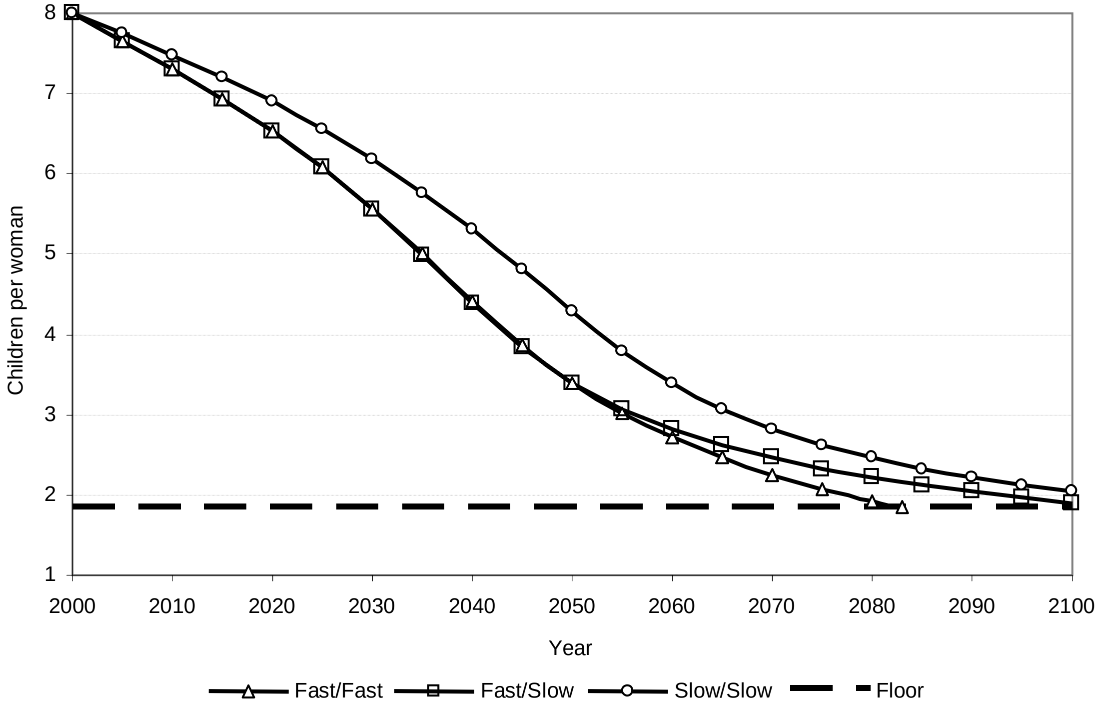
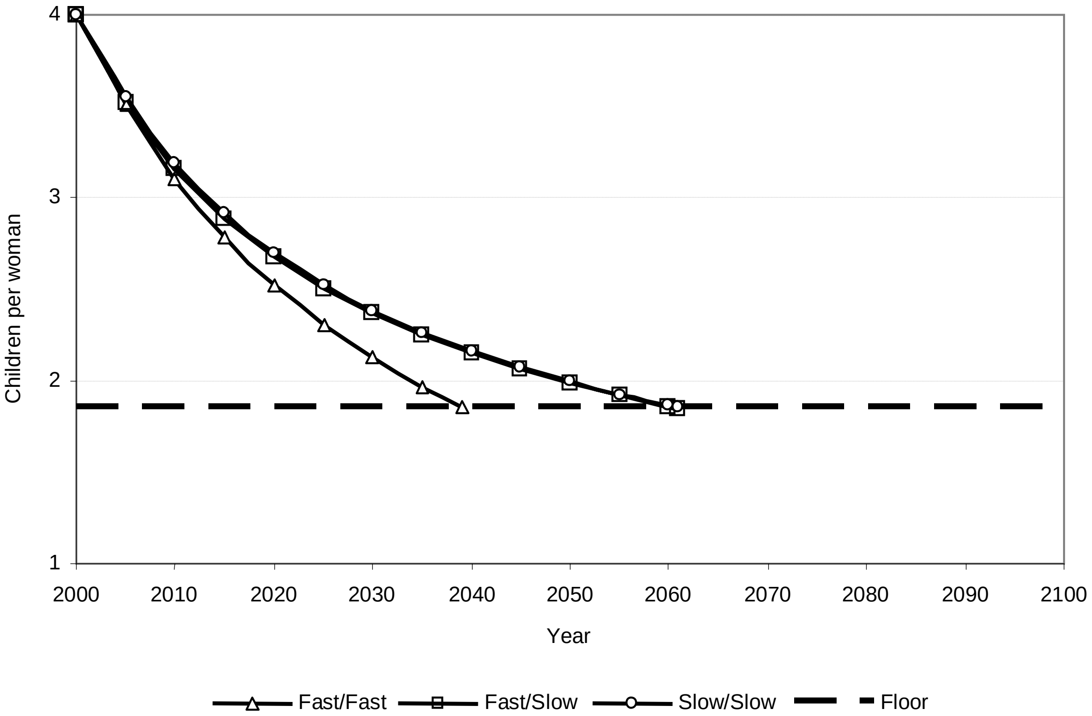


Figure VI.1b. Medium fertility: Models of fertility decline



countries, a much simpler model of fertility change was adopted. In general, it is assumed, as in the previous *Revision*, that fertility now below replacement level will also converge to the fertility floor of 1.85 children per woman, just as the high and medium fertility countries.

While all low-fertility countries were projected to have fertility of 1.85 in the long term in the *2002 Revision*, the short-term projection of total fertility for each country was accomplished taking into account the most recent trends in annual total fertility. For those countries where total fertility was below 1.85 and declining in the 1990s, the annual trend between 1990 and the most recent estimate was generally extrapolated to 2005 or 2010. Then, for most countries a pause period was projected until 2010 or 2015, at which time fertility would begin to rise at a pace of 0.07 children per woman per quinquennium Several low-fertility countries experienced a levelling-off of fertility decline or a slight increase in fertility in the late 1990s. For these countries, total fertility was generally projected to stay constant near its most recent level until 2005 or 2010.

For countries where total fertility in the 2000-2005 was above 1.85 but below 2.1, fertility was projected to decline to 1.85 during the projection period.

b. Age patterns of fertility

The projected total fertility levels were converted into age-specific fertility rates by using age patterns of fertility derived by interpolating between the most recent age pattern of fertility available and a model age-specific pattern. For the countries of Europe, model patterns were to be reached by 2025 in the market economy countries and by 2035 in the countries with economies in transition. Linear interpolation was used to move from the current fertility pattern to a model pattern. Once the model pattern was reached, it was assumed to remain constant until the end of the projection period. The model age patterns of fertility used are shown in annex table VI.4 for the market economies of Europe and in annex table VI.5 for the countries with economies in transition. They were derived from the experience of low-fertility countries by fitting a simple Beta distribution to the age-specific fertility patterns typical of market-economy countries (e.g., the Netherlands) and of countries with economies in transition (e.g., Slovenia). By varying the parameters of the Beta distribution in a manner similar to that implied by past trends, a set of model age-specific fertility patterns was generated with different mean ages of childbearing.

The model age patterns of fertility developed for Europe were also used for several of the low-fertility countries outside of Europe. In certain other low-fertility countries, the proportionate age pattern of fertility was assumed to remain constant over the projection period.

C. THE PROJECTION OF MORTALITY

In contrast with the assumptions made about future fertility trends, only one variant of future mortality trends was used for each country for the standard variants (high, medium and low variants). It must be noted, however, that the estimation and projection of HIV/AIDS related mortality, documented below, made it necessary to prepare an additional mortality variant with No-AIDS mortality for all countries with significant HIV prevalence (see chapter IV).

Assumptions are made in terms of life expectancy at birth by sex and, in most cases, an underlying model life table. As in previous *Revisions*, life expectancy was generally assumed to rise over the projection period for most countries. The major exceptions are the countries affected by the HIV/AIDS epidemic (explained below), and countries with economies in transition.

1. General approach

The often dramatic decline of mortality was - and is - a driving force behind the profound changes to population trends observed during the past two centuries. While first limited to a small number of countries in the world, the decline of mortality and rise in life expectancy has now become a global phenomenon.

For countries where mortality was assumed to follow a declining trend starting in 2005, the pace of change of life expectancy was set according to a chosen model. This *Revision* added two new models of gains in life expectancy to the existing three that were available in previous *Revisions*, namely a very slow and a very fast model of mortality improvement. Altogether, a total of five mortality models are now available: the new very fast model, the established fast, medium and slow models, and a new very slow model. The addition of the two additional models was necessary because some countries have experienced smaller gains in life expectancy than the traditional slow model suggested, while other countries have experienced consistently faster increases in life expectancy than the fast improvement model envisioned.

All five models are based on a broad empirical basis of increasing life expectancy during the period 1950 to 2005, covering life expectancies between 50 and about 85 years. The models represent the average experience of this historical period grouped according to the 90th percentile (very fast, modelled on Japan), the 75th percentile (fast model) the arithmetic mean (medium model), the lowest 25th percentile (slow model), and the lowest 10th percentile (the very slow model).

In order to be useful for the projection of life expectancy for advanced countries such as Japan, the models needed to be extended to cover levels of life expectancy not yet achieved. Such an extension was carried out to levels of life expectancy of 92.5 years by extrapolating the trends in the given models using the Lee-Carter model (United Nations 2004b). Because an unconstrained extrapolation of declining gains in life expectancy would ultimately yield zero and even negative increments, it was assumed that after the gains in life expectancy reached a certain low level, future gains would stay constant at this lowest increment. While there is no strong empirical basis for such an assumption, there is also no evidence today of a particular upper limit to life expectancy (Oeppen, Vaupel 2002). Indeed, as Tuljapurkar and Li (2004) have shown, under certain conditions, a linear trend in life expectancy, and thus a constant increment, is possible.

