

## Review



**Cite this article:** Pollet TV, Stoevenbelt AH, Kuppens T. 2017 The potential pitfalls of studying adult sex ratios at aggregate levels in humans. *Phil. Trans. R. Soc. B* **372**: 20160317.  
<http://dx.doi.org/10.1098/rstb.2016.0317>

Accepted: 22 May 2017

One contribution of 16 to a theme issue 'Adult sex ratios and reproductive strategies: a critical re-examination of sex differences in human and animal societies'.

**Subject Areas:**

behaviour

**Keywords:**

methodology, aggregation, inference, cross-national analyses

**Author for correspondence:**

Thomas V. Pollet

e-mail: [t.v.pollet@fsw.leidenuniv.nl](mailto:t.v.pollet@fsw.leidenuniv.nl)

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.3807307>.

## The potential pitfalls of studying adult sex ratios at aggregate levels in humans

Thomas V. Pollet<sup>1</sup>, Andrea H. Stoevenbelt<sup>2</sup> and Toon Kuppens<sup>2</sup>

<sup>1</sup>Department of Social and Organisational Psychology, Leiden University, Room number 2A37, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands

<sup>2</sup>Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands

TVP, 0000-0002-4210-935X

Human adult sex ratios have been studied extensively across the biological and social sciences. While several studies have examined adult sex ratio effects in a multilevel perspective, many studies have focused on effects at an aggregated level only. In this paper, we review some key issues relating to such analyses. We address not only nation-level analyses, but also aggregation at lower levels, to investigate whether these issues extend to lower levels of aggregation. We illustrate these issues with novel databases covering a broad range of variables. Specifically, we discuss distributional issues with aggregated measures of adult sex ratio, significance testing, and statistical non-independence when using aggregate data. Firstly, we show that there are severe distributional issues with national adult sex ratio, such as extreme cases. Secondly, we demonstrate that many 'meaningless' variables are significantly correlated with adult sex ratio (e.g. the max. elevation level correlates with sex ratio at US state level). Finally, we re-examine associations between adult sex ratios and teenage fertility and find no robust evidence for an association at the aggregate level. Our review highlights the potential issues of using aggregate data on adult sex ratios to test hypotheses from an evolutionary perspective in humans.

This article is part of the themed issue 'Adult sex ratios and reproductive decisions: a critical re-examination of sex differences in human and animal societies'.

## 1. Introduction

## (a) Adult sex ratios effects on human behaviour

Adult sex ratios (ASRs) have long been studied not just in evolutionary biology (e.g. [1–6]), but also in the social sciences (e.g. [7–10]). However, following the seminal book by Guttentag & Secord, entitled '*Too Many Women? The sex ratio question*' [11], there has been a resurgence in the study of the influence of ASR on human behaviour. Since its publication in 1983, there has been a growing literature across a broad range of disciplines, including epidemiology, political science, criminology, sociology, economics and demography, dealing with the role of ASR for behaviour (e.g. [12–18]), such as marriage patterns ('The marriage squeeze': e.g. [17,19]) and crime (e.g. [20]). Pedersen [21] reframed Guttentag & Secord's work into a sexual selection framework, based on Bateman [22] and Trivers [23], and also tied it to evolutionary psychology [24,25]. Individuals are predicted to alter their decisions and behaviour in response to local sex ratios in an adaptive fashion [26]. Local 'market dynamics' should influence individual behaviour in line with predictions from sexual selection theory. For example, marital stability is predicted to be lower as the 'availability of a pool of young sexually active women means that men who marry are closing off a reproductive opportunity' ([27]: 260). Similarly, teenage pregnancy and fertility rates are expected to be higher when there is an oversupply of women.

Currently, there is a broad range of empirical studies examining the effect of ASR by using outcomes at the individual level (e.g. [28–34]) or by

**Table 1.** Some examples of work on ASRs, the topic and the level of outcome and the level of ASR.

reference no.	year	topic	outcome level	sex ratio level
[16]	2010	sexual behaviour and risk	individual	county or equivalent (China)
[21]	1991	family behaviour	country (USA)	country (USA)
[30]	2015	mating strategies (sociosexuality)	individual	community (Guyana)
[32]	2010	romantic relationships/sexual behaviour	individual	university campus (USA)
[36]	2012	career aspiration	individual	individual (perception in vignette)
[40]	2000	violent crime	country	country
[41]	2009	violent crime	country	country
[42]	2003	single parenthood	country	country
[43]	2009	sociosexuality, sex drive, height	country	country
[44]	2013	premature gestation/birth weight	county (USA)	county (USA)
[45]	2014	assault rates	census tract (USA)	census tract (USA)
[46]	2017	violence against women	societies (standard cross-cultural sample)	societies (standard cross-cultural sample)
[47]	2015	violence and crime	county (USA)	county (USA)
[48]	2015	preventable death	individual	ward (Northern-Ireland)
[49]	2009	marriage	metropolitan statistical areas (USA)	metropolitan statistical areas (USA)
[50]	2013	birth rate	ward (UK)	ward (UK)
[51]	2012	mate competition	hunter–gatherer populations	hunter–gatherer populations
[52]	2008	homicide	country	country
[53]	2008	marriage	individual	state (USA)
[54]	2015	sociosexuality	individual	state (USA)
[55]	2016	mating	individual	individual (perception in vignette)

mathematical modelling (e.g. [35]). Some psychological studies have also used vignettes to describe imbalances in ASR (e.g. [36–38]) or manipulated cues to sex ratios [39] and examining how individuals respond. Table 1 gives an incomplete list of examples of the levels at which ASR have been studied. Table 1 shows that there are also many papers relying solely on analyses with aggregate data (e.g. [21,40,41]). Here we focus on those papers relying (exclusively) on aggregate data. By aggregate data, we refer to data pooled to some group or area level, such as nations, wards, provinces, districts or cities. These analyses on aggregate data are routinely used to support the argument that ASR influence human behaviour (e.g. [42–46,56,57]). The majority of studies with aggregate data rely on national level data, but some rely on smaller spatial units [47–50,58–60], such as states, regions, counties, wards, cities, metropolitan areas or even hunter–gatherer societies [51]. We focus on how known problems with country-level data affect the conclusions we can draw about the effect of ASR on human behaviour. Importantly, we extend this analysis to aggregation at lower levels such as US states and counties, and demonstrate that these problems exist in the same way for data aggregated at these lower levels. Finally, we provide new analyses on the relationship between ASR and teenage fertility. Before we proceed, we will review some of the findings on ASR and human behaviour derived from analyses on aggregate data.

### (b) Aggregate-level analyses of adult sex ratio

Several studies have now investigated the role of ASR at the national level with a range of psychological and behavioural variables, mostly in a cross-sectional design. Perhaps the most extensively studied relationship, not just in evolutionary psychology, is the relationship between ASR and various indicators of violent crime, such as homicide [40,41,61] and rape [62]. Surprisingly, ASR have been claimed to be positively *and* negatively related to crime (reviews in [60,63]). Hudson & den Boer [20] argued for a positive relationship between ASR and crime via use of case studies covering Asia. However, the majority of studies have actually suggested negative associations between ASR and crime at the national level [40,41,52,64–66] (reviews in [63,67]). For example, using a sample of a 100 countries, Antonaccio & Tittle [68] reported  $r = -0.32$  between homicide (logged) and ASR (logged). More importantly, a meta-analysis covering all analyses on ASR and crime at the national level found no noteworthy relationship [69]. This meta-analysis also highlighted that substantial heterogeneity exists in the effect of ASR. Such heterogeneity could be partially due to the extensive methodological differences between studies. Firstly, the sample sizes vary considerably between studies, with some using samples of around 30 countries (e.g. [66]), while others used 100 or more (e.g. [68]). Secondly, the meta-analysis also indicated that the source of the data could contribute to the heterogeneity, even to the extent

that the sign of the sex ratio effect could be reversed depending on the source of data used (WHO: negative, UN and Interpol: positive). Finally, there is considerable variation in analytical choices across these papers (also see [63,67]). For example, some authors chose to transform the homicide data but not ASR (e.g. [70]), others transform both the homicide data and ASR (e.g. [67]), others weigh the data (e.g. [71]), or perform weighted least-squares analysis [52]. Even within studies analytical decisions can matter. To illustrate: one study reported a *positive* but non-significant zero-order correlation ( $r = 0.094$ ) between the natural logarithm of homicide and ASR, but found a *negative* and significant association in a weighted least-squares analysis when controlling for other factors (e.g. the Human Development Index) [70]. Analytical decisions such as the choice of technique, the sample, or the inclusion of covariates thus matter dramatically for the conclusions one draws.

