PAA 2014 Annual Meeting

Bayesian Reconstruction of Two-Sex Populations by Age: Estimating Sex Ratios at Birth and Sex Ratios of Mortality*

Mark C. Wheldon
Auckland University of Technology

Adrian E. Raftery, Samuel J. Clark
University of Washington

Patrick Gerland
United Nations Population Division

May 2014

Abstract

The original version of Bayesian reconstruction (Wheldon et al., 2013b), a method for estimating age-specific fertility, mortality, migration and population counts of the recent past with uncertainty, produced estimates for female-only populations. Here we show how two-sex populations can be similarly reconstructed and probabilistic estimates of various sex ratio quantities obtained. We demonstrate the method by reconstructing the populations of India from 1971 to 2001, Thailand from 1960 to 2000, and Laos from 1985 to 2005. The posterior probability that sex ratios at birth (SRBs) in India exceeded 1.06 are found to be greater than 0.9, as is the probability that there was an increase in this parameter over the period of study. In both Thailand and Laos, we found strong evidence that life expectancy at birth (e_0) was greater for females and, in Thailand, that the difference increased over the period of study. In India, the probability that female e_0 was lower was 0.8 but there was strong evidence for a narrowing of the gap through to 2001.

Keywords: Bayesian hierarchical model; Two-sex model; Population projection; Vital Rates; Sex ratio at birth; Sex ratio of mortality

* *Address for Correspondence:* Mark C. Wheldon, Biostatistics Unit, Faculty of Health and Environmental Sciences, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand; mwheldon@aut.ac.nz.

1 Introduction

The past, present and future dynamics of human populations at the country level are highly relevant to the work of social scientists in many disciplines as well as planners and evaluators of public policy. These dynamics are driven by population counts, fertility and mortality rates (vital rates), and net international migration. Demographers at the United Nations Population Division (UNPD) are tasked with producing detailed information on these quantities, published biennially in *World Population Prospects* (WPP) (e.g., UN, 2011). Estimates for each country are provided, for periods stretching back from the present to about 1950. Currently, however, WPP estimates are not accompanied by any quantitative estimate of uncertainty. Uncertainty should also be measured because the availability, coverage and reliability of data used to derive the estimates differs greatly among countries. Developing countries, in particular, often lack the extensive registration and census-taking systems developed countries maintain, meaning estimates in these cases are subject to greater uncertainty. Estimates are not error-free even for developed countries. While estimates of population counts and vital rates are likely to be very, accurate uncertainty about net international migration can be quite substantial. Beer et al. (2010) found this to be the case for Europe.

In response to the need for uncertainty quantification in estimates of the key parameters driving human population dynamics, Wheldon et al. (2013b) proposed Bayesian population reconstruction (Bayesian reconstruction for short), a method of simultaneously estimating population counts, vital rates and net international migration at the country level, by age, together with uncertainty. The original formulation was able to reconstruct female-only populations. In this article, we describe a major extension to two-sex populations. This allows us to estimate age- and time-specific indicators of fertility, mortality and migration separately for females and males and, importantly, sex ratios of these quantities, all with probabilistic measures of uncertainty. In addition, we also show how Bayesian reconstruction can be used to derive probabilities of change over time in these quantities. To demonstrate the method, we reconstruct the full populations of India from 1971–2001, Thailand from 1960–2000 and Laos from 1985–2005. These countries were selected because, in all cases, the available data are fragmentary which makes population reconstruction challenging.

Bayesian reconstruction embeds a standard demographic projection model in a hierarchical statistical model. As inputs, it takes bias-reduced initial estimates of age-specific fertility rates, survival proportions (a measure of mortality), net international migration and census-based population counts. Also required is expert opinion about the measurement error of these quantities, informed by data if available. The output is a joint posterior probability distribution on the inputs, allowing all parameters to be estimated simultaneously, together with fully probabilistic posterior estimates of measurement error. Wheldon et al. (2013b) showed that marginal credible intervals were well calibrated. They demonstrated the method by reconstructing the female population of

Burkina Faso from 1960 to 2000. Wheldon et al. (2013a) extended Bayesian reconstruction to countries with censuses at irregular intervals and showed that it works well across a wide range of data quality contexts by reconstructing the female populations of Laos, Sri Lanka, and New Zealand. Laos is a country with very little vital registration data where population estimation depends largely on surveys, Sri Lanka has some vital registration data, and New Zealand is a country with high-quality registration data. In this paper we focus on countries which lack good vital registration data or for which there are gaps in, or inconsistencies among, the available data sources.

Global sex ratio in the total population (SRTP), defined as the ratio of the number of males per female, has risen slightly from about 1.00 in 1950 to 1.02 in 2010. There is a great deal of variation among regions, however. For instance, SRTP in the more developed regions ranged from 0.91 to 0.95 over this period, while in less developed regions it remained constant at about 1.04. In Eastern Asia, which includes China, and in Southern Asia, which includes India, SRTPs ranged from 1.05 to 1.06 and from 1.09 to 1.06, respectively (UN, 2011). Guilmoto (2007b) claims that the population of Asia underwent "masculinization" during the latter half of the twentieth century, with one likely consequence being a "marriage squeeze" (Guilmoto, 2009, 2012) wherein many males will be unable to find a marriage partner. Imbalances in population sex ratios are caused by imbalances in sex ratios at birth (SRBs) and sex ratios of mortality (SRMs) (Guillot, 2002). These quantities have received considerable attention in the literature on the demography of Asia (e.g. Bhat, 2002b,c; Bongaarts, 2001; Coale, 1991; Das Gupta, 2005; Guilmoto, 2009; Mayer, 1999; Sen, 1990). Sawyer (2012) called for further work to quantify uncertainty in estimates of SRMs. Estimates of SRB are subject to a large amount of uncertainty, especially in India (Bhat, 2002b,c; Guillot, 2002; Guilmoto, 2009). Here, we respond by quantifying uncertainty in these parameters.

The article is organised as follows. In the remainder of this section we provide some background on existing methods of population reconstruction and the demography of sex ratios in Asia. In Section 2 we describe the two-sex version of Bayesian reconstruction. In Section 3.2 we present results from our case studies of India, Thailand and Laos. We focus mainly on posterior distributions of total fertility rate (TFR), SRBs and the sex difference in life expectancy. Certain sex ratios in India are widely believed to be atypical so we devote more attention to this case. We end with a discussion in Section 4 which provides further demographic context and an overall conclusion. Selected mathematical derivations are given in the appendix . Bayesian reconstruction is implemented in the "popReconstruct" package for the R environment for statistical computing (R Core Team, 2013).

1.1 Methods of Population Reconstruction

In human demography, population reconstruction is often referred to simply as "estimation" to distinguish it from forecasts of future population counts and vital rates. We use "reconstruction" merely to avoid ambiguity. Nevertheless, use of "estimation" agrees with its usage in statistics; namely the estimation of the values of some unknown parameters from data. Reviews of existing methods of population reconstruction are given by Oeppen (1993a), Barbi et al. (2004), and Wheldon et al. (2013b). Many were developed for the purpose of reconstructing populations of the distant past from data on births, deaths and marriages recorded in parish registers (e.g,. Bertino and Sonnino, 2003; Walters, 2008; Wrigley and Schofield, 1981), or in counter-factual exercises to estimate the excess of mortality due to extreme events such as famine or genocide (e.g., Boyle and Ó Gráda, 1986; Daponte et al., 1997; Goodkind and West, 2001; Heuveline, 1998; Merli, 1998) or the number of "missing women" due to male-dominated sex ratios in Asia (Coale, 1991; Das Gupta, 2005; Sen, 1990). Purely deterministic reconstruction methods used in some of these studies include "inverse projection" (Lee, 1971, 1974), "back projection" (Wrigley and Schofield, 1981) and "generalized back projection" (Oeppen, 1993b). Bertino and Sonnino (2003) proposed "stochastic inverse projection". This is a non-deterministic method but the only form of uncertainty is that which comes from treating birth and death as stochastic processes at the individual level. Counts of births and deaths are assumed to be known without error and age patterns are fixed. In the cases we treat, accurate data on births and deaths of the parish register kind are often unavailable and uncertainty due to stochastic vital rates is likely to be small relative to uncertainty due to measurement error (Cohen, 2006; Lee, 2003; Pollard, 1968). Moreover, it is designed to work with the kind of data commonly available for developing countries and does not rely on the existence of detailed births and deaths registers, although this information can be used when available (Wheldon et al., 2013a).

Daponte et al. (1997) took a fully Bayesian approach to constructing a counterfactual history of the Iraqi Kurdish population from 1977 to 1990. They constructed prior distributions for fertility and mortality rates using survey data and expert opinion about uncertainty based on historical information and knowledge of demographic processes. Measurement error in the available, fragmentary data was accounted for. However, there were some restrictions such as holding the age pattern of fertility fixed and allowing for mortality variation only through infant mortality. Rural to urban migration was accounted for by treating these populations separately; international migration was assumed negligible. Bayesian reconstruction is similar but no model age patterns are assumed to hold and international migration is explicitly estimated along with fertility, mortality and population counts.

Most methods of population reconstruction in demography employ the cohort component model of population projection (CCMPP) in some form. Population projection uses vital rates and migration to project a set of age-specific population counts in the baseline year, denoted t_0, forward in time to the end year, denoted T. In its simplest form, the population in year $t+\delta$, $t_0 \leq t \leq T-\delta$, equals the population in year t plus the intervening births and net migration, minus the intervening deaths (e.g., Preston et al., 2001; Whelpton, 1936). This is known as the demographic balancing relationship. Population projection is distinct from population forecasting since it merely entails evolving a population forward in time from some given baseline under assumptions about prevailing vital rates and migration (Keyfitz and Caswell, 2005). The period of projection may be in the future or in the past. Wheldon et al. (2013a,b) employed single-sex population projection in this way to reconstruct female-only populations, taking account of measurement error. In this article we show how two-sex population projection can be used to reconstruct full populations, thereby providing estimates of population sex ratios, SRBs and SRMs, simultaneously while accounting for measurement error.

1.2 Estimating Sex Ratios

Methods of reporting sex ratios are not standardised. Here we adopt the convention that all ratios are "male-per-female"; in the Indian literature the inverse is more common. Hence, SRTP is the total number of males per female in the population and SRB is the number of male births per female birth. The SRM can be expressed using various mortality indicators. We will use the under-five mortality rate (U5MR) exclusively (see the Appendix for a formal definition). A low SRM means that mortality is lower among males than among females. All-age mortality is summarised by life expectancy at birth, for which the standard demographic abbreviation is e_0. Comparison of e_0 by sex is more commonly done using the difference rather than the ratio and we adopt that convention here. Male e_0 is subtracted from female e_0 to obtain the difference.

Under typical conditions, SRBs for most countries are in the range 1.04–1.06 (UNFPA, 2010). Estimates of SRB in some regions in Asia are higher; the National Family Health Survey in India estimated SRB over 2000–2006 to be 1.09, for example (Guilmoto, 2009). For almost all countries, e_0 is higher for females than males. This is thought to be due to a range of biological and environmental factors, with the relative contribution of each class of factor varying among countries (Waldron, 1985). Age-specific SRMs are more variable as they are affected by sex-specific causes of death such as those associated with child birth. The preferred way of estimating SRBs and SRMs at the national level is from counts of births and deaths recorded in official registers (vital registration) together with total population counts from censuses. In many countries where such registers are not kept, surveys such as the Demographic and Health Surveys and World Fertility Surveys must be used. These typically ask a sample of women about their birth histories. Full birth histories collect information about the times of each birth and, if the child subsequently died, the time of the death. Summary birth histories ask only about the total number of births and child deaths the respondent has ever experienced (Preston et al., 2001; UN, 1983).

Estimates based on both vital registration and surveys are susceptible to systematic biases and non-systematic measurement error. Counts of births or deaths from vital registration may be biased downward by the omission of events from the register or under-coverage of the target population. Full birth histories are susceptible to biases caused by omission of births or misreporting the timing of events. Some omissions may be deliberate in order to avoid lengthy subsections of the survey (Hill, You, et al., 2012). Fertility and mortality estimates from summary birth histories are derived using so-called indirect techniques such as the Brass P/F ratio method (Brass, 1964; Feeney, 1996; UN, 1983). In addition to the biases affecting full birth histories, estimates based on summary birth histories can also be affected when the assumptions behind the indirect methods are not satisfied. These assumptions concern the pattern of mortality by age and the association between mother and child mortality. They often do not hold, for example, in populations experiencing rapid mortality decline (Silva, 2012).

In the absence of vital registration, estimates of adult mortality may be based on reports of sibling survival histories collected in surveys. Often however, the only data available are on child mortality collected from surveys of women. In such cases, estimates of adult mortality are extrapolations based on model life tables (Preston et al., 2001; UN, 1983). Model life tables are families of life tables generated from mortality data collected from a wide range of countries over a long period of time. They are indexed by a summary parameter such as e_0 or U5MR and are grouped into regions. The Coale and Demeny system (Coale et al., 1983; Preston, 1976) and the United Nations (UN) system for developing countries (UN, 1982) both have five families. Errors in estimates of adult mortality derived in this way come from errors in the survey-based estimates of under five mortality and the inability of the model life table family to capture the true mortality patterns in the population of interest.

