



## ORIGINAL ARTICLE

# Missing women in colonial India

James Fenske | Bishnupriya Gupta | Cora Neumann 

Department of Economics, University of Warwick, Coventry, UK

**Correspondence**

Bishnupriya Gupta, Department of Economics, University of Warwick, Coventry, UK.

Email: [B.Gupta@warwick.ac.uk](mailto:B.Gupta@warwick.ac.uk)

**Summary**

We construct novel data on female population shares by age, district, and religion in South Asia from 1881 to 1931 that show the sex ratio to be more male biased in the North compared with the South and East and its correlation across religious groups within districts. We find evidence of persistence over a century.

**Tweet**

Novel data on female population shares by age, district, and religion in South Asia from 1881 to 1931 show male bias in the North and correlation across religions within a district. The regional male bias is persistent over a century.

**Abstract**

We construct novel data on female population shares by age, district, and religion in South Asia from 1881 to 1931. Sex ratios skew male in Northern India and are more balanced in Southern and Eastern India, including Burma. Male-biased sex ratios emerge most visibly after age 10, and this is not specific to any one region, religion, or time period. Sikhs have the most male-biased sex ratios, followed by Hindus, Muslims, and Jains. The female share correlates across religious groups within districts. Evidence that sex ratios correlate with suitability for wheat and rice is weaker than suggested by the existing literature.

**KEYWORDS**

missing women, sex ratio, South Asia

In 1990, Amartya Sen contrasted the low ratios of women to men in South Asia, West Asia, and China with those in Europe and North America and estimated that 100 million women were ‘missing’.<sup>1</sup> Sen’s concern with sex ratios was not a new one. In the report of the Indian census of 1881, officials noted with concern the difference in the sex ratio in the population with the ratio that was found in most European countries. Females were in deficit in India compared with Europe. Where are the missing women? The census enumerators wrote that ‘[i]t was in the North of India the greatest divergence had been observed; whilst in the South especially, and less so in the East this variation became less and less conspicuous’.<sup>2</sup> They attributed the high

<sup>1</sup> Sen, ‘Missing women’.

<sup>2</sup> Report of Census of 1881, p. 51.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). The *Economic History Review* published by John Wiley & Sons Ltd on behalf of Economic History Society.

mortality among female children to lack of care rather than infanticide.<sup>3</sup> The literature has estimated the proportion of missing women for several Asian countries,<sup>4</sup> by region for India<sup>5</sup> and by age groups.<sup>6</sup> Evidence from recent decades suggests large increases in the ratio of male to female at birth, particularly in certain regions.<sup>7</sup>

The literature offers several explanations for both levels and increases in the male bias in the sex ratio, in South Asia and elsewhere. One explanation is a preference for sons that existed historically and has become accentuated with a decline in fertility and the availability of pre-natal sex selective technology.<sup>8</sup> The male bias in the sex ratio can also arise from lower bargaining power of women in post-reproductive ages and women's disadvantage in intra-household resource allocation.<sup>9</sup> Explanations of the preference for sons in South Asia and lower social status of women within the family include the relative productivity of women in certain types of agriculture,<sup>10</sup> and the roles of kinship norms and marital residence patterns.<sup>11</sup>

The validity of these explanations depends on both where women go missing and when women go missing, both over the life cycle and over time. [Anderson and Ray](#) compute the numbers of missing women by age cohort and show that 11 per cent of the deficit of women in India in 2000 was perinatal.<sup>12</sup> More than two-thirds of missing women went missing after the age of 15, and a large number went missing after the age of 70. Therefore, neither sex selection at birth nor mistreatment of young girls can be a complete explanation of India's current male-biased sex ratios.

Historical data allow us to look for explanations of missing women on the basis of variables that are historically persistent. These variables could include regional norms, religious norms, or the component of women's participation in economic activity that is determined by geographical factors.

In this paper, we introduce a new dataset on the basis of the censuses of colonial India and document both where and when women go missing. The censuses cover the years 1881, 1891, 1901, 1911, 1921, and 1931, and include the territories of present-day Bangladesh, Burma, India, and Pakistan. The dataset reports the percentage of the population that is female in each district in each census year, by religion, and for four broad age categories: 0–10, 10–20, 20–40, and over 40. The dataset reports these percentages at the level of districts between 1881 and 1931. To our knowledge, ours are the largest and most granular data that have yet been assembled on sex ratios by age, religion, and location for colonial South Asia.

