Projecting sex imbalances at birth at global, regional, and national levels from 2018 to 2100: scenario-based Bayesian probabilistic projections of the sex ratio at birth and missing female births

Corresponding author

Fengqing Chao; fengqing.chao@kaust.edu.sa

4700 King Abdullah University of Science and Technology, Thuwal, 23955-6900, Saudi Arabia

Authors:

Fengqing Chao, PhD

Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, 23955-6900, Saudi Arabia

Patrick Gerland, PhD

United Nations Population Division, DESA, United Nations, New York, NY 10017, USA

Alex R. Cook, PhD

Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, 117549, Singapore

Christophe Z. Guilmoto, PhD

CEPED/IRD, Université de Paris, Paris, 75006, France

Leontine Alkema, PhD

School of Public Health and Health Sciences, University of Massachusetts, Amherst, MA 01003, USA

ABSTRACT

Introduction: Skewed levels of the sex ratio at birth (SRB) due to sex-selective abortions have been observed in several countries since the 1970s. They will lead to long-term sex imbalances in more than a third of the world's population with yet unknown social and economic impacts on affected countries. Understanding the potential evolution of sex imbalances at birth is therefore essential for anticipating and planning for changing sex structures across the world.

Methods: We produced probabilistic SRB projections from 2018 to 2100 based on different scenarios of sex ratio transition and assessed their implications in terms of missing female births at global, regional, and national levels. Based on a comprehensive SRB database, we project the skewed SRB and missing female births with a Bayesian hierarchical time series mixture model. The SRB projections under reference scenario S1 assumed SRB transitions only for countries with strong statistical evidence of SRB inflation, and the more extreme scenario S2 assumed a sex ratio transition for countries at risk of SRB inflation but with no or limited evidence of ongoing inflation.

Results: Under scenario S1, we projected 8.1 (95% uncertainty interval [2.3; 19.6]) million additional missing female births to occur by 2100. Countries affected will be those already affected in the past by imbalanced SRB, such as China and India. If all countries at risk of SRB inflation experience a sex ratio transition as in scenario S2, the projected missing female births increase to 24.8 [13.1; 44.1] million with a sizeable contribution of sub-Saharan Africa.

Conclusion: The scenario-based projections provide important illustrations of the potential burden of future prenatal sex discrimination and the need to monitor SRBs in countries with son preference. Policy planning will be needed in the years to come to minimize future prenatal sex discrimination and its impact on social structures.

Keywords: Mathematical modelling; Public Health; Medical demography

KEY QUESTIONS

What is already known?

- No prior study has constructed scenario-based projections for the sex ratio at birth and missing female births for all countries.
- Approaches used for SRB projections include the assumption of constant country-specific levels from 2017 to 2100 in the Global Burden of Disease study and projections based on expert opinion and linear extrapolation by the UN World Population Prospects and the US Census Bureau.

What are the new findings?

- We provided projections of the sex ratio at birth and missing female births based on two scenarios for all countries till 2100.
- Our projections were based on Bayesian hierarchical time series mixture models that were fitted to an extensive database on national sex ratio at birth. The hierarchical

structure of the model allowed for sharing information on sex ratio transitions across countries and for projecting the potential future inflation process under different scenarios.

- Our study provided insights into future changes in the sex ratio at birth as well as the uncertainty associated with these changes.

What do the new findings imply?

- The scenario-based projections of sex ratio at births showed that the sex ratio at birth is most likely to stabilize and decline within less than 25 years in countries currently affected by sex imbalances at birth.
- However, there will be a deficit of more than 7 million female births before 2030 under our conservative scenario.
- A less optimistic scenario envisages a rise in the sex ratio in other countries with strong prevalence of son preference, including populous countries such as Nigeria or Pakistan, unless countries take concrete steps to reduce gender bias and to end sex selection.

INTRODUCTION

Over the last forty years, prenatal gender-biased sex selection has become the most visible consequence of son preference. Along with child marriage and female genital mutilation, sex selection is one of the key harmful practices defined by the UN and targeted under the SDGs.[1] Sex-selective abortions, the main mechanism behind sex selection, have been observed across a range of various countries from Southeast Europe to South and East Asia.[2-11] They lead to a hike in the sex ratio at birth (SRB; namely the ratio of male to female live births) above its natural (biological) level and to the emergence of a surplus of male births. Along with excess female mortality, this SRB inflation is now a major contributor to the "missing women", a concept first described by Sen in 1990.[12] According to recent studies, prenatal sex selection accounts for about half of the recent deficit of females in the world during the previous decades.[13, 14]

Data limitations necessitate the use of models for estimating SRB levels and trends.[13] Population statistics derived from birth registration, census figures and demographic surveys have contributed to estimates of the SRB in different countries and of its inflation due to prenatal sex selection.[15] However, SRB series in many affected countries are spotty—for lack of reliable civil registration—and SRB rates must therefore often be re-estimated from a variety of sources. In prior work, we gathered a comprehensive database of SRB measurements and provided a new set of SRB estimates for all countries during 1950–2017 along with estimates of the natural SRB level in the absence of deliberate sex selection.[13] The total number of female births missing over 1970–2017 due to prenatal sex selection was estimated at 45.0 million with a 95% uncertainty interval [34.4; 54.8] million. More than 95% of them were missing from China or India, countries combining severely skewed SRB levels and the largest numbers of annual births in the world.

The almost simultaneous rise of the sex ratio at birth in various countries accounting for more than a third of the world's population is an unprecedented phenomenon in demographic history. It will have long-lasting repercussions on demographic and social structures due to the mounting number of missing adult females, with long-term social consequences on affected societies.[16-18] It is therefore crucial to be able to anticipate the future dynamics of SRB inflation across the world, in already affected countries and in those at risk. The main challenge is to understand whether birth masculinity will stay indefinitely skewed in countries affected by sex-selective abortions and whether new countries may be affected in the future.

No rigorous attempt has been made so far to outline and project the future evolution of the SRB imbalances in the world and to estimate its implications on the potential deficit of future female births. The most recent set of world population projections from the Global Burden of Disease Study assumes for instance birth masculinity to remain constant through 2100 at their 2017 levels, an unlikely feature in view of the rapid changes in SRB levels observed in countries with skewed SRBs.[19] The population projections published every two years by the United Nations derives SRB trends for the next decades from expert opinion and linear extrapolation.[20] The International Data Base developed by the US census Bureau similarly projects the sex ratio at birth by linear extrapolation.[21]

The main reason for this absence of evidence-based SRB projections stems from the unique character of the human-made SRB imbalances observed over the last forty years and the corresponding lack of historical references from which to infer its future evolution. A more technical factor relates to the apparently non-linear trends observed in several countries. Mortality and fertility tend to follow monotonic and near-linear patterns such as long-term declines that are easier to model and project. In contrast, the SRB trends observed so far appear to follow a more complex and non-monotonic cycle, characterized by a three-stage transformation: an increase, stabilization and an ultimate decrease of the SRB, referred to as the *sex ratio transition*.[22] The SRB transition process has been observed in several countries in the past decades.[13] The curved bell-shaped pattern cannot be modeled by a simple linear regression model and requires additional hypotheses or covariates to recreate. However, the experience of countries already affected by SRB inflation can now help us delineate the expected patterns of SRB transformations in relation to its original drivers. We can use for this purpose the existing database of country-level SRB measurements from 1970 to 2017 and an extended modeling approach to develop a set of projection scenarios. This model will also make possible the estimation of the probability of a future SRB rise and subsequent decline in countries that have so far not been affected by SRB imbalances in spite of pronounced son preference.