Evolutionary psychologists have predominantly studied ASR and variables relating to mate preferences, such as sociosexuality [72], the degree to which individuals are willing to engage in sexual activities without being committed to a relationship (e.g. [56,73]). However, other research in this area suggested that national ASR are *not* significantly related to sociosexuality and sex differences therein [43]. Some studies moved beyond mate preferences and related national ASR to indices more directly relevant to behaviour, in particular marriage and fertility. In line with Pedersen's predictions [21], Barber showed that ASR were related to teenage pregnancies, indices of single parenthood and non-marital fertility [27,40,42].

Theoretically, the reasoning is that when one sex is scarcer, it will be driving 'the demand' for certain traits (e.g. [2,11,53,65,74]). For example, if there is an oversupply of men, a market logic would suggest that women would desire higher status in potential marriage partners. Conversely, following Guttentag & Secord [11], a low sex ratio, i.e. relatively more women, would lead to a decline in marital stability, such as higher divorce rates and greater non-marital fertility [21,27,56]. Barber extended this logic and in a series of studies showed that teenage fertility rates were negatively associated with ASR in aggregated cross-national analyses [27,56]. He argued that a low ASR, i.e. relatively more women, is associated with a 'certain level of social disintegration', and would be associated with higher teenage pregnancy rates. While he employed several statistical controls (e.g. latitude), the analyses did not account for spatial clustering of countries within regions (e.g. [75]). In this paper, we will present analyses that do control for clustering of countries. If the relationship between ASR and teenage fertility is substantially altered after accounting for such a spatial dependence, then the dynamics attributed to ASR are probably an artefact of broader regional dynamics (or unmeasured confounds). We elaborate on this point below. We revisit the relationship between sex ratio and teenage fertility and perform analyses at cross-national, US state, and county levels to illustrate the potential issues with such aggregate-level analyses.

While some research in this area has relied on multilevel models [54] or has presented both aggregated and multilevel analyses (e.g. [48]), analyses based exclusively on aggregate levels remain fairly common, at least in evolutionary psychology (e.g. [44,57,76–78]). Importantly, such findings based on aggregated data then form the basis for experimental studies [37,55] and theorizing [26,79,80]. More recent analyses have sometimes used a lower level of aggregated analysis (such as

states or metropolitan areas). One aim of this paper is to investigate whether problems with nation-level analyses also apply to aggregated analyses at a lower level of aggregation (US states), or whether these analyses avoid such problems.

### (c) Issues with aggregated, cross-cultural data

Before moving on to our analyses on ASR we discuss some other well-established issues with cross-cultural analyses [75,81–85]. These issues concern, but are not limited to, reversal-like phenomena after data pooling and cross-cultural measurement (equivalency). Several of the points we argue for below do not solely apply to ASR but are broader issues [81]. Nonetheless it is worthwhile reviewing them again here.

#### (i) Data pooling

One issue is that pooling data to aggregate levels can completely change relations between variables. It is well-established that reversal-like phenomena can occur when pooling data (e.g. [81,86–91]): a statistical relationship is found to be reversed when analysed at an aggregate instead of at a disaggregated level. A famous example demonstrated that one could find negative associations between the proportion of immigrants and illiteracy rates at US state level: states with higher proportions of immigrants had lower levels of illiteracy [87] (also see [92,93]). However, at the individual level the relationship was reversed: immigrants were more likely to be illiterate than 'native borns'. If we were to infer based on the aggregated data that a similar relationship would exist at a lower level, we would commit 'the ecological fallacy' [75,81,87,94]. The key point, as demonstrated multiple times across a variety of disciplines (e.g. [90,95–98]), is that a pattern observed at a lower analytical level can run opposite to that of a higher analytical level (or vice versa).

This issue straightforwardly applies to analyses on ASR at an aggregate level: there appears to be a fundamental mismatch between the level of statistical analysis and the level of inference based on theory. Note that while ASR is an aggregated area-level variable, seldom the variable of theoretical interest that we want to relate to ASR is an aggregate. Researchers typically have predictions based on how *individuals* modify their decisions or behaviour in response to ASR shifts (e.g. [27]), not necessarily about how country-level indices modify in response to sex ratio shifts. Rather the changes in a country's measure are seen as reflective of changes at the individual level in line with the proposed theory. As a case in point, some researchers have argued for a positive association between measures of aggression and ASR at a national level (e.g. [20]). However, the theoretical arguments for such a relationship seem to have been largely based on *individual-level* data: given that men are primarily the aggressors [60,99], we could expect that this relationship would extend to cities, districts, states, provinces and nations. Thus, given the existence of the relationship at an individual level, we might therefore assume that at aggregated levels (such as the national level) a similar positive relationship should exist, with more men equalling more violence, as some authors have indeed done (e.g. [20], review in [63]). However, much evidence shows that one should *not* generalize from the individual to aggregate level and that the relationship could be different at different levels of analysis. For example, while some studies at state level found a positive relationship

between ASRs and violence (in India: [100]), a study by Schacht and colleagues found a *negative* relationship at county level [47]. This would run counter a ‘more men, more violence’ explanation (e.g. [20]), which is derived at the individual level.

In order to fully understand the impact of ASR on behaviour at the individual level, researchers thus truly need a multilevel approach, incorporating data that matches the level at which theories are formulated. If a researcher is primarily interested in behaviour of individuals, e.g. how does a surplus of men influence mate preferences (e.g. [73]) or reproductive timing (e.g. [27]), then this analytical level should be present in the data analysis. Perhaps for some questions even longitudinal data within individuals are required [81,101]. For example, if we theorize that shifting sex ratios should lead to changes in psychological variables relating to mate preferences, then country-level associations, as they are based on aggregates, are less informative than a design that tests intra-individual change [75,101].

## (ii) Measurement issues

When conducting any study across cultures, issues with cross-cultural measurement should be considered [82,102–104]. Is the variable measured in the same way across these cultures? Intuitively, it might seem that an ‘objective’ measure such as ASR should not be substantially influenced by cross-cultural differences in measurement. Yet, it is possible that even in ‘objective’ government data one sex is consistently underreported in a country’s vital statistics. For example, systematic deviations in sex ratio at birth statistics have been observed in some former U.S.S.R. countries following its collapse [105,106]. Importantly, if there is a systematic deviation in registration of births/deaths and this is related to another variable of interest, say for example, life expectancy [107], then this could lead to a spurious association between ASR and that variable. In short, even in ‘objective’ data, such as vital statistics, there could be profound and systematic measurement issues.

In this paper, we demonstrate further potential issues with analysing aggregate-level data, specifically applied to the case of ASR. We present analyses at the national level but importantly also present US state-level and county-level data in order to investigate whether at these lower levels similar issues exist as at the national level. The issues we examined were the following. Firstly, we examine the distribution of the ASR variable and the existence of extreme values in ASR. Next, we explore the benchmark of statistical significance often employed in this research. Specifically, we examine how many statistical associations would be significant if we naively correlated ASR with a large number of other variables at national and US state level. Finally, we revisit a previously studied variable, teenage fertility rates and present results that challenge previous findings.

## 2. Methods

Our analyses consist of secondary analyses of country- and state-level databases.

### (a) Country-level database

We compiled a database with a broad range of variables measured at country level. These measures are as diverse as for

example: CO<sub>2</sub> emissions, number of McDonalds restaurants, the number of Olympic medals won, the number of psychiatrists per 100 000 people, the percentage of households with a fixed telephone line, the percentage uptake of contraception among women (15–49), the number of endangered mammals, the number of people incarcerated per 100 000 citizens, the percentage of Christians, when the country officially started, et cetera. These 110 variables are summarized in electronic supplementary material, table S1. The variables in our dataset were compiled from various sources, as we explicitly wanted to collect a diverse range of variables. The dataset we used is far from perfect, but its imperfections reflect those found in the kind of data used in published papers in this field. For example, several of our variables have a large number of missing cases, and some are measured at different time points or are collapsed across periods. These issues are common in the field (e.g. [27,40,66]).

The key variable of interest is the ASR and is derived from the *CIA World Factbook 2012* [108], though we also included two 2016 sex ratio measures in our database. In a previous version of this paper, we used the sex ratio measure as based on the CIA’s definition at the time, using a 15–64 age cut-off, with 1 being an equal ratio and a higher score indicating more men than women. This variable is fairly complete in the 2012 *Factbook* and is temporally close to most of the variables in our set, which is why we preferred it to the more recent measures. However, such ratio measures tend to have peculiar statistical properties [109,110], such as right skew and leptokurtosis. We therefore calculated the sex ratio as a proportion ( $N_{\text{men}} / (N_{\text{men}} + N_{\text{women}})$ ), as recommended by Ancona *et al.* in this issue [110] (15–64 year olds). This was calculated based on the population measures reported in [108].