Concerns about the accuracy of SRB estimates, particularly for periods between 1950 and 1970, have led some authors to suggest using age-specific population sex ratios as a proxy for SRBs. Guilmoto (2009) suggests the male-to-female ratio among those aged 0–4 (the "child sex ratio") and Bhat (2002b) suggests the ratio among those aged 0–14 for India (the "juvenile sex ratio"). Such ratios must be estimated from census data which is probably more reliable than survey data, but still subject to age misreporting and underreporting of certain groups. For example, there appears to have been under counting of females in censuses of India (Bhat, 2002b,c; Guillot, 2002). In our case study, we estimate the child sex ratio for India between 1971 and 2001.

Estimates of fertility, mortality, migration and population counts and the implied sex ratios for successive five-year periods are all related to one another via the demographic balancing equation that underlies the CCMPP. The estimates of these quantities published in WPP must be "projection consistent" in that the age-specific population counts for year t must be the counts one gets by projecting the published counts for year $t-5$ forward using the published fertility, mortality and migration rates.

Bias reduction techniques are source- and parameter-specific. For birth history data, these might involve omitting responses of very old women, or responses pertaining to events in the distant past. For census counts, adjustments may be made to compensate for well-known undercount in certain age-sex groups. In other cases, parametric models of life tables, or specially constructed life tables, may be used if available. For this reason we do not propose a generally applicable method of bias reduction, one which would work well for all parameters and data sources, since many specialized ones already exist (e.g., Alkema et al., 2012; Murray, Ferguson, et al., 2003; Murray, Rajaratnam, et al., 2010; UN, 1983). Bayesian reconstruction takes as input bias-reduced initial estimates of age-specific fertility rates and age- and sex-specific initial estimates of mortality, international migration and population counts. Measurement error is accounted for by modelling these quantities as probability distributions. Projection consistency is achieved by embedding the CCMPP in a Bayesian hierarchical model. Inference is based on the joint posterior distribution on the input parameters, which is sampled from using Markov chain Monte Carlo (MCMC).

Under the current UN procedure, all available representative data sources for a given country are considered and techniques to reduce bias are applied where UN analysts deem them appropriate. Projection consistency is achieved through an iterative "project-and-adjust" process. SRBs and SRMs are inputs to the procedure, while population sex ratios are calculated using estimate population counts, which are an output.

2 Method

2.1 Notation and Parameters

The parameters of interest are age- and time-specific vital rates, net international migration flows, population counts and SRB. We are concerned only with international migration which we will refer to as simply "migration". The symbols n, s, g and f denote population counts, survival (a measure of mortality), net migration (immigrants minus emigrants) and fertility, respectively. All of these parameters will be indexed by five-year increments of age, denoted by a, and time, denoted by t. The parameters n, s and g will also be indexed by sex, denoted by $l = F, M$, where F and M indicate female and male, respectively. SRB is defined as the number of male births for every female birth. It will be indexed by time. Reconstruction will be done over the time interval $[t_0, T]$. The age scale runs from 0 to $A > 0$; in our applications A is 80. The total number of age-groups/time-periods is denoted K. To model fertility, we define $a_L^{[\text{fert}]} \leq a_U^{[\text{fert}]}$ where fertility is assumed to be zero at ages outside the range $[a_L^{[\text{fert}]}, a_U^{[\text{fert}]}+5)$. Throughout, a prime indicates vector transpose. We will use boldface for vectors and a "$\cdot$" to indicate the indices whose entire range is contained therein. Multiple indices are stacked in the order a, t, l. For example, $\boldsymbol{n}_{\cdot,t_0,F}$ is the vector of age-specific female population counts in exact year t_0, $[n_{0,t_0,F}, \cdots, n_{A,t_0,F}]'$ and $\boldsymbol{n}_{\cdot,\cdot,\cdot} = [n_{0,t_0,F}, \cdots, n_{A,t_0,F}, n_{0,t_0+5,F}, \cdots, n_{A,t_0+5,F}, \cdots n_{0,T,F}, \cdots, n_{A,T,F}, \cdots, n_{0,T,M}, \cdots, n_{A,T,M}]'$.

The parameters are the standard demographic parameters used for projection. The fertility parameters, $f_{a,t}$, are age-, time-specific occurrence/exposure rates. They give the ratio of the number of babies born over the period $[t, t+5)$ to the number of person-years lived over this period by women in the age range $[a, a+5)$. If a woman survives for the whole five-year period she contributes five person-years to the denominator; if she survives only for the first year and a half she contributes 1.5 person-years, and so forth. The survival parameters, $s_{a,t,l}$, are age-, time-, and sex-specific proportions. They give the proportion of those alive at time t that survive for five years. The age subscript on the survival parameters indicates the age-range the women will survive *into*. For example, the number of females aged $[15, 20)$ alive in 1965 would be the product $(n_{10,1960,F})(s_{15,1960,F})$ (ignoring migration for clarity). It also means that $s_{0,1960.F}$ is the proportion of female births born during 1960–1965 that are alive in 1965, hence aged 0–5. The oldest age-group is open-ended and we must allow for survival in this age group. Thus, the proportion aged $[A, \infty)$ at time t that survives through the interval $[t, t+5)$ is denoted $s_{A+5,t,l}$. Migration is also expressed as a proportion. The net number of male migrants aged $[a, a+5)$ over the interval $[t, t+5)$ is $(n_{a,t,M})(g_{a,t,M})$.

2.2 Projection of Two-Sex Populations

The CCMPP allows one to calculate the number alive by age and sex, at any time, $t = t_0 + 5, \ldots, T$ using $\boldsymbol{n}_{\cdot,t_0,\cdot}$, the vector of age- and sex-specific female and male population counts at baseline t_0, and the age-, time-, sex-specific vital rates and migration up to time t. $\boldsymbol{n}_{\cdot,t,\cdot}$ is simply $\boldsymbol{n}_{\cdot,t_0,\cdot}$ plus the intervening births, minus the deaths, plus net migration. Projection is a discrete time approximation to a continuous time process, and several adjustments are made to improve accuracy. The form we use has two-steps; projection is done first for those aged 5 and above, $\boldsymbol{n}_{5^+,t,l}$ and then for those under five, $n_{0,t,l}$. To this end, let us write

$$\boldsymbol{n}_{\cdot,t,\cdot} = \left[\begin{array}{cc|cc} n_{0,t,F}, & \boldsymbol{n}_{5^+,t,F} & n_{0,t,M}, & \boldsymbol{n}_{5^+,t,M} \end{array}\right]', \tag{1}$$

where vectors and matrices are partitioned according to sex for clarity.

The number alive at exact time $t+5$ aged 5 and above is then given by the following matrix multiplication:

$$\left[\begin{array}{c} \boldsymbol{n}_{5^+,t+5,F} \\ \hline \boldsymbol{n}_{5^+,t+5,M} \end{array}\right] = \left[\begin{array}{cc} \mathbf{L}_{5^+,t,F} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{L}_{5^+,t,M} \end{array}\right] \left[\begin{array}{c} \boldsymbol{n}_{\cdot,t,F} + \boldsymbol{n}_{\cdot,t,F} \circ (\boldsymbol{g}_{\cdot,t,F})/2 \\ \hline \boldsymbol{n}_{\cdot,t,M} + \boldsymbol{n}_{\cdot,t,M} \circ (\boldsymbol{g}_{\cdot,t,M})/2 \end{array}\right] + \left[\begin{array}{c} (\boldsymbol{g}_{5^+,t,F})/2 \\ \hline (\boldsymbol{g}_{5^+,t,M})/2 \end{array}\right]. \tag{2}$$

The symbol "$\circ$" indicates the Hadamard (or element-wise) product; $\boldsymbol{n}_{5^+,t,l}$, $l = F, M$, are $(K-1)\times 1$ vectors containing the age-specific female and male population counts at exact time t; and $\boldsymbol{g}_{5^+,t,l}$, $l = F, M$, are $(K-1) \times 1$ vectors of age-specific female and male net migration expressed as a proportion of the population. $\mathbf{L}_{5^+,t,F}$ and $\mathbf{L}_{5^+,t,M}$ are $(K-1) \times K$ matrices of survival proportions for females and males at ages 5 and above, and $\mathbf{0}$ is a $(K-1) \times K$ matrix of zeros (the "L" is for Leslie 1945, 1948). The female and male survival matrices have the same form:

$$\mathbf{L}_{5^+,t,l} = \begin{bmatrix} s_{5,t,l} & 0 & & 0 & 0 \\ 0 & s_{10,t,l} & \ddots & 0 & 0 \\ 0 & 0 & & 0 & 0 \\ 0 & 0 & \cdots & s_{A,t,l} & s_{A+5,t,l} \end{bmatrix}, \; l = F, M \tag{3}$$

Splitting migration in half and adding the first half at the beginning of the projection interval and the second half at the end is a standard approximation to improve the discrete time approximation (Preston et al., 2001).

The number of females and males alive aged $[0, 5)$ in exact year $t + 5$ is derived from the total number of births over the interval, b_t, where

$$b_t = \sum_{a=a_L^{[\text{fert}]}}^{a_U^{[\text{fert}]}} 5 f_{a,t} \left\{ \frac{n_{a,t,F} + (n_{a-5,t,F})(s_{a,t,F})}{2} \right\} \tag{4}$$

The term in braces is an approximation to the number of person-years lived by women of child-bearing age over the projection interval. The total number of each sex aged $[0, 5)$ alive at the end of the interval is computed from b_t using under five mortality, migration and SRB:

$$n_{0,t+5,F} = b_t \frac{1}{1 + SRB_t} \left\{ s_{0,t,F} \left(1 + (g_{0,t,F})/2\right) + (g_{0,t,F})/2 \right\} \tag{5}$$

$$n_{0,t+5,M} = b_t \frac{SRB_t}{1 + SRB_t} \left\{ s_{0,t,M} \left(1 + (g_{0,t,M})/2\right) + (g_{0,t,M})/2 \right\}. \tag{6}$$

Note that the $f_{a,t}$ in (5) have only two subscripts; they are the age-specific (female) fertility rates introduced above. Thus the total number of births in the projection interval is a function of the number of females of reproductive age, but not the number of males of any age, or females of other ages. This is called "female dominant projection". This approach is preferred to alternatives, such as basing fertility on the number of male person-years lived, because survey-based fertility data is often collected by interviewing mothers, not fathers. All-sex births are computed first and then decomposed because SRB is often a parameter of interest to demographers, as it is to us here (Preston et al., 2001).

2.3 Modelling Uncertainty

In many countries, the available data on vital rates and migration are fragmentary and subject to systematic biases and non-systematic measurement error. Wheldon et al. (2013b) proposed Bayesian reconstruction as a way of estimating past vital rates, migration and population counts for a single-sex population which accounts for measurement error. Systematic biases are treated in a pre-processing step which yields a set of bias-reduced "initial estimates" for each age-, time-specific fertility rate, survival and migration proportion and population counts. We use an asterisk ("*") to denote initial estimates. Hence $f^*_{a,t}$ is the initial estimate of $f_{a,t}$. At the heart of Bayesian reconstruction is a hierarchical model which takes the initial estimates as inputs. Here, we present a substantial development of the model given in Wheldon et al. (2013b) which allows estimation of two-sex populations.