Our paper provided probabilistic scenario-based projections of the SRB for 212 countries ("countries" or "areas" as per United Nations classification with population size greater than 90,000 in 2017) until 2100 based on the trends observed until 2017. We focused on 29 countries where son preference has been documented and for which we already have the SRB baseline levels used to assess the extent of the SRB inflation. They include 12 countries where SRB inflation has already happened in the past as well as 17 additional countries at risk of SRB inflation, i.e. countries where the SRB may rise in the future. We produced Bayesian projections of the SRB for each country as well as the number of female births missing due to prenatal sex selection under different scenarios. The results are presented at global, regional, and national levels from 2017 to 2100 for two scenarios.

DATA AND METHODS

We summarized here the data and methods used. A more detailed description of the model specifications, statistical computation and model validation is available in the methodology manuscript.[23]

Data

We started from an extensive database of 10,835 observations from 202 countries. The description of the construction of the original database and preprocessing steps are given elsewhere;[13] the database is publicly available.[24] The database was used for estimating the SRB from 1950 to 2017 and projecting the SRB till 2100.

We also used the latest United Nations demographic estimates and projections for fertility and birth trends from 1950 to 2100. The United Nations World Population Prospects (WPP) were released in 2019; we used the annual medium-variant projections.[20] We used 1,000 TFR trajectories to account for uncertainty for each country-year.[25, 26]

Sex Ratio Transition Model

The findings rest on the Bayesian model used to project future SRB change based on past SRB measurements and fertility trends.[23] The four guiding assumptions of this SRB transition model used here can be summarized as follows:

1. Countries with past, ongoing, or potential future SRB inflation can be identified using criteria related to son preference intensity and the timing of fertility decline.
2. National baseline SRB estimates can be used as reference for the period under study.
3. SRB inflation trajectories follow three successive stages: rise, plateau, and decline.
4. SRB trajectories in countries affected by SRB imbalances prior to 2018 can be used as a template for Bayesian projections of the future SRB levels.

The first assumption is based on the analysis of the three preconditions of the emergence of prenatal sex selection and the subsequent rise of the SRB: strong preference for male children; fertility decline constraining parents' desired number of children; and access to modern sex-selection technology.[22] The first precondition is used to identify the propensity to sex select based on various qualitative and quantitative indicators of son preference. It led to the identification of 29 countries with actual or potential prenatal sex selection.

The second assumption rests on the stability of natural (biological) SRB level and of the preexisting regional differentials.[27-30] SRB series in industrialized countries with reliable registration system and no prenatal sex selection show that the SRB rarely change by more than 1 per cent over a century. In addition, the presence of specific regional differentials—such as the lower natural SRB among populations of Sub-Saharan Africa or of African descent in the USA—is well-established.[15, 31] Baseline estimates can therefore be used as a benchmark to model SRB inflation and missing female births. The SRB is modeled as the sum of baseline SRB level (with natural year-by-year fluctuations) and the product of an indicator to capture the presence or absence of inflation and a non-negative SRB inflation factor,

Modeled SRB = Baseline SRB + (Presence of inflation x Inflation factor).

The third assumption, that SRB trajectories successively rise, plateau, and decline, is based on the non-linear patterns of SRB changes observed in countries affected by sex selection until today. The first stage is the initial increase of the SRB from its baseline level. This rise continues over several years and is followed by stabilization at a higher SRB level, which constitutes the second stage of the cycle. The third stage begins after several years at a plateau level and corresponds to the final decline of the SRB back to its baseline natural level. Several countries (Hong Kong (SAR of China), Georgia, and Republic of Korea) already illustrate these three completed phases of SRB transition. Other countries are currently going through the second stabilization phase (Armenia, Azerbaijan, China, India, and Vietnam) or the third declining phase (Albania, Montenegro, Taiwan (Province of China), and Tunisia). The parameterization of the inflation factor follows therefore a trapezoid (/‾\ ) function to capture the three stages of the sex ratio transition: initial increase, subsequent plateau, and final decrease. The model also captures regularities across countries in the patterns of the sex ratio transition and it makes use of the total fertility rate (TFR) as a covariate of the SRB inflation. It is used in particular for estimating the start year of the sex ratio transition process.

The fourth assumption, that SRB trajectories in countries affected by SRB imbalances prior to 2018 can be used as a template for Bayesian projections of the future SRB levels, is at the core of the Bayesian projection exercise. Conditioning on the third assumption, past observations are used to determine future trends in a probabilistic manner. The length of each transition stage and the maximum extent of the SRB inflation is estimated with uncertainty by using Bayesian hierarchical models. The hierarchical structure of the model allows for sharing of information about sex ratio transitions across countries. The implementation of the model is described in the methodology manuscript.[23]

Scenario-based Projections of SRB

Among the 29 countries at risk of sex selection, we identified 12 countries with a strong statistical evidence of SRB inflation when the posterior probability of presence of SRB inflation is more than 95%. For the 12 countries with strong evidence of existing SRB inflation, the sex ratio transition model is used to produce projections of future inflation.

For the remaining 17 countries at risk but without strong evidence of past/ongoing SRB inflation, we project SRB under scenarios with varying assumptions regarding the occurrence of a sex ratio transition (see SI Appendix pp1 for the explanation of scenario names and SI

Appendix Table 1 for the type of SRB projections for each country). The two scenarios are the following:

- In the first scenario, S1, the 17 countries' SRBs are projected without SRB inflation.
- In the second scenario, S2, projections assume that the 17 countries will experience an SRB inflation with 100% probability.

S2 scenario accounts for the uncertainty of the TFR projections. The uncertainties in these projections become larger as we draw further from the current period. No other factors are used in projections under scenario S2.

Missing Female Births

To project the effect of SRB imbalances under different scenarios, we calculated the annual number of missing female births (AMFB) and the cumulative number of missing female births (CMFB) from 2018 to 2100. We define missing female births as female births prevented by sex selection, i.e. to the number of sex-selective abortions, as defined in a recent study.[32] The AMFB is computed as the difference between the numbers of female live births based on the baseline SRB and of those based on the projected SRB. The CMFB is the cumulated sum of the AMFB from 2018 till a given year. The number of live births is drawn from the 2019 United Nations WPP for 2018–2100.[20]

Patient and Public Involvement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

RESULTS

We produced scenario-based probabilistic projections of the SRB and number of missing female births on global, regional, and national levels from 2018 to 2100. Using the reference year for the Sustainable Development Goals, results are presented for 2018–2030 and 2031–2100. Estimates for 1970–2017 are in the Appendix. Detailed scenario-based projection results by country are also included in Appendix.