The models are expressed as annual increments of life expectancy for a given level of life expectancy, but are presented in the annex for the ease of use as quinquennial increments (table VI.6). Although all models differ regarding the amount of change or increments of life expectancy during a given period of time, they all share a general feature of the evolution of life expectancy: very low life expectancy is associated with small gains in life expectancy, as was the case before the onset of the demographic transition. Increments in life expectancy increased during the early stages of the demographic transition and reached a peak when life expectancy was between 50 and 60 years. As mortality is further reduced, annual gains in life expectancy tend to become smaller.

For any given country, the appropriate model was chosen by taking into account the observed pace of mortality decline in the recent to medium-term past. The selected model of improvement in life expectancy was generally followed until 2025 and, if deemed appropriate, a switch was made as of that date to the medium-pace model.

In countries with economies in transition that experienced a long period of stagnating or even increasing mortality, life expectancy was assumed to increase only very little until 2010; it was then assumed to follow one of the models just described.

Once the path of future expectation of life was determined, survival ratios by five-year age group and sex consistent with the expectation of life at birth for each quinquennium were calculated. For countries with recent empirical information on the age patterns of mortality, survival ratios for the projection period were obtained by extrapolating the most recent set of survival ratios by the rates of change of an underlying model life table. In other words, under such a procedure the empirical or estimated age pattern of mortality converges towards the underlying model pattern as life expectancy changes over time. For countries lacking recent or reliable information on age patterns of mortality, survival ratios were directly obtained from an underlying model life table. A choice could be made among nine model life table systems, four proposed by Coale and Demeny (1966; Coale, Demeny, Vaughn 1983; Coale, Guo 1989) and five model systems for developing countries produced by the United Nations (1982). These nine model life tables have been updated and extended by the Population Division in order to cover the whole age range up 100 years, and a range of life expectancies from 20 to 92.5 years (for more details, see Buettner, 2002). It must be noted that the last available entry in the revised system of model life tables of 92.5 year of life expectancy, for both males and females, are not meant to represent a ceiling for human longevity.

The general approach to the projection of mortality just described is not appropriate for countries significantly affected by the HIV/AIDS epidemic. A detailed description of assumptions made and models used to estimate and project the demographic impact of HIV/AIDS is given in the next section.

2. Modelling HIV/AIDS mortality

This *Revision* incorporates explicitly the impact of the HIV/AIDS epidemic for 60 countries, most of which had an adult HIV prevalence of at least one per cent in 2003. Brazil, China, India and the United States of America, countries where HIV prevalence is still low but which had a large number of infected persons, were also included. Among the 60 countries considered, 40 countries are in Africa, 5 are in Asia and 12 in Latin America and the Caribbean (see chapter IV, table 5). For those countries, a different approach for the estimation and projection of mortality must be used. Unlike other infectious diseases, HIV/AIDS has a very long incubation period in which an infected person is mostly symptom-free and infectious. Also unlike many other infectious diseases, individuals do not develop immunity, but, in the absence of treatment, almost always die as a consequence of their compromised immune system. Another reason for an explicit modelling of the HIV/AIDS is the avalanche-like process of the infection spreading through a population and the particular age pattern exhibited by HIV/AIDS. The additional deaths due to HIV/AIDS, predominantly adults in their reproductive age, are consequently distorting the usual U-shaped age-specific age profile of mortality, a feature which cannot be found in the model life tables that are available to demographers (Heuveline, 2003). Thus the particular dynamic of this disease and the severity of its outcome require an explicit modelling of the epidemic.

As a consequence, instead of an overall mortality process that can be captured by standard age patterns of mortality and smooth trends of changing life expectancy, for countries highly affected by HIV/AIDS, two separate mortality processes must be modelled: the mortality due to the HIV/AIDS epidemic itself and the mortality that prevails among the non-infected population. The latter is often called “background mortality”. The estimation of it is described in the next section.

a. Establishing background mortality

For countries severely affected by HIV/AIDS, hypothetical mortality paths for what is often called “background” or “No-AIDS mortality” needed to be constructed first. The background mortality is the mortality experienced by those not infected with HIV in a given country and at a given period of time. It is not the mortality that would have been observed in the complete absence of HIV/AIDS in that country, however. The distinction between background mortality and mortality in the absence of HIV/AIDS is of little importance during the first years the epidemic develops in a given country. But once a sizeable number of people is infected and ultimately dies of AIDS, the consequences of the epidemic are likely to affect severely the capacity of a country to provide health care services to its population, including the uninfected people. It is for this reason that the assumed trends in background mortality are generally less optimistic than in a similar country that is not affected by HIV/AIDS.

The background mortality can be estimated from data on causes of death (deaths caused by HIV and deaths due to other causes). Such detailed account of deaths by causes of deaths and by age and sex, however, is rarely available in countries in less developed regions that are severely affected by the epidemic. Therefore, background mortality needs to be estimated by assuming plausible levels and trends based on other information or assumptions. Often the process is one of iterative refinement: beginning with an assumed trend of the background mortality, the HIV/AIDS epidemic is modelled and the results are then compared with overall mortality estimates, if they are available (see, for example, Feeney, 2001), or with results provided by a census or survey. If necessary, background mortality is adjusted, and the procedure is repeated until a reasonable agreement between the model output and available evidence is achieved.

b. Modelling the overall dynamic of the epidemic

The approach of the Population Division to model the dynamics of the HIV/AIDS epidemic follows that suggested by the UNAIDS Reference Group on Estimates, Modelling and Projections (2002). UNAIDS has implemented this model in a software package called Epidemiological Program Package or EPP, described in Ghys et al (2004). In the following, both the epidemiological model and its software implementation are called EPP for short.

The first stage in modelling the epidemic is to derive estimates of the yearly probability of being infected by HIV (annual incidence) from available estimates of HIV prevalence. In countries of sub-Saharan Africa these prevalence estimates are derived mainly from data on the proportion of sero-positive females among pregnant women attending antenatal clinics that belong to the system of sentinel surveillance sites in each country. Consequently, available estimates of prevalence refer to the HIV prevalence among pregnant women only. It has been shown, however, that prevalence levels among pregnant women aged 15-49 provide reasonable estimates of prevalence levels among all women in the same age group (Gregson and Zaba, 1998; Glynn et al, 2001; Gregson, Zaba and Hunger, 2002). There is scant information on how well prevalence levels among pregnant women represent those among men. Only recently have nationally representative surveys of HIV sero-prevalence begun to be taken in countries of sub-Saharan Africa; their results will inform modelling specification in future *Revisions*. In the absence of more information, the models presented here assume that available estimates of prevalence among pregnant women aged 15-49 are adequate proxies of HIV prevalence among both women and men.

The EPP model divides the total population of persons over 15, denoted by N, into three groups:

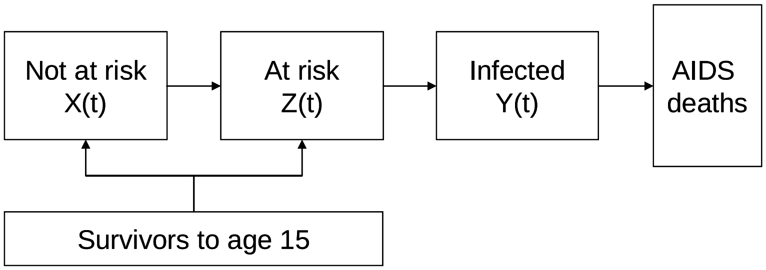
- Persons who, at time t, are not at risk of being infected by HIV, denoted by $X(t)$.
- Persons already infected by HIV at time t, denoted by $Y(t)$;
- Persons at risk of being infected by HIV at time t, denoted by $Z(t)$,

The model of the HIV/AIDS epidemic is described by a system of three differential equations (see annex), with the following four parameters to be estimated:

- *The parameter* f_0 is the fraction of individuals who entered the at-risk population at age 15 at the time the HIV epidemic started. This parameter determines the endemic level of the epidemic.
- The parameter r represents the force of infection or reproductive potential of the epidemic. It is the probability that interactions between an infected and a non-infected individual results in the infection of the latter. This parameter governs to a large extend how the epidemic grows: If r is larger than one, the epidemic grows, if it is smaller than one, the epidemic will disappear over time.
- Φ *(Phi)* captures the recruitment of people into the at-risk population and is therefore also called the behavioural or response parameter. If Φ is positive, more people are entering the at-risk population than die of the epidemic. As a consequence, the epidemic is sustained at a higher level. If Φ is negative, less people are entering the at-risk population than die of the epidemic. With Φ negative, the epidemic declines.
- The parameter t_0, the year the epidemic started in a particular country

In order to keep the number of parameters to be estimated to a minimum, other information necessary to formulate the epidemiological model are set to predetermined values, such as the rate of Mother-To-Child Transmission (MTCT), the incubation period (time from infection to death caused by HIV/AIDS), and the fertility reduction

Figure VI.2. Structure of the EPP model



associated with a HIV positive status of women. Demographic parameters such as the crude birth and death rates as well as the probability of survival from birth to age 15, the age of entry into the adult population, are taken the No-AIDS scenario.

In the original implementation of the model, the parameters r, f_0, Φ and t_0 are kept constant over time, assuming that they will not change significantly during the relatively short period of 20 to 30 years covered by the model. However, over longer periods, as in this *Revision*, the impact of the epidemic itself on the demographic variables needs to be explicitly taken into account. In addition, behavioural change as well as medical treatment of infected people is poised to alter the dynamics of the epidemic. For those reasons, the implementation of the epidemiological model by the Population Division allows for changes over time for most parameters.

In addition, because of the inclusion of treatment with Anti-Retroviral Treatment (ART) in this *Revision*, two new parameters - the coverage rate of ART treatment and annual survival for people under treatment - were included. These additional parameters are currently set outside the model and not obtained from fitting the model to empirical data.

The system of differential equations of the EPP model can be solved numerically by using, for instance, the Runge-Kutta method, provided values of all relevant parameters are known. Population-based estimates of population size at the start of the epidemic, births and mortality risks over time are available. In addition, assumptions are made about the probabilities of dying of AIDS among those infected, about the probability of mother-to-child transmission and about the extent to which the fertility of HIV-positive women is reduced. Then it is possible to estimate, via numerical approximation methods, the values of r, f_0 and Φ that minimize the distance between the HIV prevalence generated by the model and the HIV prevalence estimated on the basis of data from antenatal clinics at various points in time. More specifically, a non-linear iterative optimization procedure is used to obtain estimates of the parameters r, f_0, Φ and t_0.