For the additional analyses on teenage fertility, we also examined the effect of sex ratio (15–17) on teenage fertility (15–19) from the United Nations World Population Prospects (2010–2015) [111], as proportions. Note however that these measures are less complete than the 2012 CIA measure and are based on (estimated) averages for 2010–2015. The methodology is described at length in [112].

### (b) State-level database and county level database for United States

We also compiled a database with a broad range of variables measured at US state level. As with the country-level data, we aimed for a diverse set, and our set has variables as diverse as, for example, the number of rollercoasters per state, the maximum elevation level, the number of Amish, and the number of building permits. The 50 variables and their sources are listed in electronic supplementary material. ASR (15–64) were calculated based on the US Census of 2010 [113], as described above [110]. The county level database was composed to specifically examine the relationship between ASR and teenage fertility.

### (c) Analyses

All analyses were run in R v. 3.3.1 [114] and we included both the data and script with our paper (electronic supplementary material). The key analyses consist of Pearson correlations and a series of OLS regressions, as these are the analyses commonly used with these aggregate data (e.g. [27,77,78]). The purpose of these analyses is to illustrate multiple limitations of such analyses on ASR. Previously use of *p* value testing with aggregate data has been criticized [115]. We are fully aware of the limits of a null hypothesis significance testing approach (also see [116]). However, for the purpose of our paper, we will report *p* values as they remain commonly reported in this research area, and we want to empirically examine how ‘easy’ it is to obtain statistically significant results when analysing ASR at national and state level.



### 3. Results

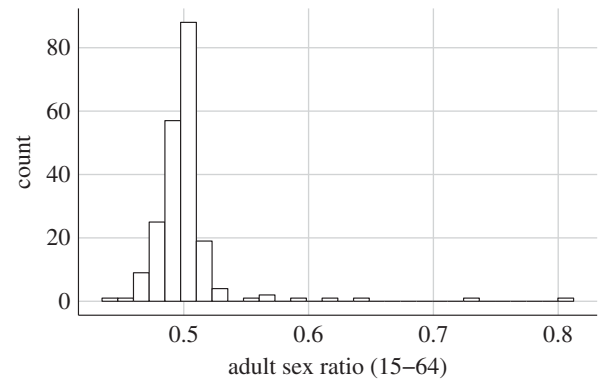
#### (a) Distributional issues: adult sex ratios at the national level have issues

Figure 1 shows the histogram for ASR at country level. We find evidence for deviations in skew and kurtosis (5.58 and 44.59) and thus non-normality (Shapiro–Wilk  $W = 0.52$ ,  $p < 0.00001$ ; Jarque–Bera test = 16372,  $p < 0.00001$ ). While several authors do not transform the data (e.g. [43]), some authors have suggested to take the natural logarithm (e.g. [40,41]) to combat these issues. However, there is still evidence for substantial deviations from normality after such a transformation (skewness: 4.53, kurtosis: 33.21; Shapiro–Wilk  $W = 0.61$ ,  $p < 0.00001$ ; Jarque–Bera test = 8785.3,  $p < 0.00001$ ). At US state level, there are only limited deviations from normality in ASR (skew and kurtosis: 0.55 and 4.27; Shapiro–Wilk  $W = 0.964$ ,  $p = 0.13$ ; Jarque–Bera test = 6.05,  $p = 0.049$ ). The degree to which distributional issues matter could thus be dependent on the level at which aggregation occurs.

Next to deviations from a normal distribution, outliers could influence any statistical conclusions we draw. We examined the distribution of ASR using Tukey's approach [117], relying on the interquartile range (IQR), with outliers defined according to the  $1.5 \times \text{IQR}$  rule and extreme values defined via the  $3 \times \text{IQR}$  rule. We found six outliers on the left side of the distribution (Antigua and Barbuda, Chad, Curaçao, Djibouti, Northern Mariana Islands, and the Virgin Islands (US) and 10 outliers on the right of the distribution (Bahrain, Faeroe Islands, Greenland, Kuwait, Maldives, Oman, Palau, Qatar, Saudi Arabia and the United Arab Emirates). Nine of these 16 outliers were also extreme values, eight on the right side of the distribution (Bahrain, Kuwait, Maldives, Oman, Palau, Qatar, Saudi Arabia and the United Arab Emirates) and one on the left side of the distribution (Djibouti). The two most extreme cases, Qatar and the United Arab Emirates, were respectively 9.03 and 6.86 standard deviations away from the mean. Note that if we systematically deleted extreme cases, this would be distinctly non-random and we would then remove several 'oil states' in the Middle East. Any inclusion or exclusion thus has severe consequences on how representative the data are. Figure 2*a,b* illustrates the dramatic effect of excluding these extreme values on the global distribution of ASR. We will therefore report analyses on national ASR with and without these 9 extreme values. The dataset with US states showed two outliers in ASR with the  $1.5 \times \text{IQR}$  rule (Alaska (ASR: 0.53) and the District of Columbia (ASR: 0.48)). However, there were no extreme cases at US state level with the  $3 \times \text{IQR}$  rule.

#### (b) It is easy to find aggregate-level correlations with ASR

We correlated all variables in our national database with ASR, and then calculated the absolute median correlation and all  $p$  values (see figure 3). The (absolute) median correlation with ASR was 0.13. More importantly, 39 correlations out of 110 were statistically significant at  $p = 0.05$  (35.45% of tests significant). This would imply that if we are conservative, we should not pay too much attention to statistical significance because there are many such statistically significant



**Figure 1.** Histogram for adult sex ratio (15–64) at national level (CIA World Factbook data 2012).

correlations. Moreover, this suggests that any correlation with ASR roughly below 0.13 in absolute strength does not necessarily warrant our attention: many such correlations will exist. If we exclude the extreme values, we obtain qualitatively similar results (median (absolute)  $r = 0.09$ , 28 out of 110 correlations statistically significant at  $p = 0.05$ ).

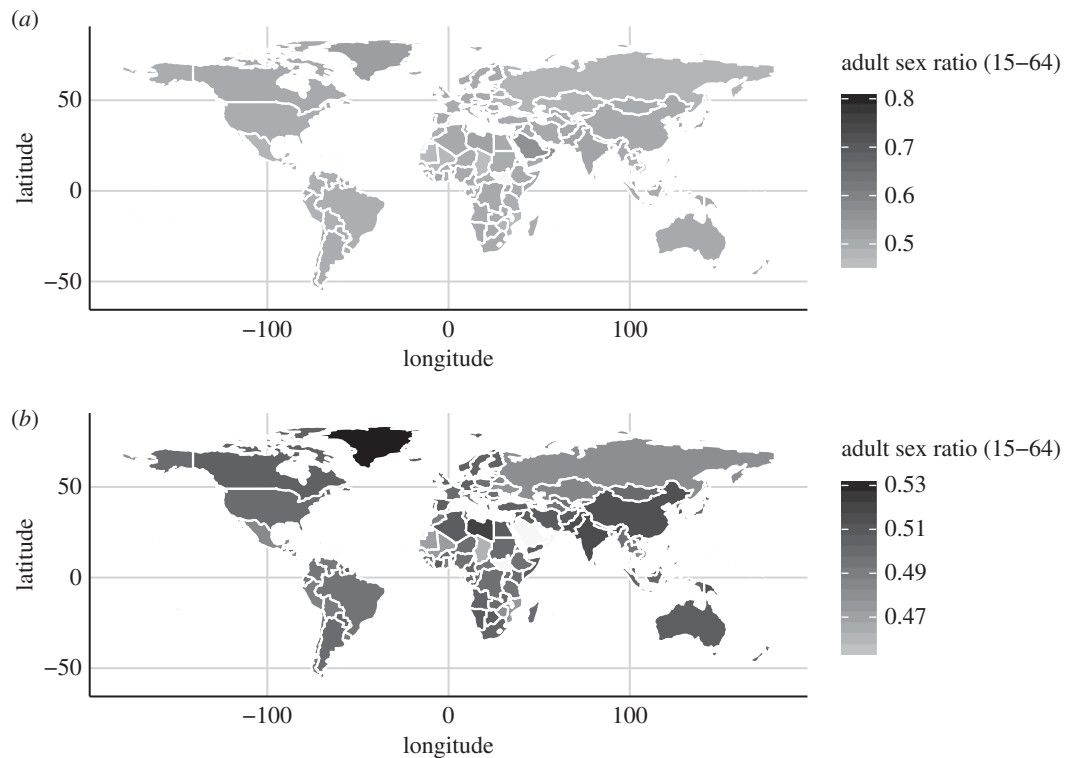
The top five correlations in terms of (absolute) strength were: CO<sub>2</sub> production in 2010, Oil use in 2011, Chocolate consumption in 2004, Military spending (% GDP) in 2013, and Population aged between 15 and 64 in 2013. However, exclusion of extreme values ( $3 \times \text{IQR}$  rule) gives a rather different top 5: Chocolate consumption in 2004, Contraception prevalence in 2010, Research & development expenditure in 2010, the % of malnutrition in 2010, and the % of teenage mothers in 2012. The fact that this top five looks quite different, points to the importance of outliers in ASR and exemplifies how volatile these statistical relationships are. For example, Military spending, ranked fourth without exclusion of extreme values, is now ranked 98th in terms of strength when excluding extreme values.

