Gender differences in infant and child mortality: Estimation and identification of countries with outlying levels or trends*

Leontine Alkema
National University of Singapore

Fengqing Chao
National University of Singapore

Cheryl C. Sawyer
United Nations

Extended abstract
for
XXVII IUSSP International Population Conference 2013,
August 16, 2013

*Leontine Alkema and Fengqing Chao, Department of Statistics and Applied Probability and Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117546; Email: alkema@nus.edu.sg and fengqing_chao@nus.edu.sg. Cheryl C. Sawyer, Mortality Section, United Nations Population Division, New York, NY 10017; Email: sawyerc@un.org. This work was supported by a research grant from the National University of Singapore and UNICEF. The views expressed in this article are those of the authors and do not necessarily reflect the views of the United Nations. Its contents have not been formally edited and cleared by the United Nations. The authors are grateful to the members of (the technical advisory group of) the Inter-Agency Group for Child Mortality Estimation for helpful comments and discussions. Additional thanks to Philip Bastian, Jin Rou New, Jing Liu, Jon Pederson, Stella Ting Hai Yin, Jingxian Wu and Danzhen You for additional assistance on research and database construction.

Abstract

Under natural circumstances, the ratio of the male to female under-five mortality rate is greater than one. However, deprivation of girls' access to health care or proper nutrition could lead to distorted ratios of under-five mortality. Monitoring of mortality by sex is challenging because of issues with data availability and quality. Moreover, the sex ratio is expected to vary with under-five mortality, which makes it challenging to define "expected levels". We present a Bayesian model to estimate the sex ratio of under-five mortality for all countries. In addition, we estimate the relative difference between national sex ratios and expected sex ratios based on the global relation between mortality and sex ratios. All estimates include an uncertainty assessment to enable assessments of whether differences between countries or within countries over time are significant or highly uncertain.

1 Introduction

After birth, girls have biological advantages over boys with respect to survival for many causes of death. Under natural circumstances, the ratio of the male to female under-five mortality rate (U5MR, the probability of dying before age 5), is greater than one. However, deprivation of girls access to health care or proper nutrition could lead to distorted ratios of under-five mortality (E. J. Croll 2000). Current international monitoring initiatives call for disaggregation of under-five mortality rates by sex.

Recent assessments that synthesized data from various sources have been in agreement that a number of countries in southern Asia experience higher mortality for girls compared to boys from one to four years old (Sawyer 2012; Wang et al. 2012). However, findings on whether higher female than male mortality at ages 1-4 exists in scattered countries in other parts of the world, including northern Africa/western Asia, parts of sub-Saharan Africa, and Latin America, are inconsistent.

Unfortunately, monitoring of sex differences in under-five mortality is challenging because of issues with data availability and quality. This is illustrated in Figure 1, which shows observations on the ratio of male to female U5MR in Jordan (denoted by SR5). The data, which come from a variety of surveys, censuses and vital registration, do not provide a coherent story on levels and trends in SR5. Smoothing or modeling techniques must be applied to arrive at estimates that can be used for international comparisons. Estimation methods should convey the uncertainty in the underlying data.

The monitoring of trends in sex differentials is further complicated by lack of a clear definition of how large we should expect girls' advantage to be. Boys and girls have different probabilities of death due to biological factors, and these differences vary between infancy and early childhood. If sex-disaggregated estimates are to be used for monitoring or advocacy purposes, it must be clearly explained to users (1) what the expected differences are, (2) when a given difference might indicate excessive disadvantage for one sex or the other, and (3) how to understand changes. The SR5 tends to vary with U5MR as shown in Figure 2. As under-five mortality decreases, the share of causes of death that more typically affect boys tends to increase and girls' advantage tend to increase as well. Given that girls are expected to have lower mortality under natural circumstances, it is possible for girls to have lower absolute mortality than boys but still be prevented from enjoying their full biological advantage