We use these new data to understand where and when women went missing in the population a century ago. We describe variation in sex ratios by region, religion, geography, and age. We

<sup>3</sup> *Ibid.*, p. 52.

<sup>4</sup> [Das Gupta](#), 'Missing women'.

<sup>5</sup> [Agnihotri et al.](#), 'Missing women'.

<sup>6</sup> [Anderson and Ray](#), 'Missing women'.

<sup>7</sup> Censuses 1991–2011.

<sup>8</sup> [Das Gupta](#), 'Missing women'; [Jayachandran and Pande](#), 'Indian children'; [Pande](#), 'Gender differences'; [Jayachandran](#), 'Missing women'; [Basu](#), 'Fertility decline'; [Das Gupta and Bhat](#), 'Fertility decline'.

<sup>9</sup> [Calvi](#), 'Older women'.

<sup>10</sup> [Alesina et al.](#), 'Traditional agricultural practices'; [Bardhan](#), 'Life and death'; [Boserup](#), *Woman's role*; [Carranza](#), 'Soil endowments'; [Miller](#), *Endangered sex*; [Rosenzweig and Schultz](#), 'Market opportunities'; [Qian](#), 'Missing women'.

<sup>11</sup> [Dyson and Moore](#), 'Kinship structure'.

<sup>12</sup> [Anderson and Ray](#), 'Missing women'.



begin with the question of where women go missing by region. Broadly, the literature has classified India into four broad regions: the North, which encompasses regions such as the Punjab and Uttar Pradesh; the East, which includes regions such as Bengal, Assam, and also Burma, which was ruled as part of British India until 1937; the West, which includes Gujarat, Maharashtra, and the territory of the former Bombay Presidency; and the South, which includes the territory of the former British-ruled Madras Presidency, and the princely states of Hyderabad, Mysore, Travancore, and Cochin, today's Andhra Pradesh, Karnataka, and Kerala.

Using these data we have four key findings. First, by region, the well-known North–South divide in sex ratios is visible in the earliest census, and within-region variations in sex ratios persist to the present day. Sex ratios in colonial India as a whole are male biased, but there is significant regional variation. If we divide India along a diagonal that connects Surat on the western coast to Almora in the Himalayas, the Northern region has more male-biased sex ratios. The South and East are more balanced in their sex ratios. This pattern is consistent with the data from recent censuses. Late colonial sex ratios correlate strongly with sex ratios in the present, not only across broad regions, but even looking within states of modern India. Evidence of this persistence is strongest among children and adolescents. We thus confirm the regional divide recognized in past and recent studies and show persistence at a granular level.

Second, by religion, we show that sex ratios are most male biased for Sikhs and still skew male among Hindus, Muslims, and Jains, and that the share of the population that is female correlates across religious groups within a district. For example, if the Hindu cohort under age 10 in a district is more male, the Muslim cohort in the same district is more likely to skew male. In colonial India, sex ratios show similarities across religions both within regions and over the life cycle. This pattern differs from the evidence on son preference from recent decades and suggests that the difference between Hindu and Muslim sex ratios in the under 5 age group in the present is a result of greater use of sex selective abortions in Hindu communities.<sup>13</sup> The technology of sex selective abortion was not available in the colonial period.

Third, by geography, there are geographic variables that predict sex ratios, but these variables are not always those stressed by the literature. Past work has sought to explain the North–South divide in terms of factors such as female labour participation and the greater involvement of women in rice cultivation rather than wheat cultivation.<sup>14</sup> Spatial variation in the economic value of women has been attributed in the literature to women's participation in agricultural activity, such as hoeing, weeding, and planting. According to this view, in rice-growing regions, women's labour force participation is high and their economic value makes son preference less important. In the drier areas where millet is grown and in the wheat-growing regions, women's participation in economic activity is low. In these regions, preference for sons prevails. The broad regional divide between North, South, and East may therefore reflect a divide between rice and wheat cultivation. The literature provides additional reasons to anticipate that geography will correlate with sex ratios. In the industrial towns, the sex ratios were often male biased due to migration of men working in industrial jobs. Docklands were yet another preserve for male employment, as women were rarely employed in these areas. Men migrated, leaving their families in the village.<sup>15</sup> Therefore, coastal areas with ports may be expected to have a deficit of women in the population.

<sup>13</sup> Bhalotra et al., 'Religion and abortion'.

