Demographic Techniques: Data Adjustment and Correction

Heather Booth, Australian Demographic and Social Research Institute, The Australian National University, Canberra, Australia
Patrick Gerland, Population Division, United Nations, NY, USA

The views and opinions expressed in this article are those of the authors and do not necessarily represent those of the United Nations.
This article is a revision of the previous edition article by H. Booth, volume 5, pp. 3452–3456, © 2001, Elsevier Ltd.

Abstract

Demographic data suffer from sampling errors and from biases arising from coverage and content errors that may be systematic and noncompensating. Common and problematic errors for demographic estimation are those affecting the reporting of age, parity, and deaths. Age misreporting affects population counts and vital rates. Techniques of data evaluation and correction relying on individual-level analysis include the postenumeration survey, imputation, capture–recapture methods, and statistical analysis. Techniques using aggregate-level data employ digit preference indices, sex and age ratios, and smoothing; demographic accounting and internal consistency; and parametric functions, relational models, statistical modeling, and time-based methods.

Demographic data suffer from sampling errors and from nonsampling errors related to coverage and content. Nonsampling errors are often systematic and noncompensating, introducing biases in the same direction rather than random. Such errors are more common or more serious in populations with lower levels of education and economic development. Errors in the reporting and classification of age, parity (number of live-born children), and deaths are the most common and most problematic for demographic measurement and estimation. Techniques of data evaluation and correction may rely on re-enumeration and matching of individual records, data editing, and individual-level statistical analysis and modeling. Alternatively, analytical methods based on population-level aggregated data may be applied, using mathematical formulae, statistical modeling, and the relationship between demographic processes (birth, death, and migration) and age distribution.

Types of Errors

Sampling Errors

Issues such as sampling errors, sample size, and representativity apply to data collected through sample surveys. This increasingly includes data traditionally collected through the census, because many countries (e.g., USA, France, China) are turning to sampling techniques to avoid the cost of complete coverage.

One of the critical factors in determining sampling error is the sample size: the larger the sample, the smaller the sampling error. But the overall size of the population surveyed, cost, and logistical issues often limit the sample size. In a simple random sample, doubling of the sample size will decrease sampling error by 30%. To double the accuracy requires quadrupling the sample size, incurring associated costs. Stratification and more complex sample designs can also help to improve the accuracy of the results (ICF International, 2012b; United Nations, 2005; United Nations, 2008a).

Sampling errors need to be accounted for when analyzing the data. Differences in means or proportions between subgroups, for instance, should always take into account standard errors to determine whether differences are statistically significant within a margin of error. The increasing availability of statistical software able to handle complex sample designs makes it easier to undertake analyses that appropriately take standard errors into account (Deaton, 1997).

The overall size of the population under study is also relevant for the computation of standard errors of vital rates by age (and sex). Random fluctuations in events in small populations (often due to disaggregation by particular characteristics or by small areas) can be addressed by pooling multiple years to ensure a sufficient number of person-years of observation. This is often necessary when computing life tables for small areas or subpopulations (Eayres and Williams, 2004; Scherbov and Ediev, 2011; *see* Life Table; Demographic Techniques: Small Area Estimates and Projections). Pooling of data is also useful for sample data when sample sizes are small (Dwyer-Lindgren et al., 2013; Pedersen and Liu, 2012).

More general issues related to sampling theory and sample design, such as the computation of sampling errors and design effects, are beyond the scope of this section and are addressed in sampling statistics textbooks (Kish, 1965; Levy and Lemeshow, 2013; Lohr, 2009; Som, 1996).

Errors of Coverage

Coverage errors are nonsampling errors and result in bias, affecting the representativity of the data. Coverage errors may occur in data collected through both censuses and sample surveys.

In censuses, errors of coverage comprise the underenumeration or (less commonly) the overenumeration of individuals in the population. Underenumeration occurs when areas, settlements, or households are erroneously omitted from the census, and when individuals within (otherwise covered) households are not enumerated. Outdated or incomplete maps or housing listings, a shortage of trained and experienced interviewers, inadequate field supervision, extreme remoteness or inaccessibility, political or security problems, conflict zones,

 http://dx.doi.org/10.1016/B978-0-08-097086-8.31011-X

difficulties in locating nomadic people and those living in slums or refugee camps, refusals to be enumerated, and loss of census records after enumeration may all contribute to the underenumeration of households. Short-term migrants and infants are among household members most likely to be omitted from enumeration. Overenumeration may arise from double counting, which can occur as a result of poor planning and management of fieldwork leading different interviewers to enumerate the same household or as a result of migrants and more mobile persons (e.g., household members who are temporarily absent, traveling, studying or working elsewhere) being reported both at their household and at their place of visit, work, or study (United Nations, 2008b). Overenumeration also occurs when ineligible persons (e.g., visitors or nonresident aliens) are enumerated as if eligible. Sex-selective underreporting of girls, especially when reporting is by proxy, may also occur in settings favoring sons (e.g., parts of South Asia), especially in the context of restrictive child policy (e.g., China).

Surveys are affected by similar problems. Sample underrepresentation of the population may arise from use of an outdated or inadequate sampling frame for the selection of primary sampling units (usually households), particularly in countries lacking a census in the past decade or where population change has been rapid. Similar sampling bias occurs when a sampling frame systematically excludes some population members (e.g., when a reproductive health survey samples only family housing units, thereby excluding women living in institutions or collectives) or where the probability of selection is lower for certain population members (e.g., because of cultural differences in defining who is a household member (Randall et al., 2013, 2011)). In reproductive health surveys, these issues can lead to single women being omitted (Hull and Hartanto, 2009), or to the overrepresentation of women with more children (Avery et al., 2013), upwardly biasing fertility. The extent of the resulting bias will depend on the difference in the variable of interest between those omitted from and included in the sample, and on their fractions of the total population.

Moreover, the representativity of the sample can be compromised by selection biases during fieldwork (United Nations, 2008a). For example, interviewers may predominantly visit households that are easier to reach (along main roads, for instance) especially in conflict zones (Johnson et al., 2008), or they may miss households that have been recently dissolved due to mortality or selective out-migration of the population at risk in zones experiencing conflicts or crises (Walker et al., 2012), leading to biased mortality estimates, *see* Selection Bias, Statistics of. In addition, many surveys restrict some questions to subgroups of individuals based on eligibility criteria such as sex and age (e.g., females aged 15–49 for questions on reproduction) or duration (e.g., vital events in the last 5 years). Interviewers often need to assess eligibility when respondents either do not know (even after probing) or when values are close to the eligibility limits (ICF International, 2012a), because of the implications in terms of extra workload (i.e., additional sets of questions to collect on supplementary topics). The interpretation by interviewers of selection filters can lead to the omission of potential respondents or events and to shifting of dates or ages (ICF Macro 2009; Pullum, 2006; Silva, 2012; Sullivan, 2008).

Incomplete coverage is also a common deficiency of vital registration data. Further, since year of registration rather than year of event is the usual basis of compilation, late reporting adds to the inaccuracies involved. Births are more affected by late reporting because registration is often not required until children reach school age. Death registration, where it occurs, is usually fairly timely since registration is often a legal requirement for burial; though in some populations such requirements are often ignored, notably in rural areas, leading to underregistration rather than late registration.

Nonresponse is a further source of bias and nonrepresentativity (Groves and Couper, 1998). Item nonresponse occurs when only partial information is collected for an individual. For most variables, nonresponses are omitted from analysis, only their extent being reported; this implicitly assumes that respondents and nonrespondents do not differ with respect to the item in question. For age, nonresponses are often distributed on a *pro rata* basis for each sex because of the need to maintain population size in demographic measurement; this also assumes no difference in the age distributions of respondents and nonrespondents. These assumptions may be erroneous and care is needed in the interpretation of results, especially if nonresponse is substantial.

Errors of Content

Misclassification and content error may affect all demographic variables. These errors arise from incorrect substantive reporting on the part of the respondent and/or interviewer, or from mistakes during data processing. Some errors may be caused by specification error, including poorly phrased questions or instructions, ambiguity of interpretation, and respondent misunderstanding or inability to distinguish key concepts, for example, respondent's biological children from children of brothers or sisters living in the same household. Other errors may be related to the mode of data collection (e.g., length and place of interview, order of questions, whether through interview or self-administered questionnaire, etc.) and interviewer effects (e.g., interviewer errors in phrasing questions or transcribing responses). Errors on the part of the respondent may include digit preference, telescoping and memory lapse related to long recall periods, deliberate misreporting, and errors due to proxy responses. Finally, data processing errors may occur during coding, data entry, editing, analysis, or dissemination (United Nations, 1982; United Nations, 2005).

Assessment and Adjustment of Individual-Level Data

The evaluation and adjustment of individual-level data may be achieved through the use of postenumeration checks, data editing, and individual-level statistical analysis and modeling.