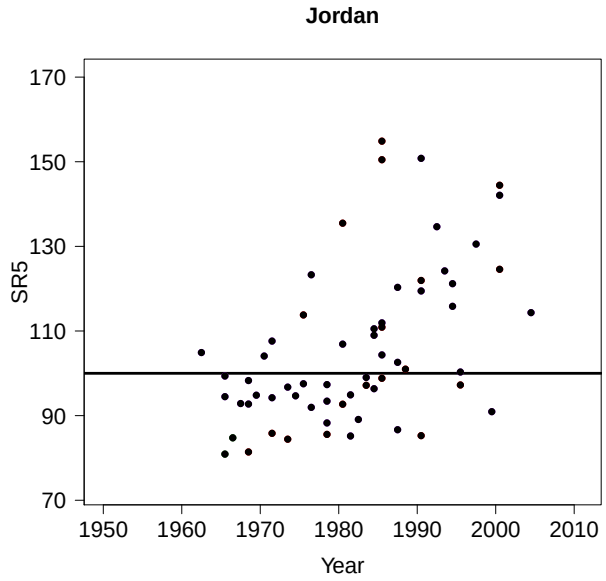

Figure 1: Illustration of SR5 observations for Jordan.

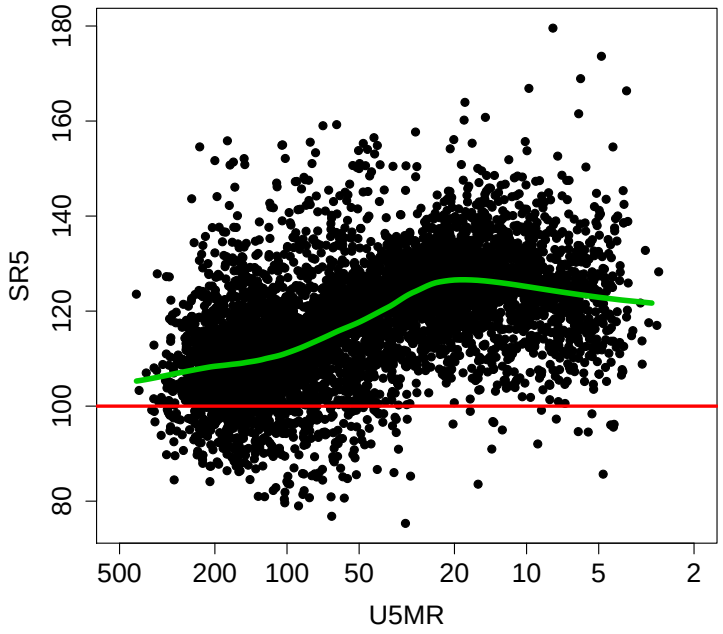

Figure 2: Illustration of the empirical relation between SR5 and U5MR. SR5 observations from all countries are plotted in black versus U5MR (on log-scale), a loess smoother is added in green.

due to some level of disadvantaging treatment. Some studies have used standards based on data from selected Western countries to assess female disadvantage (Hill and Upchurch 1995; Monden and Smits 2013) or calculate excess female deaths (World Bank 2011). Other studies (Sawyer 2012; United Nations 2011) have noted this difficulty but avoided the selection of a comparative reference level, highlighting only absolute excess of female mortality. Still others (Wang et al. 2012) have published estimates of sex-specific child mortality but refrained from analysis of sex differences.

We developed a new modeling approach to estimate SR5 levels and trends for all countries. The approach takes into account data quality issues, includes an uncertainty assessment and that improves upon existing approaches by providing information on whether the estimated SR5 is higher or lower than expected, based on an estimated global relations between mortality levels and sex ratios. In this extended abstract we describe the data and model used to construct the sex-specific estimates and identify countries with outlying sex ratios, and present illustrative results.

2 Data

The data used in this study are observed sex ratios since 1950 for infant, child and under-five mortality from vital registration systems (VR), sample registration/surveillance systems (SRS), surveys and censuses.

Data from VR/SRS were used to obtain sex ratios for age groups 0 and 1–4 through standard life table methods. Sex ratios for the age group 0–4 were not used to avoid using the same data twice.

For estimates from full birth histories collected in Demographic and Health Surveys, World Fertility Surveys and selected surveys from the Pan-Arab Programme on Family Health (PAPFAM), sex ratios for infant mortality from full birth histories were calculated for periods of varying lengths, according to the method in Pedersen and Liu (2012). For sex ratios for ages 1-4, 5-year estimates were used. As for VR/SRS data, sex ratios for the age group 0–4 are not used to avoid using the same data twice.

For surveys and censuses in which only summary birth histories were collected, sex-specific indirect estimates of U5MR were calculated using the Brass method either from microdata or from published tabulations of children ever born and children living.

In cases where microdata or tabulations were not available, direct or indirect estimates were taken from published survey or census reports.