<sup>14</sup> Bardhan, 'Life and death'; Miller, 'Female labor'; Boserup, *Woman's role*.

<sup>15</sup> Ojha, 'Regulating work'; Chandavarkar, *Industrial Capitalism*.

In our data, we correlate the percentage of the population that is female with several observable geographic variables. Basic correlations show that, in various age ranges, the share of the population that is female is greater in Eastern districts and in districts suited to the growing of cotton, dryland rice, and wetland rice, whilst it is lower at higher latitudes and in districts suitable for the growing of wheat. However, once we perform regressions that account for the correlations between these variables, and include fixed effects that focus our comparisons within colonial provinces, these patterns are less pronounced. The evidence that wheat suitability predicts more male populations is weak, and wetland rice suitability is rarely a significant predictor of more female populations. Suitability for dryland rice, by contrast, is a robust predictor of more female populations in later census years. In sum, we provide systematic evidence of how geography may predict women's status in society, as reflected in the sex ratio. The rice–wheat distinction that is discussed widely in the literature is not statistically robust, a conclusion that is also apparent using data on actual crop cultivation.

Next we show when women went missing in the population by age. We find a drop in the share female among adolescents that is shared across regions and across religions. We compute age-specific sex ratios, which capture when women go missing over the life cycle. Recent literature finds a large number of missing girls at birth in some regions in present-day India. It is more difficult to obtain data on sex ratios at birth for districts in colonial India. Province-level evidence from the colonial censuses points to a male deficit before age 1 that becomes a female deficit by age 10.<sup>16</sup> In our data, we show that sex ratios, whilst already slightly male-biased in the 0–10 age cohort, become much more strikingly male in the 10–20 cohort before becoming more balanced again in the 20–40 age range. The dip in the share of females in the age group 10–20 is similar in all regions, and suggests that child marriage and maternal mortality may be a factor everywhere. It is also consistent with under-reporting and misreporting of girls in this age cohort over all four regions. This age profile is apparent in each census wave between 1881 and 1931, though it is more modest in 1931 than in earlier years. Despite the regional differences in sex ratios, we find that variation in the sex ratio by age shows similar patterns across all regions.

## I | CONTRIBUTION

We contribute to a literature on gender in economics,<sup>17</sup> in particular to a branch of this literature that has focused on South Asia.<sup>18</sup> More specifically, we contribute to the literature on the origins of gender norms, which includes work on the prevalence of plough-dependent agriculture<sup>19</sup> and cotton spinning,<sup>20</sup> among other factors. Because of data limitations, many of these papers have been compelled to 'compress' history, ignoring the evolution of the outcomes they

<sup>16</sup> Bhaskar and Gupta, 'Missing girls'.

<sup>17</sup> de Moor and van Zanden, 'Girl power'; Goldin, 'U-shaped'; Bordalo et al., 'Beliefs'; Croson and Gneezy, 'Gender differences'; Niederle and Vesterlund, 'Competition'; Behrman et al. 'Women's schooling'; BenYishay et al., 'Gender gaps'; Luke and Munshi, 'Female income'; Giuliano, 'Female labor'.

<sup>18</sup> Boserup, *Woman's role*; Carranza, 'Soil endowments'; Miller, 'Female labor'; Ashraf et al., 'Bride price'; Dyson and Moore, 'Kinship structure'; Lowes, 'Matrilineal'.

<sup>19</sup> Alesina et al., 'Gender roles'.

<sup>20</sup> Xue, 'Cotton revolution'.



study between the distant past and the present day.<sup>21</sup> By introducing new data that cover several points in time dating back to 1881, we are able to evaluate the degree to which one gender outcome – the sex ratio – has been persistent over more than a century within South Asia despite changes in life expectancy, the economic position of women, and the technology of birth control.

We also contribute to the literature on women's relatively poor health outcomes and disproportionate mortality over the life cycle in developing countries, including within South Asia.<sup>22</sup> Research on this theme in South Asia has found, for example, that girls receive fewer prenatal investments,<sup>23</sup> they are breastfed less,<sup>24</sup> a preference for eldest sons constrains investments in daughters' health,<sup>25</sup> and that the survival of girls is more sensitive to rainfall shocks.<sup>26</sup> We introduce historic data on sex ratios at a much finer geographic level than has been available previously for South Asia. This permits us to quantitatively re-examine several hypotheses about the determinants of male-biased sex ratios in the region that have been tested only in more recent data, or supported only by coarse regional comparisons. Our findings suggest two revisions to the literature on gender norms in India. First, differences in sex ratios across religions often vary substantially across regions, sometimes even changing rank order. Second, differences in son preference arising from rice and wheat agriculture are less pronounced than is generally claimed in the literature.

