Sex Ratio at Birth: Estimation and Projection using Bayesian Hierarchical Time Series Model

Fengqing Chao[1], Leontine Alkema[2], Patrick Gerland[3]

[1](Email: ephchf@nus.edu.sg), Saw Swee Hock School of Public Health, **National University of Singapore**.
[2]Department of Epidemiology and Biostatistics, **University of Massachusetts, Amherst**.
[3]Mortality Section, **United Nations Population Division**

Background

Current situation

- Under normal circumstance, the sex ratio at birth (SRB; i.e. number of male live births over number of female live births) varies between 1.04 and 1.07 (Chahnazarian, 1988).
- However, the observed SRB from countries like India and China have been well above the biological level for decades (Guilmoto, 2012; Goodkind, 2011).
- Characteristics for countries with SRB inflation (Guilmoto, 2012): 1) strong son preference; 2) widely available sex selection technology; 3) fertility squeeze, i.e. low total fertility rate (TFR, approximately the number of births per woman).
- So far, no assessment of SRBs for all countries over time has been carried out using all available data and reproducible estimation methods.

Objective

- To construct a database for SRB including data from vital registration and surveys;
- To estimate and project country-specific SRBs for 221 countries from 1990 (or the earliest year of data collection) to 2100, and assess the uncertainty in SRBs;
- To identify countries with outlying SRBs.

Data

Data Types	# obs.
Census	56
DHS	1,976
Other DHS	857
Others	148
VR	6,115
total	**9,152**

Table: **Database for modeling.** DHS: Demographic and Health Surveys; VR: Vital Registration.

- data are from 192 countries;
- 14,403 country-years of data;
- 67.9 country-years of data from each of the 221 countries estimated.

Method part 1: SRB model

We modeled the true SRB $R_{c,t}$ for country c (from region $r[c]$), year t as:

$$R_{c,t} = N_{r[c]} \cdot P_{c,t} + \alpha_{c,t}.$$

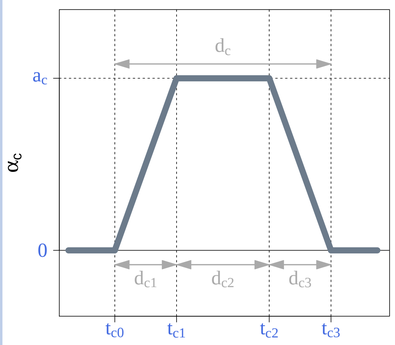
- N_r – the regional biological norm for SRB: estimated using data from country-years without potential prenatal discrimination, and assumed to be constant over time. We used independent non-informative priors to estimate N_r:

$$N_r \sim N(1, 1.1), \text{ for } r = 1, \ldots, R.$$

- $P_{c,t}$ – the divergence from N_r for the respective country-year under natural circumstances: modeled with an autoregressive time series model of order one (AR(1)):

$$\log(P_{c,t}) \sim N_{(0.9,1.1)}(\rho \cdot \log(P_{c,t-1}), \sigma^2_\epsilon).$$

- $\alpha_{c,t}$ – the upward adjustment factor for selected countries to capture inflated SRB levels that may be due to potential son preference and sex-selective abortion.



- $t_{c,0}$: starting year of SRB inflation period;
- $d_c (= d_{c,1} + d_{c,2} + d_{c,3})$: total length of the inflation period;
- $t_{c,3} (= t_{c,0} + d_c)$: ending year of inflation period;
- a_c: maximum value that $\alpha_{c,t}$ could reach.

A Bayesian hierarchical model (Gelman A. et al., 2004; Lindley and Smith, 1972) is used to model country-level parameters:

$$a_c \sim N_{(0,+\infty)}(\mu_a, \sigma^2_a), \text{ for selected } c$$
$$d_{c,g} \sim N_{(0,+\infty)}(\mu_{d_g}, \sigma^2_{d_g}), \text{ for } g = 1, 2, 3.$$

We used the fertility squeeze criterion (3) to inform the starting year of past or future upwards adjustments $t_{c,0}$. The year when TFR hits 2.5 is used for all countries except India, based on empirical evidence from other countries:

$$t_{c,0} \sim N_{(1970,2100)}(t_{c,tfr.target}, \sigma^2_{tfr.target}), \text{ all except India,}$$
$$t_{c,0} \sim U(1970, 2100), \text{ for India.}$$

Method part 2: Selection of Countries to Model SRB Inflation

Selection of countries with potential son preference that leads to inflated SRB:

- Observed SRB is suspected to be beyond biological norm as supported by literature; OR
- Desired sex ratio at birth > 120 and/or sex ratio of last birth > 130 (Bongaarts 2013); OR
- Outlying female under-5 mortality rate (Alkema et al. 2014).