SRB Projections for Global and Regional Aggregates

Figure 1 illustrate the projected SRB for the world and by region for the two scenarios. At the global level, the SRB projections are similar. Under this scenario S1, in which sex imbalances are restricted to countries with strong statistical evidence of SRB inflation, the global SRB is projected to decline from 1.067 [1.058; 1.076] in 2017 to 1.052 [1.045; 1.062] by 2030 and to 1.045 [1.040; 1.050] by 2100. The SRBs under this scenario are slightly lower than those under scenario S2 in which additional countries register a rise in their SRB. Under S2, the global SRB would decline to 1.054 [1.046; 1.064] by 2030 and 1.048 [1.041; 1.058] in 2100.

The situation is more diverse across regions. The SRBs for S1 are projected to converge towards their aggregated regional baseline levels by 2100, ranging from 1.033 [1.026; 1.040] in sub-Saharan Africa to 1.067 [1.045; 1.089] in Oceania. Most of this convergence will take place around 2030, notably for the three largest contributors to world sex imbalances at birth

(China, India and Vietnam). The decline is especially rapid in China, which displayed the highest SRB level in the world in 2010.

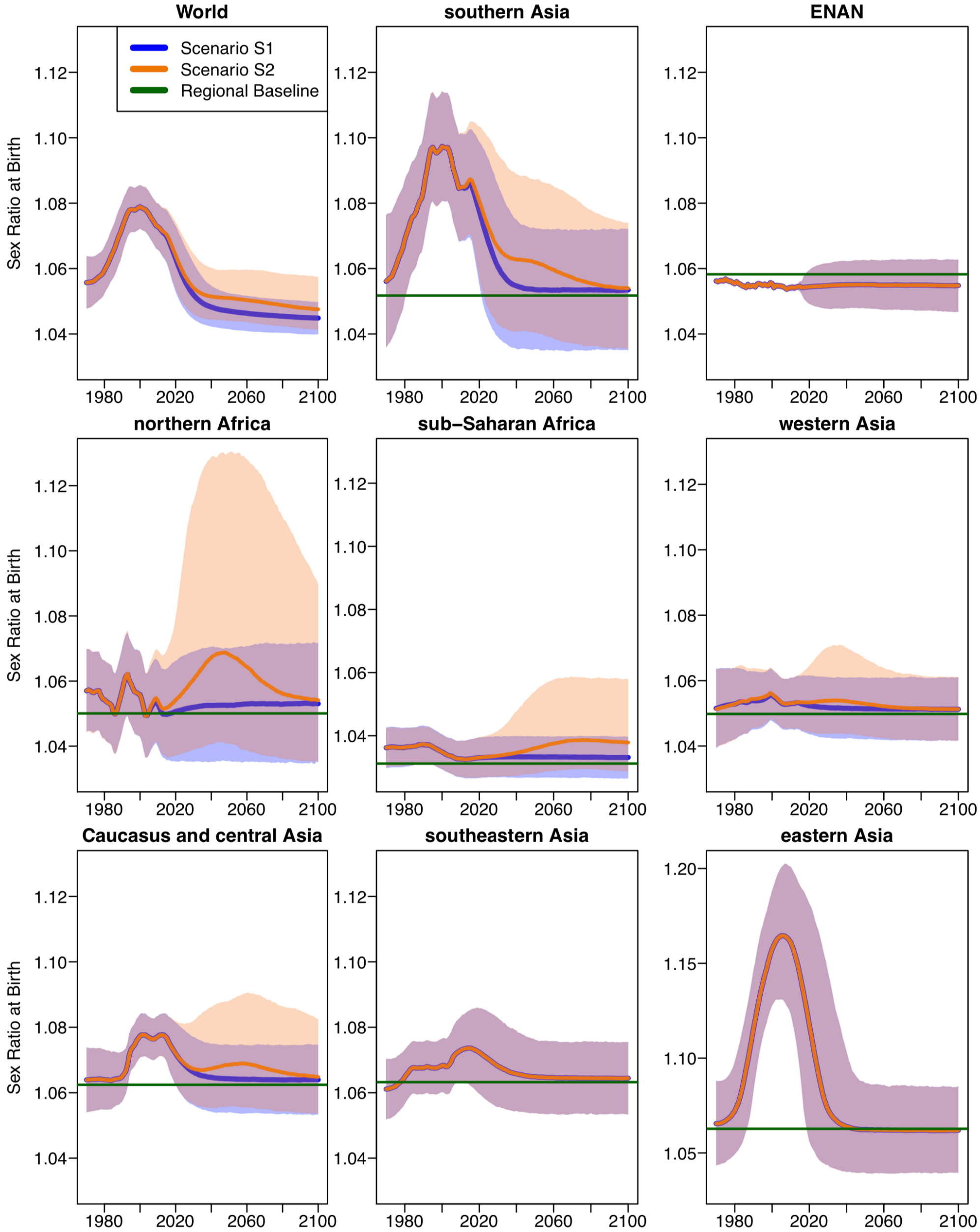

Figure 1: Global and regional SRB estimates and projections 1970–2100. Point estimates (solid lines) and 95% uncertainty intervals (shaded areas) for scenario-based projections. The green horizontal lines are regional baseline SRB median estimates. Selected regions contain at least one country at risk of sex imbalance. ENAN: the combination of countries in Europe, North America, Australia, and New Zealand.

The projected SRBs for scenario S2 differ from those for S1 in regions where additional countries at risk of sex imbalances at birth are located: northern Africa, sub-Saharan Africa, Caucasus and central Asia, and southern Asia. In the Asian regions, under S2, the SRB is estimated to have peaked in southern Asia before 2017 and is projected to first decline in Caucasus and central Asia until 2040 and then rebound up to 1.069 [1.056; 1.090] in around 2060 before an ultimate return to natural level at the end of the century. The onset of the SRB inflation is expected to take place after 2030 in two countries of southern Asia (Afghanistan and Pakistan). In the African regions under scenario S2, the SRB is projected to peak in the mid-2040s in northern Africa at 1.069 [1.041; 1.129] and in the 2080s in sub-Saharan Africa at 1.039 [1.029; 1.058]. This later rise of the SRB in sub-Saharan Africa corresponds to a later timing of fertility decline. This is notably the case for populated countries such as Nigeria and Tanzania where the SRB is expected to rise only after 2060 and to remain elevated till the end of the century in scenario S2 because the span of the projected SRB inflation process is 37 [15; 64] years.

By 2100, the differences in projected SRB under S1 and S2 are almost negligible in all regions, with the exception of sub-Saharan Africa where the SRB inflation is expected to take place mostly during the second half of the century. The projected rise in the SRB from S1 to S2 in 2100 is projected to be modest at 0.004, but with substantial uncertainty [0.000; 0.024].