Both coverage and content of surveys and, in particular, censuses are assessed through the postenumeration survey

(PES). The PES involves the recollection of data for a representative selection of the original population or sample. The two sets of data are matched for each individual enumerated in the PES (Alho and Spencer, 2006; United Nations, 2010b; United States. Bureau of the Census. 1985). Similarly, event capture–recapture and matching methods, using two or more independent data sources for a representative sample of cases, are used to estimate the completeness of birth and death registration or of retrospective reports from censuses or surveys (Amstrup et al., 2005; Krótki, 1978; United Nations, 1991). This general approach is regarded as the gold standard, but its cost and complexity often limit its usage. Dual-registration systems (e.g., India's sample registration system (Mahapatra, 2010) and China's disease surveillance point system (Yang et al., 2005)) rely on this approach to derive adjusted sample vital statistics using the Chandrasekaran–Deming formula (1949).

Methods employed to address coverage issues due to unit nonresponse include the use of longitudinal data. Survival analysis and other longitudinal methods such as random- and fixed-effects modeling can deal with issues such as loss to follow-up and censoring (*see* Survival Analysis: Whether, and If So, When?; Panel Surveys: Uses and Applications). In addition, even with cross-sectional data, various multivariate statistical methods can be used to correct for sample selection biases (Stolzenberg and Relles, 1997; Winship and Mare, 1992).

Item nonresponses affecting content can be addressed during data editing or analysis. Missing responses, for instance, may be imputed based on responses from other respondents with similar known characteristics (United Nations, 2010a). In recent years, the greater availability of microdata sets and more advanced multivariate statistical software have made it increasingly easy to apply multiple imputations based on extensive sets of household or individual characteristics (Little and Rubin, 2002; Schafer, 1997), thus allowing missing values to represent multiple sets of plausible values and to analyze data without assuming that nonresponses are missing at random (*see* Multiple Imputation).

Content errors such as digit preference can also be analyzed at the individual level through the use of multinomial logistic regression. An example is Pullum's (2005, 2006) adaptation of the Myers' blended method (Hobbs, 2004; see also below section Evaluation and Adjustment of Population Distribution by Age) for the estimation of interpretable misreporting parameters based on covariates (with standard errors and confidence intervals corrected for sample design). In addition, general patterns of digit preference can be modeled statistically, and the age distribution can potentially be adjusted without imposing any other constraint than a smoothness assumption for the underlying true distribution. For example, Impicciatore and Billari (2011) proposed a general method to estimate smoothed age profiles and relative risks for time-fixed and time-varying covariates using generalized additive models; and Camarda et al. (2008) proposed a method to estimate the unobserved latent distribution, free of the effects of misreporting, using a composite link model with a penalized likelihood.

Until recently, the evaluation of individual-level data had been the prerogative of national statistical offices or other agencies charged with data collection and custody. The increasing access to microdata enables a wider range of users to directly apply and develop methods of evaluation and correction, greatly improving capability in this area.

Assessment and Adjustment of Aggregate-Level Data

Many analytical methods of assessment and adjustment are based on aggregated data. These methods use mathematical formulae, statistical modeling, and the relationship between demographic processes (birth, death, and migration) and the age distribution.

Assessment and Adjustment for Completeness

The completeness of a census enumeration may be assessed, at the aggregate level, through analytical demographic methods using successive censuses and intercensal estimates of the components of population change (Alho and Spencer, 2006; United Nations, 2010b; United States Bureau of the Census, 1985). Similarly, the completeness of birth and death registration (or of retrospective reporting in censuses and surveys) may be assessed at the population level using demographic relationships (*see* Demographic Techniques: Indirect Estimation). For example, vital rates can be compared across independent data sources for similar periods or cohorts, and analyzed for consistency in age-specific trends in rates and sex ratios; such analyses may also be disaggregated by the characteristics of the deceased or, in the case of fertility, of the mother, or of the household (United Nations, 1991). Adjustment factors are usually based on the reciprocal of the estimate of the completeness of vital registration for the total or disaggregated population.

In addition to these approaches, the completeness of adult death registration (or reporting in censuses or surveys) can be evaluated using death distribution or growth-balance methods. These methods require the availability of death data and at least one census, and preferably two or more censuses (Hill et al., 2009; Moultrie et al., 2013; United Nations, 2002). Death distribution methods assume that age is accurately reported, and therefore do not correct for age misreporting, which at older ages usually involves exaggeration in age at death leading to downwardly biased mortality rates (prior to the open-ended age group). These methods address only relative under- and overreporting of registered age-specific deaths against intercensal estimates, which are potentially biased by differential accuracy of census enumeration, and by migration in the intercensal period (Bhat, 2002). In addition, they assume that both the proportion of deaths registered and the completeness of the census enumerations are constant over adult ages. More robust results are usually obtained by combining the results of growth-balance methods and synthetic extinct generations methods (Hill et al., 2009), and using various age combinations (e.g., 20+/50+, 5+/50+, 5+/65+) to minimize the perturbating effects of violations of underlying assumptions in each method (Murray et al., 2010). The resulting estimates of completeness are used to adjust registered deaths.

These estimates of mortality and others obtained from indirect methods (e.g., intercensal cohort survival, orphanhood, siblings, widowhood – *see* Demographic Techniques: Indirect Estimation) can be compared across methods using different data sources. The most reliable and robust estimates can be used as a baseline reference to potentially adjust more deficient data sources (Hill et al., 2005; Moultrie et al., 2013; Timaeus, 1991; United Nations, 2002).

Evaluation and Correction of Parity Data

Parity data, or number of live-born children a woman has borne, are collected through censuses and surveys or at birth registration. Errors in the reporting of parity often involve understatement due to the omission of births, especially births of girls. Analysis of sex ratios of births by parity (of women) can reveal systematic sex-selective understatement due to gender preference (*see* Missing Girls: A Globalizing Issue). Where adoption is common, overstatement may occur due to both the natural and adoptive mother reporting the birth (Moultrie, 2013a). The inclusion of stillbirths as live births also results in slight parity overstatement.

Omissions of births may be detected and corrected by indirect estimation (*see* Demographic Techniques: Indirect Estimation). Further error arises when childless women are recorded as 'parity not stated' due to the ambiguous use of the dash symbol (−) by interviewers. If these apparent not-stated cases are excluded, average parity is overestimated. If all not-stated cases are included and treated as childless, average parity is underestimated. The El-Badry correction estimates the proportion of women with parity not stated, who are in fact childless (El-Badry, 1961; Moultrie, 2013b). Based on the observed high correlation between true proportions of childless, $Z^*(i)$, and proportions of parity not stated, $NS(i)$ where $i = 1$ to 8 represents the age groups 10–14 to 45–49, El-Badry assumes $NS(i) = \alpha\ Z^*(i) + \beta$, where α is the proportion of truly childless women who are erroneously recorded with parity not stated and β is the constant true proportion of parity not stated. Since the reported proportions of childless, $Z(i)$, equal $(1 - \alpha)\ Z^*(i)$, then $NS(i) = [\alpha/(1 - \alpha)]Z(i) + \beta$, and $\alpha/(1 - \alpha)$ and β can be estimated by ordinary least squares based on the first four or five age groups. Then, β is used to estimate the number of women in each age group for analytical purposes, and true proportions childless are estimated by $Z^*(i) = Z(i) + NS(i) - \beta$. This correction is based on all women. If data on parity are available only for ever-married women (in which case all never-married women are generally assumed childless), the procedure can be applied on the assumption of a linear relation between the proportions of ever-married women who are childless and who have parity not stated.

Evaluation and Adjustment of Population Distribution by Age

Errors in the reporting of age can affect population counts and vital events. Age misreporting includes digit preference (or age heaping) and age under- and overstatement, which may occur together. Digit preference occurs when real age is not known and its estimation involves the 'rounding' of age (or year of birth) to end in a preferred digit such as 0 or 5. Digit preference is best detected graphically. Figure 1, for example, shows persistent age heaping across the last five censuses of India.