Method part 3: Data Model

We take into account sampling and non-sampling variances in our data model as follows:

$$\log(r_i) \sim N(\log(R_{c[i],t[i]}), \sigma^2_i + \omega^2_{s[i]}).$$

- r_i: the i-th observed SRB input data;
- σ^2_i: the i-th sampling/stochastic variance (calculated);
- $\omega^2_{s[i]}$: the i-th non-sampling variance (to be estimated), where $s[i]$ refers to the source type s for the i-th SRB observation. For VR data, we assume their non-sampling variances are zero.

Computing We obtained posterior samples of all the model parameters and hyper parameters using a Markov chain Monte Carlo (MCMC) algorithm, implemented in open source softwares `R 3.0.2` (R Core Team, 2016) and `JAGS 4.0.1` (Just Another Gibbs Sampler) (Plummer, 2015).

Result 1: SRB Country Estimates

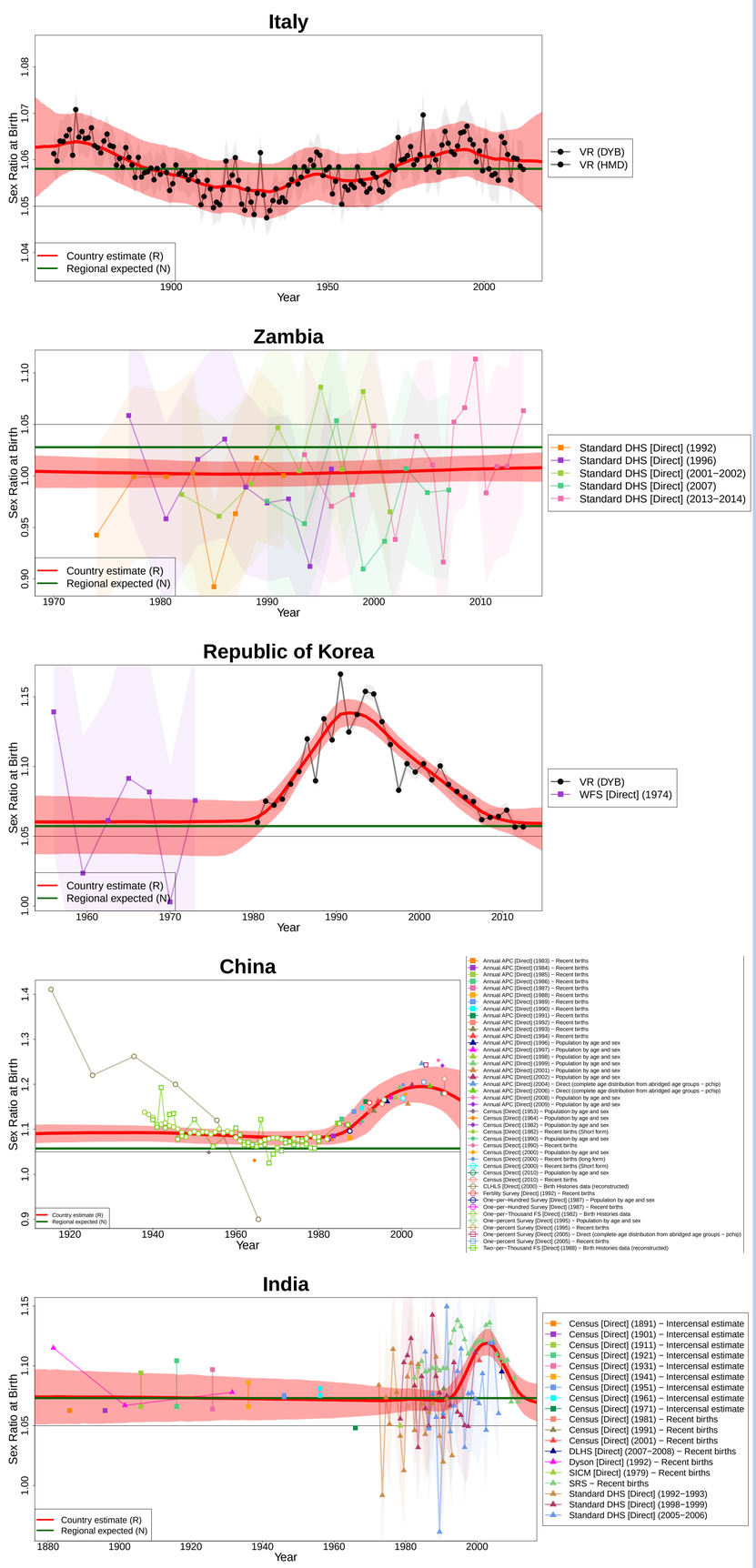


Figure: **SRB over time for Italy, Zambia, Republic of Korea, China, and India.**

Result 2: SRB Country Projections

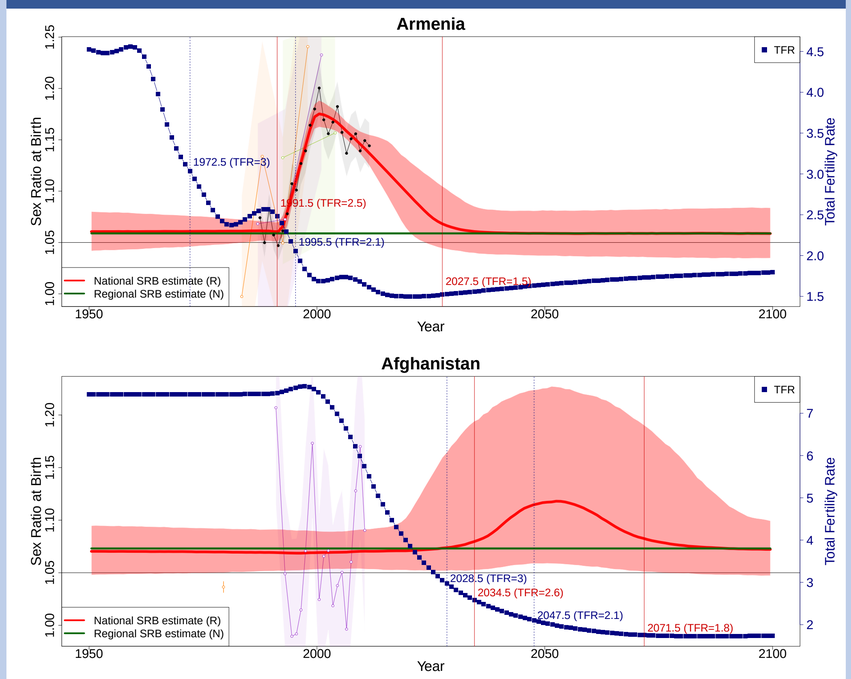


Figure: **SRB over time for Armenia, and Afghanistan.** Illustrations of projections for countries with ongoing and potential future SRB inflations, with TFR estimates and projections from the United Nations (2015 version).