Take t_0 and T to be the years for which the earliest and most recent bias adjusted census-based population counts are available (henceforth, we refer to these simply as census counts). Years following t_0 for which census counts are also available are denoted $t_0 < t_L^{[\text{cen}]}, \ldots, t_U^{[\text{cen}]} < T$. Let $\boldsymbol{\theta}$ be the vector of all age-, time- and sex-specific fertility rates, survival and migration proportions over the period $[t_0, T)$, the SRBs, and the age- and sex-specific census counts in year t_0. These are the inputs required by the CCMPP, which we abbreviate by $M(\cdot)$. Let $\boldsymbol{\psi}_t$ be the components of $\boldsymbol{\theta}$ corresponding to time t, excluding $\boldsymbol{n}_{t_0}$. Therefore, $\boldsymbol{\theta} = [\boldsymbol{n}'_{\cdot,t_0,\cdot},\ \boldsymbol{f}'_{\cdot,\cdot,\cdot},\ \boldsymbol{s}'_{\cdot,\cdot,\cdot},\ \boldsymbol{g}'_{\cdot,\cdot,\cdot},\ \boldsymbol{SRB}']'$ and $\boldsymbol{\psi}_t = [\boldsymbol{f}'_{\cdot,t,\cdot},\ \boldsymbol{s}'_{\cdot,t,\cdot},\ \boldsymbol{g}'_{\cdot,t,\cdot},\ \boldsymbol{SRB}'_t]'$. Reconstruction requires estimation of $\boldsymbol{\theta}$ which we do using the following hierarchical model:

$$\text{Level 1}: \quad \log n^*_{a,t,l} \mid n_{a,t,l}, \sigma^2_n \sim \text{Normal}\left(\log n_{a,t,l}, \sigma^2_n\right) \qquad t = t_L^{[\text{cen}]}, \ldots, t_U^{[\text{cen}]} \tag{7}$$

$$\text{Level 2}: \quad n_{a,t,l} \mid \boldsymbol{n}_{\cdot,t-5,\cdot},\ \boldsymbol{\psi}_{t-5} = M\left(\boldsymbol{n}_{\cdot,t-5,\cdot},\ \boldsymbol{\psi}_{t-5}\right) \qquad t = t_0 + 5, \ldots, T \tag{8}$$

$$\text{Level 3}: \quad \log SRB_t \mid SRB^*_t, \sigma^2_{SRB} \sim \text{Normal}\left(\log SRB^*_t, \sigma^2_{SRB}\right) \tag{9}$$

$$\log f_{a,t} \mid f^*_{a,t}, \sigma^2_f \sim \begin{cases} \text{Normal}\left(\log f^*_{a,t}, \sigma^2_f\right), & a = a_L^{[\text{fert}]}, \ldots, a_U^{[\text{fert}]} \\ \text{undefined}, & \text{otherwise} \end{cases} \tag{10}$$

$$\log n_{a,t_0,l} \mid n^*_{a,t_0,l}, \sigma^2_n \sim \text{Normal}\left(\log n^*_{a,t_0,l}, \sigma^2_n\right) \tag{11}$$

$$\text{logit}\, s_{a,t,l} \mid s^*_{a,t,l}, \sigma^2_s \sim \text{Normal}\left(\text{logit}\, s^*_{a,t,l}, \sigma^2_s\right), \quad a = 0, 5, \ldots, A+5 \tag{12}$$

$$g_{a,t,l} \mid g^*_{a,t,l}, \sigma^2_g \sim \text{Normal}\left(g^*_{a,t,l}, \sigma^2_g\right) \tag{13}$$

(where $a = 0, 5, \ldots, A$; $t = t_0, t_0 + 5, \ldots, T$; $l = F, M$ in (7)–(13) unless otherwise specified)

$$\text{Level 4}: \qquad \sigma^2_v \sim \text{InvGamma}(\alpha_v, \beta_v),\ v = n, f, s, g, SRB. \tag{14}$$

For $x < 0 < 1$, $\text{logit}(x) \equiv \log(x/(1-x))$. The joint prior at time t is multiplied by

$$I\left\{M\left(\boldsymbol{n}_t, \boldsymbol{\psi}_t\right) > 0\right\} \equiv \begin{cases} 1 & \text{if, for all } a = 0, \ldots, A \text{ and } l = F, M,\ n_{a,t+5,l} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

to ensure a non-negative population. Wheldon et al.'s (2013) female-only model had SRB fixed at 1.05 and $l = F$. SRB can be interpreted as the odds that a birth is male, so (9) is a model for the log-odds that a birth is male.

The parameters $\alpha_v, \beta_v,\ v = n, f, s, g, SRB$ define the distribution of the standard deviation parameters that represent measurement error in the initial estimates. We set these parameters based on the expert opinion of UNPD analysts by eliciting liberal, but realistic, estimates of initial estimate accuracy. We elicit on the observable marginal quantities, $f_{a,t}$, $s_{a,t,l}$, $g_{a,t,l}$, and $n_{a,t,l}$, which have Student's t distributions centred at the initial point estimates and with variance and degrees of freedom dependent on α and β. We set $\alpha_v = 0.5$, $v = f, s, n, g$, which gives the initial estimates a weight equivalent to a single data point. The β_v are then determined by specifying the limits of the central ninety percent probability interval of the marginal distributions. Population counts and fertility rates are modelled on the log scale so this amounts to making a statement of the form "the probability that the true parameter values are within p-percent of the initial point estimates is 90 percent". Migration is explicitly modeled as a proportion so this interpretation is direct for this parameter. The survival parameters are also proportions but they are modelled on the logit scale. We set β_s such that the untransformed $s_{a,t,l}$ lie within the elicited intervals. In all cases, we call p the *elicited relative error*.

Bayesian reconstruction defines a joint prior distribution over the input parameters (9)–(15) which induces a prior on the population counts after the baseline via CCMPP. This is "updated" using the census counts for which a likelihood is given in (7). Some methods of estimating migration rely on "residual" counts; projeted counts based only on vital rates are compared with census counts and the difference attributed to international migration. Methods of adjusting vital rates and census counts to ensure mutual consistency have also been proposed that use a similar approach (e.g., Luther, Dhanasakdi, et al., 1986; Luther and Retherford, 1988). Initial estimates of $f^*_{a,t}$,

$s^*_{a,t,l}$, $g^*_{a,t,l}$ should not be based on such methods since this would amount to using the data twice and uncertainty will be underestimated in the posterior.

3 Application

We apply two-sex Bayesian reconstruction to the populations of India from 1971–2001, Thailand from 1960–2000 and Laos from 1985–2005. The periods of reconstruction are determined by the available data. Laos has no vital registration data. Initial estimates of fertility are based on surveys of women and the only mortality estimates are for ages under five derived from these same surveys. Thailand and India have acceptable vital registration data for these periods which provide information about fertility and mortality at all ages. Nevertheless adjustments are necessary to reduce bias due to undercount of certain groups. For example, vital registration is thought to have underestimated U5MR in Thailand, for example (Hill, Vapattanawong, et al., 2007; Vapattanawong and Prasartkul, 2011) and in India 50–60 percent of children are born at home which increases the likelihood of omission from the register (UNFPA, 2010).

Estimates of population sex ratios in India have been relatively high throughout the twentieth century. Prior to the late 1970s, these were thought to have been caused by an excess of female mortality (high SRMs), and from the late 1970s onwards by high SRBs. Both of these phenomena have been linked to cultural preferences for sons over daughters which were intensified by a rapid fall in fertility rates (Bhat, 2002b,c; Das Gupta, 2005; Guilmoto, 2007a; Visaria, 1971). Concern over the accuracy of certain estimates of SRB has led some authors to suggest using the SRTP and sex ratios for young age groups as proxies for SRB and SRMs (Bhat, 2002b,c; Guilmoto, 2009). We use Bayesian reconstruction to derive credible intervals for the SRTP and the sex ratio among those aged 0–5 for India.

Thailand experienced an even more rapid decline in TFR between 1960 and about 1980 (Kamnuansilpa et al., 1982). Estimates of Thailand's SRBs between 1960 and about 1970 are relatively high, but are within the typical range from about 1970 to 2000. Surveys of Thai families in the 1970s found that girls and boys were desired about equally (Guilmoto, 2009; Knodel, Ruffolo, et al., 1996).

Fertility rates in Laos have fallen since 1985 but remain high relative to other Asian countries. Very little has been written about sex ratios for this country (but see Frisen, 1991).

In the remainder of this section, we briefly describe the data sources for each country and the method used to derive initial estimates. These are followed by results for selected parameters. All computations were done using the R environment for statistical computing (R Core Team, 2013); Bayesian reconstruction is implemented in the package "popReconstruct". The method of Raftery and Lewis (1996) was used to select the length of MCMC chains.

3.1 Data Sources and Initial Estimates

3.1.1 India, 1971–2001

Censuses have been taken roughly every 10 years in India since 1871. We begin our period of reconstruction with 1971. This is the first census year for which vital rate data independent of the censuses are available, collected by the Indian Sample Registration System. Subsequent censuses were taken in 1981, 1991 and 2001 (sufficiently detailed results from the 2011 census were not available at the time of writing). Counts in WPP 2010 were used as these were adjusted to reduce bias. Estimates of SRB, fertility and survival were based on data from the Sample Registration System (The Registrar General and Census Commissioner of India, 2011), the National Family Health Surveys conducted between 1992 and 2006 (The Registrar General and Census Commissioner of India, 2009) and the 2002–04 Reproductive Child Health Survey. Weighted cubic splines were used to smooth estimates of SRB and fertility in the same manner as for Thailand. The same initial estimates for migration were used for India as for Laos and Thailand and the elicited relative errors were also the same. The elicited relative error of 10 percent for the vital rates is consistent with independent assessments of the coverage of the Sample Registration System (Bhat, 2002a; Mahapatra, 2010).

3.1.2 Thailand, 1960–2000

Censuses were conducted in 1960, 1970, 1980, 1990 and 2000 (detailed results from the census conducted in 2010 were not available at the time of writing). We used the counts in WPP 2010 which were adjusted for known biases such as undercount. Initial estimates of sex ratio at birth were taken from current fertility based on vital registration. The elicited relative error was set to 10 percent. Initial estimates of age-specific fertility were based on direct and indirect estimates of current fertility and children ever born (CEB) based on the available data including surveys and vital registration. Each data series was normalized to give the age pattern and summed to give TFR. These were smoothed separately using weighted cubic splines and the resulting estimates combined to yield a single series of initial estimates of age-specific fertility rates. The weights were determined by UN analysts based on their expert judgment about the relative reliability of each source. The elicited relative error was set to 10 percent. Initial estimates of survival for both sexes were based on life tables calculated from vital registration, adjusted for undercount using data from surveys. We used the same initial estimates of international migration as for Laos.

3.1.3 Laos, 1985–2004

National censuses were conducted in 1985, 1995 and 2005, hence we reconstruct the whole population between 1985 and 2005. We used Wheldon et al.'s (2013) initial estimates for fertility, female mortality, migration and population counts. In these, migration was centred at zero for all sexes, ages and time periods, with a large relative error of 20 percent. Initial estimates for males were derived in an analogous manner. There was very little information about the sex ratio at birth, so initial estimates were set at 1.05, a demographic convention (Preston et al., 2001), with a large elicited relative error of 20 percent.

3.2 Results

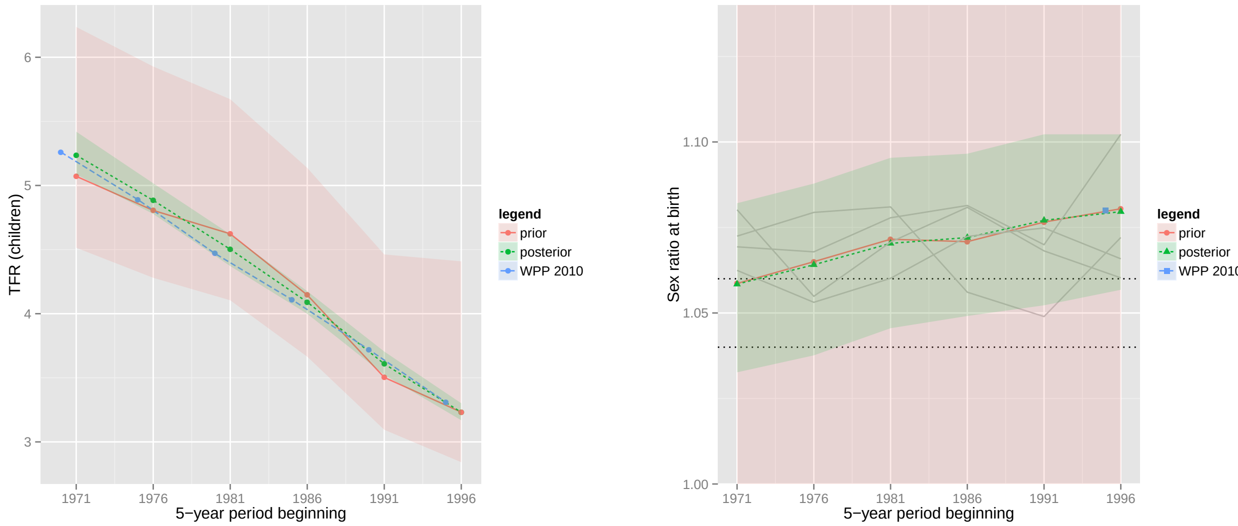
Key results are given by country. We show the limits of central 95 percent credible intervals for the marginal prior and posterior distributions of selected parameters. The magnitude of uncertainty will be summarized using half-widths of these intervals, averaged over sex and time. We compare our results to those published in WPP 2010 for years with comparable estimates. WPP 2010 did not use Bayesian reconstruction but it is based on the same data, hence the comparison is useful.

3.2.1 India, 1971–2001

Figures 1a and 1b show posterior 95 percent intervals for TFR and SRB for India. Median TFR decreased consistently and the posterior intervals have half-width 0.11 children per woman. The marginal posterior for SRB is centered above the range 1.04–1.06 from 1976–2001, which suggests that SRBs might have been atypically high over this period. There also appears to have been an increase in SRB over the period of reconstruction. Under Bayesian reconstruction, the posterior probabilities of these events can be estimated in a straightforward manner from the poster sample. The posterior probabilities that SRB exceeded 1.06 in each of the five-year periods are in Table 1a. Strong evidence for high SRB was found for the period 1991–2001.