3 Estimating sex ratios

3.1 Overview and notation

Given differences in sex ratios for infants and children, we estimated the relation between mortality levels and sex ratios separately for both age groups (as opposed to directly estimating the relation between the U5MR and SR5).

For infants, we estimated a global relation between IMR and the sex ratio for infants (denoted by SR1) and used that relation to construct the expected SR1 for each country-

year. These expected sex-ratios were multiplied by a second component, the country-specific multiplier for the infant sex ratio, that represents the relative advantage or disadvantage of girls to boys compared to other countries at similar mortality rates. This multiplier was modeled for each country by a time series model and estimated using the available data in the country. The country-specific levels that these multipliers fluctuate around were estimated using a Bayesian hierarchical model.

The global relation between child mortality (CM) and its sex ratio (SR4) is estimated in a similar fashion. Finally, for each country-year, the expected level of SR5 is derived from the expected level of SR1 and SR4 (based on the estimates for IMR and CM).

In the model description, the sex ratios for infant, child and under-five mortality are denoted by $S_{a,c,t}$ for country c, year t for $a = 1, 4, 5$ respectively; index a refers to age groups $[0, 1)$, $[1, 5)$ and $[0, 5)$ respectively. The infant, child and under-five mortality rates are denoted by $Q_{a,c,t}$ for the corresponding age group and country-year. The j-th observed ratio of male to female mortality is denoted by $s_{a,j}$ in country $c[a, j]$, year $t[a, j]$ for $a = 1, 4, 5$.

3.2 Model details for estimating SR1 and SR4

$S_{a,c,t}$ for $a = 1, 4$ is modeled as follows:

$$\begin{aligned} S_{a,c,t} &= W_{a,c,t} \cdot P_{a,c,t}, \\ W_{a,c,t} &= f^{(a)}(Q_{a,c,t}), \end{aligned}$$

where $f^{(a)}(\cdot)$ represents the relation between the level of mortality for both sexes combined and the expected sex ratio on a global level, and $P_{a,c,t}$ represents the relative advantage or disadvantage of girls to boys compared to other countries at similar mortality rates, as indicated by the data in the country.

The country multipliers $P_{a,c,t}$ are estimated with a time series model:

$$\begin{aligned} \log(P_{a,c,t}) &= \beta_{a,c} + \varepsilon_{a,c,t}, \\ \varepsilon_{a,c,t} &\sim N(\rho \cdot \varepsilon_{a,c,t-1}, \sigma_\varepsilon^2), \end{aligned}$$

where the multiplier fluctuates around country-specific level $\beta_{a,c}$. The fluctuations $\varepsilon_{a,c,t}$ are modeled with an autoregressive time series model of order 1.

Country-specific levels $\beta_{a,c}$ for $a = 1, 4$, representing the average level difference in $P_{a,c,t}$ across countries, are estimated using a hierarchical model (Lindley and Smith 1972; Gelman, Carlin, Stern, and Rubin 2004):

$$\beta_{a,c} \sim t_3(\mu = 0, \sigma^2 = \sigma_{a,\beta}^2, \nu = 3)T(, \log(1.6)),$$

where a t-distribution with 3 degrees of freedom was chosen to allow for countries with outlying levels, and the truncation was imposed to exclude the possibility of extreme (and unrealistic) median country-specific levels (here the median levels are restricted to be smaller than 1.6).

Specification of global relation between mortality and sex ratios We use a flexible penalized B-splines regression model (Eilers and Marx 2010, 2011) to estimate the global relation between mortality and sex ratios, which is represented by function $f^{(a)}(\cdot)$, for age groups $a = 1, 4$. The function $f^{(a)}(q)$ for some value q for both-sexes mortality is specified as follows:

$$\log(f^{(a)}(q)) \quad = \quad \sum_{k=1}^{K_a} B_k^{(a)}(q)\alpha_k^{(a)}, \tag{1}$$

where $B_k^{(a)}(q)$ refers to the k-th B-spline evaluated at q and $\alpha_k^{(a)}$ to the k-th spline coefficients. B-splines are symmetric third-order polynomials that add up to one at any level of mortality. In this application, the splines are equally spaced on log-transformed Q and knots are set to be 0.3 apart. To avoid extreme extrapolations, splines are added up for $q < 0.005$ for $a = 1, 4$ and for q greater than the 95-th percentile of $Q_{a,c,t}$'s for country-years included in the data set. When fitting the splines model to observations, second-order differences in adjacent splines coefficients are penalized to guarantee smoothness of the resulting global norm. Spread out prior distributions are used for the splines model parameters. The expected sex ratio for country c, year t with mortality $Q_{a,c,t}$ is given by $W_{a,c,t} = f^{(a)}(Q_{a,c,t})$ for $a = 1, 4$ (where $Q_{a,c,t}$'s are rounded to 3 decimals to reduce the number of splines evaluations).