Missing Female Births

The global burden of sex imbalances at birth during 2018–2100 is projected to be 8.1 [2.3; 19.6] million missing female births (CMFB) under S1. This number increases to 24.8 [13.1; 44.1] million under S2 (Figure 2, Table 1, Figure 3). Based on the first scenario, missing female births are projected to occur mostly before 2030, with the projected CMFB at 7.3 [2.2; 14.1] million during 2018–2030 and 0.8 [0.0; 6.3] million during 2031–2100. Under scenario S2, the projections for 2018–2030 are very similar to the first scenario at 8.4 [3.1; 15.8] million. However, the number of CMFB projected for 2031–2100 is much larger under S2 at 16 .4 [7.3; 32.6] million.

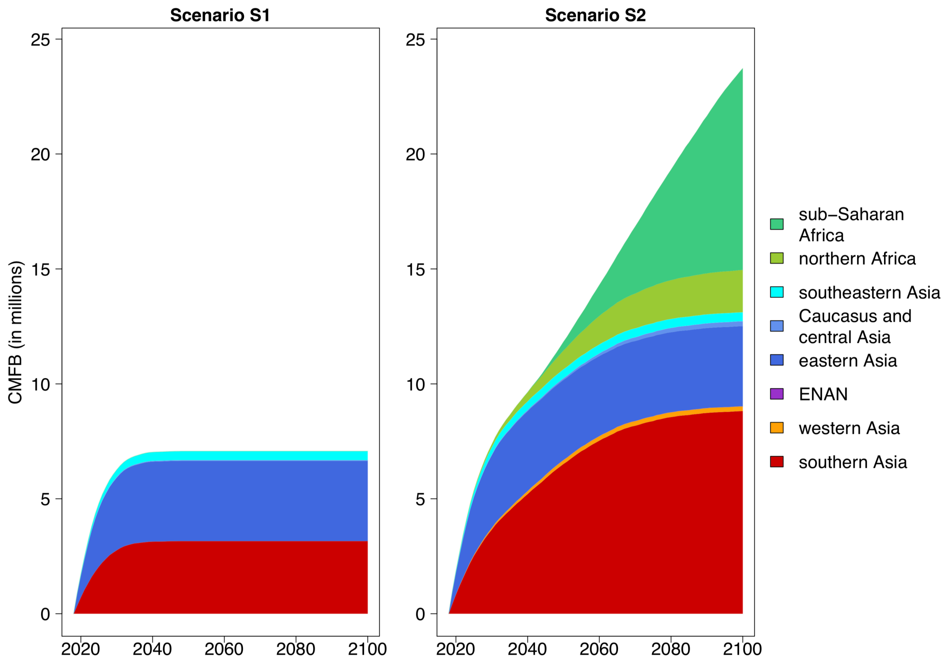

Figure 2: Global and regional CMFB projections 2018–2100, by scenario. CMFB: cumulative number of missing female births. Regional CMFBs are stacked to obtain global CMFB. Regions are sorted by the timing of the imbalance (i.e. the year in which the annual number of missing female births peaks). ENAN: the combination of countries in Europe, North America, Australia, and New Zealand.

World/ Region		Scenario S1				Scenario S2			
	CMFB in ,000 (1970–2017)	**CMFB in ,000 (2018–2030)**	**CMFB in ,000 (2031–2100)**	**CMFB in ,000 (2018–2100)**	**CMFB % total (2018–2100)**	**CMFB in ,000 (2018–2030)**	**CMFB in ,000 (2031–2100)**	**CMFB in ,000 (2018–2100)**	**CMFB % total (2018–2100)**
World	46,449	7,288	809	8,096	100	8,412	16,371	24,784	100
	[37,826; 55,736]	[2,209; 14,142]	[1; 6,257]	[2,330; 19,590]	–	[3,131; 15,823]	[7,292; 32,561]	[13,116; 44,147]	–
southern Asia	21,198	3,191	397	3,589	44.3	4,104	5,154	9,257	37.4
	[15,819; 26,876]	[494; 7,024]	[0; 3,194]	[494; 9,735]	[7.5; 90.8]	[1,023; 8,316]	[1,051; 11,983]	[3,511; 17,725]	[15.1; 63.6]
ENAN	27	2	0	2	0	2	0	2	0
	[18; 44]	[0; 7]	[0; 2]	[0; 8]	[0.0; 0.2]	[0; 7]	[0; 2]	[0; 8]	[0.0; 0.0]
northern Africa	90	3	0	4	0	133	1,710	1,842	7.4
	[43; 385]	[0; 19]	[0; 3]	[0; 22]	[0.0; 0.4]	[0; 1,000]	[50; 4,949]	[140; 5,199]	[0.6; 21.1]
sub-Saharan Africa	0	0	0	0	0	0	8,783	8,783	35.4
	[0; 2]	[0; 0]	[0; 0]	[0; 0]	[0.0; 0.0]	[0; 609]	[1,435; 22,769]	[1,456; 22,868]	[7.8; 63.4]
western Asia	149	0	0	0	0	79	136	215	0.9
	[0; 457]	[0; 0]	[0; 0]	[0; 0]	[0.0; 0.0]	[0; 388]	[2; 1,072]	[14; 1,287]	[0.1; 5.1]
Caucasus and central Asia	202	37	2	39	0.5	41	181	221	0.9
	[174; 233]	[7; 78]	[0; 33]	[7; 105]	[0.1; 2.3]	[8; 108]	[5; 564]	[36; 607]	[0.1; 2.7]
southeastern Asia	531	360	72	432	5.3	360	72	433	1.7
	[283; 814]	[0; 970]	[0; 874]	[0; 1,757]	[0.0; 28.8]	[0; 970]	[0; 874]	[0; 1,757]	[0.0; 7.6]
eastern Asia	24,252	3,694	337	4,030	49.8	3,694	337	4,030	16.3
	[17,389; 31,677]	[8; 9,784]	[0; 5,042]	[8; 14,111]	[0.2; 86.8]	[8; 9,784]	[0; 5,042]	[8; 14,111]	[0.0; 44.6]

Table 1: Global and regional CMFB for periods 1970–2017, 2018–2030, 2031–2100 and 2018–2100, by scenario. CMFB: cumulative number of missing female births. Median estimates are the numbers before the brackets. Numbers in brackets are 95% uncertainty intervals. Proportions may not sum up to 100%, due to rounding. ENAN refers to the combination of countries in Europe, North America, Australia, and New Zealand.