We repeated this exercise at US state level (see figure 4). The median (absolute) correlation with sex ratio was 0.22. Seventeen out of 50 correlations with sex ratio were statistically significant (34%). The top five correlations in terms of (absolute) strength were: Date of statehood, the Gini Coefficient (2010), mean travel time to work (2009–2013), the maximum level of elevation, and the land area of a state. Again, this suggests that many correlations with ASR will be statistically significant.

#### (c) Revisiting the link between adult sex ratios and teenage fertility at national level

Previous research has found a negative relationship between adult sex ratio and teenage fertility [27,56]. We focus here on teenage fertility, as another relevant variable, teenage pregnancy variable (from the World Bank) in our dataset only has data available on 14 countries. The teenage fertility variable we used is part of the World Development Indicators and is defined as the number of births per 1000 women ages 15–19. Teenage fertility was ranked 35th out of 110 and 25th out of 110 correlations, depending on inclusion or exclusion of extreme values. There is thus a wide range of variables showing stronger associations with ASR than teenage fertility did.

Our OLS regression results including all available data suggest a significant negative effect of ASR on teenage fertility (table 2: Model 1;  $F_{1,190} = 7.347$ ,  $p = 0.007$ ). The effect is



**Figure 2.** (a) Adult sex ratio (15–64) across the world (CIA World Factbook data 2012) (see electronic supplementary material for further description). (b) Adult sex ratio (15–64) across the world (CIA World Factbook data 2012) (after exclusion of extreme values, >3IQR rule) (see electronic supplementary material for further description).

upheld when extreme values are excluded (table 2: Model 3;  $F_{1,182} = 7.034$ ,  $p = 0.009$ ). However, after adding a crude control variable for spatial clustering (continent) there is no evidence for an effect ( $p$ -values for ASR greater than 0.11; table 2: Model 2 and 4).

The above analyses are based on ASR with a range of 15–64 years old, while the range of teenage fertility are based on those aged 15–19. Therefore, following the suggestion of a reviewer, we also analysed data from the UN (table 3). We tested the effect of sex ratios (15–17) on teenage fertility (15–19). While there was a strong association in a base model, there was again little suggestion of an association after accounting for continent (table 3; model 1–2). After excluding extreme values in sex ratios (3\*IQR rule), we obtained similar results (table 3; model 3–4). Using sex ratio based on an age range of 15–24, leads to similar conclusions (see electronic supplementary material). In sum, accounting for spatial clustering, even if in a very crude way, renders the association between sex ratio and teenage fertility non-significant. This means that the statistical relationship between teenage fertility and ASR is probably due to mechanisms at a higher level (regions or continents) or to unmeasured confounding variables.

#### (d) ASR and teenage fertility at US state and county level

State level analyses showed that there is no evidence for a significant association between sex ratio and teenage fertility in a base model ( $p = 0.838$ ; table 4: Model 1). County level analyses also showed no evidence of an association between sex ratio and teenage fertility in a base model ( $p = 0.109$ ; table 4: Model 2). There is thus no evidence that sex ratios (15–19) are related to teenage fertility at either the state or county

level in OLS regressions (both  $p > 0.1$ ). Interestingly, the effects in the respective base models, i.e. state and county, are *opposite* in sign. While the state-level analyses were in the same direction as the pattern in cross-national analysis, a negative relationship between sex ratio and teenage fertility, the county-level analyses showed the opposite, a positive relationship. Depending on the level of analysis, one would thus reach opposing conclusions on the direction of the effect of sex ratio. We also ran a multilevel model where counties were nested in states via a random intercept model fitted with Restricted Maximum Likelihood (lme4 [118]), but this also did not support a robust effect for ASR (95% CI based on 10 000 replicates in a percentile bootstrap:  $-7.85$  to  $95.39$ ; electronic supplementary material, R script for further details).

## 4. Discussion

In this paper, we discussed and illustrated some pitfalls to analysing ASR at aggregated levels and drawing inferences from such analyses. Firstly, we showed that, at least at the national level, ASR have a very peculiar non-normal distribution. Even after a logarithmic transformation the data would remain distinctly non-normal. There were also many extreme cases, with for example some countries being an order of magnitude more male biased (for example, the United Arab Emirates was around 7 s.d. away from the mean). The inclusion or exclusion of such extreme cases can have profound effects on any analysis and even a single extreme case could overturn a statistical relationship. Our recommendation is to report the results both with and without those extreme cases. For the US state level, there were no severe issues and therefore distributional issues should be reviewed on a case-by-case basis.

**Table 2.** The effect of adult sex ratio (15–64) on teenage fertility in OLS regression models. Model 1 contains all available data, Model 2 controls for continent. Model 3, excludes extreme values in adult sex ratio (>3IQR rule) and Model 4 further controls for continent. Reference category for continent is Europe.

teenage fertility				
	Model 1	Model 2	Model 3	Model 4
sex ratio (15–64)	–228.356 (84.249) $p = 0.008$	–93.707 (61.548) $p = 0.130$	–643.956 (242.804) $p = 0.009$	–273.621 (173.628) $p = 0.117$
North America		6.684 (20.404) $p = 0.744$		7.342 (20.302) $p = 0.719$
Latin America		44.668 (6.516) $p < 0.001$		43.509 (6.565) $p < 0.001$
Africa		78.555 (5.968) $p < 0.001$		79.388 (5.986) $p < 0.001$
Asia		18.612 (6.135) $p = 0.003$		19.391 (6.190) $p = 0.003$
Oceania		19.397 (9.020) $p = 0.033$		20.489 (9.025) $p = 0.025$
constant	164.551 (42.391) $p = 0.0002$	61.080 (31.030) $p = 0.051$	371.403 (120.677) $p = 0.003$	150.822 (86.723) $p = 0.084$
$N$	192	192	184	184
$R^2$	0.037	0.545	0.037	0.555
adjusted $R^2$	0.032	0.531	0.032	0.540
residual std. error	40.403 (d.f. = 190)	28.141 (d.f. = 185)	40.598 (d.f. = 182)	27.988 (d.f. = 177)
$F$ statistic	7.347** (d.f. = 1; 190)	36.970*** (d.f. = 6; 185)	7.034** (d.f. = 1; 182)	36.789*** (d.f. = 6; 177)

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

Second, our correlational analyses suggested that we should be wary of relationships between aggregated variables, as many variables might correlate with ASR without implying any substantive, theoretically meaningful relationship. For example, the level of elevation correlates significantly with ASR at US state level. Around a quarter to a third of the variables in our datasets had significant associations with ASR at national level. Importantly, also at the much smaller level of US states we found around a third of our variables to be associated with ASR. This means that problems with nation-level analyses are not necessarily resolved by moving to a lower level of aggregation. Therefore, we suggest that any finding of an association with ASR at an aggregate level should be treated with great caution.

Third, we argue that a reason for why we find many significant associations is probably due to the substantial (spatial) clustering in aggregated data. Such clustering violates the statistical assumption of non-independence of data points. Analytical units such as countries, states and counties

are non-independent data points and this fact is often not fully appreciated [75]. Such clustering is often not (fully) accounted for (e.g. [40,72,73,77]). We illustrated the importance of clustering by controlling for region (continent) in an analysis of the relation between sex ratio and teenage fertility. While the analyses of country-level UN data suggest a statistically significant association between sex ratios (15–17) and teenage fertility, after controlling for region we no longer found any evidence for such a significant association. It is important to indicate here that we are not claiming that the dummy covariate approach used here is novel (e.g. [71,119]), or that is the optimal way for dealing with spatial clustering. In fact, many analytical options exist, such as, for example, multilevel modelling [120–122] or a phylogenetic approach [123–125]. What this dummy approach demonstrates is that, if controlling for clustering, even in a very crude way by using a division into continents, substantially reduces the effect we should be wary of the conclusions we can draw. In our case, the effects were strongly reduced

**Table 3.** The effect of sex ratio (15–17) on teenage fertility in OLS regression models. Model 1 contains all available data, Model 2 controls for continent. Model 3, excludes extreme values in adult sex ratio ( $>3$ IQR rule) and Model 4 further controls for continent. Reference category for continent is Africa.