The remainder of this paper is as follows. In section II, we explain our underlying sources of census data and describe the new datasets we produce. In section III, we describe regional patterns in sex ratios and how they changed over time. In section IV, we discuss how sex ratios varied by religion across colonial India. In section V, we explore the geographic correlates of sex ratios. In section VI, we explore age-specific sex ratios to understand where women go missing over the life cycle. Section VII concludes.

## II | DATA

Where are the missing women in colonial India and when do they go missing over the life cycle? In this section, we describe our data sources.

We introduce a dataset based on the colonial Indian censuses of 1881, 1891, 1901, 1911, 1921, and 1931. The coverage of districts by the censuses increased over time. Therefore, our dataset shows an increasing number of districts over the census waves. We report at the district level the percentage of the population that is female in a given census year, including by sub-categories defined by four broad age bins – 0 to 10, 10 to 20, 20 to 40, and over 40 – and six religious categories – Buddhist, Hindu, Jain, Muslim, Sikh, and Tribal. Rather than working with the female to male ratio in a given age bin, we use the percentage of the population that is female in our analyses. We use percentages, rather than the sex ratio, because of the tendency of sex ratios to take extreme

<sup>21</sup> Austin, 'Compression of history'.

<sup>22</sup> Anderson and Ray, 'Missing women'; Sudha and Rajan, 'Female demographic'; Bhaskar and Gupta, 'Missing girls'.

<sup>23</sup> Bharadwaj and Lakdawala, 'Discrimination begins'.

<sup>24</sup> Jayachandran and Kuziemko, 'Mothers breastfeed'.

<sup>25</sup> Jayachandran and Pande, 'Indian children'.

<sup>26</sup> Rose, 'Female mortality'.



values, especially among small populations, creating skewed distributions. Readers interested in converting percentages to sex ratios can simply transform them as  $SexRatio = \frac{PercentFemale}{1-PercentFemale}$ .

Whilst the sex ratio in a given district, age bin, and religion depends on relative female mortality, and hence is an indicator of the degree to which women are missing, labour migration in some situations can also lead to sex imbalance when the work force in an industry has a strong sex imbalance. Workers in cotton mills in Bombay, for example, were disproportionately male.<sup>27</sup> Migration, including migration for marriage, is, however, unlikely to have a large influence on sex ratios in our sample. In 1881, at least 97 per cent of persons in India lived in the provinces in which they were born. In 1901, 91.3 per cent of Indians in India lived in their district of birth.<sup>28</sup> It is similarly unlikely that, except in a small number of identifiable cases of major cities, mines and penal colonies, urbanization, and structural change drive the patterns we observe in sex ratios. Urbanization was slow in India – whilst 8.7 per cent of the population lived in towns with more than 5000 persons in 1872, by 1931 this had risen only to 11.1 per cent.<sup>29</sup> Similarly, the dominance of agriculture changed little. In the 1881 Census Report (p. 350), approximately 77 per cent of workers were in agriculture.<sup>30</sup> The 1931 Census Report (p. 276) noted that 67 per cent of workers worked in the ‘exploitation of animals and vegetation’ – the category most closely corresponding to agriculture. Following the recent public health literature, we take sex ratios at the earliest ages as those that are least likely to be affected by other factors.<sup>31</sup>

The first attempt by the British to systematically count the population of India was undertaken between 1867 and 1872, but this census left out large portions of the country and was imprecise due to different regions being enumerated in different years.<sup>32</sup> Following this inadequate first attempt, British authorities created a system to accurately count the population once every 10 years, starting in 1881. Over time, more territories were censused.

To improve the process, in 1901 the country was divided into blocks and a large workforce was recruited, which was split into three levels: charge superintendents, supervisors, and enumerators. Enumerators were responsible for counting people in one block, or roughly 30–50 houses. They then reported their tables to the supervisors, who in turn combined these tables to send them to their charge superintendents, who then combined these tables and sent them to the census commissioner. The number of people involved in collecting the information was large. For example, for the 1901 Census, 1.3 million enumerators, 122 000 supervisors, and 9800 superintendents were employed. To count people by different categories, for example by religion, age, or marital status, in the first two census waves each person had been ticked into all their categories and then all the ticks were added up to calculate the population counts. Starting in 1901, however, each counted person’s categories were recorded on a slip, and then slips were counted together by category. The slips were colour coded for religion, and shapes and symbols were used to record sex and civil condition. This system reduced the frequency of making errors. During January and February of 1901, information from households was collected by the

<sup>27</sup> Morris, *Industrial labor*, p. 65.