3.3 Derivation of SR5

SR5 for country c in year t, $S_{5,c,t}$, is derived from $S_{1,c,t}$ and $S_{4,c,t}$ through standard cohort equations (by calculating the sex-specific mortality rates for infants and children). The expected sex ratio $W_{5,c,t}$ for the country-year of interest is derived in a similar fashion, using $W_{1,c,t}$ and $W_{4,c,t}$ instead of $S_{1,c,t}$ and $S_{4,c,t}$. Finally, country multiplier $P_{5,c,t} = S_{5,c,t}/W_{5,c,t}$.

Contrary to age groups 0 and 1–4, there is no function that describes the global relation between mortality for both sexes combined and the expected sex ratio for ages 0–4 (because the expected SR5 depends on the expected SR1 and SR4, which depend on the age-group specific mortality levels). To visualize the global relation between U5MR and the expected sex ratio, a Loess curve is fitted to all estimates of combinations $(Q_{5,c,t}, W_{5,c,t})$. The resulting relation will be denoted by $\tilde{f}^{(5)}(\cdot)$.

3.4 Data model

For most observations, observed sex ratios for age groups 1 and 4 are used. To avoid using data twice, observed under-5 sex ratios are not included if information on age groups 1 and 4 is included. There are two exceptions: (1) for observations from summary birth histories only sex ratios for age group 5 are used (unless these are missing while the sex ratios for the infant and children are available) and (2) for a small number of observations, information on sex ratios for infant mortality is missing. For these observations, information on the under-5 sex ratio is used instead.

For observations on age groups 1 and 4, the data model is given by

$$\log(s_{a,j}) \quad \sim \quad N(\log(S_{a,c[a,j],t[a,j]}), \sigma_{a,j}^2 + \omega_{a,x[a,j]}^2), \text{ for } a = 1, 4,$$

where $s_{a,j}$ is the j-th observed ratio of male to female mortality for age group a in country $c[a,j]$, year $t[a,j]$ for $a = 1, 4$, and $x[a,j]$ is the source type of that observation. The variance σ^2 is the sum of sampling variance $\sigma^2_{a,j}$ and non-sampling variance $\omega^2_{a,x[a,j]}$ (explained further below).

For observations from summary birth histories, and for a small number of observations where information on sex ratios for infant mortality is missing, the data model for observations from age group 5 is given by:

$$\log(s_{5,j}) \quad \sim \quad t(\mu = \log(S_{5,c[5,j],t[5,j]}), \sigma^2 = \sigma^2_{5,j} + \omega^2_{5,x[5,j]}, \nu = \nu_5),$$

which is based on a t-distribution because additional analysis suggested that more outliers are present. The degrees of freedom ν_5 is assigned a spread out prior distribution.

Sampling variance is given for a large subset of DHS and MICS observations. For observations from vital registration systems, a Monte Carlo simulation was used to approximate the stochastic variances. For all other observations with missing standard errors, the standard error on the log-scale was set at 15%, approximately equal to the median standard error in the data set of non-VR observations. Non-sampling variance parameter $\omega^2_{a,x[a,j]}$ is estimated by source type and set to 0 for observations from VR/SRS.

3.5 Computing

A Markov Chain Monte Carlo (MCMC) algorithm was used to obtain samples from the posterior distribution of the parameters, implemented in the software JAGS (Plummer 2003). We used 24 parallel chains with a total of 150,000 iterations in each chain. Of these, the first 5,000 iterations in each chain were discarded as burn-in and after additional thinning, the resulting chains contained 8,640 samples each. Standard diagnostics checks (using trace plots and the Gelman and Rubin diagnostic (Gelman and Rubin 1992)) were used to check convergence.

Estimates of relevant quantities are given by the posterior medians while 90% credible intervals were constructed from the 5% and 95% percentiles of the posterior sample.