This simple epidemiological model is capable of producing a large number of different epidemics and thus can be applied to countries with varying levels of severity of the epidemic (see figure IV.10 in the chapter about mortality).

This versatility is illustrated below by comparing epidemiological curves produced by varying one parameter while keeping all others constant. Each of the four parameters shapes the epidemiological curve in a particular way, allowing to attribute certain characteristics of the epidemic to one particular parameter.

The fraction of new entrants to the at-risk population, f_0, largely determines the endemic level of the epidemic (figure VI.3). If, at the beginning of the epidemic, a large fraction of people is already in the at-risk category, the epidemic will level off at higher level than in cases where initially the fraction of the at-risk population is smaller.

The force of infection (parameter r) determines the growth of the epidemic (figure VI.4). A higher force of infection results in a faster growth of the epidemic, with a higher endemic level after the peak prevalence. A lower force of infection, on

Figure VI.3. Initial fraction of people in at-risk population

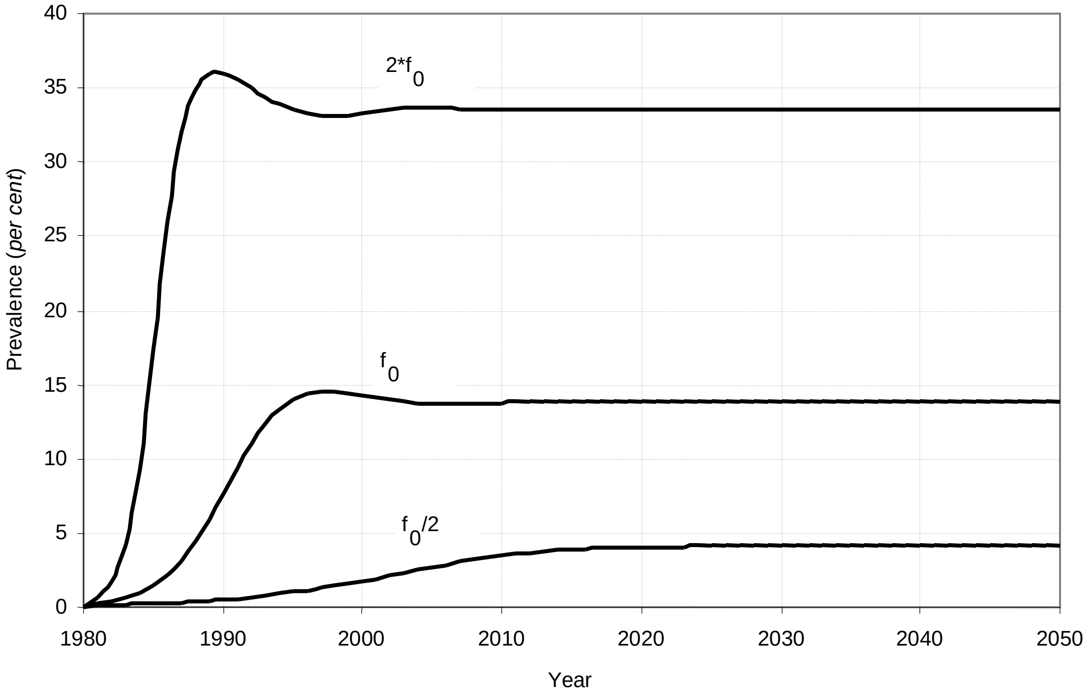
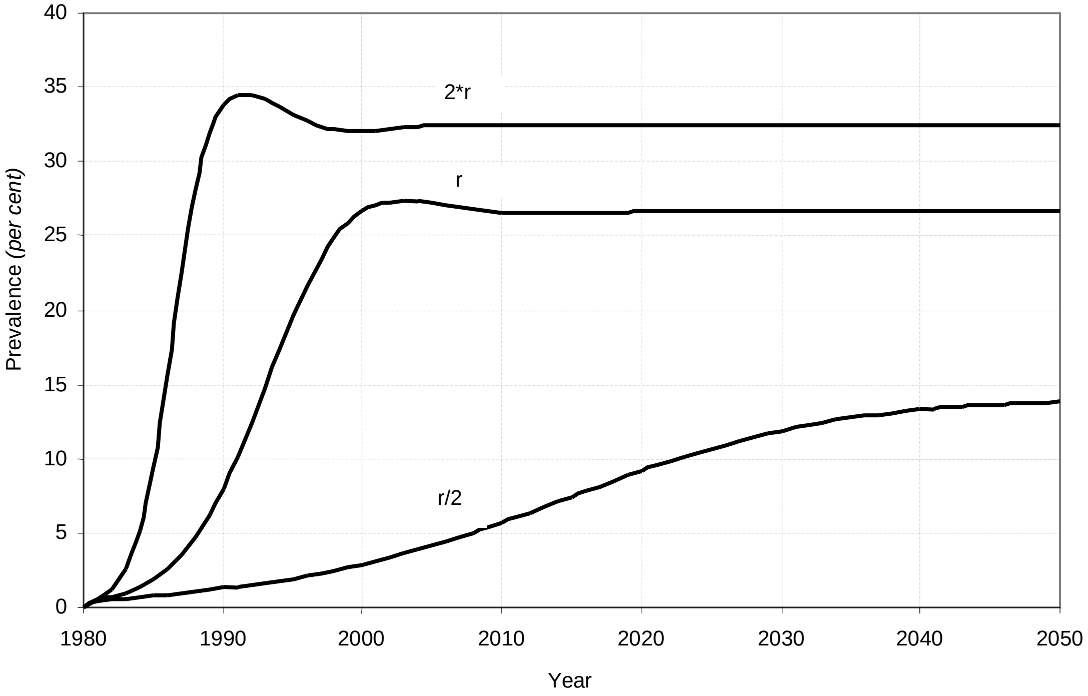


Figure VI.4. Force of infection



the other hand, produces an epidemic curve that grows slower, reaches its peak prevalence later and also has a lower endemic level.

The parameter Φ models the recruitment of people into the at-risk population (figure VI.5). If Φ is zero, then the at-risk population maintains its initial fraction of the population, that is f_0. A positive Φ means that more people are recruited into the at-risk population, resulting in an epidemic with higher endemic level. If Φ is negative, the prevalence of the epidemic declines more rapidly because the number of people dying of AIDS is larger than the number of people entering the at-risk population.

The parameter t_0 determines when the epidemic started in a given country. This parameter simply shifts a given epidemic curve horizontally on the time axis. It is not shown as a chart.

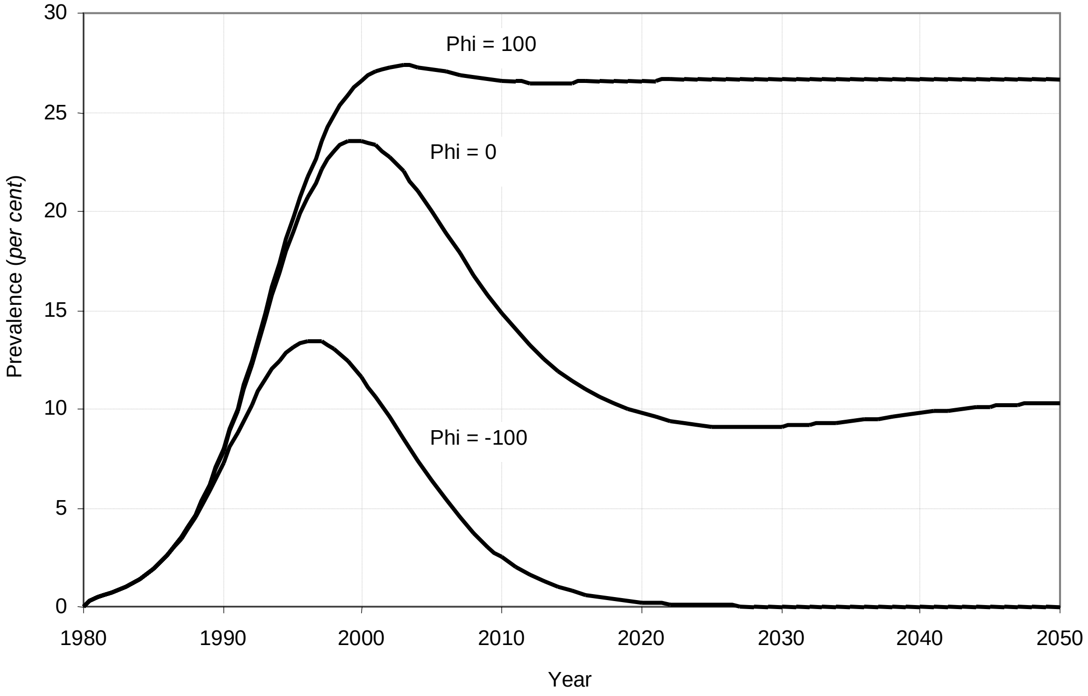
Once the values of all parameters are obtained, the mathematical model is used to calculate the number of adult persons living with HIV, the number in the at-risk group and the number who are not susceptible, as well as the number of newly infected individuals for each year t ranging from the start of the epidemic to 2003, the most recent year with data on prevalence available at the time of this *Revision*. Also calculated is the incidence rate for the total population at risk.

However, in order to estimate the effect of the HIV/AIDS epidemic on mortality and population dynamics, it is necessary to derive estimates of the infected population by age and sex. The procedures followed in such derivation are described in the next sections.

c. Estimating the demographic impact of HIV/AIDS

The simple epidemiological model EPP just outlined captures the overall dynamics of the epidemic for three sub-population groups. It does not, however, allow for a detailed account of the demographic impact of the epidemic. As mentioned, no provision is made for a disaggregation by age or sex, two key elements of a demographic analysis of the epidemic. Modelling only adult populations, it does not include explicitly the paediatric dimension of the epidemic. In order to

Figure VI.5. Recruitment into at risk population



generate a full account of the demographic impact of HIV/AIDS, the Population Division developed abcDIM[1], a software package that combines the EPP model with a full multistate demographic projection model. In it, the epidemiological dynamics captured in EPP are translated into age-and sex specific values, thus providing a full demographic account. Figure VI.6 shows the schematic structure of the model used in abcDIM.