teenage fertility (2010–2015)				
	Model 1	Model 2	Model 3	Model 4
sex ratio (15–17)	–1309.521 (296.264) $p = 0.00002$	–257.612 (227.266) $p = 0.259$	–1452.711 (321.356) $p = 0.00002$	–281.114 (250.403) $p = 0.264$
Asia		–60.750 (5.673) $p < 0.001$		–61.174 (5.775) $p < 0.001$
Europe		–77.145 (6.147) $p < 0.001$		–77.137 (6.276) $p < 0.001$
Latin America		–33.077 (6.134) $p < 0.001$		–33.193 (6.210) $p < 0.001$
North America		–70.223 (20.364) $p = 0.001$		–70.175 (20.545) $p = 0.001$
Oceania		–57.449 (9.029) $p < 0.001$		–57.376 (9.183) $p < 0.001$
constant	715.502 (150.555) $p = 0.00001$	223.143 (114.512) $p = 0.053$	787.874 (163.217) $p = 0.00001$	235.160 (126.135) $p = 0.064$
$N$	194	194	191	191
$R^2$	0.092	0.550	0.098	0.548
adjusted $R^2$	0.088	0.535	0.093	0.533
residual std. error	39.393 (d.f. = 192)	28.112 (d.f. = 187)	39.487 (d.f. = 189)	28.316 (d.f. = 184)
$F$ statistic	19.537*** (d.f. = 1; 192)	38.061*** (d.f. = 6; 187)	20.436*** (d.f. = 1; 189)	37.215*** (d.f. = 6; 184)

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

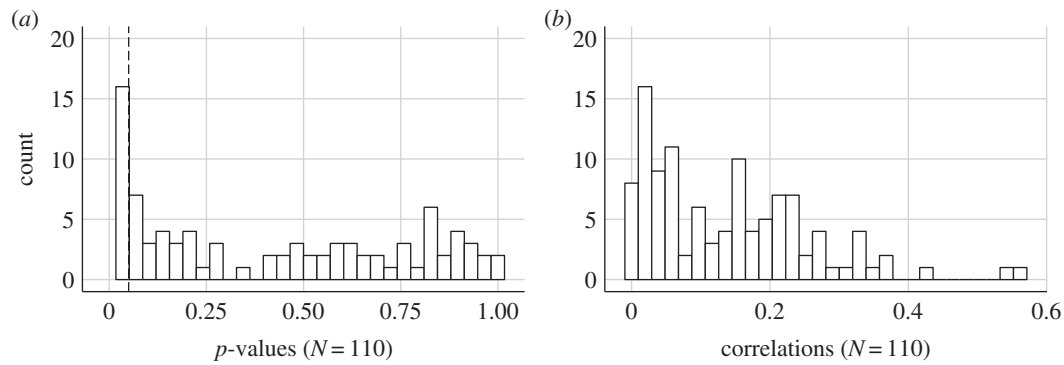
and no longer statistically significant after accounting for clustering. Without accounting for clustering, we are ignoring that a statistical relationship could be (partially) driven by non-independence. The acknowledgement of clustering based on geography/history also has profound implications for theory formation based on aggregate data. In the context of national data, it implies that at least part of the observed variation is not due to differences in national-level sex ratios, but rather differences in broader geographical units (or confounds at that level) [75].

Finally, we briefly revisited the relationship between ASR and teenage fertility rates at the national, US state and US county level. Across all levels, we found little evidence for a substantive association. While some of the associations at the national level could be seen as suggestive, once we controlled for continent, there no longer was any indication of a significant association. At US state level, there was no effect of ASR on teenage fertility in a regression model. At US county level, there was also no robust evidence for an ASR effect. Interestingly, the sign at county level was positive and thus at odds with previous

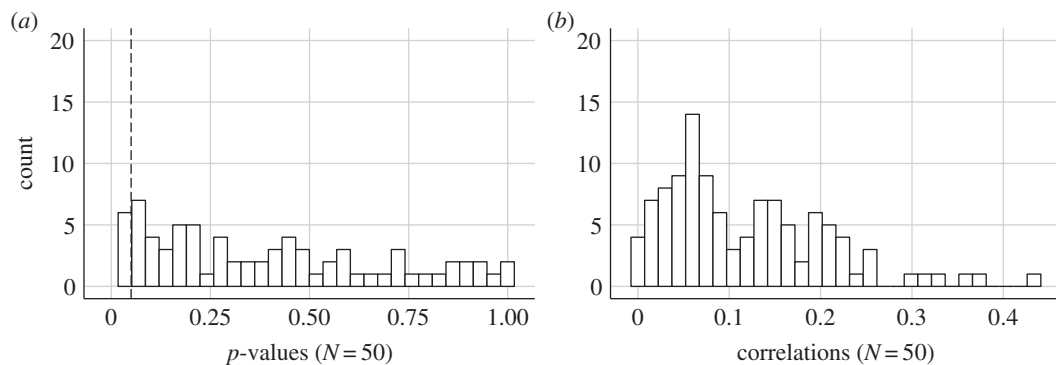
findings. There could be multiple reasons for the discrepancy between our findings and the literature. For example, it could be due to the measures we used, or it could be that the relationship no longer holds 10 years later. For now, we conclude that aggregate analyses offer no substantial support for predictions on sex ratios and teenage fertility, as derived in evolutionary psychology (e.g. [27,56]). Moreover, if anything, the data at a lower, county level suggest the opposite pattern. Rather than a low sex ratio, ‘too many women’, being positively associated with teenage fertility rates, the overall pattern was that high sex ratios were associated with high teenage fertility rates at US county level.

The analyses we presented have many limitations. Firstly, there are different ways to operationalize ASRs and age cut-offs in the calculation of sex ratios could matter a great deal [126]. We used an age range of 15–64 for the sex ratio variable from the *CIA Factbook* which has been previously used (e.g. [43,77]) for most of our analyses, but other operationalizations are perhaps somewhat narrower (e.g. 15–45 as used in [47] or 15–49 in [72]) and better map on to





**Figure 3.** (a)  $p$ -Values of 110 Pearson correlations with adult sex ratio at national level, all data (dashed line at 0.05). (b) Absolute value of 110 Pearson correlations with adult sex ratio at national level, all data.



**Figure 4.** (a)  $p$ -Values of 110 Pearson correlations with adult sex ratio at state level, all data (dashed line at 0.05). (b) Absolute value of 50 Pearson correlations with adult sex ratio at state level, all data.

**Table 4.** The effect of sex ratio (15–19) on teenage fertility in OLS regression models for US states (Model 1) and for US counties (Model 2).

	teenage fertility	
	Model 1	Model 2
sex ratio	–39.090	42.885
(15–19)	(189.666)	(26.760)
	$p = 0.838$	$p = 0.110$
constant	47.588	9.714
	(97.113)	(13.934)
	$p = 0.627$	$p = 0.486$
$N$	51	3138
$R^2$	0.001	0.001
adjusted $R^2$	–0.020	0.0005
residual std. error	8.559 (d.f. = 49)	41.318 (d.f. = 3136)
$F$ statistic	0.042 (d.f. = 1; 49)	2.568 (d.f. = 1; 3136)

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

‘operational sex ratio’. Worrisome for this area of research is that sex ratios might only correlate moderately with each other when using different age cut-offs (see [126] for a discussion with US census data). The recommendation, albeit based on US census data, is to use broader rather than narrower measures [126]. Therefore, we opted for the 15–64 cut-off,

at least for most of our analyses. However, many of the points we illustrated should similarly apply to other cut-offs or operationalizations of sex ratios. Finally, it is also important to reiterate that we focused on certain issues with analyses on aggregated ASR and did not illustrate other issues, such as cross-cultural measurement. This does not imply that we believe that these issues are of lesser importance and should therefore be ignored. Instead, we argue that issues such as cross-cultural equivalency in measurement further compound the issues we outlined here.

Our results add to existing knowledge that associations found at an aggregated level need not correspond to associations at a lower level [81]. We have demonstrated this with the analyses on teenage fertility: while across nations and across US states there is a negative association between sex ratio and teenage fertility, the relationship is entirely reversed and positive at county level. Our recommendation is that researchers pay close attention to the analytical level their theory is most relevant to and analyse the data at that level, avoiding inferences to other levels. For researchers in the field of evolution and human behaviour, it would seem that the individual level, or within-individual level, should be of more interest than aggregate units, such as nations [81,101]. We believe it could be time to abandon aggregate analyses on sex ratios altogether, in favour of multilevel analyses (e.g. [28,30,33,34,54]). Furthermore, in this context it is important to highlight that the purported effect of local sex ratio might be heavily influenced by perception, and that this perception could be of greater importance than the actual sex ratio (e.g. [36,39]). Interestingly, a recent study showed that there was no correspondence between actual neighbourhood sex ratio and perceived neighbourhood sex ratio [127]. A challenge for

the future will be to examine the psychological mechanism via which sex ratio effects occur and whether these are more or less important than other cues in one's local ecology.