<sup>28</sup> Collins, ‘Labor mobility’, p. 252.

<sup>29</sup> Visaria and Visaria, ‘Population’, p. 519.

<sup>30</sup> To arrive at this figure, we subtract 48 794 195 persons with unspecified or unstated occupations (a large portion of which are children) from 115 417 956 persons to reach an estimate of 66 623 761 workers. We divide 51 089 021 agriculturalists by this number.

<sup>31</sup> Guilmoto et al., ‘Female mortality’.

<sup>32</sup> We base this account on the Census Reports from 1881 to 1931.



enumerator in charge of the block. On March 1, this was checked and updated, and on March 2, the totals were calculated at the level of the block. Due to the time it took to combine all tables and count the different sub-groups, the whole process took roughly 6 months.

The colonial censuses are not perfect records of the population, disrupted as they were by non-cooperation, political instability, and even physical attacks on enumerators by both respondents and wild animals.<sup>33</sup> Most importantly for our analysis, census administrators and later demographers have both noticed errors in the reporting of women around adolescence: some had their ages misstated and some were unenumerated.<sup>34</sup> We note this particular problem below as one possible explanation of differences we find in sex ratios over the life cycle.

Despite these limitations, we do not believe data quality invalidates our comparisons by region and geography. First, we make no use of the 1941 census. That census was disrupted due to the Second World War, so much so that the census commissioner wrote that ‘the great population regions of the Indus and Ganges systems in which nearly half the total population of India lies have only a limited representation in the census figures’.<sup>35</sup> Past researchers have not used the 1941 census for precisely this reason.<sup>36</sup> Second, the persistence we document across census waves helps validate the data from earlier censuses, as sex ratios are correlated over time at both the province and district levels. Third, because major differences in enumeration methods were across provinces rather than within them, we show regression results that control for province fixed effects.

The dataset covers the census years between 1881 and 1931 and takes as the unit of analysis the set of districts that exists in any given census year, reflecting both changes in the set of districts that were censused and the set that existed. The census coverage of districts becomes more complete over time. Much of Northern Burma, for example, is not covered in 1881 and 1891. Similarly, the entirety of the Central India Agency is reported as a single unit in the 1881 census, whilst the separate parts of the agency that have their own entries in later years differ across years. Coverage of the population of the Andaman and Nicobar Islands prior to 1921 includes only the penal colony of Port Blair.

Adding to these differences in coverage, districts were sometimes split (as with the division of Khandesh into East and West Khandesh after 1906) and other times reorganized (as with the districts of Berar after 1905). Whilst this shifting composition of districts complicates any use of the data in a panel setting, it preserves the granularity of the underlying information and reduces the scope for measurement error due to the aggregation of districts together. In this dataset, we have 313 districts in 1881, 327 in 1891, 366 in 1901, 380 in 1911, 408 in 1921, and 506 in 1931.

To show more general trends and to account for the geographical divide in the percentage female that has been identified in the literature, we use four regions also used by Gupta – East, North, South, and West.<sup>37</sup> The districts are assigned to these regions through their provinces, which vary across census years. In the East, these are the Andaman and Nicobar Islands, Assam, Behar, Bengal, Burma, Orissa, and, where they are separately reported, Calcutta and Mayurbhanj State. In the North, these are Ajmer-Merwara, Baluchistan, Delhi, Jammu and Kashmir, the North-West Frontier Province, Punjab (including the Native States), the Rajputana Agency, and

<sup>33</sup> Census Report of 1931, p. x.

<sup>34</sup> Census Report of 1901, p. 110

<sup>35</sup> Census Report of 1941, p. 11.

<sup>36</sup> Bharadwaj et al., ‘Big march’.

<sup>37</sup> Gupta, *Economic history*.



the United Provinces. In the South, these are Cochin, Coorg, Hyderabad, the Madras Presidency, Mysore, and Travancore. In the West, these are Baroda, Berar, Bombay, the Central India Agency, the Central Provinces, Gwalior, and the Western India States Agency.<sup>38</sup>

There are some limitations of the original data that must be accounted for so that we can draw comparisons between districts and over time. In particular, there are differences both across census waves and across provinces within any given census wave in how religions and age categories are reported.