To investigate the trend further we looked at the posterior distributions of two measures of linear increase: 1) the difference between SRBs in the first and last five-year sub-intervals; 2) the slope coefficient in the ordinary least squares (OLS) regression of SRB on the start year of each time period. Each quantity was calculated separately for each SRB trajectory in the posterior sample. Some actual trajectories are shown in Figure 1b. These measures merely summarize the posterior obtained from the reconstruction in simple ways; linear regression models were not used to obtain the sample from the posterior. The probabilities that the simple difference and slope coefficient were greater than zero are 0.92 and 0.93 respectively (Table 2).

Results for e_0 are shown in Figure 2. The mean interval half-width for the sex difference in life expectancies is 1.7 years. Life expectancy at birth increased for both sexes over the period of reconstruction (Figure 2a) but the sex difference suggests that female e_0 might have increased more rapidly than male e_0 and even exceeded it in the period 1996–2001 (Figure 2c). The posterior probabilities that female e_0 exceeded male e_0 support this (Table 1b). The possibility of an increase in the female−male difference in e_0 was investigated using the same method applied to SRB. The



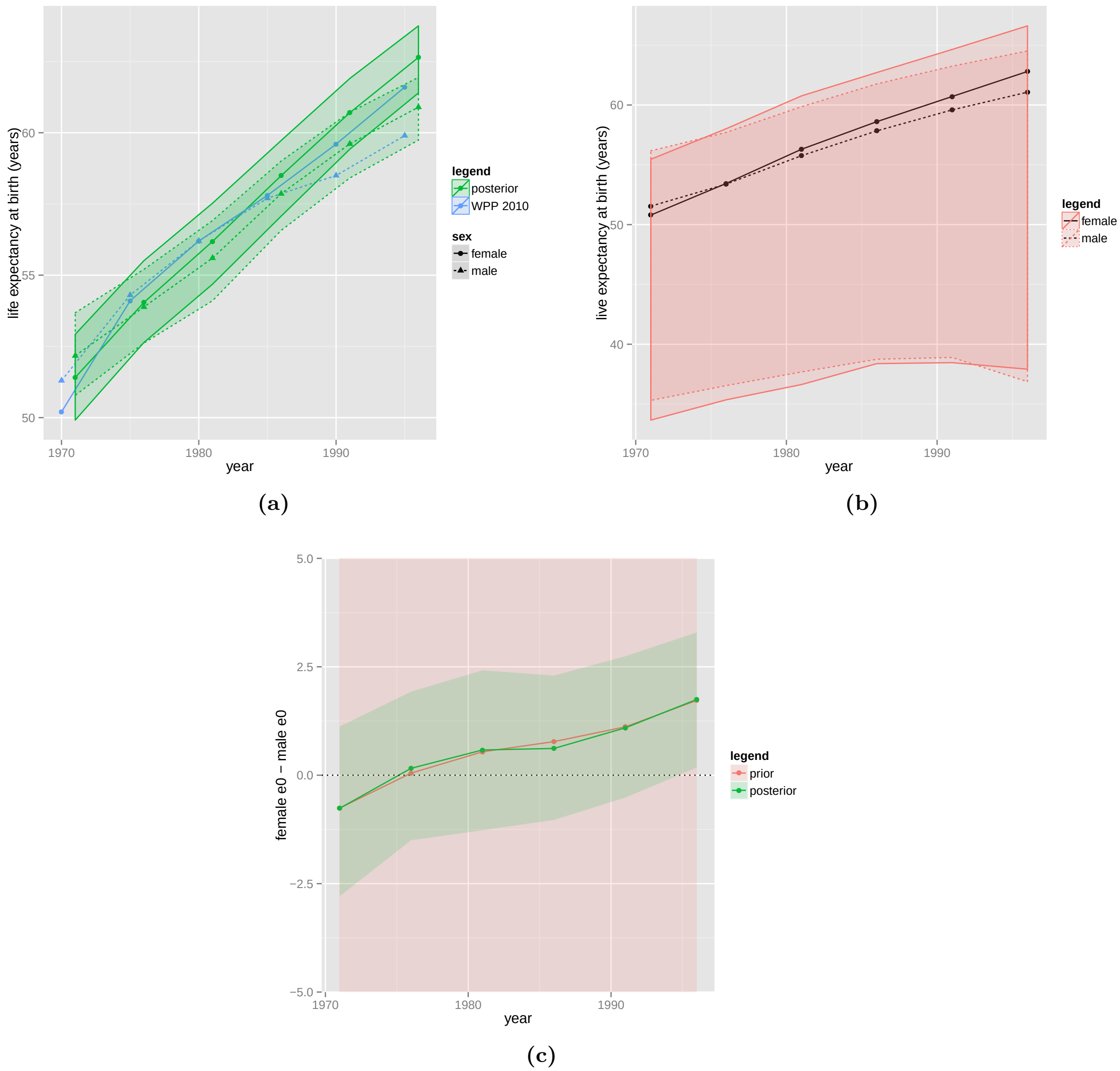
(a) **(b)**

Figure 1. Prior and posterior medians and 95 percent Bayesian confidence intervals for the reconstructed population of India, 1971–2001: (a) total fertility rate (TFR); (b) sex ratio at birth (SRB) (four trajectories from the Markov chain Monte Carlo (MCMC) sample are also shown).

probability of a simple increase between 1971 and 2001 is 0.98 and the probability that slope is less than zero is 0.98 (Table 2); strong evidence of a positive time trend.

While the posterior distribution of the sex ratio of under-five mortality rate (SRU5MR) is centred below unity, it is quite dispersed. One is contained within the 95 percent credible intervals for all time periods. The probability that female U5MR exceeded that of males ranged from 0.72 to 0.89 over the period of reconstruction. Evidence of a linear trend was not found.

Population sex ratios are shown in Figure 3a. The probability of a simple decrease in SRTP is 1 and the probability that the OLS slope is less than zero is 1; very strong evidence for a decline over the period of reconstruction. Sex ratios in the population under five (SRU5s) increased in the WPP population counts, but our posterior median remained relatively constant after an initial decline. The probability that the difference between the sex ratio in the last period minus that in the first was negative is 0.8. The probability that the OLS coefficient was negative is 0.71. Mean half-widths of the intervals for SRTP and the sex ratio in ages 0–5 are 0.01 and 0.027 respectively. Uncertainty is higher in years without a census.



(a) (b) (c)

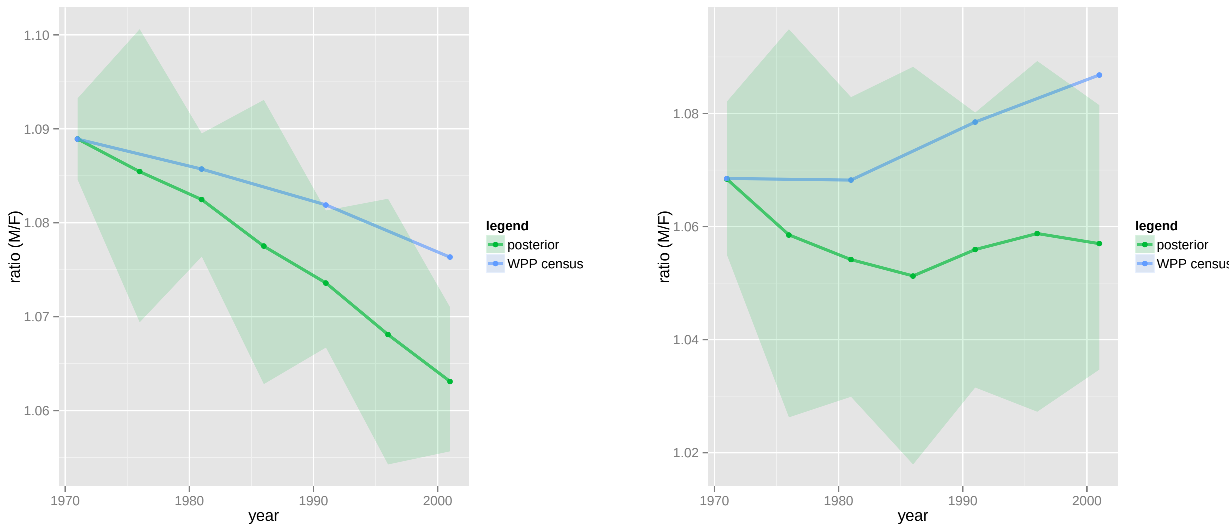
Figure 2. Prior and posterior medians and 95 percent Bayesian confidence intervals for life expectancy at birth (e_0) for the reconstructed population of India, 1971–2001. (a): Female and male posterior quantiles with *World Population Prospects* (WPP) 2010 estimates; (b): Female and male prior quantiles only; (c): Male-to-female difference.

Table 1. Probability that sex ratios and differences exceeded certain thresholds for the reconstructed population of India, 1971–2001, by five-year time period.

1971	1976	1981	1986	1991	1996
(a) Pr($SRB > 1.06$)					
0.44	0.66	0.83	0.86	0.93	0.96
(b) Pr($female\ e_0 - male\ e_0 > 0$)					
0.21	0.58	0.74	0.78	0.91	0.99

Table 2. Probabilities of an increasing linear trend and 95 percent confidence intervals for sex ratio at birth (SRB), sex difference in life expectancy at birth (sex diff. e_0), sex ratio in the total population (SRTP), and sex ratio in the population under five (SRU5) for the reconstructed population of India, 1971–2001. Two measures of trend are used: the difference over the period of reconstruction and the slope coefficient from the ordinary least squares (OLS) regression on the start year of each time period.

Measure of trend	95% CI	Prob > 0
Sex ratio at birth (SRB)		
$\mathrm{SRB}_{1996} - \mathrm{SRB}_{1971}$	[−0.011, 0.054]	0.92
OLS slope (SRB ∼ year)	[−0.00034, 0.0021]	0.93
Sex difference in life expectancy at birth (sex diff. e_0)		
$(\text{sex diff. } e_0)_{1996} - (\text{sex diff. } e_0)_{1971}$	[0.12, 5]	0.98
OLS slope (sex diff. e_0 ∼ year)	[0.0066, 0.17]	0.98
Sex ratio in the total population (SRTP)		
$\mathrm{SRTP}_{1996} - \mathrm{SRTP}_{1971}$	[−0.034, −0.017]	0.000027
OLS slope (SRTP ∼ year)	[−0.0013, −0.00045]	0.00035
Sex ratio in the population under five (SRU5)		
$\mathrm{SRU5}_{1996} - \mathrm{SRU5}_{1971}$	[−0.037, 0.017]	0.2
OLS slope (SRU5 ∼ year)	[−0.0011, 0.00062]	0.29



(a) **(b)**

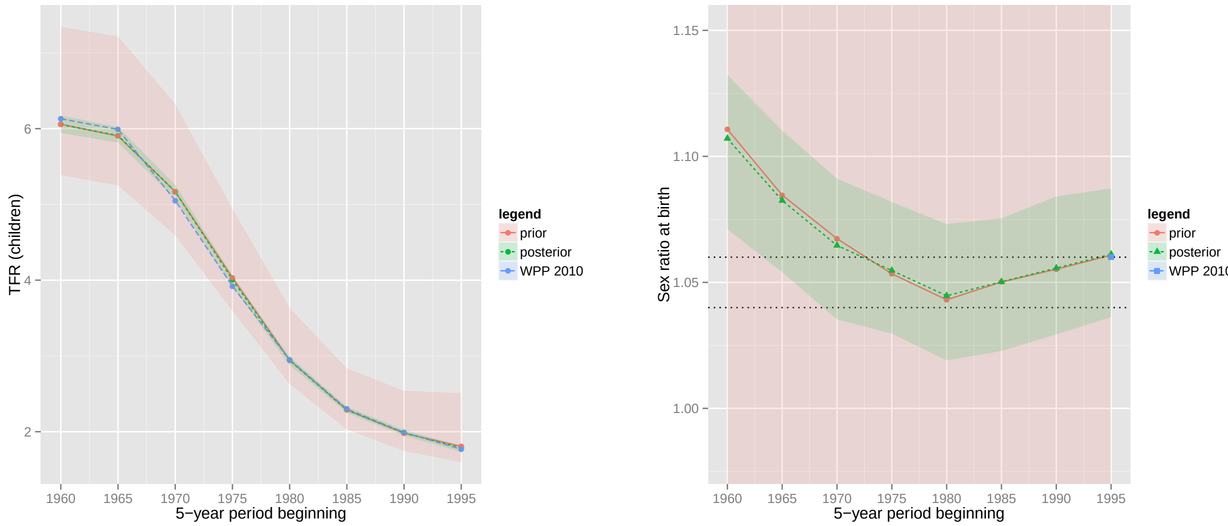
Figure 3. Prior and posterior medians and 95 percent Bayesian confidence intervals for sex ratios in the reconstructed population of India, 1971–2001. (a): Total population; (b): Population aged 0–5.