Various indices have been devised to measure the extent of digit preference in population counts, but they do not offer a means of correction (Hobbs, 2004). Whipple's index measures the preference or avoidance of the digits 0 and 5 over

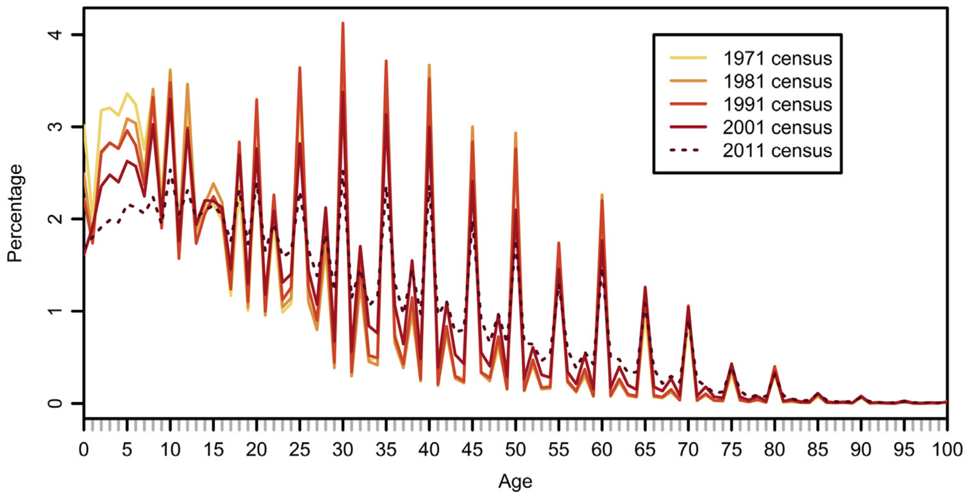

Figure 1 Percentage distribution by single age of the population (both sexes combined) enumerated in the 1971, 1981, 1991, 2001, and 2011 censuses of India. Source: Data from (a) United Nations Statistics Division, 2014. Demographic Yearbook. Population Censuses' Datasets, Population by Age, Sex and Urban/Rural Residence Tabulation. http://unstats.un.org//unsd/demographic/products/dyb/dybcensusdata.htm (accessed 20.01.14.) and (b) Government of India, Office of the Registrar General & Census Commissioner, 2014. 2011 Census Final Results, India. Single Year Age Data – C13 Table (India-States-UTs). http://www.censusindia.gov.in/2011census/Age_level_data/Age_level_data.html (accessed 09.01.14.). Computations by authors. Figure made using R 3.0.2.

the age interval 23–62, though other multidecennial intervals can be used. Myers' blended index measures preference or avoidance of all 10 digits. Since neither index is theoretically precise, Bachi derived a composite index without this deficiency. Whipple's and Myers' indices assume rectangularity or linearity within quinquennial age groups. Later indices, developed by Carrier and by Ramachandran, take account of age structure and variable digit preference over the age range. In practice, the Bachi, Carrier, and Ramachandran indices are laborious and are used less frequently. Given the focus of Whipple's index on 0 and 5, its value in reflecting improvements in the quality of age reporting diminish as digit preference becomes less common (Spoorenberg, 2009). The Myers' blended index and the total modified Whipple's index provide better assessments of the overall spatial and temporal variability of age reporting (Spoorenberg, 2007).

Age misstatement generally involves a directional tendency often related to cultural factors (Ewbank, 1981; Pullum, 2006). Age understatement may occur among women of lower parity than average for their age, especially in cultures with high fertility. Age overstatement often occurs at older ages, particularly among males, in cultures where status is gained with age. Other sex-specific age misstatement may also occur. For example, young female children generally are reported as younger than their male peers. Inaccuracies in the reporting of age tend to be more accentuated in vital events than in population counts. In the case of deaths, age exaggeration may partly be related to proxy reporting. Overstatement in age at death can lead to substantial underestimation of mortality at older ages (Coale and Li, 1991; Condran et al., 1991; Preston et al., 1999) and even to abnormal crossovers and paradoxes between population subgroups when such errors affect more systematically some subgroups than others (Elo and Preston, 1997; Elo et al., 2004; Preston et al., 1996; *see* Mortality Crossover).

In many cases, it is impossible to distinguish between age misstatement, differential coverage by age, and digit preference. Further, one error may offset another. For population counts, the pattern of overall error may be seen by examination of age ratios and sex ratios for 5- or 10-year age groups (Hobbs, 2004; Pullum, 2006). **Figure 2** illustrates these deviations for the 1991 census of India. Panel A shows the age distributions by sex. Panel B shows how sex ratios by age differ from sex equality (sex ratio = 100), indicating distortion due to age misreporting (given sex differences in migration and mortality). Panel C shows age ratios by sex, contrasting over- and understatement of age by age group. These sex and age ratios may also be compared with ratios from suitable model life table populations or smoothed age distributions (as seen in **Figure 2**).

An indication of the overall accuracy of age reporting is given by the United Nations age-sex accuracy index (or 'joint score') (United Nations, 1955). This index assumes no significant irregularities by age (i.e., limited change in past birth cohort size) and sex parity at all ages (implying equal numbers of male and female births and equal mortality). Since the index assumes that all irregularities are errors, it should be interpreted with caution in populations affected by historical events (e.g., baby booms or baby busts, age–sex-specific migration, or mortality shocks or crises).

Errors may also be assessed through comparison with expected counts and distributions based on external data using demographic relationships (Pollard et al., 1990; Preston et al., 2001b). When accuracy permits, a previous enumeration of a population can be 'updated' using data on births, deaths, and net migration to derive an expected age distribution using a cohort-component population projection (Himes and Clogg, 1992; United States Bureau of the Census. 1985). Consistency between the enumerated population at older ages (e.g., age 80 and over) and recorded deaths can be enforced by using the (almost) extinct generation method when death registration records are considered more reliable than census information (Andreev, 2004; Machemedze and Dorrington, 2011; Wilmoth et al., 2007). Comparison of the cumulative age distribution with model stable distributions may identify distortions due to misreporting, in the absence of substantial migration and assuming no significant changes in past fertility and mortality. A suitable model stable population can also be used to derive the structure of the open-ended age group when it is unknown (United Nations, 1983; *see* Population Dynamics: Theory of Stable Populations). When fertility and mortality have been declining in recent decades (as in many countries), the availability of two or more censuses can help to assess the accuracy of age distributions by examining cohort consistency between censuses, and by allowing the assumption of stability to be relaxed by iteratively fitting projected population structures to the observed numbers (Bhat, 1990). Using successive censuses, cohorts can be compared at different points in time: allowing for mortality, the (smoothed) age distributions enable the detection of possible enumeration inconsistencies. **Figure 3** shows the detection of missing infants and young adults in the 1991 census of India.

The effects in demographic measurement and estimation of both digit preference and age misstatement can be reduced by collecting exact date (or at least year) of event rather than age and by using wider (than standard 5-year) age intervals at the analysis stage, though this is not always practicable and involves greater approximation. Similarly, the effects on standard quinquennial age groups can be reduced by adopting nonstandard age groups, the optimum choice of which is based on Myers' index (Hobbs, 2004), but these are generally inconvenient for analytical and comparative purposes.

For many purposes it may be appropriate to smooth or graduate the age distribution. Methods of smoothing include the use of graphs (manual smoothing), moving averages, curve fitting, and statistical modeling. Mathematical graduation is used to smooth narrower age groups within broader age groups. The Carrier–Farrag and Karup–King–Newton formulae treat all age groups the same, whereas the Arriaga's formulae treat the first and last differently (Arriaga et al., 1994). In the following formulae, ${}_nP_x$ represents the population aged x to $x + n - 1$, x and n being integers. The Carrier–Farrag ratio method defines the relationship of a 5-year age group to the 10-year age group of which it is a constituent as