4 Illustrative results

4.1 Global relation between sex ratios and mortality levels

Figure 3 shows the estimated expected sex ratios (the $W_{a,c,t}$'s) based on the global relation between mortality levels and sex ratios for ages 0, 1–4 and 0–4.

For age group 0, the estimated expected sex ratio increases from 1.15 to 1.26 as mortality decreases from around 150 deaths per 1,000 births to around 20 deaths per 1,000 births. This increase is followed by a decrease in the estimated sex ratio from 1.26 to 1.20 as total IMR decreases from 20 to 5 deaths per 1,000 births.

For ages 1–4, estimated sex ratios are close to 1 for total mortality above 30 deaths per 1000 survivors up to age 1. The ratio increases as mortality decreases, the maximum of 1.2 is reached for mortality of 5 per 1,000.

The estimated sex ratio for ages 0–4 is driven by the estimates for the sex ratios for age 0 and ages 1–4 and increases from 1.06 to 1.26 as U5MR decreases from 390 to 19 deaths per 1,000 live births. This increase is followed by a decrease in the estimated sex ratio from 1.26 to 1.20 as U5MR decreases from 19 deaths per 1,000 live births to zero.

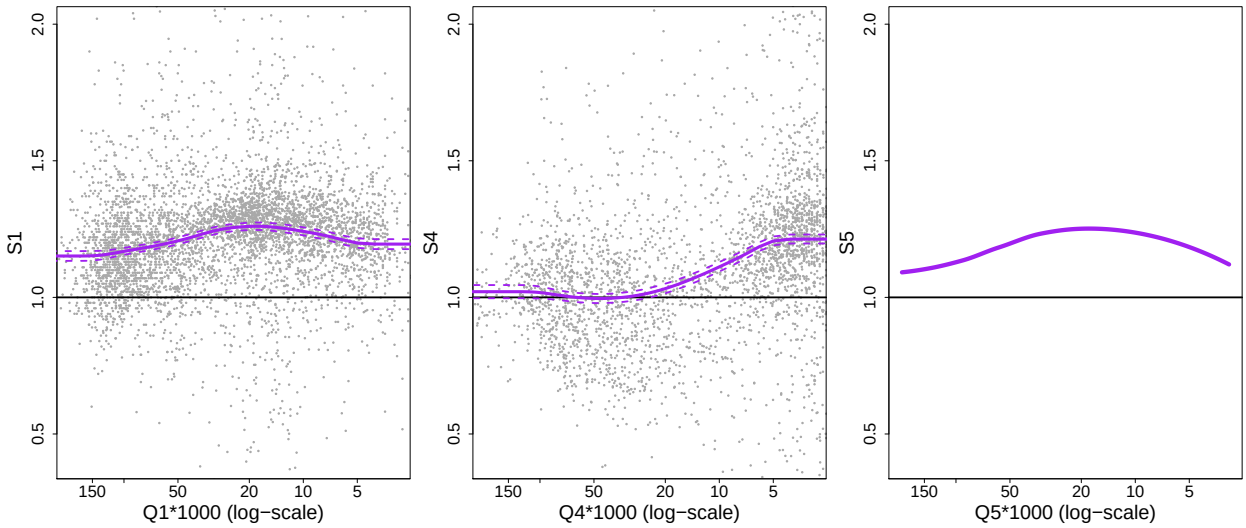

Figure 3: Estimated expected sex ratios ($W_{a,c,t}$) based on the global relation between sex ratios and mortality levels (purple, mortality level is plotted on the log-scale). Observed sex ratios (gray dots) are plotted against estimated mortality rates (for ages 0–4, no observation is displayed since $W_{5,c,t}$ is constructed based on IMR and CMR only).

4.2 Countries with outlying sex ratios

Figure 4 shows country-specific estimates for the year 2012 of the country multiplier $P_{a,c,t}$, which represents the relative advantage or disadvantage of girls to boys in the country of interest as compared to other countries at similar levels of mortality. For the countries displayed in the figure, the U5MR in 2012 was greater than 10 deaths per 1,000 births and the multiplier $P_{a,c,t}$ is significantly different from 1 (the 90% credible interval does not include 1), which means that the ratio of female to male mortality is higher or lower than expected based on the global relation between mortality levels and sex ratios. For countries displayed in the left column, the multiplier is smaller than 1, indicating higher female mortality than expected while for the countries displayed in the right column, the multiplier is greater than 1, indicating lower male mortality than expected.