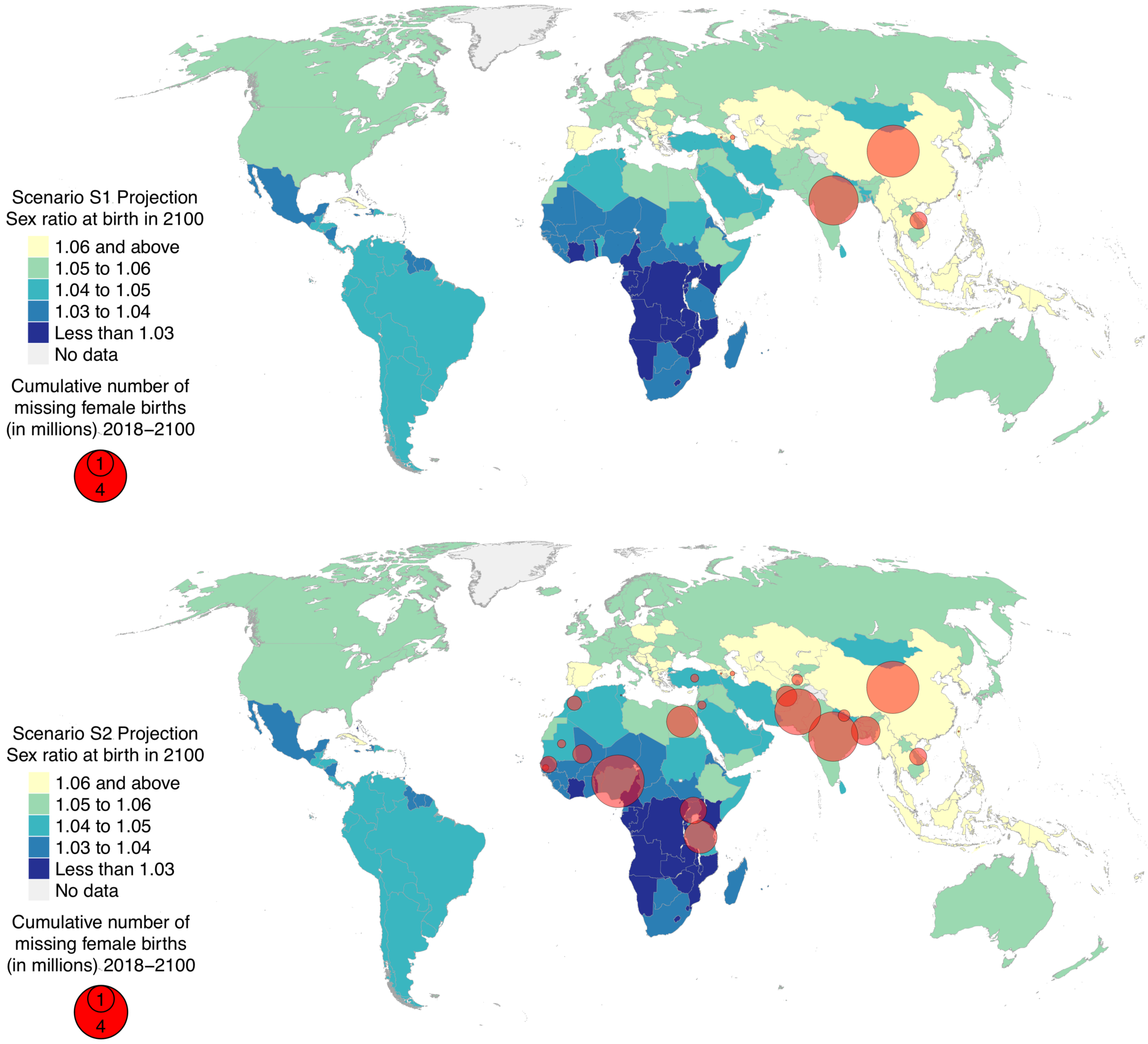

Figure 3: Projected SRB in 2100 and the CMFB during 2018–2100 for S1 and S2, by country. CMFB: cumulative number of missing female births. Countries are coloured by the levels of their SRB median estimates. Radii of circles are proportional to CMFB for countries.

Under S2, we project that from 2030 till 2100, the cumulative deficit of female births in southern Asia and eastern Asia will be 5.2 [1.1; 12.0] million and 0.3 [0.0; 5.0] million CMFB respectively, while sub-Saharan Africa may experience 8.8 [1.4; 22.8] million CMFB (Table 1). In sub-Saharan Africa, sex imbalances at birth are projected to emerge only during the second half of the century, except for Uganda where the rise may take place earlier. As a result, almost all of the missing female births in sub-Saharan Africa until 2100 are projected to occur after 2050. The annual number of missing female births (AMFB) is projected to peak during the 2080s at 132 [1; 676] thousand, declining to 110 [0; 658] thousand by 2100 (Figure 2). This implies that sub-Saharan Africa would become a third locus of missing girls by 2100 (Figure 3, Bottom), accounting for more than a third of female births missing in the world during 2018–2100 under scenario S2.

Table 2 summarizes the distribution of CMFB during 2018–2100 by scenario at country level. Among the 12 countries with strong evidence of existing SRB inflation, three countries are estimated to have completed sex ratio transitions before 2017: Republic of Korea in 2006 (95% uncertainty interval [1997; 2011]), Hong Kong (SAR of China) in 2013 [2012; 2014], and possibly Georgia in 2016 [2008; 2027]. The end years for the remaining nine countries are projected to occur in 2020s (Albania, Armenia, Montenegro, Taiwan (Province of China), and Tunisia) and in 2030s (Azerbaijan, India, China, and Vietnam). During 2018–2100, the projected CMFB in India and China are 3.2 [0.5; 9.7] million and 3.4 [0.0; 1.8] million respectively. Under scenario S1, China and India are projected to contribute the majority of the CMFB around the world during 2018–2100, representing 49.7% [0.0%; 86.8%] and 44.3% [7.5%; 90.8%] of the total number respectively.

Table 2: Scenario-based CMFB projections for period 2018–2100, by country. CMFB: cumulative number of missing female births. *: countries at risk of future SRB inflation. Countries without * are those with past/ongoing SRB inflation. 2100+: years beyond 2100. Countries are presented by region. Median estimates are the numbers before the brackets. Numbers in brackets are 95% uncertainty intervals. Proportions may not sum up to 100% due to rounding. Region ENAN: the combination of countries in Europe, North America, Australia, and New Zealand.

				Scenario S1 (2018–2100)		Scenario S2 (2018–2100)	
Country (Region)	**Inflation probability (in %)**	**Inflation start year**	**Inflation end year**	**CMFB (in ,000)**	**Proportion of total CMFB (in %)**	**CMFB (in ,000)**	**Proportion of total CMFB (in %)**
India	100	1975	2033	3,589	44.3	3,589	15.5
(southern Asia)		[1970; 1981]	[2021; 2050]	[494; 9,735]	[7.5; 90.8]	[494; 9,735]	[2.0; 36.4]
Albania	100	1988	2024	2	0	2	0
(ENAN)		[1973; 1997]	[2016; 2043]	[0; 8]	[0.0; 0.1]	[0; 8]	[0.0; 0.0]
Montenegro	100	1980	2024	0	0	0	0
(ENAN)		[1971; 1990]	[2014; 2043]	[0; 1]	[0.0; 0.0]	[0; 1]	[0.0; 0.0]
Tunisia	100	1982	2021	4	0	4	0
(northern Africa)		[1976; 1989]	[2012; 2039]	[0; 22]	[0.0; 0.4]	[0; 22]	[0.0; 0.1]
Armenia	100	1992	2029	5	0.1	5	0
(Caucasus and central Asia)		[1990; 1993]	[2020; 2042]	[1; 15]	[0.0; 0.3]	[1; 15]	[0.0; 0.1]
Azerbaijan	100	1991	2031	33	0.4	33	0.1
(Caucasus and central Asia)		[1988; 1993]	[2019; 2049]	[2; 99]	[0.0; 2.1]	[2; 99]	[0.0; 0.5]
Georgia	100	1992	2016	0	0	0	0
(Caucasus and central Asia)		[1979; 1994]	[2008; 2027]	[0; 2]	[0.0; 0.0]	[0; 2]	[0.0; 0.0]
Vietnam	100	2001	2036	432	5.3	432	1.9
(southeastern Asia)		[1991; 2005]	[2017; 2061]	[0; 1,757]	[0.0; 28.8]	[0; 1,757]	[0.0; 7.6]
China	100	1980	2030	4,024	49.7	4,024	17.4
(eastern Asia)		[1972; 1988]	[2017; 2051]	[0; 14,106]	[0.0; 86.8]	[0; 14,106]	[0.0; 44.6]
Hong Kong, SAR of China	100	2004	2013	0	0	0	0
(eastern Asia)		[2002; 2005]	[2012; 2014]	[0; 0]	[0.0; 0.0]	[0; 0]	[0.0; 0.0]
Republic of Korea	100	1982	2006	0	0	0	0
(eastern Asia)		[1978; 1984]	[1997; 2011]	[0; 0]	[0.0; 0.0]	[0; 0]	[0.0; 0.0]