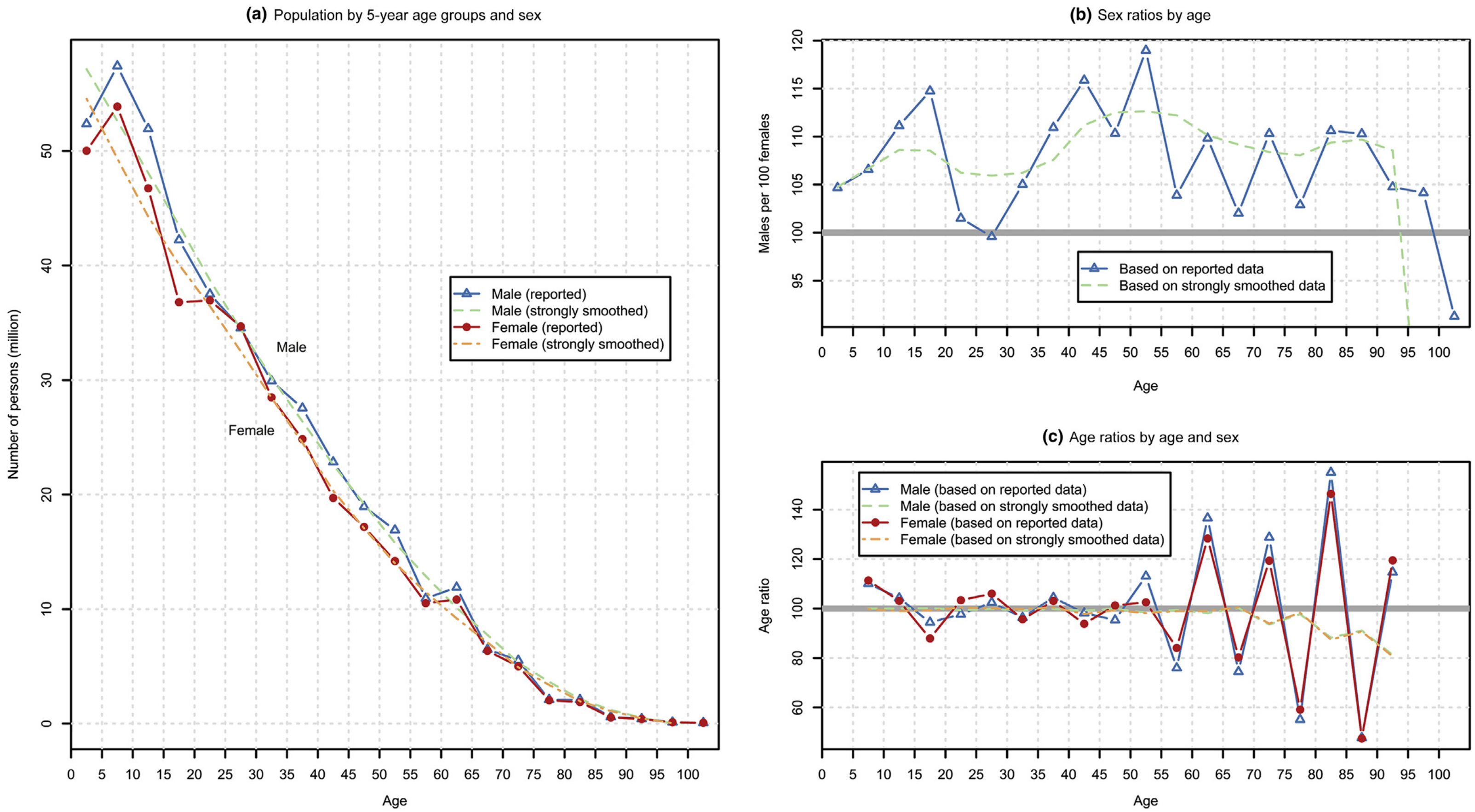

Figure 2 Analysis of the distribution by 5-year age group and sex of the 1991 population census of India. Panel (a) presents for each sex and 5-year age group the reported (solid line with symbol) and heavily smoothed (dashed line) distributions. Panel (b) shows corresponding sex ratios by 5-year age group in comparison with assumed sex equality (sex ratio = 1). Panel (c) shows corresponding age ratios for each sex, calculated as the ratio of one age group to the average of the two adjacent age groups. Source: Data from United Nations Statistics Division, 2014. Demographic Yearbook. Population Censuses' Datasets, Population by Age, Sex, and Urban/Rural Residence Tabulation. http://unstats.un.org//unsd/demographic/products/dyb/dybcensusdata.htm (accessed 20.01.14.). Computations by authors based on U.S. Bureau of the Census, International Programs Center, 2013. Population Analysis Spreadsheets (PAS): AGESEX and AGESMTH. http://www.census.gov/population/international/software/pas/ (accessed 30.01.14.). Figure made using R 3.0.2.

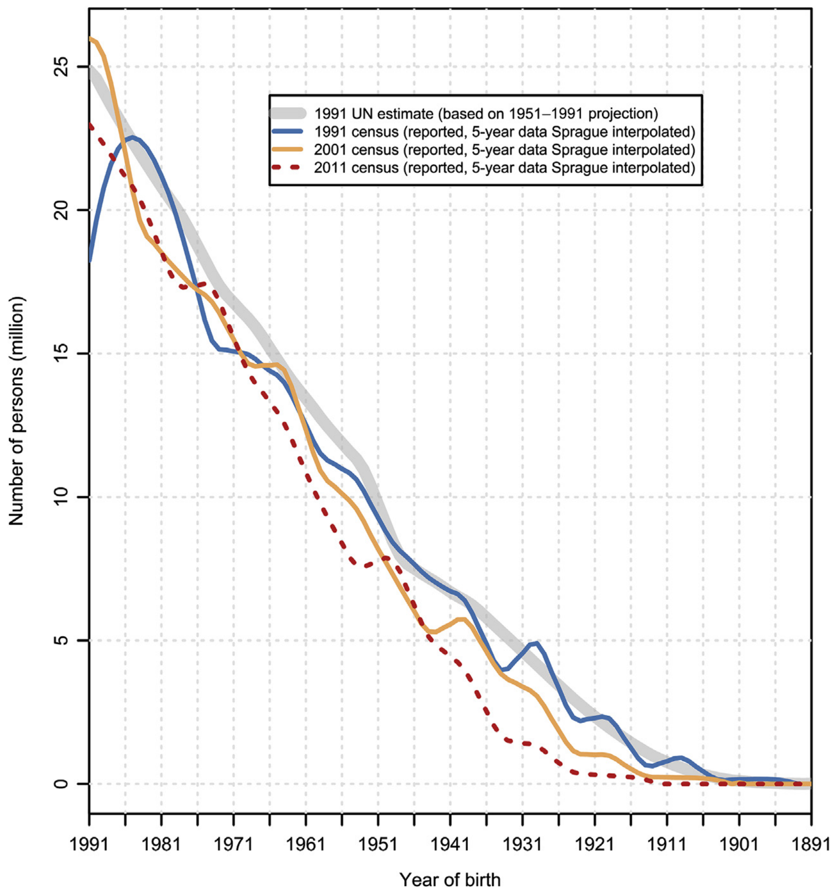

Figure 3 Evaluation of the distribution by age of the 1991 census population of India (both sexes combined). Census data are smoothed using Sprague interpolation with 5-year age groups. The 1991 age distribution (blue solid line) is compared with the UN estimate for 1991 based on a 1951–1991 cohort-component population projection (gray bold solid line), and with the corresponding birth cohorts enumerated in the 2001 census (orange solid line) and 2011 census (red dotted line). Source: Data from (a) United Nations Statistics Division, 2014. Demographic Yearbook. Population Censuses' Datasets, Population by Age, Sex, and Urban/Rural Residence Tabulation. http://unstats.un.org//unsd/demographic/products/dyb/dybcensusdata.htm (accessed 20.01.14.) and (b) Government of India, Office of the Registrar General & Census Commissioner, 2014. 2011 Census Final Results, India. Single Year Age Data – C13 Table (India-States-UTs). http://www.censusindia.gov.in/2011census/Age_level_data/Age_level_data.html (accessed 09.01.14.). Computations by authors. Figure made using R 3.0.2.

$$ {}_5P_{x+5} = {}_{10}P_x \Big/ \left[1 + ({}_{10}P_{x-10}/{}_{10}P_{x+10})^{1/4}\right] \quad \text{and} $$
$$ {}_5P_x = {}_{10}P_x - {}_5P_{x+5} $$

The Karup–King–Newton formula assumes a quadratic relationship between three consecutive 10-year age groups, that is, ${}_5P_x = \frac{{}_{10}P_x}{2} + \frac{({}_{10}P_{x-10} - {}_{10}P_{x+10})}{16}$. The Arriaga's formulae are based on second-degree polynomials passing through the midpoints of three consecutive 10-year age groups. Then

$$ {}_5P_{x+5} = \frac{(8\,{}_{10}P_x + 5\,{}_{10}P_{x+10} - {}_{10}P_{x+20})}{24} $$
when the 10-year age group is the first,
$$ {}_5P_x = \frac{(-\,{}_{10}P_{x-20} + 5\,{}_{10}P_{x-10} + 8\,{}_{10}P_x)}{24} $$
when it is the last, and
$$ {}_5P_{x+5} = \frac{(-\,{}_{10}P_{x-10} + 11\,{}_{10}P_x + 2\,{}_{10}P_{x+10})}{24} \quad \text{otherwise} $$

In all cases, the second 5-year age group is derived by subtraction with ${}_5P_x = {}_{10}P_x - {}_5P_{x+5}$

When age misreporting is more severe, the totals of all age groups need potential modification. The United Nations formula lightly smooths 5-year age groups, based on five consecutive 5-year groups, and gives the smoothed population aged x to $x+4$ as

$$ {}_5P'_x = \frac{(-\,{}_5P_{x-10} + 4\,{}_5P_{x-5} + 10\,{}_5P_x + 4\,{}_5P_{x+5} - {}_5P_{x+10})}{16} $$

For heavier smoothing, three consecutive 10-year age groups are used to derive the central smoothed age group as

$$ {}_{10}P'_x = \frac{({}_{10}P_{x-10} + 2\,{}_{10}P_x + {}_{10}P_{x+10})}{4} $$

Neither formula smooths the youngest and oldest age groups. Since smoothing changes overall population size, the smoothed age groups are adjusted proportionately to achieve