Explanation for these outlying results could include (i) biological factors (e.g. an unusual cause of death distribution in the country as compared to other countries with similar levels of mortality), (ii) data quality issues (in particular, sex differentials in the reporting of birth and/or deaths) or (iii) gender discrimination. Additional information on biological factors, data quality and/or the occurrence of gender discrimination is necessary to further investigate which explanation(s) applies to which countries.

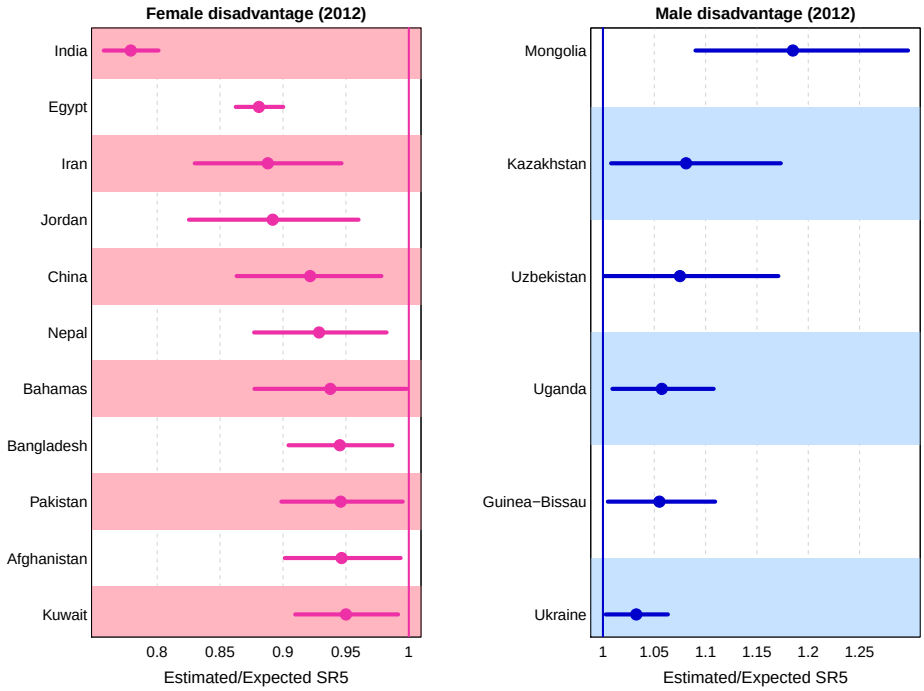

Figure 4: Estimates of the relative advantage or disadvantage of girls to boys compared to other countries at similar levels of mortality for age group 0–4 ($P_{5,c,t}$). Right: Countries where $P_{5,c,t}$ is significantly less than one in 2012, based on the global relation between SR5 and the U5MR, suggesting a female disadvantage. Left: Countries where $P_{5,c,t}$ is significantly greater than one in 2012, suggesting a male disadvantage.

Figure 5 shows estimates of sex ratios and country multipliers for Bangladesh to illustrate that a relative female disadvantage does not necessarily result in higher female mortality, given the general biological advantage of girls over boys w.r.t. survival. As illustrated in the figure, estimated sex ratios for the under-5 age group are greater than 1, indicating that male mortality is higher than female mortality, while country multipliers are below 1 and indicate a female disadvantage. In this situation, the country multipliers help to pinpoint to country-periods with unusual sex ratios, that would go unnoticed when focusing solely on sex ratios that are smaller than 1.

References

E. J. Croll (2000). *Endangered Daughters: Discrimination and Development in Asia*. London: Routledge.

Eilers, P. H. C. and B. D. Marx (2010, November). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics 2*(6), 637–653.

Eilers, P. H. C. and B. D. Marx (2011). Flexible Smoothing with B-splines and Penalties. *11*(2), 89–102.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004). *Bayesian Data Analysis* (2nd ed.). Boca Raton, Fl.: Chapman & Hall/CRC.

Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science 7*, 457–511.

Hill, K. and D. Upchurch (1995). Gender differences in child health: evidence from the Demographic and Health Surveys. *Population and Development Review 21*(1), 127–151.

Lindley, D. V. and A. F. M. Smith (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B 34*, 1–41.

Monden, C. and J. Smits (2013). Maternal education is associated with reduced female disadvantages in under-five mortality in sub-Saharan Africa and southern Asia. *International Journal of Epidemiology 42*, 211–218.