Whilst the original data report religion by 17 distinct religious groups, we focus our analysis on the 6 main religious groups in Colonial India: Buddhist, Hindu, Jain, Muslim, Sikh, and Tribal (i.e. Animist). Because a broad religious group can be reported in several distinct ways across regions and the Christians and several other religious groups are only recorded in a few districts, we have focused on these six groups to draw meaningful comparisons for the whole country. Whilst Hindu and Muslim populations, the two largest groups, are recorded in almost every district and age group, population counts for the other four religious groups are missing in some districts because either there were no inhabitants of that religious group in the district or the number was small, and thus they were grouped together with other religions to form an 'other' category in the original data.

There is a regional component to the distribution of religious groups, for example, Buddhists are mainly present in the East and North, whilst Jains are mainly present in the North and West. Hindus and Muslims together made up 89.8 per cent of the population in 1931, whilst adding Tribal populations brings this to 92.2 per cent. The reporting of the number of Christians by district is not systematic and they are often included in the 'other' category. Therefore, we do not report the share of women in this religious group, which was 1 per cent of the population in 1901.<sup>39</sup>

Across years and between provinces, population counts were reported in different age bins. For example, in 1911 the United Provinces, Travancore, and Burma reported population counts in 5-year bins for early ages (0–5, 5–10, 10–15, and 15–20) and in 20-year bins for later ages (20–40, 40–60), whilst Coorg, Cochin, and the Central India Agency reported population counts in 5-year bins until age 70. To make our analysis comparable across provinces and years, we choose age bins that can be reported consistently across provinces and years whilst still allowing for meaningful interpretation. These age bins are 0–10, 10–20, 20–40, and over 40. The coarseness of the early census years and for certain provinces in later censuses compels us to use a 0–10 age category, rather than computing under-5 sex ratios.

To our data, we have added a small number of additional control variables. We added a number of district geographic characteristics that we have computed using geographic information systems (GIS) data. These include, for example, measures of the suitability of a district for the cultivation of certain crops.

To compute geographic covariates for the districts in our data, we begin by creating a one-to-many correspondence between the historic districts mapped for each census and the third-level administrative units (e.g. tehsils) for which the Database of Global Administrative

<sup>38</sup> Some of these provinces are often separately considered as a 'Central' region in the literature. However, many districts of these regions, as given by colonial borders, would be part of present-day Gujarat or Maharashtra.

<sup>39</sup> Calculated from pp. 572–3 of the Census of India, 1931, Vol. I.—India. Part II.—Imperial Tables. The 1931 Census Report of India (p. 200) provides all-India sex ratios for both Christians and Parsis. The Christian population had 1002 girls per boy in the 0–1 age bin and 952 females per 1000 males of all ages. These are similar to the Hindu figures of 1017 and 953, respectively. For Parsis, the sex ratios were lower, at 981 and 940 in the 0–1 and all ages categories.



Areas (GADM) provides shapefiles.<sup>40</sup> We join these shapefiles to several sources of geographic data. For continuous variables such as latitude, we compute an area-weighted average of the variable over the polygons that together constitute the historic district. When the underlying data are in raster format, we average over raster cells within a sub-district. Sub-districts too small to contain a raster cell when the data have a coarse underlying resolution are assigned the value of the nearest raster cell. We will follow this procedure for all raster data. For dummy variables such as presence of a coastline, we take the maximum over this set of sub-district polygons. For example, if any part of a district has a coastline, then the district as a whole has a coastline.

We select covariates on the basis of their common use in the economics literature, on their importance to Indian agriculture, and on their relevance to broad hypothesis in the literature about the determinants of women's roles in South Asian economies.

In our data, *Coastal* is an indicator equal to 1 if the district has a coastline. We compute this variable using the polygon map of oceans provided by Natural Earth Data. Whilst this variable has not featured heavily in the literature on sex ratios, it correlates with past exposure to trade and predicts the presence of port cities.<sup>41</sup> *Latitude* is the latitude of a district's centroid, which we compute ourselves in ArcMap. *Longitude* is the longitude of a district's centroid, analogously defined. Whilst neither of these variables will be of interest in and of themselves, it has become conventional to control for them to reduce the influence of omitted variables that vary continuously over space.<sup>42</sup> These will also proxy, broadly, for more Northern and more Eastern areas of India.