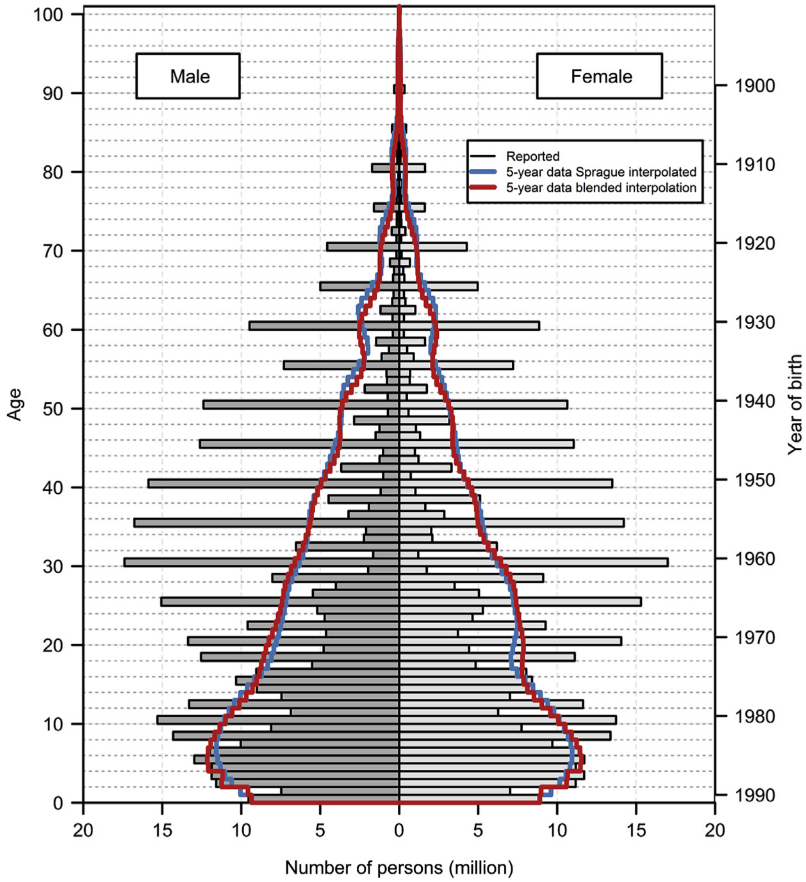

Figure 4 Analysis of the distribution by single age and sex of the 1991 census of India. Reported data by age and sex (solid gray bars) show digit preference. Smoothed data based on 5-year age groups using Sprague interpolation (solid blue line), and a blended osculatory interpolation using the five possible age groupings (red solid line) are included as reference. Source: Data from (a) United Nations Statistics Division, 2014. Demographic Yearbook. Population Censuses' Datasets, Population by Age, Sex, and Urban/Rural Residence Tabulation. http://unstats.un.org//unsd/demographic/products/dyb/dybcensusdata.htm (accessed 20.01.14.) and (b) Government of India, Office of the Registrar General & Census Commissioner (2014). 2011 Census Final Results, India. Single Year Age Data – C13 Table (India-States-UTs). http://www.censusindia.gov.in/2011census/Age_level_data/Age_level_data.html (accessed 09.01.14.). Computations by authors. Figure made using R 3.0.2 and Pyramid-package by Tim Riffe. https://github.com/timriffe/Pyramid/ (accessed on 30.01.14.)

the enumerated total. The Arriaga's (or other such) formulae are then applied to smoothed and adjusted 10-year age groups to derive smoothed 5-year age groups. An example of heavier smoothing is seen in Figure 2.

A single-year age distribution may be derived from grouped data by adopting the within-group distribution of a suitable unabridged life table population, but discontinuities at age-group junctions are inherent in this approach. A smoother distribution is obtained by mathematical interpolation, such as osculatory interpolation, which is especially useful for irregular data because of the constraint of continuity at junctions. Sprague multipliers employ fourth- and fifth-difference osculatory interpolation to smooth single-year data within 5-year age groups without changing the totals. They are less satisfactory for the first and last two age groups and where there is marked curvature (see Figure 4). Sprague multipliers can be employed with any continuous data including rates, for any age partition (e.g., tenths), and with unequal age intervals. A blended distribution can be derived by averaging smoothed distributions obtained by osculatory interpolation within the five possible age groupings with terminal digits 0–4, 1–5, 2–6, 3–7, 4–8 (Figure 4).

Spline interpolation is a more sophisticated type of osculatory interpolation that has become more widely used with increasing computing power. Spline interpolation involves extensive iteration resulting in a highly smoothed distribution (Judson and Popoff, 2004). The approach is very flexible, can be used on counts, proportions, rates, and ratios, and is applicable to any age groupings including those of unequal length. Monotonicity can be imposed using penalized splines (Ligges, 2013; Schmertmann, 2012; Smith et al., 2004); and weighted splines allow for unequal variance by age (Booth et al., 2014). More robust results may be obtained by application to cumulative distributions, with differencing of the interpolated results (Wilmoth et al., 2007).

Adjustment of Age Patterns in Demographic Processes

In addition to spline interpolation of rates by age (Section Evaluation and Adjustment of Population Distribution by

Age), specific analytical methods are available for the smoothing and adjustment of rates or age distributions. These methods include relational models and predefined mathematical functional forms.

Relational models are used when the data suffer from biases due to reporting error. An empirical standard or reference distribution is used to adjust the data. The observed distribution is related to the standard through a two-parameter linear model after applying a suitable transformation (e.g., Logit) to both distributions. Thus, the adjusted distribution is the best-fitting (over the age range employed) linear function of the standard.

For mortality, the Brass Logit life-table system involves fitting observed data to a standard model life table (Brass et al., 1968; *see* Demographic Models; Stochastic Models; Population Forecasts). A larger number of parameters help to address mortality deviations from the standard at young and old ages (Ewbank et al., 1983; Rachet et al., 2008; Zaba, 1979). For fertility, the Brass Relational Gompertz model is used with an appropriate standard, such as the Booth Standard for high fertility (Booth, 1984; Brass, 1974; Moultrie, 2013c; *see* Demographic Models). The adjustment of immigration and emigration data can also be achieved through relational models (Zaba, 1985).

Relational models are flexible and easy to apply, but the adjusted distribution is influenced by the choice of the reference standard (Moultrie, 2013d). Ideally, a more context-specific standard is used (De Beer, 2012; INDEPTH Network. 2004; Murray et al., 2003; Wang et al., 2012; Wilmoth et al., 2011). To reduce dependence on the chosen standard, the relationship between observed and standard fertility or mortality may be modeled by splines (De Beer, 2011, 2012).

Mathematical functions are used in demographic modeling and data smoothing. A parametric function is fitted to the empirical observations, imposing a functional form on the data. This approach is advantageous when data availability is limited to certain ages, or when the data are too noisy for nonparametric smoothing to provide a plausible functional form. Adult mortality rates, especially at older ages, are often smoothed and potentially adjusted using the Gompertz or Makeham functions (Heuveline and Clark, 2011), typically fitted to adult mortality rates and used to smooth and extrapolate mortality at older ages. For observed mortality at the oldest ages (e.g., 85 and over), additional smoothing or adjustment may be achieved using the Kannisto model of old-age mortality (Thatcher et al., 1998) fitted to data at age 75 and over. Adjustments to old-age mortality ensure consistency (Condran et al., 1991) by age (e.g., monotonic increase), over time (e.g., monotonic decline) and between the sexes (e.g., male rates greater than or equal to female rates). Where mortality rates must be smoothed and adjusted at all ages, additional terms are required to model infancy and childhood, and young adulthood. Several parametric functions serve to model the whole age pattern of mortality (e.g., Heligman and Pollard (1980), Kostaki (1992), Siler (1983), Carriere (1992); see reviews by Gage and Mode (1993) and Booth and Tickle (2008)).

Parametric functions are also used to model the age patterns of nuptiality and fertility (Preston et al., 2001c); see review by Booth (2006). The Coale–McNeil (Coale and McNeil, 1972) parameterization of nuptiality applies to first marriage (Coale, 1971; Hernes, 1972) and can be used to model first births (Bloom and Trussell, 1984). The parsimony of these models and the functional shape they impose on the data are useful to smooth and adjust poor-quality data. However, their ability to appropriately model empirical observations depends on their flexibility to capture the range of situations experienced by various countries over time (e.g., the emergence of bimodal fertility distributions by age in low fertility populations (Bermúdez et al., 2012)).

Migration is modeled by the multiexponential function (Rogers and Little, 1994), which is useful in imposing a smooth age pattern on data that are often highly defective; see also Booth (2006). Rogers et al. (2010) demonstrate how to combine smoothing to remove randomness with the use of the multiexponential and empirical schedules to adjust immigration and emigration data.

Adjustments over Time

Demographic series for two or more points in time can be used to evaluate trends, growth, and structural population change. Time series facilitate interpolation and extrapolation. In particular, time series of population counts and the demographic processes of birth, death, and migration jointly enable the derivation of robust estimates.

One of the most common adjustments in demography is shifting the population to refer to a specific date, such as 30 June. This may be achieved in two ways: first by linear or exponential interpolation of each age group between two time points (Arriaga et al., 1994); and second by adjusting the total population exponentially (the growth rate is required) and assuming the same age–sex structure as at the nearest enumeration. This last option is valid only for shifts of short duration (i.e., a few years). For longer extrapolations, a cohort-component population projection with assumptions about fertility, mortality, and migration trends should be used.

Age-specific growth rates can also be employed to assess and potentially adjust age-specific data through demographic relationships in nonstable populations using the variable-r approach (Cai, 2008; Preston and Coale, 1982; Preston et al., 2001a; Schmertmann, 2002; *see* Population Dynamics: Theory of Nonstable Populations). This approach, which may also use duration-specific data, assumes no significant changes in patterns of age misreporting and completeness of coverage between enumerations.