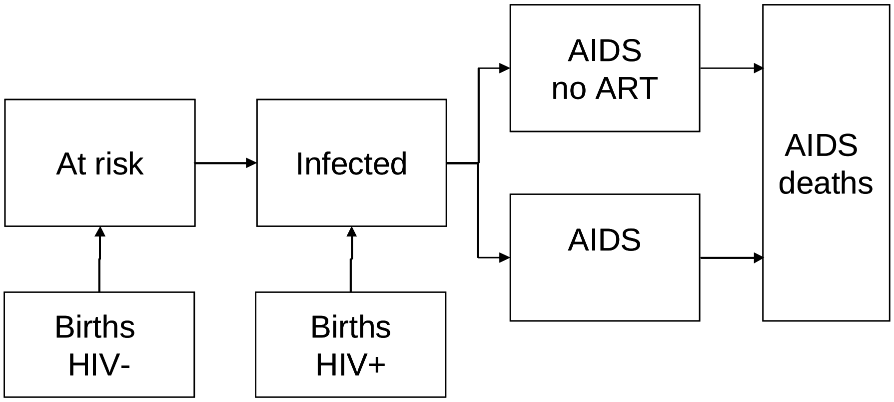
The estimation of the demographic impact of HIV/AIDS is carried out in several steps. First, the EPP epidemiological model needs to be extended until 2050, the final projection year in this *Revision*. While the simple epidemiological model used by UNAIDS makes the implicit assumption of constant parameters for the last 25 years, for long range projections such a simplification can not be maintained. In particular the parameters of the model that capture behavioural elements, such as Φ and r, cannot be assumed to be constant throughout this period. Instead, in order to incorporate the effects of intervention, such as treatment, or prevention, such as increased condoms use, reasonable assumptions about future trends of these parameters are needed.

For this *Revision*, Φ, r and the Mother-to Child Transmission (MTCT) are assumed, after 2005, to decline over time, reflecting the impact of interventions and behavioural change. The declining trend in these parameters can be specified in abcDIM by assuming a halving period, that is the number of years it take to reduce the value of the parameter to half the value it had at the beginning of the projection period (see table VI.10). The default values for the halving times are 10 years for MTCT, 30 years for r, and 20 years for Φ.

The parameters just described influence the number of people that are being newly infected, but they do not alter the chances of survival once a person is infected. Anti-retroviral treatment, on the other hand, results in longer survival of infected people. In this *Revision*, the effects of ART have been included into the abcDIM model by adding a new stage (labelled ART in figure VI.6) to the demographic projection model, and by prolonging the survival times in EPP accordingly. Consequently, depending on the proportion of HIV positive people receiving treatment and the percentage of people under treatment surviving annually, the overall survival time in EPP is dynamically adjusted in order to reflect the impact of treatment on survival of infected people.

Then, the estimates of annual HIV incidence derived from the epidemiological model with all sexes and ages combined are converted into age and sex-specific estimates of newly infected individuals and the population that was initially free from the epidemic is projected using a multi-state approach that tracks the transitions of people from at-risk to AIDS and finally deaths. Note that the abcDIM model also estimates the number of children by infection status and follows them similarly.

Figure VI.6. Structure of the abcDIM model



---

[1] Its name is derived from Demographic Impact Model (abcDIM).

All sub-populations are projected by single years of age while the infected population is further classified by duration since infection in single years. The exact steps followed and the assumptions made in recreating the dynamics of a population affected by the HIV/AIDS epidemic are described in detail below.

Step 1: *Derivation of the number of new infections by sex.*

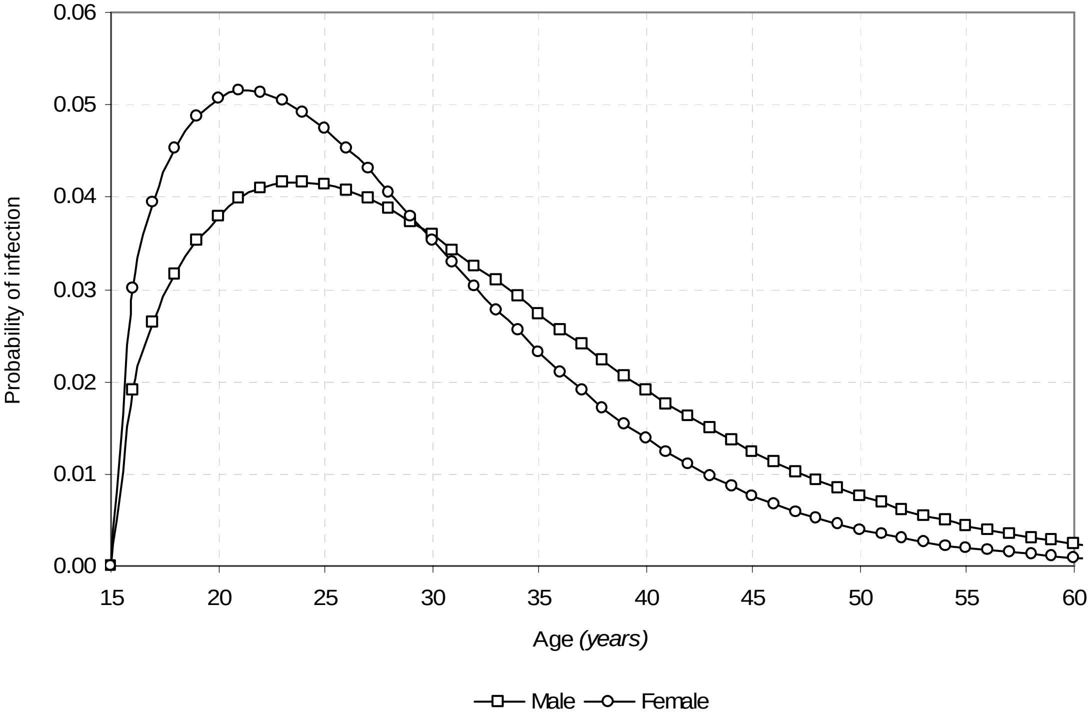
As noted above the model used to derive the parameters *r*, f_0 and Φ does not take into account the age or sex of the population infected. To derive estimates of the impact of HIV/AIDS by age and sex, it is first necessary to distribute by sex the yearly number of newly infected individuals, as yielded by the general epidemiological model EPP. Although data on the distribution by sex of newly infected individuals are rare, there is some evidence suggesting that when HIV/AIDS is spread mainly by heterosexual transmission, the proportion of males among the newly infected is high at first but declines rapidly in the years following the start of the epidemic to proportions closer to those of women.

On the basis of this observation, the proportion of males among the newly infected is assumed to decline from 80 per cent or so at the start of the epidemic to 45 per cent after a few years and to remain constant at that level for an extended period. However, in regions or countries where HIV/AIDS is not spread primarily by heterosexual contact (e.g. homosexual contact, intravenous drug use, etc.) sex patterns of newly infected individuals differ, with higher proportions being attributed to men. Under these assumptions, the annual number of newly infected individuals per year is distributed by sex.

Step 2: *Derivation of the number of newly infected men and women by age.*

Once estimates of the newly infected people by sex are available, they are distributed by single-years of age according to model age distributions derived from empirical data that were fitted to a Weibull distribution (figure VI.7, table VI.7, annex), with a mean age at infection of 29.1 years for males and 26.1 years for females. Figure VI.7 shows the density functions by age for males and females.

Figure VI.7. Age specific HIV infection probabilities by sex



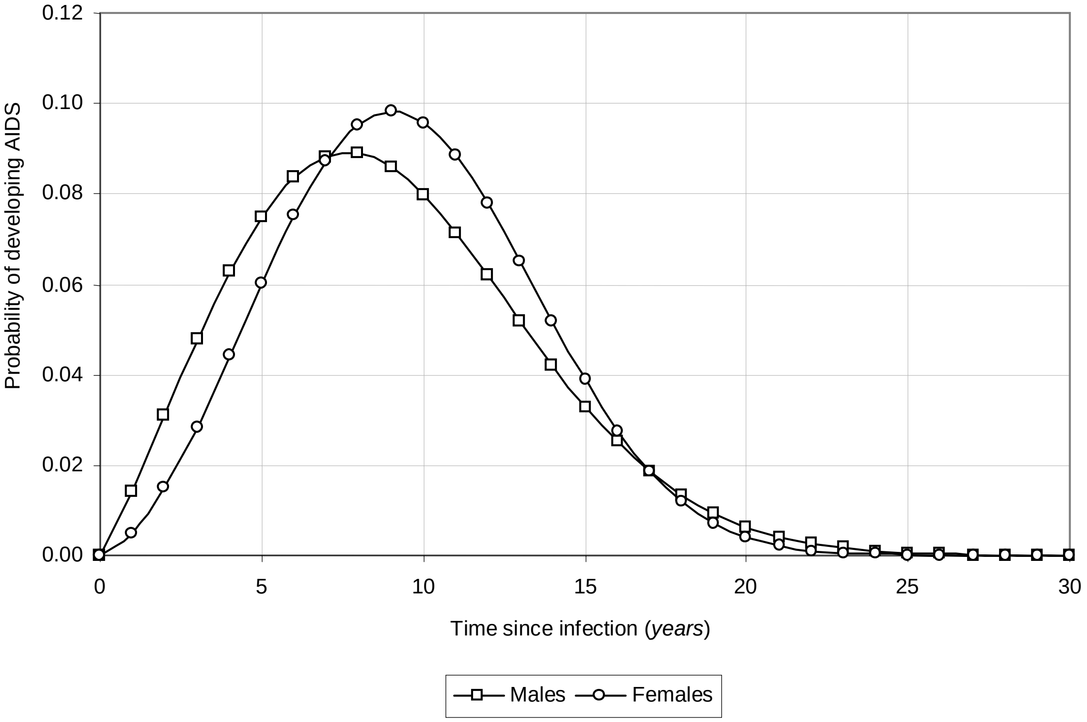
Step 3: *Estimation of the number of deaths caused by AIDS among HIV positive persons.*

As in the previous *Revision*, infected people are passing first through a stage equivalent to stages 1 through 3 of the WHO staging system (WHO, 2004) in which they are infected but, at least in stages 1 and 2, asymptomatic. Deaths that occur during those stages are not caused by HIV/AIDS. In figure VI.6, these three clinical stages are aggregated into the stage labelled “Infected”. Infected people can enter two possible final stages, associated with full-blown AIDS (stage 4 by WHO classification). In the stage labelled AIDS/noART, people do not receive ART, and it is therefore assumed that they have a further average life expectancy of just one year. In the stage labelled AIDS/ART, new in this *Revision*, they receive life-prolonging treatment, and their average remaining life time will be increased well beyond the one year associated to the stage without treatment.