In conclusion, we put forward that for researchers interested in how ASRs influence human behaviour, analyses on aggregate levels are not very informative. We therefore argue that at best these analyses on aggregates might be indicative of a relationship but ultimately these relationships are neither informative, nor do they allow for adequate tests of evolutionary theories on effects of sex ratios. Given the many serious issues with aggregate analyses on ASRs, some of which we illustrated here, we conclude that the study of ASRs exclusively at the aggregate level probably does not generate reliable findings for the field.

## References

- Mayr E. 1939 The sex ratio in wild birds. *Am. Nat.* **73**, 156–179. (doi:10.1086/280824)
- Emlen ST, Oring LW. 1977 Ecology, sexual selection, and the evolution of mating systems. *Science* **197**, 215–223. (doi:10.1126/science.327542)
- Hardy ICW. 2002 *Sex ratios concepts and research methods*. Cambridge, UK: Cambridge University Press.
- Kokko H, Jennions MD. 2008 Parental investment, sexual selection and sex ratios. *J. Evol. Biol.* **21**, 919–948. (doi:10.1111/j.1420-9101.2008.01540.x)
- Kokko H, Klug H, Jennions MD. 2012 Unifying cornerstones of sexual selection: operational sex ratio, Bateman gradient and the scope for competitive investment. *Ecol. Lett.* **15**, 1340–1351. (doi:10.1111/j.1461-0248.2012.01859.x)
- Liker A, Freckleton RP, Székely T. 2013 The evolution of sex roles in birds is related to adult sex ratio. *Nat. Commun.* **4**, 1587. (doi:10.1038/ncomms2600)
- Akers DSS. 1967 On measuring the marriage squeeze. *Demography* **4**, 907–924. (doi:10.2307/2060328)
- Cox OC. 1940 Sex ratio and marital status among Negroes. *Am. Sociol. Rev.* **5**, 937–947. (doi:10.2307/2084528)
- Schoen R. 1983 Measuring the tightness of a marriage squeeze. *Demography* **20**, 61–78. (doi:10.2307/2060901)
- Divale WT, Harris M. 1976 Population, warfare, and the male supremacist complex. *Am. Anthropol.* **78**, 521–538. (doi:10.2307/674415)
- Guttentag M, Secord PF. 1983 *Too many women? The sex ratio question*. London, UK: Sage.
- Angrist J. 2002 How do sex ratios affect marriage and labor markets? Evidence from America's second generation. *Q. J. Econ.* **117**, 997–1038. (doi:10.1162/003355302760193940)
- Abramitzky R, Delavande A, Vasconcelos L. 2011 Marrying up: the role of sex ratio in assortative matching. *Am. Econ. J. Appl. Econ.* **3**, 124–157. (doi:10.1257/app.3.3.124)
- Amuedo-Dorantes C, Grossbard S. 2007 Cohort-level sex ratio effects on women's labor force participation. *Rev. Econ. Househ.* **5**, 249–278. (doi:10.1007/s11150-007-9014-1)
- Hesketh T, Xing ZW. 2006 Abnormal sex ratios in human populations: causes and consequences. *Proc. Natl Acad. Sci. USA* **103**, 13 271–13 275. (doi:10.1073/pnas.0602203103)
- South SJ, Trent K. 2010 Imbalanced sex ratios, men's sexual behavior, and risk of sexually transmitted infection in China. *J. Health Soc. Behav.* **51**, 376–390. (doi:10.1177/0022146510386789)
- Trent K, South SJ. 2011 Too many men? Sex ratios and women's partnering behavior in China. *Soc. Forces* **90**, 247–267. (doi:10.1093/sf/90.1.247)
- Zhu WX, Lu L, Hesketh T. 2009 China's excess males, sex selective abortion, and one child policy: analysis of data from 2005 national intercensus survey. *BMJ* **338**, b1211. (doi:10.1136/bmj.b1211)
- Lichter DT, Anderson RN, Hayward MD. 1995 Marriage markets and marital choice. *J. Fam. Issues* **16**, 412–431. (doi:10.1177/019251395016004001)
- Hudson VM, Den Boer A. 2002 A surplus of men, a deficit of peace: security and sex ratios in Asia's largest states. *Int. Secur.* **26**, 5–38. (doi:10.1162/016228802753696753)
- Pedersen FA. 1991 Secular trends in human sex ratios. *Hum. Nat.* **2**, 271–291. (doi:10.1007/BF02692189)
- Bateman AJ. 1948 Intra-sexual selection in *Drosophila*. *Heredity* **2**, 349–368. (doi:10.1038/hdy.1948.21)
- Trivers RL. 1972 Parental investment and sexual selection. In *Sexual selection and the descent of man* (ed. B Campbell), pp. 136–179. New York, NY: Aldine de Gruyter.
- Symons D. 1979 *The evolution of human sexuality*. Oxford, UK: Oxford University Press.
- Buss DM. 1989 Sex differences in human mate preferences: evolutionary hypotheses tested in 37 cultures. *Behav. Brain Sci.* **12**, 1–49. (doi:10.1017/S0140525X00023992)
- Stone E. 2015 Do women compete for mates when men are scarce? In *The Oxford handbook of women and competition* (ed. ML Fisher). Oxford, UK: Oxford University Press.
- Barber N. 2001 On the relationship between marital opportunity and teen pregnancy: the sex ratio question. *J. Cross. Cult. Psychol.* **32**, 259–267. (doi:10.1177/0022022101032003001)
- Adimora AA, Schoenbach VJ, Taylor EM, Khan MR, Schwartz RJ, Miller WC. 2013 Sex ratio, poverty, and concurrent partnerships among men and women in the United States: A multilevel analysis. *Ann. Epidemiol.* **23**, 716–719. (doi:10.1016/j.annepidem.2013.08.002)
- Guzzo KB. 2006 How do marriage market conditions affect entrance into cohabitation vs. marriage? *Soc. Sci. Res.* **35**, 332–355. (doi:10.1016/j.ssresearch.2004.05.005)
- Schacht R, Borgerhoff Mulder M. 2015 Sex ratio effects on reproductive strategies in humans. *R. Soc. open sci.* **2**, 140402. (doi:10.1098/rsos.140402)
- South SJ, Lloyd KM. 1992 Marriage opportunities and family formation: further implications of imbalanced sex ratios. *J. Marriage Fam.* **54**, 440. (doi:10.2307/353075)
- Uecker JE, Regnerus MD. 2010 BARE MARKET: campus sex ratios, romantic relationships, and sexual behavior. *Sociol. Q.* **51**, 408–435. (doi:10.1111/j.1533-8525.2010.01177.x)
- Ugla C, Mace R. 2016 Local ecology influences reproductive timing in Northern Ireland independently of individual wealth. *Behav. Ecol.* **27**, 158–165. (doi:10.1093/beheco/arv133)
- Ugla C, Mace R. 2017 Adult sex ratio and social status predict mating and parenting strategies in Northern Ireland. *Phil. Trans. R. Soc. B* **372**, 20160318. (doi:10.1098/rstb.2016.0318)
- Del Giudice M. 2012 Sex ratio dynamics and fluctuating selection on personality. *J. Theor. Biol.* **297**, 48–60. (doi:10.1016/j.jtbi.2011.12.004)
- Durante KM, Griskevicius V, Simpson JA, Cantú SM, Tybur JM. 2012 Sex ratio and women's career choice: does a scarcity of men lead women to choose briefcase over baby? *J. Pers. Soc. Psychol.* **103**, 121–134. (doi:10.1037/a0027949)