Table 2 continued on next page

Table 2 – continued from previous page

Country (Region)	Inflation probability (in %)	Inflation start year	Inflation end year	Scenario S1 (2018–2100) CMFB (in ,000)	Scenario S1 (2018–2100) Proportion of total CMFB (in %)	Scenario S2 (2018–2100) CMFB (in ,000)	Scenario S2 (2018–2100) Proportion of total CMFB (in %)
Taiwan, Province of China	99.7	1982	2023	5	0.1	5	0
(eastern Asia)		[1972; 1987]	[2012; 2041]	[0; 22]	[0.0; 0.4]	[0; 22]	[0.0; 0.1]
Afghanistan*	63.3	2033	2071	0	0	624	2.7
(southern Asia)		[2013; 2063]	[2041; 2100+]	[0; 0]	[0.0; 0.0]	[36; 1,848]	[0.1; 7.9]
Bangladesh*	37.4	2009	2046	0	0	1,213	5.2
(southern Asia)		[1987; 2038]	[2015; 2085]	[0; 0]	[0.0; 0.0]	[2; 3,787]	[0.0; 15.7]
Nepal*	62.6	2009	2047	0	0	201	0.9
(southern Asia)		[1989; 2036]	[2018; 2083]	[0; 0]	[0.0; 0.0]	[1; 666]	[0.0; 3.0]
Pakistan*	63.0	2030	2068	0	0	3,122	13.5
(southern Asia)		[1995; 2058]	[2027; 2100+]	[0; 0]	[0.0; 0.0]	[34; 9,685]	[0.2; 34.7]
Egypt*	61.9	2030	2068	0	0	1,472	6.3
(northern Africa)		[2006; 2058]	[2034; 2100+]	[0; 0]	[0.0; 0.0]	[0; 4,784]	[0.0; 19.5]
Morocco*	40.8	2003	2041	0	0	312	1.3
(northern Africa)		[1982; 2033]	[2010; 2080]	[0; 0]	[0.0; 0.0]	[0; 949]	[0.0; 4.2]
Gambia*	63.1	2053	2091	0	0	58	0.3
(sub-Saharan Africa)		[2027; 2082]	[2057; 2100+]	[0; 0]	[0.0; 0.0]	[0; 193]	[0.0; 0.8]
Mali*	63.4	2061	2099	0	0	528	2.3
(sub-Saharan Africa)		[2036; 2089]	[2066; 2100+]	[0; 0]	[0.0; 0.0]	[0; 1,957]	[0.0; 8.1]
Mauritania*	63.8	2065	2100+	0	0	101	0.4
(sub-Saharan Africa)		[2038; 2093]	[2068; 2100+]	[0; 0]	[0.0; 0.0]	[0; 387]	[0.0; 1.6]
Nigeria*	63.2	2065	2100+	0	0	4,050	17.5
(sub-Saharan Africa)		[2038; 2093]	[2069; 2100+]	[0; 0]	[0.0; 0.0]	[0; 17,214]	[0.0; 47.9]
Senegal*	63.0	2061	2099	0	0	399	1.7
(sub-Saharan Africa)		[2034; 2089]	[2064; 2100+]	[0; 0]	[0.0; 0.0]	[0; 1,422]	[0.0; 6.0]
Tanzania*	62.7	2068	2100+	0	0	1,589	6.9
(sub-Saharan Africa)		[2041; 2096]	[2071; 2100+]	[0; 0]	[0.0; 0.0]	[0; 6,460]	[0.0; 23.9]
Uganda*	62.6	2042	2080	0	0	1,043	4.5
(sub-Saharan Africa)		[2020; 2070]	[2049; 2100+]	[0; 0]	[0.0; 0.0]	[26; 3,251]	[0.1; 13.5]
Jordan*	56.3	2019	2057	0	0	102	0.4
(western Asia)		[2000; 2048]	[2028; 2095]	[0; 0]	[0.0; 0.0]	[5; 309]	[0.0; 1.4]
Turkey*	34.9	1993	2031	0	0	97	0.4
(western Asia)		[1974; 2028]	[2001; 2074]	[0; 0]	[0.0; 0.0]	[0; 1,158]	[0.0; 4.6]
Tajikistan*	62.2	2038	2076	0	0	178	0.8
(Caucasus and central Asia)		[2016; 2067]	[2044; 2100+]	[0; 0]	[0.0; 0.0]	[3; 564]	[0.0; 2.5]
Singapore*	44.9	1975	2014	0	0	0	0
(southeastern Asia)		[1970; 2012]	[1990; 2055]	[0; 0]	[0.0; 0.0]	[0; 3]	[0.0; 0.0]

For the 17 countries at risk of inflation but without strong evidence today, only five countries in this group have already reached low fertility before 2018 (Bangladesh, Singapore, Morocco, Nepal, and Turkey). Even though the start years for the five countries are estimated to be before 2018, the model also suggests low probability of inflation among them (Bangladesh at 37.4%, Morocco at 40.8%, Nepal at 62.6%, Singapore at 44.9%, and Turkey at 34.9%). For the remaining 12 countries, the start years of inflation under scenario S2 are projected, in general, to be around 2030s for most Asian countries and around 2050 and later for all countries in sub-Saharan Africa. These projected start years correspond to the projected fertility decline profiles, with Asian countries expected to achieve low fertility before sub-Saharan Africa countries. Under S2, the CMFB distribution becomes more spread out and shift towards sub-Saharan African countries. The annual number of missing female births in Nigeria and Tanzania will be among the largest in the world during the second half of the century. In this scenario, China and India account in this scenario only for 17.4% [0.0%; 44.6%] and 15.5% [2.0%; 36.4%] respectively of the global CMFB during 2018–2100. In contrast, Nigeria contributes 17.5% [0.0%; 47.9%] of the total and Pakistan 13.5% [0.2%; 34.7%].