To estimate the number of deaths due to AIDS by age and sex, the infected population is projected over time using a multi-state approach that takes account of the competing risks of moving from being uninfected to being infected (HIV-positive) and from being HIV-positive to developing full blown AIDS versus the probability of dying of a cause other than AIDS. The probability schedules used to reflect the chances of developing full blown AIDS after *x* years of infection (the incubation period) are assumed to follow a Weibull distribution (see table VI.8). Different schedules were used for each sex, with a mean incubation period of about 9.3 years for both sexes combined, a slightly longer mean incubation period for females (9.6 years) and a shorter one for males (9 years). The schedules are shown in figure VI.2.

These survival functions originally suggested to cover the whole period from infection with HIV to deaths of AIDS were based on cohort studies that included background mortality. Removing the background component of mortality results in a net survival time of approximately 10 years (see Porter and Zaba, 2004). The procedure used for this *Revision* therefore increased the originally

Figure VI.8. Annual probability of transition from HIV infection to AIDS



recommended mean survival time of about 9.3 years since infection by one year. This extension was achieved by adding the one year survival in the additional AIDS stage (without treatment) to the model.

The probability of progressing from HIV infection to full-blown AIDS was assumed to be age-neutral, that is no allowance was made for systematic differences in the incubation period related to age at infection.

Competing mortality risks for causes other than AIDS were estimated on the basis of mortality estimates for the whole population. It was assumed that among HIV positive persons, the risk of dying of a cause other than AIDS was independent from the risk of dying of AIDS.

Once an infected person reaches stage 4, or full blown AIDS, he or she is in need for treatment with ART. In resource-poor settings such as in most developing countries, treatment is not available for the whole group of infected people reaching stage 4. The abcDIM program therefore allows setting a time varying parameter, the coverage rate, to reflect this situation. Table VI.10 lists country specific values of assumed current and future coverage rates for all the countries concerned.

It was mentioned earlier that the default mean survival time of people entering stage 4 or full-blown AIDS is approximately one year. With treatment, survival of infected people is extended. An exponential function is used to model the survival with or without treatment in abcDIM; it is parameterized as constant annual per cent survival. Based on early evidence on the efficiency of ART treatment in resource-poor countries, two different setting have been used, one with 80 per cent annual survival, corresponding to a mean survival time of 4.5 years after starting treatment, and another with 90 per cent annual survival, corresponding to 9.5 years mean survival time.

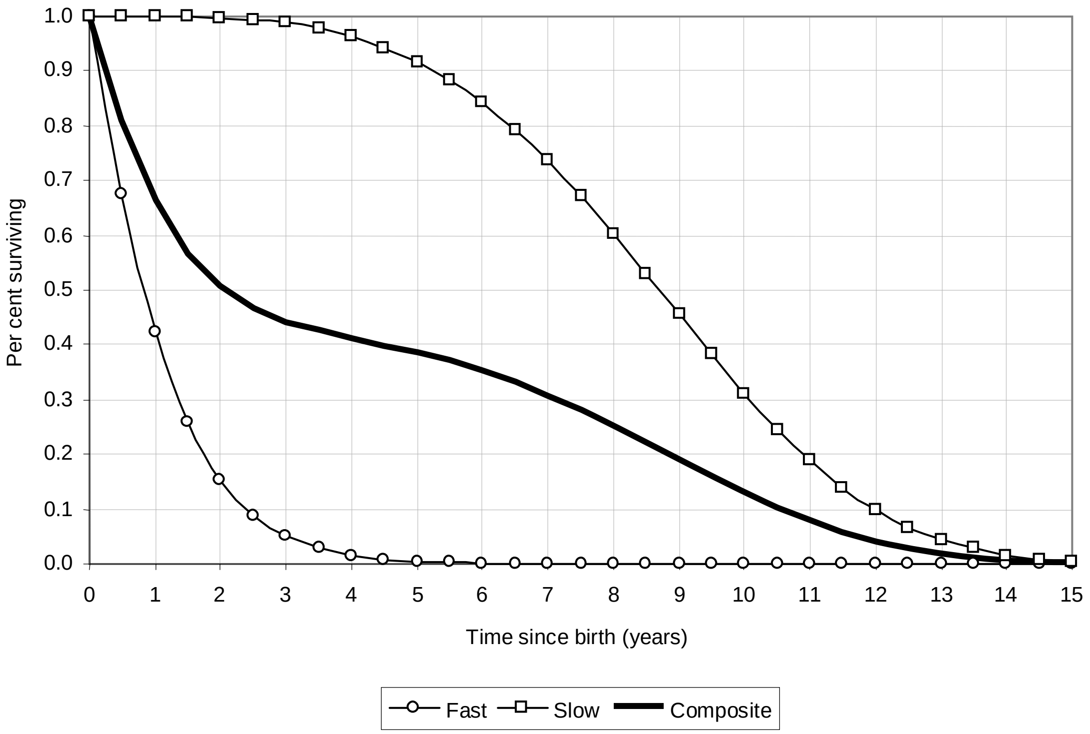
Step 4: *Calculation of the number of children infected by HIV/AIDS.*

Although HIV is primarily transmitted by sexual contact, which places adolescents and adults at risk, it also exerts a heavy toll on children. Children are infected by their HIV positive mothers passing the virus to their children in utero, at parturition or during breastfeeding. To estimate the number of children that can potentially become infected by their mothers, first the number of children born by HIV-positive women is calculated, allowing for a reduced fertility that takes into account the lower probability of conception among HIV-positive women. In this *Revision*, it was uniformly assumed that HIV positive women have a 20 per cent lower age-specific fertility than those not infected. Because most HIV-positive children acquire the disease from their infected mothers at or near the time of birth, the number of HIV-positive children is obtained by assuming a fixed rate of transmission of HIV from mother to child of 35 per cent and multiplying it by the number of children born to HIV-positive women. Such an approach produces the number of children who become HIV positive at birth or soon thereafter during each year. In addition, the age-specific fertility rates applied to non-infected women are increased in such a way that the overall fertility rates of the population as a whole (both infected and not infected women) match those estimated from available data.

Step 5: *Calculation of the number of AIDS deaths among children.*

In children, the length of infection is the same as their age. The number of surviving HIV-positive children is calculated by modelling the probability that infected children have of surviving HIV infection up to a certain age as the sum of two Weibull functions representing two subgroups of infected children: Fast progressors and slow progressors (figure VI.9). Fast progressors are those children infected *in utero*, while slow progressors are children infected at parturition or during breast feeding. In the current model, about 58 per cent of all infected children follow the fast progression schedule and will die early, on average just 1.1 years after birth. Children that follow the slow progression schedule (about 42 per cent of all infected children) have a longer incubation period and will on average die about 8.8 years after birth. Together, the life expectancy of all children infected with HIV is less than 5 years (table VI.11). The current survival model for chil-

Figure VI.9. Survival distributions for children



dren implies that none of the infected children will survive past age 15; it also assumes that the survival probabilities are the same for the male and female child.

Step 6: *Projecting the population that is not infected by HIV.*

The previous steps describe how the HIV-positive population is projected from the start of the epidemic onward. In fact, the full multi-state projection procedure projects also the non-infected population allowing for two possible and independent ways of leaving that group: (*a*) by dying from non-AIDS causes, or (*b*) by becoming infected with HIV (i.e., the yearly incidence).

Step 7: *Calculation of revised life-tables that reflect the impact of HIV/AIDS.*

The results of the multi-state projections permit the calculation of life tables that reflect both the effect of general mortality and the added impact of HIV/AIDS in a manner consistent with what is known about HIV prevalence in each country. The life tables representing average mortality for five-year periods are then used to carry out the "normal" population projections over five-year periods prepared by the Population Division for countries affected by the epidemic. That is, the mortality projection procedure ultimately used is the same for countries that are not yet affected significantly HIV/AIDS and those severely affected by the epidemic. This approach allows it to easily "splice" population projections for periods before the start of the epidemic with those after its start. It also allows to create "No-AIDS" versions of the population projections that represent estimated population dynamics in the absence of HIV/AIDS.

D. THE PROJECTION OF INTERNATIONAL MIGRATION

International migration is the component of population change most difficult to project. This is primarily due to the fact that data on past trends are often sparse or incomplete, and because the movement of people across international borders, which is often a response to rapidly changing economic, social, political and environmental factors, is a very volatile process. Not only has international migration shown drastic changes in absolute numbers, but the direction of the flows has changed as well. As discussed in more detail in chapter 5, immigration countries have in the past often become emigration countries or vice versa. Therefore, formulating assumptions of future trends must focus on dominant past trends that are then kept constant throughout the projection period.

When a person moves from one country to another, that person is an emigrant when leaving the country of origin and becomes an immigrant when entering the country of destination. Because immigration and emigration flows affect countries differently, international migration is ideally studied as the flow of people moving between countries. In practice, data on international migration flows do exist only for a small number of countries. Therefore, international migration in this *Revision*, as in previous ones, has been captured as net migration. Net migration - the difference between the number of immigrants and the number of emigrants for a particular country and period of time - shows the net effect of international migration on the respective population. It does not provide an indication about the number of immigrants and emigrants involved. In an extreme case, immigration and emigration for a country could be significant, but if the number of immigrants was equal to the number of emigrants, net migration would amount to zero.