37. Griskevicius V, Tybur JM, Ackerman JM, Delton AW, Robertson TE, White AE. 2012 The financial consequences of too many men: sex ratio effects on saving, borrowing, and spending. *J. Pers. Soc. Psychol.* **102**, 69. (doi:10.1037/a0024761)
38. Arnocky S, Ribout A, Mirza RS, Knack JM. 2014 Perceived mate availability influences intrasexual competition, jealousy and mate-guarding behavior. *J. Evol. Psychol.* **12**, 45–64. (doi:10.1556/JEP.12.2014.1.3)
39. Watkins CD, Jones BC, Little AC, DeBruine LM, Feinberg DR. 2012 Cues to the sex ratio of the local population influence women's preferences for facial symmetry. *Anim. Behav.* **83**, 545–553. (doi:10.1016/j.anbehav.2011.12.002)
40. Barber N. 2000 The sex ratio as a predictor of cross-national variation in violent crime. *Cross Cult. Res.* **34**, 264–282. (doi:10.1177/106939710003400304)
41. Barber N. 2009 Countries with fewer males have more violent crime: marriage markets and mating aggression. *Aggress. Behav.* **35**, 49–56. (doi:10.1002/ab.20291)
42. Barber N. 2003 Paternal investment prospects and cross-national differences in single parenthood. *Cross Cult. Res.* **37**, 163–177. (doi:10.1177/1069397103037002001)
43. Lippa RA. 2009 Sex differences in sex drive, sociosexuality, and height across 53 nations: testing evolutionary and social structural theories. *Arch. Sex. Behav.* **38**, 631–651. (doi:10.1007/s10508-007-9242-8)
44. Kruger DJ, Clark J, Vanas S. 2013 Male scarcity is associated with higher prevalence of premature gestation and low birth weight births across the United States. *Am. J. Hum. Biol.* **25**, 225–227. (doi:10.1002/ajhb.22369)
45. Kruger DJ, Aiyer SM, Caldwell CH, Zimmerman MA. 2014 Local scarcity of adult men predicts youth assault rates. *J. Community Psychol.* **42**, 119–125. (doi:10.1002/jcop.21597)
46. Stone EA. 2017 A test of an evolutionary hypothesis of violence against women: the case of sex ratio. *Lett. Evol. Behav. Sci.* **8**, 1–3. (doi:10.5178/lebs.2017.53)
47. Schacht R, Tharp D, Smith KR. 2016 Marriage markets and male mating effort: violence and crime are elevated where men are rare. *Hum. Nat.* **27**, 489–500. (doi:10.1007/s12110-016-9271-x)
48. Uggla C, Mace R. 2015 Effects of local extrinsic mortality rate, crime and sex ratio on preventable death in Northern Ireland. *Evol. Med. Public Health* **2015**, 266–277. (doi:10.1093/emph/eov020)
49. Kruger DJ, Schlemmer E. 2009 Male scarcity is differentially related to male marital likelihood across the life course. *Evol. Psychol.* **7**, 280–287. (doi:10.1177/147470490900700210)
50. Chipman A, Morrison E. 2013 The impact of sex ratio and economic status on local birth rates. *Biol. Lett.* **9**, 20130027. (doi:10.1098/rsbl.2013.0027)
51. Marlowe FW, Berbesque JC. 2012 The human operational sex ratio: effects of marriage, concealed ovulation, and menopause on mate competition. *J. Hum. Evol.* **63**, 834–842. (doi:10.1016/j.jhevol.2012.09.004)
52. Altheimer I. 2008 Social support, ethnic heterogeneity, and homicide: a cross-national approach. *J. Crim. Justice* **36**, 103–114. (doi:10.1016/j.jcrimjus.2008.02.002)
53. Pollet TV, Nettle D. 2008 Driving a hard bargain: sex ratio and male marriage success in a historical US population. *Biol. Lett.* **4**, 31–33. (doi:10.1098/rsbl.2007.0543)
54. Kandrik M, Jones BC, DeBruine LM. 2015 Scarcity of female mates predicts regional variation in men's and women's sociosexual orientation across US states. *Evol. Hum. Behav.* **36**, 206–210. (doi:10.1016/j.evolhumbehav.2014.11.004)
55. Moss JH, Maner JK. 2016 Biased sex ratios influence fundamental aspects of human mating. *Pers. Soc. Psychol. Bull.* **42**, 72–80. (doi:10.1177/0146167215612744)
56. Barber N. 2000 On the relationship between country sex ratios and teen pregnancy rates: a replication. *Cross Cult. Res.* **34**, 26–37. (doi:10.1177/106939710003400102)
57. Kruger DJ. 2009 When men are scarce, good men are even harder to find: life history, the sex ratio, and the proportion of men married. *J. Soc. Evol. Cult. Psychol.* **3**, 93–104. (doi:10.1037/h0099328)
58. Barber N. 2008 Explaining cross-national differences in polygyny intensity: resource-defense, sex ratio, and infectious diseases. *Cross Cult. Res.* **42**, 103–117. (doi:10.1177/1069397108314587)
59. Barber N. 2003 The sex ratio and female marital opportunity as historical predictors of violent crime in England, Scotland, and the United States. *Cross Cult. Res.* **37**, 373–392. (doi:10.1177/1069397103254011)
60. Messner SF, Sampson RJ. 1991 The sex ratio, family disruption, and rates of violent crime: the paradox of demographic structure. *Soc. Forces* **69**, 693–713. (doi:10.1093/sf/69.3.693)
61. Barber N. 2007 Evolutionary explanations for societal differences and historical change in violent crime and single parenthood. *Cross Cult. Res.* **41**, 123–148. (doi:10.1177/1069397106298926)
62. O'Brien RM. 1991 Sex ratios and rape rates: a power-control theory. *Criminology* **29**, 99–114. (doi:10.1111/j.1745-9125.1991.tb01060.x)
63. Schacht R, Rauch KL, Mulder MB. 2014 Too many men: the violence problem? *Trends Ecol. Evol.* **29**, 214–222. (doi:10.1016/j.tree.2014.02.001)
64. Chamlin MB, Cochran JK. 2006 Economic inequality, legitimacy, and cross-national homicide rates. *Homicide Stud.* **10**, 231–252. (doi:10.1177/1088767906292642)
65. Pridemore WA. 2011 Poverty matters: a reassessment of the inequality–homicide relationship in cross-national studies. *Br. J. Criminol.* **51**, 739–772. (doi:10.1093/bjc/azr019)
66. Savolainen J. 2000 Inequality, welfare state, and homicide: further support for the institutional anomie theory. *Criminology* **38**, 1021–1042. (doi:10.1111/j.1745-9125.2000.tb01413.x)
67. Trent CLS, Pridemore WA. 2012 A review of the cross-national empirical literature on social structure and homicide. In *Handbook of European homicide research*, pp. 111–135. New York, NY: Springer.
68. Antonaccio O, Tittle CR. 2007 A cross-national test of Bonger's theory of criminality and economic conditions. *Criminology* **45**, 925. (doi:10.1111/j.1745-9125.2007.00098.x)
69. Nivette AE. 2011 Cross-national predictors of crime: a meta-analysis. *Homicide Stud.* **15**, 103–131. (doi:10.1177/1088767911406397)
70. Pratt TC, Godsey TW. 2003 Social support, inequality, and homicide: a cross-national test of an integrated theoretical model. *Criminology* **41**, 611–644. (doi:10.1111/j.1745-9125.2003.tb00999.x)
71. Cole JH, Gramajo AM. 2009 Homicide rates in a cross-section of countries: evidence and interpretations. *Popul. Dev. Rev.* **35**, 749–776. (doi:10.1111/j.1728-4457.2009.00307.x)
72. Schmitt DP. 2005 Sociosexuality from Argentina to Zimbabwe: a 48-nation study of sex, culture, and strategies of human mating. *Behav. Brain Sci.* **28**, 247–275.
73. Stone EA, Shackelford TK, Buss DM. 2007 Sex ratio and mate preferences: a cross-cultural investigation. *Eur. J. Soc. Psychol.* **37**, 288–296. (doi:10.1002/ejsp.357)
74. Pollet TV, Nettle D. 2009 Market forces affect patterns of polygyny in Uganda. *Proc. Natl Acad. Sci. USA* **106**, 2114–2117. (doi:10.1073/pnas.0810016106)
75. Kuppens T, Pollet TV. 2014 Mind the level: problems with two recent nation-level analyses in psychology. *Front. Psychol.* **5**, 1110. (doi:10.3389/fpsyg.2014.01110)
76. Kruger DJ, Fitzgerald CJ, Peterson T. 2010 Female scarcity reduces women's marital ages and increases variance in men's marital ages. *Evol. Psychol.* **8**, 420–431.
77. Hassinger BE, Kruger DJ. 2013 The polygyny paradox: several male biased populations exhibit a high prevalence of polygyny. *J. Evol. Stud. Consort.* **5**, 131–137.
78. Schmitt DP, Rohde PA. 2013 The human polygyny index and its ecological correlates: testing sexual selection and life history theory at the cross-national level. *Soc. Sci. Q.* **94**, 1159–1184. (doi:10.1111/ssqu.12030)
79. Schmitt DP. 2005 Fundamentals of human mating strategies. In *Handbook of evolutionary psychology* (ed. DM Buss), pp. 258–291. New York, NY: John Wiley & Sons.
80. Cabeza de Baca T, Wahl RA, Barnett MA, Figueredo AJ, Ellis BJ. 2016 Adversity, adaptive calibration, and health: the case of disadvantaged families. *Adapt. Hum. Behav. Physiol.* **2**, 93–115. (doi:10.1007/s40750-016-0042-z)
81. Pollet TV, Tybur JM, Frankenhuys WE, Rickard IJ. 2014 What can cross-cultural correlations teach us about human nature? *Hum. Nat.* **25**, 410–429. (doi:10.1007/s12110-014-9206-3)
82. Byrne BM, Campbell TL. 1999 Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure. *J. Cross Cult.*