When demographic events in a reference time period are reported retrospectively, errors of timing may occur (e.g., displacement and recall lapse). For births, such errors are often corrected through indirect estimation (*see* Demographic Techniques: Indirect Estimation). Consistency between successive censuses and surveys for overlapping periods can be analyzed visually, as well as through robust statistical modeling of time trends (Pullum, 2006). Inconsistencies can be resolved by the adjustment of indices over age or over time using splines (or locally weighted regressions), with the optional use of covariates and hierarchical statistical modeling to take account of likely biases associated with factors such as method of data collection or length of recall period (Alkema and New, 2013; Alkema et al., 2012). The increased availability of both microdata and internationally comparable public use data sets have made it increasingly possible to perform pooled analyses of multiple datasets for a country, and to analyze all countries simultaneously to detect potential systematic biases and to adjust them if possible. Such statistical modeling approaches are being increasingly used to derive robust time series estimates (with uncertainty bounds), including for child mortality (Alkema and New, 2013; Hill et al., 1999), maternal mortality (Lozano et al., 2011; Wilmoth et al., 2012), adult mortality (Masquelier et al., 2013; Obermeyer et al., 2010; Wang et al., 2012), fertility (Alkema et al., 2012; Machiyama, 2010; Schoumaker, 2011, 2013), and contraception (Alkema et al., 2013). Furthermore, simultaneous estimation of population counts and vital rates (taking into account the underlying uncertainty in each data source) can be used to ensure internal consistency between all demographic components (Luther et al., 1987; Luther and Retherford, 1988; Wheldon et al., 2013a, 2013b; 2013c).

See also: Demographic Models; Demographic Techniques: Indirect Estimation; Demographic Techniques: Small-area Estimates and Projections; Life Table; Missing Data; Nonsampling Errors; Panel Surveys: Uses and Applications; Period and Cohort Analysis in Demography; Population Dynamics: Momentum of Population Growth; Population Dynamics: Theory of Nonstable Populations; Population Dynamics: Theory of Stable Populations; Selection Bias, Statistics of; Survival Analysis: Introduction; Survival Analysis: Whether, and If So, When?.

Bibliography

Alho, J.M., Spencer, B.D., 2006. Errors in census numbers. In: Statistical Demography and Forecasting. Springer, New York, NY, pp. 296–326.

Alkema, L., Kantorova, V., Menozzi, C., Biddlecom, A., 2013. National, regional, and global rates and trends in contraceptive prevalence and unmet need for family planning between 1990 and 2015: a systematic and comprehensive analysis. Lancet 381 (9878), 1642–1652.

Alkema, L., New, J.R., 2013. Global Estimation of Child Mortality Using a Bayesian B-Spline Regression Model. Arxiv.org.

Alkema, L., Raftery, A., Gerland, P., Clark, S.J., Pelletier, F., 2012. Estimating trends in the total fertility rate with uncertainty using imperfect data: examples from West Africa. Demographic Research 26 (15), 331–362.

Amstrup, S.C., McDonald, T.L., Manly, B.F.J. (Eds.), 2005. Handbook of Capture-Recapture Analysis. Princeton University Press, Princeton.

Andreev, K.F., 2004. A method for estimating size of population aged 90 and over with application to the U.S. census 2000 data. Demographic Research 11 (9), 235–262.

Arriaga, E.E., Johnson, P.D., Jamison, E., 1994. Age and sex composition. In: Arriaga, E.E. (Ed.), Population Analysis with Microcomputers, Presentation of Techniques, vol. I. US Bureau of the Census, Washington, DC, pp. 11–56.

Avery, C., St Clair, T., Levin, M., Hill, K., 2013. The 'own children' fertility estimation procedure: a reappraisal. Population Studies. A Journal of Demography 67 (2), 171–183.

Bermúdez, S., Blanquero, R., Hernández, J.A., Planelles, J., 2012. A new parametric model for fitting fertility curves. Population Studies 66 (3), 297–310.

Bhat, P.N.M., 1990. Estimating transition probabilities of age misstatement. Demography 27 (1), 149–163.

Bhat, P.N.M., 2002. General growth balance method: a reformulation for populations open to migration. Population Studies 56 (1), 23–34.

Bloom, D.E., Trussell, J., 1984. What are the determinants of delayed childbearing and permanent childlessness in the United States? Demography 21 (4), 591–611.

Booth, H., 1984. Transforming Gompertz's function for fertility analysis: the development of a standard for the relational Gompertz function. Population Studies 38 (3), 495–506.

Booth, H., 2006. Demographic forecasting: 1980 to 2005 in review. International Journal of Forecasting 22 (3), 547–581.

Booth, H., Hyndman, R., Tickle, L., 2014. Prospective life tables. In: Charpentier, A. (Ed.), Computational Actuarial Science with R. CRC Press, Taylor & Francis Group, Boca Raton, FL, pp. 319–344.

Booth, H., Tickle, L., 2008. Mortality modelling and forecasting: a review of methods. Annals of Actuarial Science 3 (1–2), 3–43.

Brass, W., 1974. Perspectives in population prediction: illustrated by the statistics of England and Wales. Journal of the Royal Statistical Society. Series A (General) 137 (4), 532–583.

Brass, W., Coale, A.J., Demeny, P., Heisel, D.F., Lorimer, F., Romaniuk, A., Van de Walle, E., 1968. The Demography of Tropical Africa. Princeton University Press, Princeton, NJ.

Cai, Y., 2008. An assessment of China's fertility level using the variable-r method. Demography 45 (2), 271–281.

Camarda, C.G., Eilers, P.H.C., Gampe, J., 2008. Modelling general patterns of digit preference. Statistical Modelling 8 (4), 385–401.

Carriere, J.F., 1992. Parametric models for life tables. Transactions of the Society Actuaries 44, 77–99.

Chandra Sekar, C., Deming, W.E., 1949. On a method of estimating birth and death rates and the extent of registration. Journal of the American Statistical Association 44 (245), 101–115.

Coale, A.J., 1971. Age patterns of marriage. Population Studies 25 (2), 193–214.

Coale, A.J., Li, S., 1991. The effect of age misreporting in China on the calculation of mortality rates at very high ages. Demography 28 (2), 293–301.

Coale, A.J., McNeil, D.R., 1972. The distribution by age of the frequency of first marriage in a female cohort. Journal of the American Statistical Association 67 (340), 743–749.

Condran, G.A., Himes, C.L., Preston, S.H., 1991. Old-age mortality patterns in low-mortality countries: an evaluation of population and death data at advanced ages, 1950 to the present. Population Bulletin of the United Nations 30, 23–60.

De Beer, J., 2011. A new relational method for smoothing and projecting age-specific fertility rates: TOPALS. Demographic Research 24 (18), 409–454.

De Beer, J., 2012. Smoothing and projecting age-specific probabilities of death by TOPALS. Demographic Research 27 (20), 543–592.

Deaton, A., 1997. The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. World Bank Publications, Washington, DC.

Dwyer-Lindgren, L., Gakidou, E., Flaxman, A., Wang, H., 2013. Error and bias in under-5 mortality estimates derived from birth histories with small sample sizes. Population Health Metrics 11 (1), 13.

Eayres, D., Williams, E.S., 2004. Evaluation of methodologies for small area life expectancy estimation. Journal of Epidemiology and Community Health 58 (3), 243–249.

El-Badry, M.A., 1961. Failure of enumerators to make entries of zero: errors in recording childless cases in population censuses. Journal of the American Statistical Association 56 (296), 909–924.

Elo, I.T., Preston, S.H., 1997. Racial and ethnic differences in mortality at older ages. In: Martin, L.G., Soldo, B.J., National Research Council Committee on Population (Eds.), Racial and Ethnic Differences in the Health of Older Americans. National Academy Press, Washington, DC, pp. 10–42.

Elo, I.T., Turra, C.M., Kestenbaum, B., Ferguson, B.R., 2004. Mortality among elderly hispanics in the United States: past evidence and new results. Demography 41 (1), 109–128.

Ewbank, D.C., 1981. Age Misreporting and Age-Selective Underenumeration: Sources, Patterns, and Consequences for Demographic Analysis. National Academy Press, Washington, DC.

Ewbank, D.C., Leon, J.C.G.D., Stoto, M.A., 1983. A reducible four-parameter system of model life tables. Population Studies 37 (1), 105–127.

Gage, T.B., Mode, C.J., 1993. Some laws of mortality: how well do they fit? Human Biology 65 (3), 445–461.

Groves, R.M., Couper, M.P., 1998. Nonresponse in Household Interview Surveys, Wiley Series in Probability and Statistics. Survey Methodology Section. Wiley, New York.

Heligman, L., Pollard, J.H., 1980. The age pattern of mortality. The Journal of the Institute of Actuaries 107 (434), 49–80.