In preparing assumptions about future trends in international migration, several pieces of information were taken into account: (*1*) information on net international migration or its components (immigration and emigration) as recorded by countries; (*2*) data on labour migration flows; (*3*) estimates of undocumented or irregular migration; (*4*) and data on refugee movements in recent periods.

The basic approach for formulating future international migration assumptions is straightforward. For any given country, a distinction is made between international migration flows and movement of refugees. For international migration, it is assumed that recent levels, if stable, continue throughout the projection period. Government's views on international migration as well as estimates of undocumented and irregular migration flows affecting a country are also considered (see, for example United Nations, 2003). Regarding the movements of refugees, it is assumed in general that refuges return to their country of origin within the next one or two projection periods, or within 5 to 10 years. If a country experiences both international migration and refugee movements, the two processes are added in order to capture to overall net migration during a particular period in the future.

Usually, migration assumptions are expressed in terms of net number of international migrants. Their distribution by sex is established on the basis of what is known about the participation of men and women in different types of flows for any given country (e.g. labour migration, family reunification, etc.). Given the lack of suitable information on the age distribution of migrant flows, models are generally used to distribute the overall net number of male and female migrants by age group according to the dominant type of migration flow assumed (i.e., labour migration, family migration). These age and sex profiles of the net migration flows are then used as input for the cohort-component projection model (United Nations 1988, pp 65-70). For few countries with a known age and sex distribution of international migrants, those distributions were used to determine which model is most suitable or, in some cases, they were used directly as input. The distribution of net migrants by age and sex was generally kept constant over the projection period. However, if a country was known to attract temporary labour migrants, an effort was made to model the return flow of those labour migrants accounting for aging of the migrants involved. The same idea was applied to refugee flows.

International migration has become a universal phenomenon affecting countries all over the world. For countries known not to admit international migrants and known not to be the source of a sizeable number of migrants, net migration was set to zero during 2005-2050. In fact, 15 of 192 countries were assumed to have zero net migration over the projection period. For an additional 10 countries migration was assumed to become zero some time during 2005-2025. Most countries in this category were countries affected by refugee flows. In general, refugees who had found asylum in less developed countries were assumed to return to their countries of origin by 2010-2015. Hence, net migration for host countries of refugees was assumed to differ from zero in 2010-2015 but was set to zero after 2015.

The remaining 167 countries were projected to experience non-zero net international migration during the entire projection period. Among these 167 countries 58 were projected to be receiving countries with positive net flows, while 109 countries were projected to be sending countries with negative net flows.

E. THE PREPARATION OF POPULATION PROJECTIONS

1. Projection methods

The Population Division has employed the cohort-component projection method for individual country projections since the *1963 Revision*. This method, the most commonly projection method used by demographers, provides an accounting framework for the three demographic components of change: births, deaths and international migration and relates them to the population affected. Technically, it is not a projections method, as it requires the components of change - births, deaths, migration - to be projected in advance. Rather, it is a calculation device that describes how to combine the demographic components arithmetically such that correct results are obtained. At its core, the cohort component method follows people in a certain age group at a certain point in time as they survive n years and are n years older. During a projection interval of n years, deaths occurring to that group of people are subtracted and international migration is added or subtracted, depending on the direction of the migration. Births that occur during a projection period are also exposed to the risk of deaths and then added as the youngest age group. A formal description of the mathematics of the cohort-component method can be found in Preston, Heuveline and Guillot (2001); it is not repeated here.

The cohort component method is applied for 192 countries with populations of 100,000 or more inhabitants in 2000. The 36 countries that fell below that threshold are projected assuming growth rates of their total populations. As a consequence, only total population and growth rates are available for these countries with relatively small population sizes.

2. Variants and scenarios

This *Revision* includes six projection variants in addition to the medium variant, plus three scenarios related to the HIV/AIDS epidemic. Three variants—high, low and constant-fertility—differ from the medium variant only in the projected level of total fertility. In the high variant, total fertility is projected to approach a fertility level that is 0.5 children above the total fertility in the medium variant For example, countries reaching a total fertility of 1.85 in the medium variant reach a total fertility of 2.35 in the high variant. In the low variant, total fertility is projected to remain 0.5 children below the total fertility in the medium variant. In the constant-fertility variant, total fertility remains constant at the level estimated for 2000-2005.

Three additional variants with constant-mortality, zero-migration and instant-replacement fertility have been prepared. The constant-mortality and zero migration variants have the same fertility assumption as the medium variant. Furthermore, the constant-mortality variant has the same international migration assumption as the medium variant. Consequently, the results of the constant-mortality variant can be compared with those of the medium variant to assess the effect that changing mortality has on other demographic parameters. Similarly, the zero-migration variant differs from the medium variant only with respect to the underlying assumption regarding international migration. Therefore, the zero-migration variant allows an assessment of the effect that non-zero migration has on other demographic parameters. The instant-replacement variant shares mortality and migration settings with the medium variant, but sets, beginning in 2005-2010, fertility to levels that would assure replacement of future generations.

For illustrative purposes, three scenarios related to the HIV/AIDS epidemic have been prepared: a

No-AIDS mortality scenario, a high-AIDS mortality scenario and an AIDS-vaccine scenario. The No-AIDS mortality scenario serves as the basis for the estimation of the demographic impact of the epidemic and assumes that only the background mortality applies (see VI.C.2.a). In the high-AIDS mortality scenario, all epidemiological parameters are kept constant over time. This scenario can therefore be used to gauge the impact of the assumed and projected treatment and behavioural changes contained in the medium variant. Finally, the AIDS-vaccine scenario assumes that, beginning in 2006, all additional infections can be averted by a perfect and universally available vaccine.

The various projections variants and scenarios are made available, in varying degree of completeness, on three CD-ROMs (see order form).

3. Interpolation procedures

The cohort-component method requires a uniform age format for the population and the vital events, usually single-year or five-year age groups. For the purpose of global population estimates and projections, most data are only available in five-year age groups. As a consequence, all results produced by the cohort-component method also in five year age groups and, for vital events, represent five year periods. Life expectancy, for instance, is given as the average over the five-year period from mid- 2000 to mid-2005. However, users of the estimates and projections often need to have demographic information for single calendar years or for single year-age-groups. In those cases, it is customary to apply special interpolation routines to produce such indicators. It must be noted, however, that interpolation procedures cannot recover the true series of events or the true composition of an aggregated age group. All these procedures can do is to provide the user with a smooth, reasonable and internally consistent annualized estimate of the indicator under consideration.

a. Interpolation of populations by age and sex

The basis for the calculation of interpolated population figures by single years of age and for each calendar year are estimated and projected quinquennial population figures by five-year age groups and sex. Interpolation into annual population figures is carried out by applying Beers ordinary formula (Siegel and Swanson, 2004, p.728). This interpolation procedure generates a smooth interpolated series of figures while maintaining the original values. The interpolation of five-year age groups into single year age groups is carried out by applying Sprague's fifth-difference osculatory formula (Siegel and Swanson, 2004, p. 727) for subdivision of groups into fifths. It should be noted that for ages above 80 and for age under five, the stability and reliability of the interpolation procedure is not always satisfactory.

b. Interpolation of vital events and summary statistics

For the interpolation of vital events, their rates and other measures into annualized times series, the modified Beers formula was used (Siegel and Swanson, 2004, p. 729). This formula combines interpolation with some smoothing. Beers modified methods is to be preferred over Beers "ordinary" formula as it avoids fluctuations at the beginning and the end of the series that are not typical for the variables concerned.

The time periods in the estimates and projections of this *Revision* are anchored to mid-year. Each observation or projections period starts at 1 July of a particular year and ends at mid-year five years later. Therefore, the annualized interpolated indicators refer to the period between the mid-year points of two consecutive calendar years. In order to provide annualized variables that refer to calendar years, an adjustment is made that simply assumes that the arithmetic average between two such periods will be a good representation of the calendar year based indicator.

4. Tabulations

Once the individual country projections are prepared, the results are aggregated into the world, regions, major areas, development groups and other aggregates. For a list of the aggregation units see the explanatory notes.

The aggregation of populations by age and sex and vital events by age and sex is performed by simply adding the variables according to lists that assign individual countries to the aggregates. For synthetic variables, like life expectancy, total fertility, median age or net reproduction rates, proper population weighted averages are calculated.

Finally, after estimates and projections for all countries are performed and aggregated, it is necessary to ensure that the sum of all international migration adds to zero at the global level. This is achieved by an iterative process in which individual country projections are re-visited and altered accordingly.

F. ANNEX

This appendix contains the detailed information about the various models used for estimating and projecting demographic components, namely the models of fertility decline, of mortality improvement, and especially the mathematical model of the HIV/AIDS epidemic.

1. Fertility

a. Models of fertility decline

The decline of fertility in this *Revision* is modelled using logistic functions.

A logistic function exhibits an s-shape and describes a diffusion process growing from an initial level to an upper or lower asymptote.

The general form of a logistic can be expressed as

$$P(t)=\frac{k}{1+\exp[-\alpha(t-\beta)]} \tag{1}$$

k	Saturation level or asymptote of the diffusion process
α	Growth rate of the s-curve
β	Length of time the curve takes to reach the midpoint of the growth trajectory.

For modelling purposes, a re-parameterised logistic function is sometimes used (Meyer, Young Ausubel, 1999), with easier to interpret parameters:

$$P(t)=\frac{k}{1+\exp[-\frac{Ln(81)}{\Delta t}(t-t_m)]} \tag{2}$$

t_m	Midpoint of the growth/diffusion process
Δt	Duration for the growth process to proceed from 10 per cent to 90 per cent of the asymptote (k).

This function relates to the general form by substituting

$$\beta=t_m$$
$$\Delta t=\frac{\ln(81)}{\alpha}.$$

As discussed in VI.B.1.a, the process of fertility decline consists of two phases: a first phase of accelerating rates of decline that is followed by a second phase of slowing rates of decline. Such a two-phase process can be modelled by two logistic functions, one approaching an upper limit and a second one that approaches a lower limit.