DISCUSSION

It is crucial to be able to anticipate the potential course of sex imbalances at birth in order to strengthen policies against prenatal sex selection and to plan for the impact of future changes in sex structures across the world. To do so, we projected the SRB for all countries from 2018 to 2100 under two scenarios. The SRB model used here was based on a comprehensive database of 10,835 observations from 202 countries, the experience of countries facing elevated SRB levels prior to 2018, and model assumptions regarding the sex ratio transition process. The Bayesian hierarchical time series mixture model used allows sharing information from data-rich country-years with those with limited or no data for estimating baseline levels and natural deviations. In addition, the model included an inflation probability and provides probabilistic SRB projections based on two scenarios. Based on the out-of-sample validation and simulation results, the model showed reasonable predictive performance.[23]

The strength of our projection procedures lies in the quality of the original database and the model we developed. The national SRB database is comprehensive and is based on rigorous assessment and quality checks.[13] The Bayesian hierarchical time series mixture models enable (i) the estimation of SRB baselines on regional and national levels, (ii) the use of the experience of countries affected over the last forty years to model future change, and (iii) the incorporation of fertility declines as the covariate of future SRB change in countries with established son preference.

This study rests, however, on several model assumptions and is therefore subject to limitations.[13, 23] Firstly, the only source of heterogeneity of baseline (natural) accounted for in the model is ethnicity groups approximated by region and country. Secondly, due to the absence of quality estimates of other variables that might be relevant to sex-selective abortion, such as accessibility of modern reproductive technologies and to legal abortion, we use fertility decline as the only covariate in the model. Thirdly, no reliable measurement of son preference was available for 57 out of the 212 countries included in the study and accounting for 3.2% of the global births in 1970–2017. Lastly, we selected the countries at risk of SRB rise prior to model fitting instead of incorporating the selection in the model. The pre-model selection of these countries was meant to resolve potential identifiability issues between natural fluctuations and inflation in the SRB. In addition, there are limitations specific to the scenario-based projections presented in this study. Firstly, there is substantial uncertainty associated with the inflation predictions for countries prior to observing country-specific transition data due to the variability across countries that have started their SRB transition. Secondly, information that is available for extrapolation sex ratio imbalance at birth to countries without existing SRB inflation is limited, because the sex ratio transition has been observed in only 12 countries where most inflation processes are incomplete. Subsequently, long-term projections up to 2100 are subject to additional uncertainty not captured in our approach while we took into account uncertainty in future fertility trajectories. Lastly, we used the United Nations WPP birth projections for computing missing female births when alternative demographic projections incorporating our SRB scenarios may have given slightly different figures due to the declining female population and its impact on future births.

Our findings indicate that under the conservative scenario S1, where sex ratio transitions are included only for 12 countries with strong statistical evidence of past or ongoing SRB inflation, the cumulative global female births deficits are projected at 7.3 million [2.2; 14.1] during 2018–2030 compared to 0.9 [0.0; 6.3] during the rest of the century. This projected deficit is smaller than the figure for 1970–2017—estimated at 45.0 million missing female births—and stems from the projected gradual recovery from the skewed SRB in China and India in the coming years. This projected deficit is also significantly lower than previous estimates that put the future number of missing female births in 2018–2030 at 14.8 million.[14] The use of our model signals a faster reduction of prenatal sex selection in these countries than previously anticipated.

The impact of the last phase in the sex ratio transition in the world remains nonetheless still large in absolute terms, with millions of predicted missing female births added to the already existing demographic deficit of women. The global figure is mostly driven by SRB trends in large countries, but SRB levels above 1.1 are also projected in other countries affecting a sizable share of annual births. Even under scenario S1, immediate actions are needed in countries with ongoing sex ratio transitions—notably Albania, Armenia, Azerbaijan, China, India, and Vietnam—to accelerate the return of the SRB back to normal levels and to mitigate the impact on these societies of the mounting surplus of men among adult populations and the resulting marriage squeeze.[9, 33]

We provided an additional set of scenario-based projections that quantify the effect of SRB inflation in 17 countries at risk but without strong evidence of existing SRB inflation. Most of these 17 countries are characterized by higher fertility and son preference, and their sex ratio transitions would happen later in this century, when fertility approaches replacement level. Scenario S2 leads to a cumulative number of missing female births at 24.8 [13.1; 44.1] million by 2100. The assumptions made for scenario S2 remain hypothetical as they depend on the projected effect of fertility decline during the future decades in countries such as Afghanistan, Egypt, Nigeria, Pakistan, or Tanzania. The projections under this scenario assume that these countries have the same inflation as observed so far. Yet, these projections offer useful simulations of the potential burden of future prenatal sex discrimination and highlight the need to anticipate the possible deterioration of the SRB in vulnerable countries if gender bias and son preference remain what they are today.

It should be added that projections in this paper are labeled as "scenarios" because the experience of countries with past or ongoing sex ratio transitions, mostly in eastern Asia and eastern Europe, does not have to be predictive of the future evolution of countries belonging to different social and cultural environments with different cultures, e.g., countries in sub-Saharan Africa. In this regard, we recommend using projections deriving from the conservative scenario S1—assuming no inflation for the 17 at-risk countries—for future population projections such as those conducted by the United Nations or the Global Burden of Disease study. While it is too early to project SRB inflation in countries where birth masculinity levels have so far remained constant, the scenario S2 helps visualize what might happen under the impact of prolonged fertility decline in these countries.

In view of the high level of heterogeneity in gender systems within large countries such as China, India, and Viet Nam, national-level projections still represent an imperfect tool to simulate future imbalances. The well-documented subregional diversity of SRB levels within countries should encourage the development of subnational projections in such settings to provide additional insights into the geographic concentration of missing female births and future excess male populations.[11, 34-36]

These findings underline the need to monitor SRBs in countries with son preference and to address the factors behind the persistence of gender bias in families and institutions. Measurement issues remain severe, notably in the absence of birth registration data, and regular efforts to estimate trends and differentials are required to monitor SRB dynamics in vulnerable countries. In addition, policies based on monitoring, advocacy campaigns, as well as direct and indirect measures to combat gender bias are required to slow down the rise of SRBs or to accelerate its decline.[37] A broader objective relates to the need to influence gender norms which lie at the core of harmful practices such as prenatal sex selection. This calls for broader legal frameworks to ensure gender equality.[1, 38]

Supporting information

Appendix is deposited to figshare repository. Available at:
https://doi.org/10.6084/m9.figshare.14101583.

Acknowledgements

We are grateful to Nicolas Todd for helpful comments and discussions. The views expressed herein 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.

Contributors

FC and LA developed the Bayesian statistical model. FC carried out the analysis, drafted the initial manuscript, and prepared the appendix. FC and PG oversaw database construction, assessed and compiled the database. CZG interpreted the results and its implications. All authors reviewed model results, edited the manuscript, and agreed on the final version of the manuscript.

Funding

Funding source was research grant R-608-000-125-646 from the National University of Singapore.

Competing interests

We declare no competing interests.