Pedersen, J. and J. Liu (2012, August). Child Mortality Estimation: Appropriate Time Periods for Child Mortality Estimates from Full Birth Histories. *PLoS Medicine 9*(8), e1001289.

Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20-22, Vienna, Austria. ISSN 1609-395X. http://mcmc-jags.sourceforge.net/.*

Sawyer, C. (2012). Child Mortality Estimation: Estimating Sex Differences in Childhood Mortality since the 1970s. *PLoS Med 9*(8).

United Nations (2011). *Sex Differentials in Childhood Mortality.*

Wang, H., L. Dwyer-Lindgren, K. Lofgren, J. Rajaratnam, J. Marcus, A. Levin-Rector, C. Levitz, A. Lopez, , and C. Murray (2012). Age-specific and sex-specific mortality in 187 countries, 19702010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet 380*, 2071–2094.

World Bank (2011). *World Development Report 2012: Gender Equality and Development.* Washington, D.C.

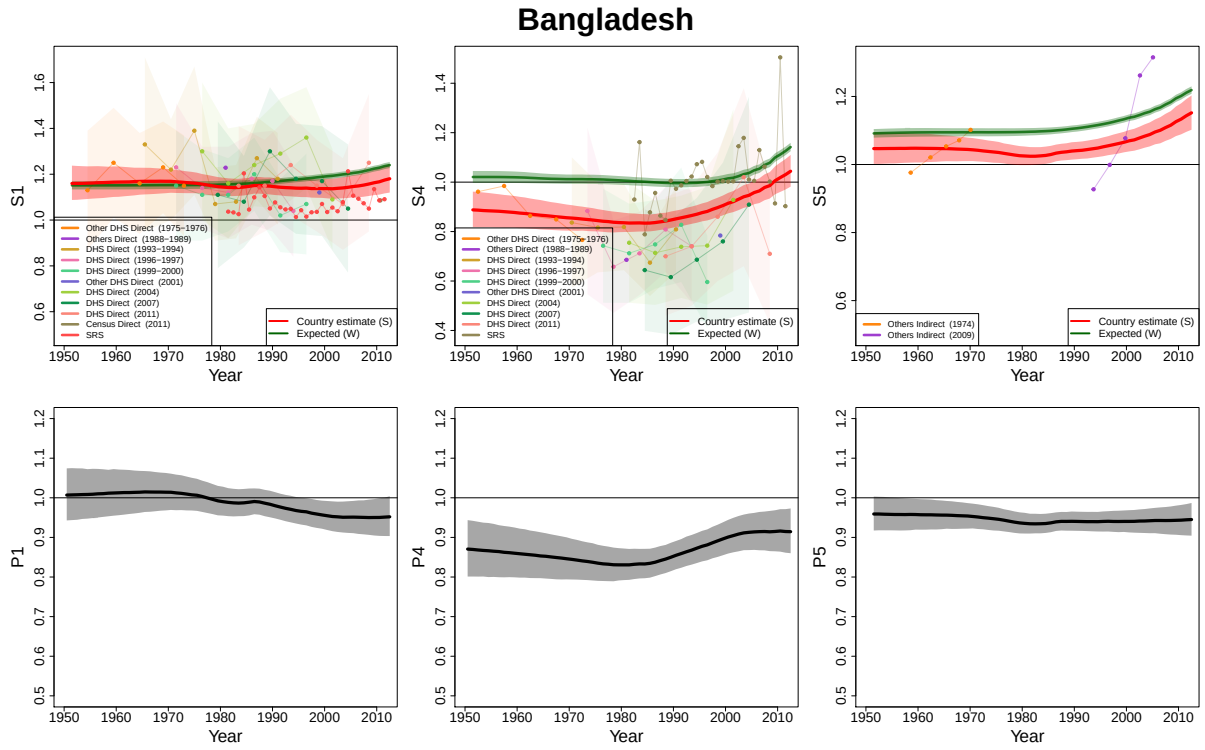

Figure 5: Estimates for Bangladesh. Top row: Estimated sex ratio $S_{a,c,t}$ for the three age groups (red) and expected sex ratio $W_{a,c,t}$ (green). Observations are displayed by dots, shaded areas around observations illustrate sampling errors (where available). Bottom row: Estimated country multipliers $P_{a,c,t,}$ for the three age groups. Shaded areas around the estimates illustrate the 90% credible bounds.