3.2.2 Thailand, 1960–2000

TFR fell very steeply in Thailand from 1960–2000 (Figure 4a). Posterior uncertainty about this parameter is small; the mean half-width of the posterior intervals is 0.07 children per woman. Posterior uncertainty about SRB is higher. In particular, 95 percent posterior intervals contain the typical range of 1.04–1.06 for all but the first two time periods (Figure 4b). The probability that SRB exceeded 1.06 in each period is given in Table 3. There is strong evidence that SRBs were atypically high in the period 1960–1969.

Assessing the time trend in SRB is complicated by the curvilinear trend in the posterior median and the probabilities in Table 3. Linear summaries across the whole period of reconstruction, such as the OLS slope parameter, are clearly not appropriate. Piece-wise linear regression models are available (e.g., Hinkley, 1969; Hinkley, 1971), but the posterior traces consist of only eight points and are quite volatile, making identification of the change point difficult. Instead we summarize the time trend with the following two differences: $\text{SRB}_{1960} - \text{SRB}_{1980}$ and $\text{SRB}_{1980} - \text{SRB}_{1995}$. The posterior joint probability of a decrease from 1960–1964 to 1980–1984 followed by an increase to 1995–1999 (i.e., $\Pr(\{\text{SRB}_{1960} - \text{SRB}_{1980} > 0\} \cap \{\text{SRB}_{1980} - \text{SRB}_{1995} < 0\})$) is 0.84.

Results for the sex difference in e_0 mortality are shown in Figure 5. The mean half-width for e_0 (across all time periods and sex) is 1.7 years. Our posterior intervals for the sex difference in e_0 lie entirely above zero in each five-year period, suggesting that female longevity was greater than that of males in Thailand from 1960–2000 (mean half-width of the difference: 2.4 years). There is



(a) **(b)**

Figure 4. Prior and posterior medians and 95 percent Bayesian confidence intervals for the reconstructed population of Thailand, 1960–2000: (a) total fertility rate (TFR); (b) sex ratio at birth (SRB).

Table 3. Probability that sex ratio at birth (SRB) was greater than 1.06 for the reconstructed population of Thailand, 1971–2001, by five-year time period.

1960	1965	1970	1975	1980	1985	1990	1995
0.99	0.95	0.66	0.32	0.11	0.19	0.35	0.54

also strong evidence for a positive trend; the probability that the simple difference (1995 period minus 1960 period) in the sex differences in e_0 was greater than 0 is 0.96 and the probability that the OLS slope is greater than 0 is 1 (Table 4).

The posterior for SRU5MR suggests mortality at ages 0–5 was similar for both sexes. In no time period is there convincing evidence for a male-to-female ratio less than one. Similarly there is not strong evidence for a linear trend in this parameter over the period of reconstruction.

Table 4. Probabilities of an increasing linear trend and 95 percent confidence intervals for sex difference in life expectancy at birth (sex diff. e_0) for the reconstructed population of Thailand, 1960–2000. Two measures of trend are used: the difference over the period of reconstruction and the slope coefficient from the ordinary least squares (OLS) regression on the start year of each time period.

Measure of trend	95% CI	Prob > 0
$(\text{sex diff. } e_0)_{1996} - (\text{sex diff. } e_0)_{1971}$	$[-0.42, 7.1]$	0.96
OLS slope (sex diff. $e_0 \sim$ year)	$[0.027, 0.18]$	1

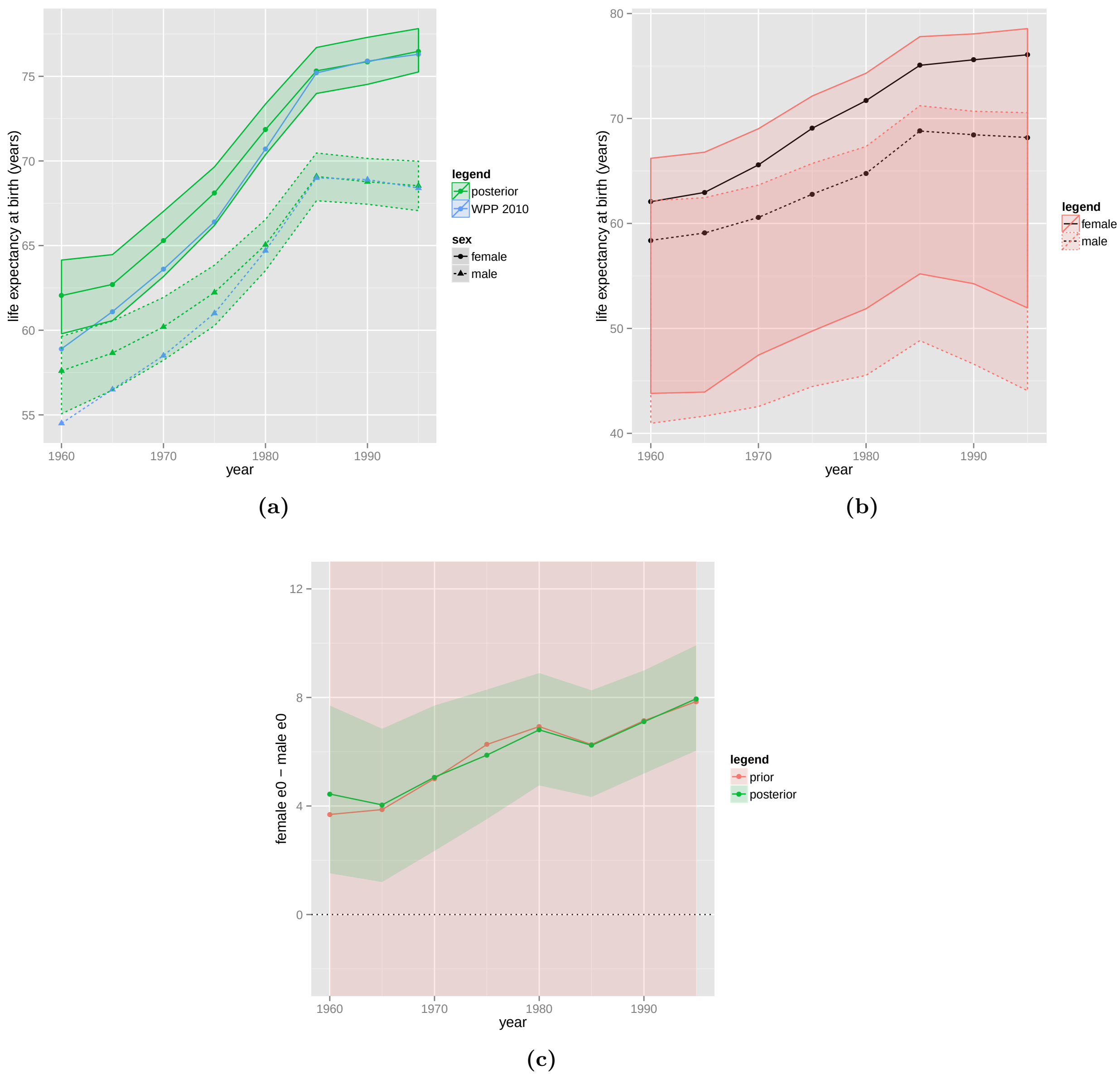
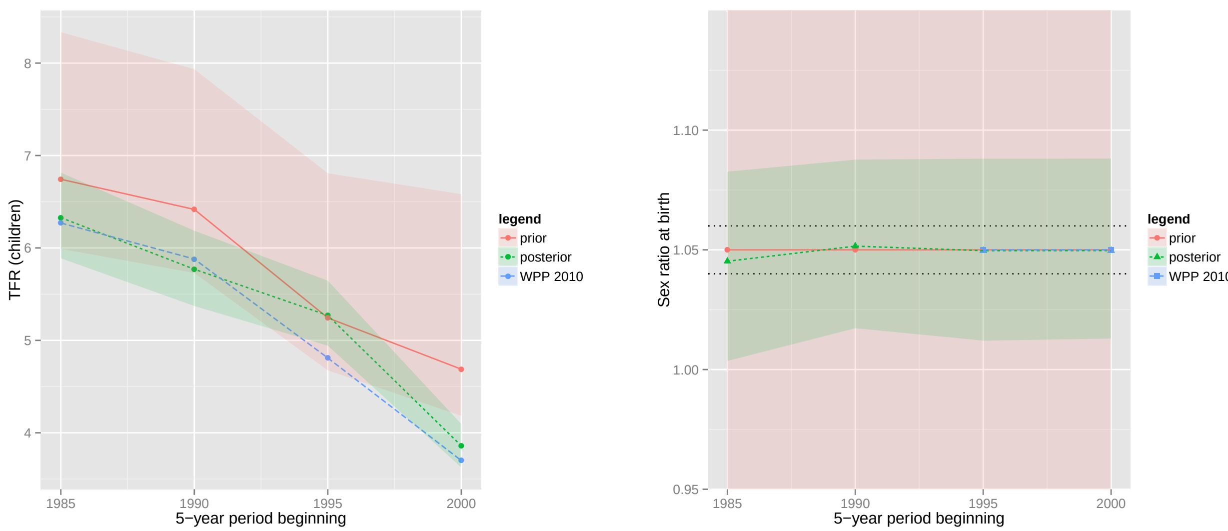


Figure 5. Prior and posterior medians and 95 percent Bayesian confidence intervals for life expectancy at birth (e_0) for the reconstructed population of Thailand, 1960–2000. (a): Female and male posterior quantiles with *World Population Prospects* (WPP) 2010 estimates; (b): Female and male prior quantiles only; (c): Male-to-female difference.

3.2.3 Laos, 1985–2004

Medians and prior and posterior credible intervals for TFR and SRB are shown in Figure 6. The posterior for TFR obtained here (Figure 6a) is very similar that obtained by Wheldon et al. (2013a) who reconstructed the female-only population and did not estimate SRB; it was kept fixed at 1.05 throughout. The value 1.05 is a demographic convention (Preston et al., 2001).

There was very little data on SRB for Laos so, in this study, the initial estimate of SRB was fixed at 1.05 in all five-year periods but a posterior distribution was estimated using the model. The posterior median SRB deviates very little from the initial estimate, although the uncertainty has been considerably reduced; the mean of the half-widths of the 95 percent credible intervals is 0.038 compared with 0.39 for the prior (Figure 6b). The probability that SRB was above 1.06 in any of the five-year periods is low (Table 5) and the evidence for a trend over time is equivocal (Table 6).



(a) (b)

Figure 6. Prior and posterior medians and 95 percent Bayesian confidence intervals for the reconstructed population of Laos, 1985–2004: (a) total fertility rate (TFR); (b) sex ratio at birth (SRB).

There is strong evidence that female e_0 was higher from 1990 through 2005 but there appears to be no evidence for a sex difference between 1985 and 1990 (Figure 7, Table 5b). The posterior distributions of both trend summaries provide strong evidence for an increase in the sex difference over the period of reconstruction (Table 6), although this is due primarily to the increase immediately following the 1985–1990 period.