REFERENCES

1 Erken A. State of World Population 2020: Defying the practices that harm women and girls and undermine equality. UNFPA; 2020. Available at: https://www.unfpa.org/sites/default/files/pub-pdf/UNFPA_PUB_2020_EN_State_of_World_Population.pdf (accessed 31st Jan 2021).

2 Park CB, Cho NH. Consequences of son preference in a low-fertility society: imbalance of the sex ratio at birth in Korea. *Popul Dev Rev* 1995;1:59–84.

3 Attané I, Guilmoto CZ. Watering the neighbour's garden: The growing demographic female deficit in Asia. Paris: Committee for International Cooperation in National Research in Demography; 2007.

4 Guilmoto CZ, Hoàng X, Van TN. Recent increase in sex ratio at birth in Viet Nam. *PLoS One* 2009;4(2):e4624.

5 Lin TC. The decline of son preference and rise of gender indifference in Taiwan since 1990. *Demogr Res* 2009;20:377–402.

6 Goodkind D. Child underreporting, fertility, and sex ratio imbalance in China. *Demography* 2011;48(1):291–316.

7 Duthé G, Meslé F, Vallin J, Badurashvili I, Kuyumjyan K. High sex ratios at birth in the Caucasus: modern technology to satisfy old desires. *Popul Dev Rev* 2012;38(3):487–501.

8 Guilmoto CZ, Ren Q. Socio-economic Differentials in Birth Masculinity in China. *Dev Change* 2011;42(5):1269–96.

9 Guilmoto C. Sex imbalances at birth: Trends, consequences and policy implications. Thailand: UNFPA, United Nation Population Fund of Asia and the Pacific Regional Office; 2012.

10 Basten S, Verropoulou G. 'Maternity migration' and the increased sex ratio at birth in Hong Kong SAR. *Popul Stud* 2013;67(3):323–34.

11 Chao F, KC S, Ombao H. Levels and trends in the sex ratio at birth in seven provinces of Nepal between 1980 and 2016 with probabilistic projections to 2050: a Bayesian modeling approach. arXiv 2007.00437 [Preprint]. Jul 1, 2020 [cited 2021 Feb 22] https://arxiv.org/abs/2007.00437.

12 Sen A. More than 100 million women are missing. *New York Rev Books* 1990;37(20):61–6.
13 Chao F, Gerland P, Cook AR, Alkema L. Systematic assessment of the sex ratio at birth for all countries and estimation of national imbalances and regional reference levels. *Proc Natl Acad Sci U S A* 2019;116(19):9303–11.
14 Bongaarts J, Guilmoto CZ. How many more missing women? Excess female mortality and prenatal sex selection, 1970–2050. *Popul Dev Rev* 2015;41(2):241–69.
15 Guilmoto CZ, Tove J. The masculinization of births. Overview and current knowledge. *Population (Paris)* 2015;70(2):185–243.
16 Kaur R, editor. Too many men, too few women: Social consequences of gender imbalance in India and China. New Delhi: Orient BlackSwan; 2016.
17 Hudson VM, Den Boer AM. Bare branches: The security implications of Asia's surplus male population. Cambridge, MA: MIT Press; 2004.
18 Guilmoto CZ. Skewed sex ratios at birth and future marriage squeeze in China and India, 2005–2100. *Demography* 2012;49(1):77–100.
19 Vollset SE, Goren E, Yuan CW, et al. Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study. *Lancet* 2020;396(10258):1285–306.
20 United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2019 Revision; 2019. Available at: http://esa.un.org/unpd/wpp/Download/Standard/Population/ (accessed 31st Jan 2021).
21 US Census Bureau. International data base: Population estimates and projections methodology. 2013. Available at: https://www2.census.gov/programs-surveys/international-programs/technical-documentation/methodology/idb-methodology.pdf (accessed 31st Jan 2021).
22 Guilmoto CZ. The sex ratio transition in Asia. *Popul Dev Rev* 2009;35(3):519–49.
23 Chao F, Gerland P, Cook AR, Alkema L. Global estimation and scenario-based projections of sex ratio at birth and missing female births using a Bayesian hierarchical time series mixture model. *Ann Appl Stat*. Forthcoming 2021 [cited 2021 Feb 22] Available at: https://arxiv.org/pdf/2006.07101 (accessed 22nd Feb 2021).
24 Chao F, Gerland P, Cook AR, Alkema L. pnas.1812593116.sd01: SRB database. *figshare*; 2019. Available at: https://figshare.com/articles/dataset/pnas_1812593116_sd01_SRB_database/12764249/1 (accessed 22 Feb 2021).
25 Ševčíková H, Alkema L, Raftery AE. bayesTFR: An R package for probabilistic projections of the total fertility rate. *J Stat Softw* 2011;43(1):1–29.
26 Ševčíková H, Alkema L, Raftery AE, Fosdick B, Gerald P. bayesTFR: Bayesian Fertility Projection. R Package and documentation version 6.4-0. R-CRAN; 2019. Available at: https://cran.r-project.org/web/packages/bayesTFR/ (accessed 31st Jan 2021).
27 Chahnazarian A. Determinants of the sex ratio at birth: review of recent literature. *Soc Biol* 1988;35(3-4):214–35.
28 Helle S, Helama S, Lertola K. Evolutionary ecology of human birth sex ratio under the compound influence of climate change, famine, economic crises and wars. *J Anim Ecol* 2009;78(6):1226–33.

29 James WH. Behavioural and biological determinants of human sex ratio at birth. *J Biosoc Sci* 2010;42(5):587–99.

30 James WH, Grech V. A review of the established and suspected causes of variations in human sex ratio at birth. *Early Hum Dev* 2017;109:50–6.

31 Garenne M. Sex ratios at birth in populations of Eastern and Southern Africa. *South Afr J Demogr* 2004;9(1) 91–6.

32 Guilmoto CZ, Chao F, Kulkarni PM. On the estimation of female births missing due to prenatal sex selection. *Popul Stud* 2020;74(2):283–9.

33 Hesketh T, Xing ZW. Abnormal sex ratios in human populations: causes and consequences. *Proc Natl Acad Sci U S A* 2006;103(36):13271–5.

34 Chao F, Yadav AK. Levels and trends in the sex ratio at birth and missing female births for 29 states and union territories in India 1990–2016: A Bayesian modeling study. *Foundations of Data Science* 2019;1(2):177–96.

35 Guilmoto CZ. Mapping the diversity of gender preferences and sex imbalances in Indonesia in 2010. *Popul Stud* 2015;69(3):299–315.

36 Jiang Q, Ge T, Tai X. Change in China's Sex Ratio at Birth Since 2000: A Decomposition at the Provincial Level. *Appl Spat Anal Policy* 2019;13:547–74.

37 Rahm L. Gender-Biased Sex Selection in South Korea, India and Vietnam: Assessing the Influence of Public Policy. Cham, Switzerland: Springer International Publishing; 2020.

38 Gupta GR, Oomman N, Grown C, et al. Gender equality and gender norms: framing the opportunities for health. *Lancet* 2019;393(10190):2550–62.