Result 3: Countries with Outlying SRB

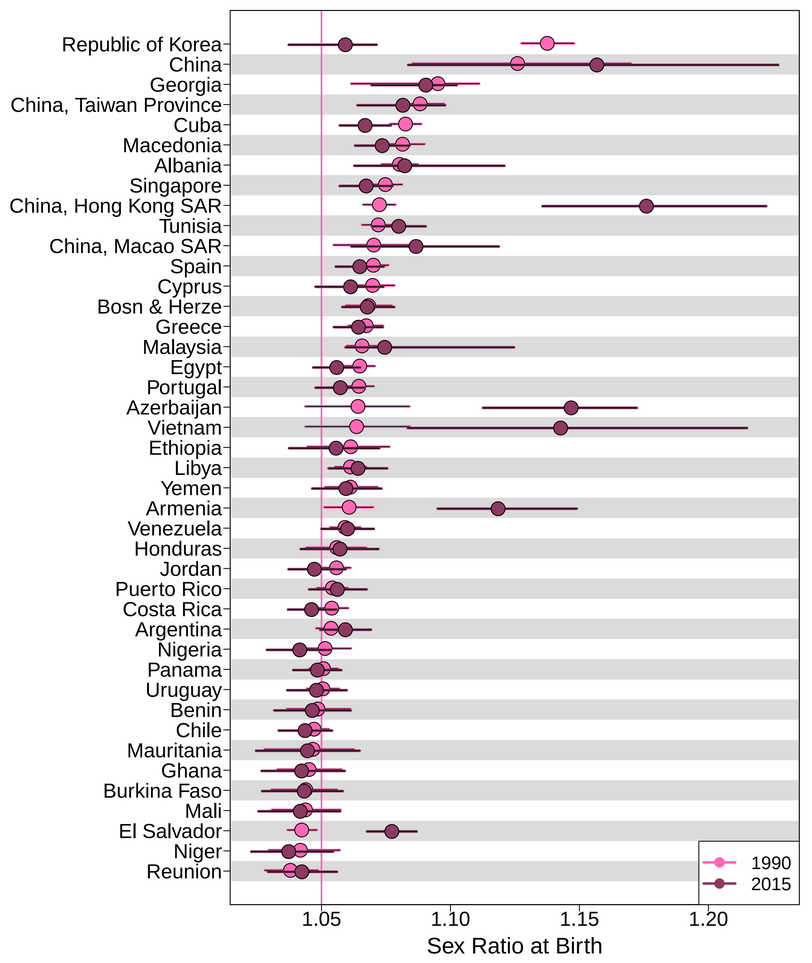


Figure: **Countries with outlying SRB estimates in 1990 and/or 2015.** "Outlying" means the 95% CI does not include its corresponding regional norm. *Note: data quality issues may partially explain outlying SRBs in some identified countries.

Result 4: Global and Regional SRB

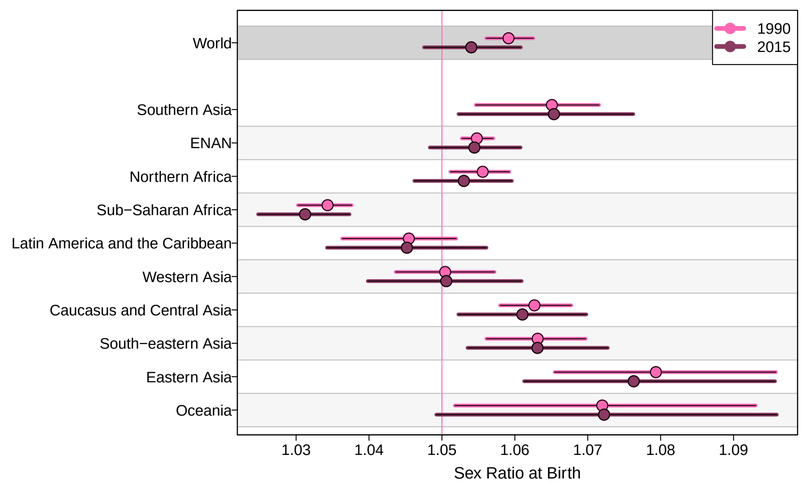


Figure: **Sex ratio at birth for the world and regions in 1990 and 2015.** Region "ENAN" refers to the combination of countries in Europe, North America, Australia, and New Zealand.

References & Acknowledgement

- Chahnazarian A., *Determinants of the sex ratio at birth: Review of recent literature*. Biodemography and Social Biology, 35(3-4):214–235, 1988.
- Goodkind D., *Child underreporting, fertility, and sex ratio imbalance in China*. Demography, 48(1):291–316, 2011.
- Guilmoto C., *Sex Imbalances at Birth: Current trends, consequences and policy implications*. UNFPA Asia and Pacific Regional Office, Bangkok, Thailand, 2012. ISBN 978-974-680-338-0.
- Bongaarts J. *The implementation of preferences for male offspring*. Population and Development Review. 2013;39(2):185208.
- Alkema L, Chao F, You D, Pedersen J, Sawyer CC. *National, regional, and global sex ratios of infant, child, and under-5 mortality and identification of countries with outlying ratios: a systematic assessment*. The Lancet Global Health. 2014;2(9):e521e530.
- Gelman A. et al., *Bayesian Data Analysis*. Chapman Hall/CRC, Boca Raton, Fl., 2nd edition, 2004.
- Lindley D. V. and Smith A. F. M., *Bayes estimates for the linear model*. 34:1–41, 1972.
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2016. Available from: `https://www.R-project.org/`.
- Plummer M. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling (version 4.0.1), 2015.

The project described is solely the responsibility of the authors and does not necessarily represent the official views of the United Nations. The authors are grateful to Christophe Z. Guilmoto for helpful comments and discussions, and Vladimira Kantorova for assistance on database.
This project is funded by research grant from the National University of Singapore: `R-608-000-125-646`.