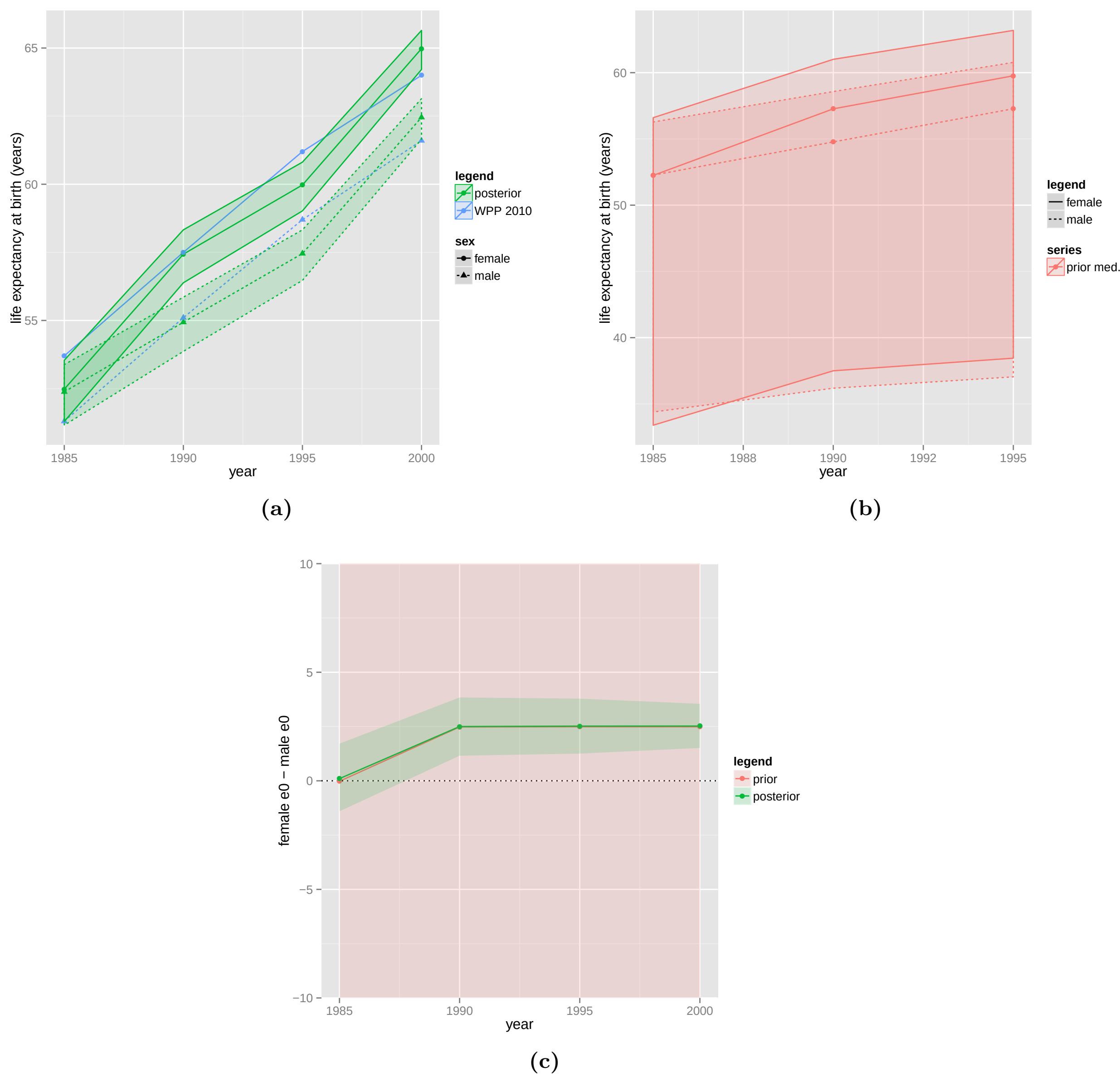


Figure 7. Prior and posterior medians and 95 percent Bayesian confidence intervals for life expectancy at birth (e_0) for the reconstructed population of Laos, 1985–2004. (a): Female and male posterior quantiles with *World Population Prospects* (WPP) 2010 estimates; (b): Female and male prior quantiles only; (c): Male-to-female difference.

Table 5. Probability that sex ratios and differences exceeded certain thresholds for the reconstructed population of Loas, 1985–2005, by five-year time period.

1985	1990	1995	2005
(a) Pr($SRB > 1.06$)			
0.20	0.30	0.27	0.27
(b) Pr($female\ e_0 - male\ e_0 > 0$)			
0.56	1.00	1.00	1.00

Table 6. Probabilities of an increasing linear trend and 95 percent confidence intervals for protectsex difference in life expectancy at birth (sex diff. e_0) for the reconstructed population of Laos, 1985–2000. Two measures of trend are used: the difference over the period of reconstruction and the slope coefficient from the ordinary least squares (OLS) regression on the start year of each time period.

Measure of trend	95% CI	Prob > 0
Sex ratio at birth (SRB)		
$\text{SRB}_{2000} - \text{SRB}_{1985}$	[−0.046, 0.061]	0.58
OLS slope (SRB ∼ year)	[−0.0034, 0.0042]	0.56
Sex difference in life expectancy at birth (sex diff. e_0)		
$(\text{sex diff. } e_0)_{2000} - (\text{sex diff. } e_0)_{1985}$	[0.54, 4.2]	0.99
OLS slope (sex diff. e_0 ∼ year)	[0.025, 0.26]	0.99

4 Discussion

We have described a major extension to Bayesian reconstruction, a method of reconstructing human populations of the recent past which yields probabilistic estimates of uncertainty (Wheldon et al., 2013a,b). We reconstructed the populations of Laos from 1985–2005, Thailand from 1960–2000 and India from 1971–2001, paying particular attention to sex ratios of fertility and mortality indicators. We estimate that the posterior probability that SRB was above 1.06 is greater than 0.9 in India between 1991 and 2001 and the probability it increased over this period is about 0.92. The SRB was also above 1.06 with high posterior probability in Thailand from 1960–1970. We estimate that the probability it decreased between 1960 and 1980, then increased from 1985 to 2000 is 0.84. We found no evidence for atypically high SRBs, or a trend over the period of reconstruction, for Laos, a country with much less available data than Thailand and India. In both Thailand and Laos, we found strong evidence that e_0 was greater for females and, in Thailand, that the difference increased over the period of reconstruction. In India, the probability that female e_0 was lower was 0.79 but there was strong evidence for a narrowing of the gap through to 2001.

In its original formulation, Bayesian reconstruction was for female-only populations; here we show how two-sex populations can be reconstructed using the same framework. The method takes a set of data-derived, bias-reduced initial estimates of age-specific fertility rates, and age-sex-specific survival proportions, migration proportions and population counts from censuses, together with expert opinion on the measurement error informed by data if available. Bayesian reconstruction updates initial estimates using adjusted census counts via a Bayesian hierarchical model. The periods of reconstruction used in our applications begin in the earliest census year for which non-census vital rate data were available, and end with the year of the most recent census. Reconstruction can be done further ahead, but without a census the results are based purely on the initial estimates.

Previous methods of population reconstruction were purely deterministic, were not designed to work with the type of data commonly available for many countries over the last sixty years or did not account for measurement error (e.g., Bertino and Sonnino, 2003; Lee, 1971, 1974; Wrigley and Schofield, 1981). Daponte et al. (1997) used a Bayesian approach to construct a counterfactual population, but age patterns of fertility were held fixed and mortality varied only through infant mortality. Bayesian reconstruction does not impose fixed age-patterns and mortality can vary at each age. Moreover, international migration is estimated in the same way as fertility and mortality.

We considered sex ratios of births and mortality because these are of interest to demographers and policy-makers, especially since they determine the SRTP (Griffiths et al., 2000; Guillot, 2002). It is convention to compare sex-specific U5MRs with a ratio but sex-specific e_0s with using a difference. We have not studied the associations among U5MR ratios, e_0 differences and population sex ratios. The SRTP is a function of life-time cohort mortality but the U5MR and e_0 presented here are period measures for which the relationship used by Guillot (2002) does not hold. Our results add to previous work on SRMs, especially that of Sawyer (2012) who studied sex ratios of U5MR and called for further work to quantify its uncertainty. Sawyer (2012) decomposed U5MR into mortality between ages 0 and 1 (infant mortality) and mortality between ages 1 and 5 (child mortality). We reconstructed populations in age- time-intervals of width five because these are the intervals for which data is most widely available across all countries.

Many methods of adjusting vital registration using census data have been proposed (e.g., Bennett and Horiuchi, 1981; Hill, 1987; Luther and Retherford, 1988), but these either ignore, or deal poorly with, intercensal migrations by essentially truncating the age groups most affected by migration (Murray, Rajaratnam, et al., 2010). Luther, Dhanasakdi, et al. (1986) and Hill, Vapattanawong, et al. (2007) applied these methods to Thailand to estimate under-counts. The aim of these methods is to produce improved point estimates of vital rates. We have not used any census data to derive our initial estimates and Bayesian reconstruction does not produce improved point estimates. For example, initial estimates of survival were not based on inter-censal cohort survival and initial estimates of migration were not based on "residual" counts. Doing so would have amounted to using the census data twice and under-estimated the uncertainty. The outputs of Bayesian reconstruction are interval estimates which quantify uncertainty probabilistically.

4.1 Sex Ratios in Asia

The SRTP is a crude measure of the balance of sexes in a population since it is not age-specific; sex ratios may be quite different among age groups, for example. Nevertheless, there is a large literature devoted to estimating SRTPs and exploring the causes and consequences, expecially in Asia where SRTPs are the highest in the world (UN, 2011). A persistently high sex ratio among younger cohorts will probably lead to a "marriage squeeze" in which many young males may not be able to marry due to a lack of eligible females (Guilmoto, 2012). Formally, high SRTPs are the result of high SRBs and life-time sex ratios of cohort mortality (Guillot, 2002). The relative effects of these two factors may vary by time and country.

A normal range for SRB is believed to be 1.04–1.06 (UNFPA, 2010); on average, slightly more than half of all newborns are male. Concerns about quality, or a complete lack of data, have made it difficult to accurately estimate SRBs in many Asian countries.

With very few exceptions, country-level e_0s for females exceed those of males. The only countries for which this is not true are in South and Central Asia (e.g., India) and countries in Africa with generalised HIV epidemics (UN, 2011). This pattern is consistent with a female survival advantage; in virtually all contemporary human populations, females appear to age slower and live longer than males (Soliani and Luchetti, 2006; Vallin, 2006). Behavioural factors appear to play an important role, especially in developed countries where the prevalence of harmful activities, such as smoking, consumption of alcohol, and risky behaviour leading to accidental death, tends to be lower among women than among men (Lalic and Raftery, 2012; Waldron, 1985, 2009). However, consistency across many very different cultural groups suggests an underlying biological cause. To-date, there is no consensus as to what this might be (Austad, 2011).

4.1.1 India

The SRTP in India has received considerable attention, particularly as an indicator of discrimination against women (UNFPA, 2010). Griffiths et al. (2000) showed that only slightly elevated SRBs and SRMs at young ages are sufficient to produce the observed SRTP in India, if they persist for a long period of time. The relative contribution of these two factors may have varied over time. Bhat (2002c) and Guilmoto (2007a, 2009) argue that India experienced a transition in the 1970s whereby high SRB replaced low SRM as the cause of the high SRTPs observed throughout the period (recall that a low SRM is a result of lower male mortality). Evidence suggested that low SRMs were due to female neglect and infanticide. In the 1970s, these practices gave way to sex selective abortion which raised the SRB instead.

The transition hypothesis is based on several pieces of evidence. Data suggested a possible rise in SRB in India above the typical range of 104–106 in the mid 1980s (Guilmoto, 2007a). In the 1970s, amniocentesis started to became available as a method for determining the sex of a foetus and abortion was legalized. Ultrasonography, a less invasive way of determining foetal sex, started to became available in many parts of India in the 1980s. In certain regions, such as the northern states and highly urbanized areas, there is a long standing tradition of preference for sons over daughters (Bongaarts, 2001; Mayer, 1999). The arrival of these new technologies in this context appears to have led to an increase in sex selective abortions in India (Bhat, 2002c; APCRSHR, 2007; Jha, Kesler, et al., 2011; Jha, Kumar, Vasa, et al., 2006; UNFPA, 2010). The steep decline in Indian TFR, which began in the early 1970s, appears to have contributed to this phenomenon. Several studies have found evidence that SRB is higher at higher parities (birth orders), both in India (Bhat, 2002c; Das Gupta and Bhat, 1997; Jha, Kesler, et al., 2011; Jha, Kumar, and Dhingra, 2006) and other Asian countries (Das Gupta, 2005). The increase appears to be greater if none of the earlier births were male. As family sizes decrease with TFR, the risk of having no sons increases (Guilmoto, 2009). Therefore, in cultures where sons play important economic and social functions, or where families benefit materially much more from the marriage of a son than of a daughter, the incentive to use sex selective abortion increases (Guilmoto, 2009; Mayer, 1999). After combining all available data and including uncertainty, we estimate that the probability SRB was greater than 1.06 was above 0.9 over the period 1971–2001 and between 0.8 and 0.9 over the period 1981–1991. After accounting for uncertainty, the probability that SRBs was atypically high between 1971 and 1981 is 0.66 or lower. The probability that there was an increase in SRB between 1971 and 2001 is estimated to be above 0.9.

Overall mortality decreased rapidly in India from about 1950 as infectious diseases were brought under control, food security increased and health services became more widely available (Bhat, 2002c). Our results suggest that, after taking account of uncertainty, there was an increase in e_0 and a continual decrease in U5MR between 1971 and 2001 for both sexes. Using Sample Registration System data, (Bhat, 2002c) noted that the decrease was greater for females than males and our analysis of the change in the sex difference of e_0 supports this; we found that the probability that the difference increased over the period of reconstruction is about 0.98. The probability of a decline in the SRTP was found to be similarly strong. For the 0–5 age group, findings were not as clear. Evidence that female U5MR exceeded male was not strong, and there was little evidence for a trend in either SRU5MR or SRU5.

India's large population makes it a very important case for the study of sex ratios in Asia (Guilmoto, 2007a) and, like other authors (e.g., Guillot, 2002; UN, 2011), we have focused on country level estimates only. However, where available, data suggest that there are large regional veriations in SRBs and population sex ratios, with estimates for urban areas and northern states being much higher than other areas (Bhat, 2002c; Guilmoto, 2009; Jha, Kesler, et al., 2011).

Currently, Bayesian reconstruction is not able to produce sub-national estimates, but could be extended to do so in the future.