Hernes, G., 1972. Process of entry into first marriage. American Sociological Review 37 (2), 173–182.

Heuveline, P., Clark, S.J., 2011. Model schedules of mortality. In: Rogers, R.G., Crimmins, E.M. (Eds.), International Handbook of Adult Mortality. Springer, New York, pp. 511–532.

Hill, K., Choi, Y., Timaeus, I., 2005. Unconventional approaches to mortality estimation. Demographic Research 13, 281–300.
Hill, K., Pande, R., Mahy, M., Jones, G., 1999. Trends in Child Mortality in the Developing World: 1960 to 1996. UNICEF, New York, NY.
Hill, K., You, D., Choi, Y., 2009. Death distribution methods for estimating adult mortality: sensitivity analysis with simulated data errors. Demographic Research 21 (9), 235–254.
Himes, C.L., Clogg, C.C., 1992. An overview of demographic analysis as a method for evaluating census coverage in the United States. Population Index 58 (4), 587–607.
Hobbs, F.B., 2004. Age and sex composition. In: Siegel, J.S., Swanson, D.A., Shryock, H.S. (Eds.), The Methods and Materials of Demography. Elsevier/Academic Press, Amsterdam; Boston, pp. 125–174.
Hull, T.H., Hartanto, W., 2009. Resolving contradictions in Indonesian fertility estimates. Bulletin of Indonesian Economic Studies 45 (1), 61–71.
ICF International, 2012a. Demographic and Health Survey Interviewer's Manual, Demographic and Health Survey's DHS Toolkit of Methodology. ICF International, Calverton, Maryland, USA.
ICF International, 2012b. Demographic and Health Survey Sampling and Household Listing Manual, Demographic and Health Survey's DHS Toolkit of Methodology. ICF International, Calverton, Maryland, USA.
ICF Macro, 2009. Training Field Staff for DHS Surveys, Demographic and Health Survey's DHS Toolkit of Methodology. ICF Macro, Calverton, Maryland, USA.
Impicciatore, R., Billari, F., 2011. MAPLES: a general method for the estimation of age profiles from standard demographic surveys (with an application to fertility). Demographic Research 24 (29), 719–748.
INDEPTH Network, 2004. INDEPTH Model Life Tables for Sub-Saharan Africa. Ashgate, Aldershot, Hants, England; Burlington, VT.
Johnson, N.F., Spagat, M., Gourley, S., Onnela, J.P., Reinert, G., 2008. Bias in epidemiological studies of conflict mortality. Journal of Peace Research 45 (5), 653–663.
Judson, D.H., Popoff, C.L., 2004. Appendix C. Selected general methods. In: Siegel, J.S., Swanson, D.A., Shryock, H.S. (Eds.), The Methods and Materials of Demography. Elsevier/Academic Press, Amsterdam; Boston, pp. 677–732.
Kish, L., 1965. Survey Sampling. Wiley, New York, NY.
Kostaki, A., 1992. A nine-parameter version of the Heligman-Pollard formula. Mathematical Population Studies 3 (4), 277–288.
Krótki, K.J. (Ed.), 1978. Developments in Dual System Estimation of Population Size and Growth. University of Alberta Press, Edmonton.
Levy, P.S., Lemeshow, S., 2013. Sampling of Populations: Methods and Applications. Wiley.
Ligges, U., 2013. Package 'signal': pchip (Piecewise cubic hermite interpolation) and interp1 (Interpolation methods, including linear, spline, and cubic interpolation). In: R Core Team (Ed.), R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Little, R.J.A., Rubin, D.B., 2002. Statistical Analysis with Missing Data, Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics, second ed. Wiley, New York, NY.
Lohr, S., 2009. Sampling: Design and Analysis. Cengage Learning, Boston, MA.
Lozano, R., Wang, H., Foreman, K.J., Rajaratnam, J.K., Naghavi, M., Marcus, J.R., Dwyer-Lindgren, L., Lofgren, K.T., Phillips, D., Atkinson, C., Lopez, A.D., Murray, C.J.L., 2011. Progress towards millennium development goals 4 and 5 on maternal and child mortality: an updated systematic analysis. Lancet 378 (9797), 1139–1165.
Luther, N.Y., Gaminiratne, K.H.W., Silva, S.D., Retherford, R.D., 1987. Consistent correction of international migration data for Sri Lanka, 1971–81. International Migration Review 21 (4), 1335–1369.
Luther, N.Y., Retherford, R.D., 1988. Consistent correction of census and vital registration data. Mathematical Population Studies 1 (1), 1–20.
Machemedze, T., Dorrington, R., 2011. Levels of mortality of the South African aged population using the method of extinct generations. Journal of African Population Studies 25 (Suppl. 1), 63–76.
Machiyama, K., 2010. A Re-examination of Recent Fertility Declines in Sub-saharan Africa. Demographic and Health Surveys Working Papers No. 68. Macro International, Calverton, MD.
Mahapatra, P., 2010. An overview of the sample registration system in India. In: Prince Mahidol Award Conference & Global Health Information Forum.
Masquelier, B., Reniers, G., Pison, G., 2013. Divergences in trends in child and adult mortality in sub-Saharan Africa: survey evidence on the survival of children and siblings. Population Studies, 1–17.
Moultrie, T.A., 2013a. Assessment of parity data. In: Moultrie, T.A., Dorrington, R.E., Hill, A.G., Hill, K., Timæus, I.M., Zaba, B. (Eds.), Tools for Demographic Estimation. International Union for the Scientific Study of Population, Paris.
Moultrie, T.A., 2013b. The el-Badry correction. In: Moultrie, T.A., Dorrington, R.E., Hill, A.G., Hill, K., Timæus, I.M., Zaba, B. (Eds.), Tools for Demographic Estimation. International Union for the Scientific Study of Population, Paris.
Moultrie, T.A., 2013c. The relational Gompertz model. In: Moultrie, T.A., Dorrington, R.E., Hill, A.G., Hill, K., Timæus, I.M., Zaba, B. (Eds.), Tools for Demographic Estimation. International Union for the Scientific Study of Population, Paris.
Moultrie, T.A., 2013d. Relational standards used in this project. In: Moultrie, T.A., Dorrington, R.E., Hill, A.G., Hill, K., Timæus, I.M., Zaba, B. (Eds.), Tools for Demographic Estimation. International Union for the Scientific Study of Population, Paris.
Moultrie, T.A., Dorrington, R.E., Hill, A.G., Hill, K., Timæus, I.M., Zaba, B. (Eds.), 2013. Tools for Demographic Estimation. International Union for the Scientific Study of Population.
Murray, C.J.L., Rajaratnam, J.K., Marcus, J., Laakso, T., Lopez, A.D., 2010. What can we conclude from death registration? Improved methods for evaluating completeness. PLoS Med 7 (4), e1000262.
Murray, C.J.L., Ferguson, B.D., Lopez, A.D., Guillot, M., Salomon, J.A., Ahmad, O., 2003. Modified logit life table system: principles, empirical validation, and application. Population Studies 57 (2), 165–182.
Obermeyer, Z., Rajaratnam, J.K., Park, C.H., Gakidou, E., Hogan, M.C., Lopez, A.D., Murray, C.J., 2010. Measuring adult mortality using sibling survival: a new analytical method and new results for 44 countries, 1974–2006. PLoS Med 7 (4), e1000260.
Pedersen, J., Liu, J., 2012. Child mortality estimation: appropriate time periods for child mortality estimates from full birth histories. PLoS Med 9 (8), e1001289.
Pollard, A.H., Yusuf, F., Pollard, G.N., 1990. Testing the accuracy of demographic data. In: Demographic Techniques. Pergamon Press, Rushcutters Bay, NSW, pp. 152–163.
Preston, S.H., Coale, A.J., 1982. Age structure, growth, attrition, and accession: a new synthesis. Population Index 48 (2), 217–259.
Preston, S.H., Elo, I.T., Rosenwaike, I., Hill, M., 1996. African-American mortality at older ages: results of a matching study. Demography 33 (2), 193–209.
Preston, S.H., Elo, I.T., Stewart, Q., 1999. Effects of age misreporting on mortality estimates at older ages. Population Studies 53 (2), 165–177.
Preston, S.H., Heuveline, P., Guillot, M., 2001a. Demographic relations in nonstable populations. In: Demography: Measuring and Modeling Population Processes. Blackwell Publishers, Malden, MA, pp. 171–190.
Preston, S.H., Heuveline, P., Guillot, M., 2001b. Methods for evaluating data quality. In: Demography: Measuring and Modeling Population Processes. Blackwell Publishers, Malden, MA, pp. 211–223.
Preston, S.H., Heuveline, P., Guillot, M., 2001c. Modeling age patterns of vital events. In: Demography: Measuring and Modeling Population Processes. Blackwell Publishers, Malden, MA, pp. 191–210.
Pullum, T.W., 2005. A statistical reformulation of demographic methods to assess the quality of age and date reporting, with application to the Demographic and Health Surveys. In: Annual Meeting of the Population Association of America. Population Association of America, Philadelphia, PA.
Pullum, T.W., 2006. An Assessment of Age and Date Reporting in the DHS Surveys, 1985–2003. Demographic and Health Surveys Methodological Reports No. 5. Macro International, Calverton, MD.
Rachet, B., Maringe, C., Ellis, L., 2008. Ewblft (creation and smoothing of life tables). In: Cancer Research UK Cancer Survival Group (Ed.), Tools for Cancer Survival Analysis. London School of Hygiene & Tropical Medicine, University of London, London, UK.
Randall, S., Coast, E., Compaore, N., Antoine, P., 2013. The power of the interviewer. Demographic Research 28, 763–792.
Randall, S., Coast, E., Leone, T., 2011. Cultural constructions of the concept of household in sample surveys. Population Studies. A Journal of Demography 65 (2), 217–229.
Rogers, A., Little, J., Raymer, J., 2010. The Indirect Estimation of Migration: Methods for Dealing with Irregular, Inadequate, and Missing Data. In: The Springer Series on Demographic Methods and Population Analysis. Springer, Dordrecht.
Rogers, A., Little, J.S., 1994. Parameterizing age patterns of demographic rates with the multiexponential model schedule. Mathematical Population Studies 4 (3), 175–195.
Schafer, J.L., 1997. Analysis of Incomplete Multivariate Data, Monographs on Statistics and Applied Probability. Chapman & Hall, New York, NY.
Scherbov, S., Ediev, D., 2011. Significance of life table estimates for small populations: simulation-based study of estimation errors. Demographic Research 24 (22), 527–550.
Schmertmann, C., 2012. Calibrated Spline Estimation of Detailed Fertility Schedules from Abridged Data. MPIDR Working Papers. Max Planck Institute for Demographic Research, Rostock, Germany.