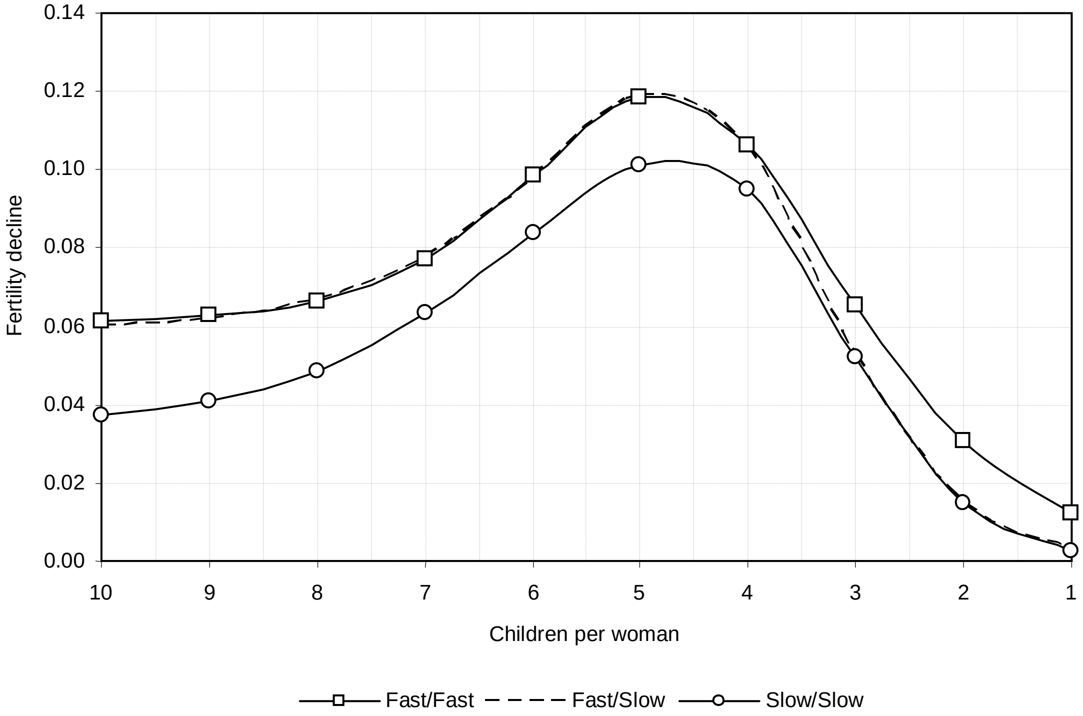
$$P(t)=\frac{k_1}{1+\exp[-\frac{Ln(81)}{\Delta t_1}(t-t_{m1})]}+\frac{k_2}{1+\exp[-\frac{Ln(81)}{\Delta t_2}(t-t_{m2})]} \tag{3}$$

Table VI.2 presents the parameters of the three models used in this *Revision* for projecting fertility decline. Figure VI.10 shows the composite curves of fertility decline for all three models.

TABLE VI.2. PARAMETERS OF THREE FERTILITY MODELS

Parameter	*Slow/Slow*	*Fast/Slow*	*Fast/Fast*
k_1	-0.11	-0.16	-0.25
Δt_1	5.03	4.34	4.01
t_{m1}............	5.77	5.06	5.17
k_2	0.15	0.22	0.31
Δt_2	2.75	3.02	4.32
t_{m2}............	3.21	3.52	3.94

Figure VI.10. Models of fertility decline



b. Models of age patterns of fertility

Model age patterns of fertility are presented as proportionate age-specific mortality, indexed by the mean age at childbirth.

2. *Models of general mortality improvement*

Trends of mortality improvement are modelled as gains in life expectancy over a five-year period for a given range of life expectancy at the previous five-year period, and are shown in table VI.6 (see chapter VI.C.1).

3. *Models of HIV/AIDS*

a. Epidemiological model

The model used to derive annual estimates of incidence from observed prevalence levels is based on three differential equations representing the dynamics of the epidemic over time (UNAIDS Reference Group on Estimates, Modelling and Projections, 2002). The model divides the total adult population of persons over 15, denoted by *N*, into three groups:

1. Persons who, at time *t*, are not at risk of being infected by HIV, denoted by *X*(*t*).
2. Persons already infected by HIV at time *t*, denoted by *Y*(*t*);
3. Persons at risk of being infected by HIV at time *t* (the susceptible population), denoted by *Z*(*t*),

The first differential equation indicates how the susceptible or at-risk population changes over time:

$$\frac{dZ(t)}{dt} = F\left(\frac{X(t)}{N(t)}\right)E(t) - \left[\mu + \frac{rY(t)}{N(t)} + \theta(t)\right]Z(t) \qquad (1)$$

The second equation shows how the non-susceptible population (not at-risk population) changes over time:

$$\frac{dX(t)}{dt} = \left(1 - F\left(\frac{X(t)}{N(t)}\right)\right)E(t) - \mu X(t) \qquad (2)$$

TABLE VI.3. MODEL FERTILITY SCHEDULES FOR HIGH AND MEDIUM-FERTILITY COUNTRIES

Model	*Percentage of total fertility by age group*							*Total*	*Mean age at childbirth*
	15-19	*20-24*	*25-29*	*30-34*	*30-39*	*40-45*	*45-49*		
Early child-bearing........................	20	40	25	10	4	1	—	100	28.3
Intermediate child-bearing	12	31	31	16	8	2	—	100	26.5
Late child-bearing..........................	4	22	40	22	10	2	—	100	24.3

TABLE VI.4. MODEL AGE PATTERNS OF FERTILITY USED FOR THE MARKET ECONOMY COUNTRIES OF EUROPE

Model	*Percentage of total fertility by age group*							*Total*	*Mean age at childbirth*
	15-19	*20-24*	*25-29*	*30-34*	*35-39*	*40-45*	*45-49*		
1	2.2	22.9	43.2	26.2	5.2	0.2	0.0	100	28.0
2	1.5	17.5	40.4	31.4	8.7	0.6	0.0	100	29.0
3	1.0	13.2	36.3	35.3	13.0	1.3	0.0	100	30.0
4	0.6	9.8	31.6	37.6	17.9	2.5	0.0	100	31.0
5	0.4	7.2	26.7	38.1	23.0	4.5	0.1	100	32.0

TABLE VI.5. MODEL AGE PATTERNS OF FERTILITY USED FOR THE COUNTRIES WITH ECONOMIES IN TRANSITION

Model	*Percentage of total fertility by age group*							*Total*	*Mean age at childbirth*
	15-19	*20-24*	*25-29*	*30-34*	*30-39*	*40-45*	*45-49*		
1...................	7.9	35.3	38.4	15.9	2.4	0.1	0.0	100	26.0
2...................	5.6	29.5	39.3	21.0	4.4	0.2	0.0	100	27.0
3...................	4.0	24.1	38.4	25.6	7.3	0.6	0.0	100	28.0
4...................	2.8	19.4	36.2	29.5	10.8	1.3	0.0	100	29.0
5...................	2.0	15.4	33.1	32.1	14.8	2.5	0.1	100	30.0

The third differential equation captures how the number of infected persons (Y) changes over time:

$$\frac{dY(t)}{dt}=\left[\frac{rY(t)}{N(t)}+\theta(t)\right]Z(t)-\int_{0}^{t}\left[\frac{rY(s)}{N(s)}+\theta(s)\right]Z(s)\,M(t-s)ds \tag{3}$$

In order to start the epidemic in this model, an external pulse is required, implemented here as parameter $\theta(t)$. It is set to a positive value when the epidemic starts, and becomes zero thereafter.

The three sub-populations in this model are adult populations (age 15 and over), so they need to be connected to births to allow for the renewal of the population. $E(t)$ in formula 1 represents the number of individuals entering the population aged 15 or over at time t. $E(t)$ is therefore the number of persons reaching exact age 15 at time t. $E(t)$ can be estimated as:

$$E(t)=l(15,t-15)\,b(t-15)\left[X(t-15)+Z(t-15)+(1-\nu)\,\xi\,Y(t-15)\right] \tag{4}$$

where $l(15, t\text{-}15)$ is the probability of surviving from birth to age 15 among persons born at time t-15, $b(t\text{-}15)$ is the birth rate at time t-15, ν is the probability of HIV transmission from mother to child, and ξ is a factor reflecting the reduction of fertility among HIV positive women. The infected population Y at time t-15 is reduced by the transmission rate from mother to child and the fertility reduction factor, which is equivalent to assuming that no child born with HIV will survive to age 15.

Not all persons reaching age 15 are susceptible to being infected with HIV. The fraction that becomes part of the susceptible population is a function of the proportion of the population that is not susceptible and is defined as:

$$F\left(\frac{X(t)}{N(t)}\right) = \frac{\Omega\left(\frac{X(t)}{N(t)}\right)}{\left(\exp\left[\Phi\left(\frac{X(t)}{N(t)} - 1 + f_0\right)\right] - 1 + \frac{1}{f_0}\right)} \qquad (5)$$

In formula 5, f_0 is the fraction of individuals who entered the susceptible group at age 15 just as the HIV epidemic started:

$$f_0 = \frac{X(0)}{N(0)} \qquad (6)$$

The parameter *Φ (Phi)* captures the recruitment of persons into the susceptible group. In addition, μ represents the mortality rate among the population not infected with HIV (the background mortality) and the rest of the expression in parenthesis represents the decrement of $Z(t)$ caused by the transfer of persons from the susceptible group to the group of those infected with HIV.

The parameter r represents the force of infection, that is, the probability that an interaction between an infected individual and a susceptible one results in the infection of the latter.