- Psychol.* **30**, 555–574. (doi:10.1177/00220222199030005001)
83. Van de Vijver FJR, Leung K. 1997 *Methods and data analysis of comparative research*. Boston, MA: Allyn & Bacon.
  84. Van de Vijver FJR, Leung K. 2000 Methodological issues in psychological research on culture. *J. Cross Cult. Psychol.* **31**, 33–51. (doi:10.1177/0022022100031001004)
  85. Apicella CL, Barrett HC. 2016 Cross-cultural evolutionary psychology. *Curr. Opin. Psychol.* **7**, 92–97. (doi:10.1016/j.copsyc.2015.08.015)
  86. Simpson EH. 1951 The interpretation of interaction in contingency tables. *J. R. Stat. Soc. Ser. B* **13**, 238–241. (doi:10.1038/168063d0)
  87. Robinson WS. 1950 Ecological correlations and the behavior of individuals. *Am. Sociol. Rev.* **15**, 351–357. (doi:10.2307/2087176)
  88. Tu Y.-K., Gunnell D, Gilthorpe MS. 2008 Simpson's Paradox, Lord's Paradox, and Suppression Effects are the same phenomenon—the reversal paradox. *Emerg. Themes Epidemiol.* **5**, 2. (doi:10.1186/1742-7622-5-2)
  89. Pollet TV, Stulp G, Henzi SP, Barrett L. 2015 Taking the aggravation out of data aggregation: a conceptual guide to dealing with statistical issues related to the pooling of individual-level observational data. *Am. J. Primatol.* **77**, 727–740. (doi:10.1002/ajp.22405)
  90. Gelman A, Shor B, Bafumi J, Park D. 2007 Rich state, poor state, red state, blue state: what's the matter with Connecticut. *Quart. J. Polit. Sci.* **2**, 345–367. (doi:10.1561/100.00006026)
  91. Pearce N. 2000 The ecological fallacy strikes back. *J. Epidemiol. Community Health* **54**, 326–327. (doi:10.1136/jech.54.5.326)
  92. Subramanian S. V, Jones K, Kaddour A, Krieger N. 2009 Revisiting Robinson: the perils of individualistic and ecologic fallacy. *Int. J. Epidemiol.* **38**, 342–360. (doi:10.1093/ije/dyn359)
  93. Te Grotenhuis M, Eisinga R, Subramanian SV. 2011 Robinson's ecological correlations and the behavior of individuals: methodological corrections. *Int. J. Epidemiol.* **40**, 1123–1125. (doi:10.1093/ije/dyn357)
  94. Freedman DA. 1999 Ecological inference and the ecological fallacy. *Int. Encycl. Soc. Behav. Sci.* **6**, 4027–4030.
  95. Yip PSF, Liu KY. 2006 The ecological fallacy and the gender ratio of suicide in China. *Br. J. Psychiatry* **189**, 465–466. (doi:10.1192/bjp.bp.106.021816)
  96. Rücker G, Schumacher M. 2008 Simpson's paradox visualized: the example of the Rosiglitazone meta-analysis. *BMC Med. Res. Methodol.* **8**, 34. (doi:10.1186/1471-2288-8-34)
  97. Connolly P. 2006 Summary statistics, educational achievement gaps and the ecological fallacy. *Oxford Rev. Educ.* **32**, 235–252. (doi:10.1080/03054980600645404)
  98. Scheiner SM, Cox SB, Willig M, Mittelbach GG, Osenberg C, Kaspari M. 2000 Species richness, species-area curves and Simpson's paradox. *Evol. Ecol. Res.* **2**, 791–802. (doi:10.1152/ajpendo.00279.2010)
  99. Daly M, Wilson M. 1988 *Homicide*. New Brunswick, NJ: Transaction Books.
  100. Dreze J, Khera R. 2000 Crime, gender, and society in India: insights from homicide data. *Popul. Dev. Rev.* **26**, 335–352. (doi:10.1111/j.1728-4457.2000.00335.x)
  101. Tybur JM, Frankenhuys WE, Pollet TV. 2014 Behavioral immune system methods: surveying the present to shape the future. *Evol. Behav. Sci.* **8**, 274–283. (doi:10.1037/ebs.0000017)
  102. Church AT. 2010 Measurement issues in cross-cultural research. In *The Sage handbook of measurement* (eds G Walford, E Tucker, M Viwanathan), pp. 151–177. London, UK: Sage.
  103. Hui CH, Triandis HC. 1985 Measurement in cross-cultural psychology. *J. Cross Cult. Psychol.* **16**, 131–152. (doi:10.1177/0022002185016002001)
  104. Poortinga YH. 1989 Equivalence of cross-cultural data: an overview of basic issues. *Int. J. Psychol.* **24**, 737–756. (doi:10.1080/00207598908247842)
  105. Duthé G, Badurashvili I, Kuyumjian K, Meslé F, Vallin J. 2010 Mortality in the Caucasus: an attempt to re-estimate recent mortality trends in Armenia and Georgia. *Demogr. Res.* **22**, 691–732. (doi:10.4054/DemRes.2010.22.23)
  106. Duthé G, Meslé F, Vallin J, Badurashvili I, Kuyumjian K. 2012 High sex ratios at birth in the Caucasus: modern technology to satisfy old desires. *Popul. Dev. Rev.* **38**, 487–501. (doi:10.1111/j.1728-4457.2012.00513.x)
  107. Dama MS. 2011 Sex ratio at birth and mortality rates are negatively related in humans. *PLoS ONE* **6**, e23792. (doi:10.1371/journal.pone.0023792)
  108. Central Intelligence Agency. 2012 *World Factbook* 2012.
  109. Atchley WR, Gaskins CT, Anderson D. 1976 Statistical properties of ratios. I. Empirical results. *Syst. Biol.* **25**, 137–148. (doi:10.2307/2412740)
  110. Ancona S, Dénes FV, Krüger O, Székely T, Beissinger SR. 2017 Estimating adult sex ratios in nature. *Phil. Trans. R. Soc. B* **372**, 20160313. (doi:10.1098/rstb.2016.0313)
  111. United Nations Department of Economic and Social Affairs Population Division. 2015 Revision of World Population Prospects.
  112. United Nations Department of Economic and Social Affairs Population Division. 2015 World Population Prospects The 2015 Revision Methodology of the United Nations Population Estimates and Projections. *Work. Pap. No. ESA/P/WP.242*.
  113. United States Census Bureau. In press. Census Explorer. <https://www.census.gov>.
  114. R Development Core Team. 2008 *R: a language and environment for statistical computing*.
  115. Pollet TV. 2013 Much ado about p. What does a p value mean when testing hypotheses with aggregated cross-cultural data in the field of evolution and human behavior? *Front. Psychol.* **4**, 734. (doi:10.3389/fpsyg.2013.00734)
  116. Anderson DR, Burnham KP, Thompson WL. 2000 Null hypothesis testing: problems, prevalence, and an alternative. *J. Wildl. Manage.* **64**, 912. (doi:10.2307/3803199)
  117. Tukey JW. 1977 *Exploratory data analysis*. Reading, MA: Addison-Wellesley.
  118. Bates D, Maechler M, Bolker B. 2012 lme4: Linear mixed-effects models using Eigen and Eigen++. *J. Stat. Softw.* **65**, 1–67. (doi:10.18637/jss.v065.b01)
  119. Hauser RM. 1970 Context and context: a cautionary tale. *Am. J. Sociol.* **75**, 645–664. (doi:10.1086/224894)
  120. Stegmüller D. 2013 How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *Am. J. Pol. Sci.* **57**, 748–761. (doi:10.1111/ajps.12001)
  121. Hox JJ. 2010 *Multilevel analysis: techniques and applications*, 2nd edn. London, UK: Taylor & Francis.
  122. Snijders TAB, Bosker RJ. 1999 *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. London, UK: Sage Publications Limited.
  123. Jordan F, Mace R. 2005 The evolution of human sex-ratio at birth: a bio-cultural analysis. In *The evolution of cultural diversity: a phylogenetic approach*, pp. 207–216. London, UK: UCL Press.
  124. Mace R, Jordan F, Holden C. 2003 Testing evolutionary hypotheses about human biological adaptation using cross-cultural comparison. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* **136**, 85–94. (doi:10.1016/S1095-6433(03)00019-9)
  125. Mace R, Pagel M. 1994 The comparative method in anthropology. *Curr. Anthropol.* **35**, 549–564. (doi:10.1086/204317)
  126. Fossett MA, Kiecolt KJ. 1991 A methodological review of the sex ratio: alternatives for comparative research. *J. Marriage Fam.* **53**, 941. (doi:10.2307/352999)
  127. Gilbert U, Juggla C, Mace R. 2016 Knowing your neighbourhood: Local ecology and personal experience predict neighbourhood perceptions in Belfast, Northern Ireland. *R. Soc. open sci.* **3**, 160468. (doi:10.1098/rsos.160468)