Schmertmann, C.P., 2002. A simple method for estimating age-specific rates from sequential cross sections. Demography 39 (2), 287–310.

Schoumaker, B., 2011. Omissions of births in DHS birth histories in sub-Saharan Africa: measurement and determinants. In: Annual Meeting of the Population Association of America. Population Association of America, Washington, DC.

Schoumaker, B., 2013. Reconstructing long term fertility trends with pooled birth histories. In: XXVII IUSSP International Population Conference. Republic of Korea, Busan. XXVII IUSSP International Population Conference.

Siler, W., 1983. Parameters of mortality in human populations with widely varying life spans. Statistics in Medicine 2 (3), 373–380.

Silva, R., 2012. Child mortality estimation: consistency of under-five mortality rate estimates using full birth histories and summary birth histories. PLoS Med 9 (8), e1001296.

Smith, L., Hyndman, R., Wood, S., 2004. Spline interpolation for demographic variables: the monotonicity problem. Journal of Population Research 21 (1), 95–98.

Som, R.K., 1996. Practical Sampling Techniques, Statistics, Textbooks and Monographs, second ed. M. Dekker, New York, NY.

Spoorenberg, T., 2007. Quality of age reporting: extension and application of the modified Whipple's index. Population (english ed.) 62 (4), 729–741.

Spoorenberg, T., 2009. Is the Whipple's index really a fair and reliable measure of the quality of age reporting? An analysis of 234 censuses from 145 countries. In: XXVI IUSSP International Population Conference,Marrakech, Morocco. XXVI IUSSP International Population Conference.

Stolzenberg, R.M., Relles, D.A., 1997. Tools for intuition about sample selection bias and its correction. American Sociological Review 62 (3), 494–507.

Sullivan, J.M., 2008. An Assessment of the Credibility of Child Mortality Declines Estimated from DHS Mortality Rates. Working Draft; Revision 1. UNICEF, New York, NY.

Thatcher, A.R., Kannisto, V., Vaupel, J.W., 1998. The Force of Mortality at Ages 80 to 120, Odense Monographs on Population Aging Series. Odense University Press, Odense, Denmark.

Timaeus, I.M., 1991. Estimation of adult mortality from orphanhood before and since marriage. Population Studies 45 (3), 455–472.

United Nations, 1955. Manual II. Methods of Appraisal of Quality of Basic Data for Population Estimates, Population Studies. Series a: Manuals on Methods of Estimating Population. United Nations, New York, NY.

United Nations, 1982. Non-sampling Errors of Household Surveys: Sources, Assessment and Control, National Household Survey Capability Programme, Prelim. version. ed. United Nations, New York, NY.

United Nations, 1983. Manual X: Indirect Techniques for Demographic Estimation, Population Studies. United Nations, New York, NY.

United Nations, 1991. Handbook of Vital Statistics Systems and Methods. In: Legal, Organization and Technical Aspects, Studies in Methods Series F, vol. I. United Nations, New York, NY.

United Nations, 2002. Methods for Estimating Adult Mortality. United Nations, New York.

United Nations, 2005. Household Surveys in Developing and Transition Countries, Studies in Methods, Series F. United Nations, New York, NY.

United Nations, 2008a. Designing Household Survey Samples: Practical Guidelines, Studies in Methods. United Nations, New York, NY.

United Nations, 2008b. Principles and Recommendations for Population and Housing Censuses, Revision 2, Statistical Papers Series M, rev. second ed. United Nations, New York, NY.

United Nations, 2010a. Handbook on Population and Housing Census Editing: Revision 1, Studies in Methods Series F. United Nations, New York,NY.

United Nations, 2010b. Post Enumeration Surveys: Operational Guidelines (Technical Report). 2010 World Population and Housing Census Programme. United Nations, New York, NY.

United States. Bureau of the Census, 1985. Evaluating Censuses of Population and Housing, Statistical Training Document. U.S. Department of Commerce, Bureau of the Census, Washington, DC.

Walker, N., Hill, K., Zhao, F., 2012. Child mortality estimation: methods used to adjust for bias due to AIDS in estimating trends in under-five mortality. PLoS Med 9 (8), e1001298.

Wang, H., Dwyer-Lindgren, L., Lofgren, K.T., Rajaratnam, J.K., Marcus, J.R., Levin-Rector, A., Levitz, C.E., Lopez, A.D., Murray, C.J., 2012. Age-specific and sex-specific mortality in 187 countries, 1970-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 380 (9859), 2071–2094.

Wheldon, M.C., Raftery, A.E., Clark, S.J., Gerland, P., 2013a. Bayesian reconstruction of past populations and vital rates by age for developing and developed countries. In: Paper Presented to the Annual Meeting of the Population Association of America, New Orleans, LA.

Wheldon, M.C., Raftery, A.E., Clark, S.J., Gerland, P., 2013b. Bayesian Reconstruction of Two-sex Populations by Age: Estimating Sex Ratios at Birth and Sex Ratios of Mortality arXiv preprint arXiv:1312.4594.

Wheldon, M.C., Raftery, A.E., Clark, S.J., Gerland, P., 2013c. Reconstructing past populations with uncertainty from fragmentary data. Journal of the American Statistical Association 108 (501), 96–110.

Wilmoth, J., Zureick, S., Canudas-Romo, V., Inoue, M., Sawyer, C., 2011. A flexible two-dimensional mortality model for use in indirect estimation. Population Studies 66 (1), 1–28.

Wilmoth, J.R., Andreev, K., Jdanov, D., Glei, D.A., Boe, C., Bubenheim, M., Philipov, D., Shkolnikov, V., Vachon, P., 2007. Methods Protocol for the Human Mortality Database. University of California, Berkeley and Max Planck Institute for Demographic Research, Rostock. http://mortality.org (version 31.05.07.).

Wilmoth, J.R., Mizoguchi, N., Oestergaard, M.Z., Say, L., Mathers, C.D., Zureick-Brown, S., Inoue, M., Chou, D., 2012. A new method for deriving global estimates of maternal mortality. Statistics, Politics and Policy 3 (2).

Winship, C., Mare, R.D., 1992. Models for sample selection bias. Annual Review of Sociology 18, 327–350.

Yang, G., Hu, J., Rao, K.Q., Ma, J., Rao, C., Lopez, A.D., 2005. Mortality registration and surveillance in China: history, current situation and challenges. Population Health Metrics 3 (1), 3.

Zaba, B., 1979. The four-parameter logit life table system. Population Studies 33 (1), 79–100.

Zaba, B., 1985. A parameterized procedure for projecting population. In: International Population Conference. International Union for the Scientific Study of Population, Florence, Italy; Liege, Belgium, pp. 137–150.

Relevant Websites

http://demographicestimation.iussp.org/ – IUSSP Tools for Demographic Estimation.

http://www.census.gov/population/international/software/pas/ – U.S. Bureau of the Census, International Programs Center: Population Analysis System (PAS).

http://unstats.un.org/unsd/demographic/standmeth/ – UN Statistics Division: Standards and Methods.

http://www.un.org/en/development/desa/population/publications/manual/ – UN Population Division: Demographic Manuals.

http://www.demog.berkeley.edu/∼eddieh/toolbox.html – Applied Demography Toolbox.