4.1.2 Thailand

Like other parts of Asia, Thailand experienced rapid economic growth and a rapid fall in TFR from 1960. The TFR decline was accompanied by an increase in the widespread use of modern contraceptive methods made available through government supported, voluntary family planning programs (Kamnuansilpa et al., 1982; Knodel, 1987; Knodel, Ruffolo, et al., 1996). Unlike India, Thailand is not considered to have had a high SRB (Guilmoto, 2009). Vital registration data, which formed the basis of our initial estimates, indicate that SRB was high during the period 1960 and 1970, but remained within the typical range thereafter. This is consistent with studies conducted from the mid-1970s onward which found that the most commonly desired family configuration in Thailand was for small families of two or three children consisting of at least one boy and one girl. TFR decline in Thailand may have intensified this preference (Knodel and Prachuabmoh, 1976; Knodel, Ruffolo, et al., 1996).

4.1.3 Laos

Fertility in Laos remained high relative to its neighbours. For example, the estimated TFR for 1985 in Laos is comparable to the 1960 estimate for Thailand (Frisen, 1991). Posterior uncertainty about SRB is high and our results provide no evidence to suggest levels were atypical between 1985 and 2005. All-age mortality as summarised by e_0 does appear to have been higher for females from 1990 onwards.

4.2 Further Work and Extensions

Our prior distributions were constructed from initial point estimates of the CCMPP input parameters, together with information about measurement error. In the examples given here, this was elicited from UN analysts who are very familiar with the data sources and the demography of each country. In cases where good data on measurement error are available, they can be used. For example, Wheldon et al. (2013a) used information from post-enumeration surveys to estimate the accuracy of New Zealand censuses. Data of this kind are rarely available in developing countries for either census or vital rate data. Nevertheless, surveys come with a substantial amount of meta-data which can be used to model accuracy. This approach was taken by Alkema et al. (2012) who developed a method for estimating the quality of survey-based TFR esitmates in west Africa by modelling bias and measurement error variance as a function of selected "data quality covariates". Wheldon et al. (2013b) used these results in their reconstruction of the female population of Burkina Faso. Extending this work to more countries could provide a source of initial estimates together with uncertainty that could be used as inputs to Bayesian Reconstruction.

Posterior uncertainty in estimates of U5MR was found to be high and we did not find strong evidence for a skewed sex ratio in any of our examples. U5MR is under-identified in the model because the only census count it affects, that for ages 0–5, is also dependent on SRB and TFR. Recent work by the UN Interagency Group for Child Mortality Estimation (IGME) has focused on producing probabilistic estimates of U5MR (Hill, You, et al., 2012). It might be possible to improve identifiability of the U5MR in Bayesian reconstruction by splitting the variance parameter σ^2 into $\sigma^2_{0\text{–}5}$ and σ^2_{5+}, say, and setting $\sigma^2_{0\text{–}5}$ to IGME estimates of uncertainty, or using them to directly inform a prior on it. σ^2_{5+} would be treated in the same way as σ^2_s.

The census counts we used were not raw census output but adjusted counts published in WPP. In some cases, these counts are adjusted by the UNPD to reduce bias due to factors such differential undercount by age. This may have led to an underestimate of uncertainty. The use of raw census outputs instead is worth investigating, but the effects of undercount would still need to be addressed through modifications to the method.

Our prior distributions for international migration were centered at zero with large variances. This is a sensible default when accurate data are not available. Further work could investigate the possibility of using stocks of foreign-born, often collected in censuses, to provide more accurate initial estimates. Data on refugee movements is another source that could be investigated where available.

We have reconstructed national level populations only, principally because this is the level at which the UN operate. We have already mentioned that sub-national reconstructions might be of interest. Subnational reconstructions could be done without any modifications to the method if the requisite data are available; national level initial estimates and population counts would just be replaced with their subnational equivalents and the method applied as above. Reconstructing adjacent regions, or regions between which there is likely to be a lot of migration would require special care, however, as there is no way of accounting for dependence among migratory flows under the current approach.

Appendices

A Derivation of Demographic Indicators

A.1 Total Fertility Rate

The TFR for the period $[t, t+5)$ is the average number of children born to women of a hypothetical cohort which survives through ages $a_L^{[\text{fert}]}$ to $a_U^{[\text{fert}]}$, all the way experiencing the age-specific fertility rates, $f_{a,t}$. It is computed as:

$$TFR_t \equiv 5 \cdot \sum_{a_L^{[\text{fert}]}}^{a_U^{[\text{fert}]}} f_{a,t}.$$

A.2 Life Expectancy at Birth

Life expectancy at birth is the average age at death for members of a hypothetical cohort which, at each age, experience age-specific mortality encoded in $s_{a,t,l}$. It is computed as:

$$e_{0,t,l} = 5 \sum_{a=0}^{A} \prod_{i=0}^{a} s_{i,t,l} + 5 \left(\prod_{i=0}^{A} s_{i,t,l} \right) \left(s_{A+5,t,l} / (1 - s_{A+5,t,l}) \right).$$

The derivation is straightforward and can be found in Wheldon et al. (2013b).

A.3 Under 5 Mortality Rate

The U5MR is constructed in the same way as the standard "Infant Mortality Rate" (e.g., Preston et al., 2001, Ch. 2), except for the age interval [0,5). It is defined as follows:

$$U5MR_t \equiv \frac{\text{No. deaths to those aged [0,5) in the interval } [t,t+5)}{\text{No. births in the interval } [t,t+5)}.$$

It is neither a true demographic rate nor a probability but, nevertheless, is in common use.

Acknowledgments

Mark C. Wheldon is Lecturer, Biostatistics Unit, Faculty of Health and Environmental Sciences, Auckland University of Technology, New Zealand, and Biostatistician, Centre for Clinical Research and Effective Practice, Middlemore Hospital, Auckland, New Zealand (e-mail: mwheldon@aut.ac.nz). Part of this work was done while he was a Ph.D. student in the Department of Statistics at the University of Washington, Seattle, WA, USA. Adrian E. Raftery is Professor of Statistics and Sociology, University of Washington, Seattle, WA, USA (e-mail: raftery@uw.edu).

Samuel J. Clark is Associate Professor of Sociology, University of Washington, and Researcher at the Institute of Behavioral Science (IBS), University of Colorado at Boulder, CO, USA and at MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa (e-mail: samclark@uw.edu). Patrick Gerland is Population Affairs Officer, United Nations Department of Economic and Social Affairs, Population Division, New York, NY, USA (e-mail: gerland@un.org). This work was supported by grants R01 HD054511, R01 HD070936, K01 HD057246 and 5R24HD042828 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). The first author received partial support from both a Shanahan Endowment Fellowship to the Center for Studies in Demography & Ecology at the University of Washington, and the Faculty of Health and Environmental Sciences, Auckland University of Technology.

References

Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., and Pelletier, F. (2012). Estimating Trends in the Total Fertility Rate with Uncertainty Using Imperfect Data: Examples from West Africa. *Demographic Research* 26.15, pp. 331–362. DOI: 10.4054/DemRes.2012.26.15. eprint: http://www.demographic-research.org/volumes/vol26/15/26-15.pdf. URL: http://www.demographic-research.org/volumes/vol26/15/.

Austad, S. N. (2011). "Sex Differences in Longevity and Aging." In: *Handbook of the Biology of Aging*. Ed. by E. J. Masoro and S. N. Austad. 7th ed. Boston: Academic Press. Chap. 23, pp. 479–495.

Barbi, E., Bertino, S., and Sonnino, E., eds. (2004). *Inverse Projection Techniques: Old and New Approaches*. Berlin: Springer-Verlag.

Beer, J. de, Raymer, J., Erf, R. van der, and Wissen, L. van (2010). Overcoming the Problems of Inconsistent International Migration Data: A New Method Applied to Flows in Europe. *European Journal of Population/Revue Européenne de Démographie* 26.4, pp. 459–481. DOI: dx.doi.org/10.1007/s10680-010-9220-z.

Bennett, N. G. and Horiuchi, S. (1981). Estimating the Completeness of Death Registration in a Closed Population. *Population Index* 47.2, pp. 207–221. URL: http://www.jstor.org/stable/2736447.

Bertino, S. and Sonnino, E. (2003). The Stochastic Inverse Projection and the Population of Velletri (1590–1870). *Mathematical Population Studies* 10.1, pp. 41–73.

Bhat, P. N. M. (2002a). Completeness of India's Sample Registration System: An Assessment Using the General Growth Balance Method. English. *Population Studies* 56.2, pp. 119–134. URL: http://www.jstor.org/stable/3092885.

— (2002b). On the Trail of 'Missing' Indian Females I: Search for Clues. English. *Economic and Political Weekly* 37.51, pp. 5105–5118. URL: http://www.jstor.org/stable/4412986.

— (2002c). On the Trail of 'Missing' Indian Females II: Illusion and Reality. English. *Economic and Political Weekly* 37.52, pp. 5244–5263. URL: http://www.jstor.org/stable/4413022.

Bongaarts, J. (2001). Fertility and Reproductive Preferences in Post-Transitional Societies. English. *Population and Development Review* 27, pp. 260–281. URL: http://www.jstor.org/stable/3115260.

Boyle, P. P. and Ó Gráda, C. (1986). Fertility Trends, Excess Mortality, and the Great Irish Famine. *Demography* 23.4, pp. 543–562. URL: http://www.jstor.org/stable/2061350.

Brass, W. (1964). "Uses of Census and Survey Data for the Estimation of Vital Rates." In: *African Seminar on Vital Statistics*. (Addis Ababa, Dec. 14–19, 1964).

Coale, A. J. (1991). Excess Female Mortality and the Balance of the Sexes in the Population: An Estimate of the Number of Missing Females. *Population and Development Review* 17, pp. 517–523.

Coale, A. J., Demeny, P., and Vaughan, B. (1983). *Regional Model Life Tables and Stable Populations*. 2nd ed. New York, New York: Academic Press.

Cohen, J. E. (2006). Stochastic Demography. In: *Encyclopedia of Statistical Sciences*. Article Online Posting Date: August 15, 2006. John Wiley and Sons, Inc. DOI: 10.1002/0471667196.ess2584.pub2.

Daponte, B. O., Kadane, J. B., and Wolfson, L. J. (1997). Bayesian Demography: Projecting the Iraqi Kurdish Population, 1977–1990. English. *Journal of the American Statistical Association* 92.440, pp. 1256–1267. URL: http://www.jstor.org/stable/2965396.

Das Gupta, M. (2005). Explaining Asia's "Missing Women": A New Look at the Data. *Population and Development Review* 31.3, pp. 529–535. DOI: 10.1111/j.1728-4457.2005.00082.x. URL: http://dx.doi.org/10.1111/j.1728-4457.2005.00082.x.

Das Gupta, M. and Bhat, P. N. M. (1997). Fertility Decline and Increased Manifestation of Sex Bias in India. *Population Studies* 51.3, pp. 307–315.

Feeney, G. (1996). *A New Interpretation of Brass' P/F Ratio Method Applicable When Fertility is Declining*. URL: http://www.heart-intl.net/HEART/HIV/Comp/InterpretationofFertiline.pdf (visited on 12/12/2013).

Frisen, C. M. (1991). Population Characteristicsin the Lao People's Democratic Republic. *Asia-Pacific Population Journal* 6.2, pp. 55–66. URL: http://scholar.googleusercontent.com/scholar?q=cache:kW4rlEwV6OUJ:scholar.google.com/+Laos+fertility+sex+ratio+at+birth&hl=en&as_sdt=0,48.

Goodkind, D. and West, L. (2001). The North Korean Famine and its Demographic Impact. *Population and Development Review* 27.2, pp. 219–238.

Griffiths, P., Matthews, Z., and Hinde, A. (4 2000). Understanding the Sex Ratio in India: A Simulation Approach. English. *Demography* 37.4, pp. 477–488. DOI: 10.1353/dem.2000.0004. URL: http://dx.doi.org/10.1353/dem.2000.0004.

Guillot, M. (2002). The Dynamics of the Population Sex Ratio in India, 1971–96. English. *Population Studies* 56.1, pp. 51–63. URL: http://www.jstor.org/stable/3092941.

Guilmoto, C. Z. (2007a). "Characteristics of Sex-Ratio Imbalance in India and Future Scenarios." In: 4th Asia Pacific Conference on Reproductive and Sexual Health and Rights. (Hyderabad, India, Oct. 29–31, 2007). URL: http://www.unfpa.org/gender/docs/studies/india.pdf.

— (2007). "Characteristics of Sex-Ratio Imbalance in India and Future Scenarios: Executive Summary." In: 4th Asia Pacific Conference on Reproductive and Sexual Health and Rights. (Hyderabad, India, Oct. 29–31, 2007). URL: http://www.unfpa.org/gender/docs/studies/summaries/india_summary.pdf.

Guilmoto, C. Z. (2007b). "Sex Ratio Imbalance in Asia: Trends, Consequences and Policy Responses." In: 4th Asia Pacific Conference on Reproductive and Sexual Health and Rights. (Hyderabad, India, Oct. 29–31, 2007). URL: http://www.unfpa.org/gender/docs/studies/summaries/regional_analysis.pdf.