TABLE VI.6. MODELS FOR MORTALITY IMPROVEMENT. QUINQUENNIAL GAINS IN LIFE EXPECTANCY AT BIRTH ACCORDING TO INITIAL LEVEL OF LIFE EXPECTANCY

Initial life expectancy level (years)	*Very fast pace*		*Fast pace*		*Medium pace*		*Slow pace*		*Very slow pace*	
	Male	*Female*	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
40.0-42.5..........	2.5	2.6	2.1	2.3	1.9	2.0	1.3	1.4	1.1	1.1
42.5-45.0..........	2.8	3.0	2.4	2.5	2.0	2.1	1.4	1.5	1.1	1.2
45.0-47.5..........	3.0	3.1	2.5	2.6	2.1	2.2	1.8	1.9	1.2	1.3
47.5-50.0..........	3.0	3.2	2.6	2.7	2.2	2.3	1.8	1.9	1.3	1.4
50.0-52.5..........	3.2	3.4	2.7	2.9	2.3	2.4	1.9	2.0	1.4	1.5
52.0.5-55..........	3.6	3.7	2.7	3.0	2.4	2.6	2.0	2.0	1.5	1.7
55.0-57.5..........	3.7	3.7	2.6	3.0	2.4	2.6	2.0	2.0	1.5	1.8
57.5-60.0..........	3.8	4.0	2.6	3.0	2.4	2.6	2.0	2.0	1.5	1.8
60.0-62.5..........	3.4	3.8	2.5	3.0	2.2	2.6	1.7	2.0	1.0	1.7
62.5-65.0..........	3.2	3.6	2.3	2.8	1.9	2.4	1.5	2.0	0.9	1.5
65.0-67.5..........	3.2	3.5	2.0	2.6	1.6	2.3	1.0	1.8	0.7	1.0
67.5-70.0..........	2.0	3.3	1.5	2.6	1.2	2.1	1.0	1.5	0.6	1.0
70.0-72.5..........	1.5	3.0	1.2	2.0	1.0	1.8	0.8	1.2	0.5	0.8
72.5-75.0..........	1.3	2.0	1.0	1.5	0.9	1.2	0.8	0.9	0.5	0.8
75.0-77.5..........	1.1	1.8	0.8	1.2	0.6	1.0	0.5	0.8	0.5	0.7
77.5-80.0..........	1.0	1.6	0.5	1.0	0.5	0.9	0.4	0.7	0.4	0.5
80.0-82.5..........	0.9	1.4	0.5	0.8	0.5	0.6	0.4	0.5	0.4	0.5
82.5-85.0..........	0.8	1.3	0.5	0.5	0.5	0.5	0.4	0.4	0.3	0.4
85.0-87.5..........	0.7	1.3	0.5	0.5	0.4	0.4	0.3	0.3	0.2	0.2
87.5-90.0..........	0.6	1.2	0.5	0.5	0.4	0.4	0.3	0.3	0.2	0.2
90.0-92.5..........	0.6	0.8	0.5	0.5	0.4	0.4	0.3	0.3	0.2	0.2

The integral in equation (3) represents the cumulative number of deaths among individuals infected by HIV since the start of the epidemic. The function $M(t)$ is the instantaneous probability of dying at time t by all causes (AIDS or other causes) and is given by:

$$M(t) = \left(\mu + \frac{\alpha t^{\alpha-1}}{\beta^{\alpha}} \right) \exp\left[-\mu t - \left(\frac{t}{\beta} \right)^{\alpha} \right] \tag{7}$$

That is, the probability of dying is modelled as a Weibull density function with shape parameter α and position parameter β. In equation (7), μ represents the force of mortality due to causes other than AIDS (background mortality).

With the model just specified, four essential parameters r, f_0, Φ, t_0 (the time of the start of the epidemic) are then estimated from the empirical prevalence data. All other information captured in the model is assumed to be known or is set to plausible values.

b. Models of adult HIV/AIDS

The age-specific infection pattern for the adult population has been parameterized as a Weibull function with the following parameters

TABLE VI.7. PARAMETERS FOR ADULT HIV INFECTION PATTERNS BY SEX

Parameter	*Males*	*Females*
Alpha	1.51	1.45
Beta.......................	17.91	14.34
Median..................	31.2	28.0
Mean.....................	29.1	26.1

The net adult survival patterns describes the survival of an infected person from infection to the onset of full blown AIDS, that is from clinical stage 1 to clinical stage 3. It was modelled by a Weibull function, for each sex separately and for both sexes combined.

TABLE VI.8. PARAMETERS FOR ADULT INCUBATION PERIOD BY SEX

Parameter	*Males*	*Females*	*Both sexes combined*
Alpha...........	2.17	2.66	2.42
Beta.............	10.18	10.81	10.48
Median	8.6	9.4	9.0
Mean	9.0	9.6	9.3

The net survival in clinical stage 4, that is from the onset of full blown AIDS to death, is modelled by a Weibull function with parameters alpha and beta as shown in table VI.9.

TABLE VI.9. PARAMETERS FOR SURVIVAL FROM FULL-BLOWN AIDS TO DEATHS, BY TREATMENT STATUS

		Annual survival with ART treatment	
Parameter	*No treatment*	*80 per cent*	*90 per cent*
Alpha..........	1.00	1.00	1.00
Beta............	1.00	4.50	9.50
Median.......	0.7	4.5	9.5
Mean..........	1	3.1	6.6

It should be noted that the above function is equivalent to a simple exponential function, since the Weibull function becomes an exponential function when alpha is set to 1.0. In the practical implementation of the model in abcDIM, Weibull functions are used throughout abcDIM.

c. Models of paediatric HIV/AIDS

Survival of children infected at birth with HIV is modelled with a double Weibull function (Marston et al. 2005):

$$s(x) = 1 - \left[p(1 - \exp(-(\beta_1 x)^{\alpha_1})) + (1-p)(1 - \exp(-(\beta_2 x)^{\alpha_2})) \right] \tag{8}$$

In (8), subscript 1 denotes the fast progressor group and subscript 2 the slow progressor group (see VI.C.2.c, step 5). Infected children that follow the fast progression schedule have a median survival time of just 0.8 years, while children that follow the slow progression schedule have a median survival time of 8.8 years. Together, this function implies a median survival time for all infected children of about 2 years; that is, about half of all infected children will not survive to their second birthday. The current survival model for children also implies that none of the infected children will survive past age 15, and that there is no significant difference between the male and female child. Table VI.11 shows the parameters of the two Weibull function for child mortality, adapted from the model fitted to all data in (Marston et al, 200, p. 225).

TABLE VI.10. PARAMETERS FOR MODELLING THE EFFECTS OF TREATMENT AND BEHAVIOURAL CHANGE

	Antiretroviral treatment coverage		*Halving of MTCT level*	*People on ART surviving annually*
	2004[1]	*2015*		
Major area, country or area	*Per cent*		*Years*	*Per cent*
Africa				
1 Angola	10	70	6	80
2 Benin	17	70	6	80
3 Botswana	50	80	4	80
4 Burkina Faso	7	70	6	80
5 Burundi	9	70	6	80
6 Cameroon	14	70	6	80
7 Central African Republic	1	40	10	80
8 Chad	…	40	10	80
9 Congo	…	40	10	80
10 Côte d'Ivoire	5	40	10	80
11 Dem. Republic of the Congo	2	40	10	80
12 Djibouti	15	70	6	80
13 Equatorial Guinea	…	40	10	80
14 Eritrea	…	40	10	80
15 Ethiopia	5	40	10	80
16 Gabon	29	80	4	80
17 Gambia	14	70	6	80
18 Ghana	4	40	10	80
19 Guinea	4	40	10	80
20 Guinea-Bissau	…	40	10	80
21 Kenya	13	70	6	80
22 Lesotho	5	40	10	80
23 Liberia	…	40	10	80
24 Madagascar	0	40	10	80
25 Malawi	8	40	10	80
26 Mali	…	40	10	80
27 Mozambique	4	40	10	80
28 Namibia	28	80	4	80
29 Niger	…	40	10	80
30 Nigeria	2	40	10	80
31 Rwanda	18	70	6	80
32 Sierra Leone	…	40	10	80
33 South Africa	7	40	10	80
34 Sudan	…	40	10	80
35 Swaziland	16	70	6	80
36 Togo	12	70	6	80

TABLE VI.10 (*continued*)

	Major area, country or area	*Antiretroviral treatment coverage*		*Halving of MTCT level*	*People on ART surviving annually*
		2004[1]	*2015*		
		Per cent		*Years*	*Per cent*
37	Uganda	52	80	4	80
38	United Republic of Tanzania	1	40	10	80
39	Zambia	13	70	6	80
40	Zimbabwe	3	40	10	80
Asia					
1	Cambodia	23	80	4	80
2	China	7	50	10	80
3	India	4	50	10	80
4	Myanmar	3	50	10	80
5	Thailand	44	80	4	90
Latin America and the Caribbean					
1	Bahamas	25	80	6	90
2	Barbados	64	85	4	85
3	Belize	39	85	4	80
4	Brazil	88	85	4	90
5	Dominican Republic	7	70	10	80
6	Guatemala	30	80	6	80
7	Guyana	28	80	6	80
8	Haiti	8	70	10	80
9	Honduras	30	80	6	80
10	Jamaica	18	80	6	80
11	Suriname	25	80	6	80
12	Trinidad and Tobago	16	80	6	80
More developed countries					
1	Russian Federation	3	50	10	90
2	Ukraine	2	50	10	90
3	United States of America[2]	98	98	-	95

[1] As of December 2004. *Source*: WHO (2005).

[2] Mother-to-child transmission declined drastically in the 1990s due to administration of zidovudine. It was set to level off at 2 per cent in 2010 and to stay constant thereafter.

TABLE VI.11. PARAMETERS FOR CHILD SURVIVAL FUNCTION

Parameter	*Fast progressors*	*Slow progressors*	*Combined*
Shape parameter α[1]	1.13	3.75	-
Position parameter β[2]	1.15	9.59	-
Proportion p[3]	0.58	0.42	1.00
Median (years)	0.8	8.7	2.1
Mean (years)	1.1	8.7	4.3

[1] The shape parameter is denoted λ in Marston et al, 2005. Note that $\alpha=1/\lambda$.

[2] The position parameter is denoted with μ in Marston et al, 2005.

[3] The proportion of children in the rapid progression group is denoted with π in Marston et al 2005.