— (2009). The Sex Ratio Transition in Asia. *Population and Development Review* 35.3, pp. 519–549. DOI: 10.1111/j.1728-4457.2009.00295.x. URL: http://dx.doi.org/10.1111/j.1728-4457.2009.00295.x.

— (2012). Skewed Sex Ratios at Birth and Future Marriage Squeeze in China and India 2005–2100. *Demography* 49, pp. 77–100.

Heuveline, P. (1998). 'Between One and Three Million': Towards the Demographic Reconstruction of a Decade of Cambodian History (1970–79). *Population Studies* 52, pp. 49–65.

Hill, K. (1987). Estimating Census and Death Registration Completeness. *Asian and Pacific Population Forum* 1, pp. 8–13.

Hill, K., Vapattanawong, P., Prasartkul, P., Porapakkham, Y., Lim, S. S., and Lopez, A. D. (2007). Epidemiologic Transition Interrupted: A Reassessment of Mortality Trends in Thailand, 1980–2000. *International Journal of Epidemiology* 36.2, pp. 374–384. DOI: 10.1093/ije/dyl257.

Hill, K., You, D., Inoue, M., Oestergaard, M. Z., and Technical Advisory Group of the United Nations Inter-agency Group for Child Mortality Estimation (IGME) (2012). Child Mortality Estimation: Accelerated Progress in Reducing Global Child Mortality, 1990–2010. *PLoS Medicine* 9.8, e1001303. DOI: 10.1371/journal.pmed.1001303. URL: http://dx.doi.org/10.1371/journal.pmed.1001303.

Hinkley, D. V. (1969). Inference About the Intersection in Two-Phase Regression. *Biometrika* 56.3, pp. 495–504. URL: http://www.jstor.org/stable/2334655?origin=JSTOR-on-page.

Hinkley, D. V. (1971). Inference in Two-Phase Regression. *Journal of the American Statistical Association* 66.336, pp. 736–743. URL: http://www.jstor.org/stable/2284220?origin=JSTOR-on-page.

Jha, P., Kesler, M. A., Kumar, R., Ram, F., Ram, U., Aleksandrowicz, L., Bassani, D. G., Chandra, S., and Banthia, J. K. (2011). Trends in Selective Abortions of Girls in India: Analysis of Nationally Representative Birth Histories from 1990 to 2005 and Census Data from 1991 to 2011. *The Lancet* 377, pp. 1921–1928.

Jha, P., Kumar, R., and Dhingra, N. (2006). Sex Ratio in India — Authors' Reply. *The Lancet* 367.9524, p. 1727. DOI: 10.1016/S0140-6736(06)68761-8. URL: http://www.sciencedirect.com/science/article/pii/S0140673606687618.

Jha, P., Kumar, R., Vasa, P., Dhingra, N., Thiruchelvam, D., and Moineddin, R. (2006). Low Male-to-female Sex Ratio of Children Born in India: National Survey of 1.1 Million Households. *The Lancet* 367.9506, pp. 211–218. DOI: 10.1016/S0140-6736(06)67930-0. URL: http://www.sciencedirect.com/science/article/pii/S0140673606679300.

Kamnuansilpa, P., Chamratrithirong, A., and Knodel, J. (1982). Thailand's Reproductive Revolution: An Update. English. *International Family Planning Perspectives* 8.2, pp. 51–56. URL: http://www.jstor.org/stable/2948088.

Keyfitz, N. and Caswell, H. (2005). *Applied Mathematical Demography*. third. Statistics for Biology and Health. Springer.

Knodel, J. (1987). Thailand's Reproductive Revolution. *Social Science* 72.1, pp. 52–56.

Knodel, J. and Prachuabmoh, V. (1976). Preferences for Sex of Children in Thailand: A Comparison of Husbands' and Wives' Attitudes. English. *Studies in Family Planning* 7.5, pp. 137–143. URL: http://www.jstor.org/stable/1964858.

Knodel, J., Ruffolo, V. P., Ratanalangkarn, P., and Wongboonsin, K. (1996). Reproductive Preferences and Fertility Trends in Post-transition Thailand. English. *Studies in Family Planning* 27.6, pp. 307–318. URL: http://www.jstor.org/stable/2138026.

Lalic, N. and Raftery, A. E. (2012). "Joint Probabilistic Projection of Female and Male Life Expectancy." In: *Annual Meeting of the Population Association of America*. New Orleans, LA. URL: http://paa2012.princeton.edu/abstracts/120140.

Lee, R. D. (2003). Reflections on Inverse Projection: Its Origins, Development, Extensions, and Relation to Forecasting. In: vol. 10. 1, pp. 1–9.

Lee, R. D. (1971). *Econometric Studies of Topics in Demographic History*. Disseration in Economics. Cambridge, Massachucetts: Harvard University.

— (1974). Estimating Series of Vital Rates and Age Structures from Baptisms and Burials: A New Technique, with Applications to Pre-Industrial England. *Population Studies* 28.3, pp. 495–512. URL: http://www.jstor.org/stable/2173642.

Leslie, P. H. (1945). On the Use of Matrices in Certain Population Mathematics. *Biometrika* 33.3, pp. 183–212. URL: http://www.jstor.org/stable/2332297.

— (1948). Some Further Notes on the Use of Matrices in Population Mathematics. *Biometrika* 35.3/4, pp. 213–245. URL: http://www.jstor.org/stable/2332342.

Luther, N. Y., Dhanasakdi, N., and Arnold, F. (1986). *Consistent Correction of Census and Vital Registration Data for Thailand, 1960–80*. Papers of the East-West Population Institute. Honolulu, Hawaii.

Luther, N. Y. and Retherford, R. D. (1988). Consistent Correction of Census and Vital Registration Data. *Mathematical Population Studies* 1, pp. 1–20.

Mahapatra, P. (2010). "An Overview of the Sample Registration System in India." In: *Prince Mahidol Award Conference & Global Health Information Forum*. URL: http://unstats.un.org/unsd/vitalstatkb/KnowledgebaseArticle50447.aspx.

Mayer, P. (1999). India's Falling Sex Ratios. English. *Population and Development Review* 25.2, pp. 323–343. URL: http://www.jstor.org/stable/172428.

Merli, M. G. (1998). Mortality in Vietnam, 1979-1989. *Demography* 35.3, pp. 345–360. URL: http://www.jstor.org/stable/3004042.

Murray, C. J. L., Ferguson, B. D., Lopez, A. D., Guillot, M., Salomon, J. A., and Ahmad, O. (2003). Modified Logit Life Table System: Principles, Empirical Validation, and Application. *Population Studies* 57.2, pp. 165–182. URL: http://www.jstor.org/stable/3595746.

Murray, C. J. L., Rajaratnam, J. K., Marcus, J., Laakso, T., and Lopez, A. D. (2010). What Can We Conclude from Death Registration? Improved Methods for Evaluating Completeness. *PLoS Med* 7.4, e1000262. DOI: 10.1371/journal.pmed.1000262. URL: http://dx.doi.org/10.1371/journal.pmed.1000262.

Oeppen, J. (1993a). Back Projection and Inverse Projection: Members of a Wider Class of Constrained Projection Models. English. *Population Studies* 47.2, pp. 245–267. URL: http://www.jstor.org/stable/2174729.

— (1993b). Generalized Inverse Projection. In: *Old and New Methods in Historical Demography*. Ed. by D. S. Reher and R. Schofield. Oxford, England: Clarendon Press. Chap. 2, pp. 29–39.

Pollard, J. H. (1968). A Note on Multi-Type Galton-Watson Processes with Random Branching Probabilities. English. *Biometrika* 55.3, pp. 589–590. URL: http://www.jstor.org/stable/2334271.

United Nations (1982). *Model Life Tables for Developing Countries*. Population Studies (Department of International, Economic and Social Affairs) 77. New York, New York: United Nations, Sales No. E.07.XIII.7.

Preston, S. H. (1976). *Mortality Patterns in National Populations*. New York, New York: Academic Press.

Preston, S. H., Heuveline, P., and Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Malden, Massachusetts: Blackwell.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Raftery, A. E. and Lewis, S. M. (1996). Implementing MCMC. In: *Markov Chain Monte Carlo in Practice*. Ed. by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. London: Chapman & Hall, pp. 115–130.

Sawyer, C. C. (2012). Child Mortality Estimation: Estimating Sex Differences in Childhood Mortality since the 1970s. *PLoS Med* 9.8, e1001287. DOI: 10.1371/journal.pmed.1001287. URL: http://dx.doi.org/10.1371/journal.pmed.1001287.

Sen, A. (1990). More Than 100 Million Women are Missing. *New York Review of Books* 20, pp. 61–66.

Silva, R. (2012). Child Mortality Estimation: Consistency of Under-Five Mortality Rate Estimates Using Full Birth Histories and Summary Birth Histories. *PLoS Medicine* 9.8, e1001296. DOI: 10.1371/journal.pmed.1001296. URL: http://dx.doi.org/10.1371/journal.pmed.1001296.

Soliani, L. and Luchetti, E. (2006). "Genetic Factors in Mortality." In: *Demography: Analysis and Synthesis*. A Treatise in Population. Ed. by G. Caselli, J. Vallin, and G. J. Wunsch. Vol. 2. Burlington, MA: Academic Press. Chap. 49, pp. 117–128.

The Registrar General and Census Commissioner of India (2009). *National Family Health Survey, India*. Misitry of Home Affairs. URL: http://censusindia.gov.in/Vital_Statistics/SRS/Sample_Registration_System.aspx (visited on 12/12/2013).

— (2011). *Vital Statistics: Sample Registration*. Misitry of Home Affairs, The Government of India. URL: http://censusindia.gov.in/Vital_Statistics/SRS/Sample_Registration_System.aspx (visited on 12/12/2013).

United Nations (1983). *Manual X: Indirect Techniques for Demographic Estimation*. Population Studies (Department of International, Economic and Social Affairs) 81. New York, NY: United Nations, Sales No. E.83.XIII.2.

— (2011). *World Population Prospects: The 2010 Revision, Standard Variants (Updated: 28 June 2011)*. Department of Economic and Social Affairs (DESA), Population Division, Population Estimates and Projections Section. URL: http://esa.un.org/unpd/wpp/index.htm.

United Nations Population Fund (2010). *UNFPA Guidance Note on Sex Selection*. URL: http://www.unfpa.org/webdav/site/global/shared/documents/publications/2010/guidenote_prenatal_sexselection.pdf (visited on 12/12/2013).

Vallin, J. (2006). "Mortality, Sex, and Gender." In: *Demography: Analysis and Synthesis*. A Treatise in Population. Ed. by G. Caselli, J. Vallin, and G. J. Wunsch. Vol. 2. Burlington, MA: Academic Press. Chap. 53, pp. 177–194.

Vapattanawong, P. and Prasartkul, P. (2011). Under-Registration of Deaths in Thailand in 2005–2006: Results of Cross-Matching Data from Two Sources. *Bulletin of the World Health Organization* 89.11, pp. 806–812. DOI: 10.2471/BLT.10.083931. URL: http://www.who.int/bulletin/volumes/89/11/10-083931/en/index.html.

Visaria, P. M. (1971). *The Sex Ratio of the Population of India*. Vol. 1. Census of India, Monographs 10. New Delhi, India: Ministry of Home Affairs, Office of the Registrar General.

Waldron, I. (1985). What Do We Know About Causes of Sex Differences in Mortality? A Review of the Literature. *Population Bulletin of the United Nations* 18, pp. 59–76.

— (2009). Gender Differences in Mortality—Causes and Variation in Different Societies. In: *The Sociology of Health and Illness: Critical Perspectives*. Ed. by P. Conrad. 8th ed. New York, New York: Worth. Chap. 4, pp. 38–54.

Walters, S. L. (2008). *Fertility, Mortality and Marriage in Northwest Tanzania, 1920–1970: a Demographic Study Using Parish Registers*. Doctoral dissertation. Cambridge, England: King's College, University of Cambridge.

Wheldon, M. C., Raftery, A. E., Clark, S. J., and Gerland, P. (2013a). "Bayesian Reconstruction of Past Populations and Vital Rates by Age for Developing and Developed Countries." In: Annual Meeting of the Population Association of America. (New Orleans, Louisiana, Apr. 11–13, 2013). Population Association of America. URL: http://paa2013.princeton.edu/abstracts/130303.

— (2013b). Reconstructing Past Populations with Uncertainty from Fragmentary Data. *Journal of the American Statistical Association* 108.501, pp. 96–110. DOI: 10.1080/01621459.2012.737729. URL: http://www.tandfonline.com/doi/pdf/10.1080/01621459.2012.737729.

Whelpton, P. K. (1936). An Empirical Method of Calculating Future Population. *Journal of the American Statistical Association* 31.195, pp. 457–473. URL: http://www.jstor.org/stable/2278370.

Wrigley, E. A. and Schofield, R. S. (1981). *The Population History of England, 1541–1871: A Reconstruction*. Studies in Social and Demographic History. London, England: Edward Arnold.