*Malaria Suitability* is an index of the stability of malaria transmission, and is originally available as raster data.<sup>43</sup> Whilst this has not featured prominently in the literature on sex ratios, it is both a variable of interest and conventional control variable in several studies in development.<sup>44</sup> Diseases such as malaria increase the infant mortality rate generally.<sup>45</sup> In recent decades, research on the effect of malaria on pregnant women find negative health outcomes for women and infants.<sup>46</sup> Pregnant women are more susceptible to malaria, especially in the first pregnancy.<sup>47</sup> Not much is known about its effect on pregnant women in early twentieth century.

*Major River* is a dummy for whether any part of the district is intersected by a river in the polyline shapefile of 'Major Rivers of the World' provided by the World Bank.<sup>48</sup> The rivers that intersect South Asia in this source are the Indus, Ganges, Brahmaputra, Salween, and Irrawaddy, the latter two of which are in Burma. This dummy will be a correlate of several possible sources of relative female survival, including land suitability for agriculture, the importance

<sup>40</sup> We base these correspondences off several scanned maps. For present-day India, we make extensive use of Chandramouli et al. 'Administrative Atlas'. We also employ the map included in the 1931 census, which Fenske and Kala, 'Linguistic distance,' have previously converted to shapefile format, maps from the Imperial Gazetteer of India made available by the Digital South Asia Library, and maps available in the census reports of the specific provinces in several census years.

<sup>41</sup> Jha, 'Ethnic tolerance'.

<sup>42</sup> See e.g. Michalopoulos and Papaioannou, 'Ethnic institutions'.

<sup>43</sup> Kiszewski et al., 'Global index'.

<sup>44</sup> Alsan, 'Tsetse fly'; Cervellati et al., 'Epidemic shocks'.

<sup>45</sup> Drevenstedt et al., 'Infant mortality'.

<sup>46</sup> Desai et al., 'Epidemiology'.

<sup>47</sup> Chima et al., 'Malaria'.

<sup>48</sup> <https://datacatalog.worldbank.org/dataset/major-rivers-world/resource/c8d8958a-51e4-4e84-88a6-75e5abc26ea4>

of irrigation, commerce, and population density. *Ruggedness* is the Nunn and Puga index of terrain ruggedness.<sup>49</sup> The underlying raster data consist of the average deviation in elevation between a given grid cell and its eight immediate neighbours. Within South Asia, larger values typically correspond to mountainous terrain in Himalayan regions. Whilst this has not featured prominently in the literature on sex ratios, it is well known that the Himalayan regions of South Asia that score highly on this index are distinctive along several cultural and institutional dimensions.<sup>50</sup> Polyandry, for example, is more common in these areas.<sup>51</sup>

*Cotton Suitability* is the district's potential output of cotton under rain-fed agriculture and with low inputs. The underlying data here are from the Food and Agriculture Organization of the United Nations' Global Agro-Ecological Zones (FAO-GAEZ) project. These are initially available as raster data, and measure exogenous potential crop output on the basis of factors beyond human control. This is measured in kilograms dry weight per hectare. Whilst the FAO computations of potential output are based on modern-day remote sensing data, they have become widely used in economic history.<sup>52</sup>

Cotton is chosen here as one of India's main export cash crops, as it may have increased rural incomes, but may also have increased food insecurity.<sup>53</sup> We compute several analogous suitability measures from the FAO data: *Dryland Rice Suitability* is the district's potential output of dryland rice, *Tea Suitability* its potential output of tea, *Wetland Rice Suitability* is the district's potential output of wetland rice, and *Wheat Suitability* captures a district's potential output of wheat. Rice and wheat are included because they have featured prominently in the literature on sex ratios. Tea, by contrast, is a crop that was largely cultivated by migrant workers who were roughly half female, however, this was not the case in other crops and was driven by female participation in plucking of leaves.<sup>54</sup>

### III | REGIONAL VARIATIONS IN SEX RATIOS IN COLONIAL INDIA

Where do women go missing? In this section, we describe regional patterns in sex ratios and their persistence over time.

From the early censuses, enumerators noted the North–South difference in sex ratios and considered the East to be somewhere in between these two regions. Bardhan, using data from the 1950s and 1960s, noted that female life expectancy was higher than male life expectancy in the Eastern states of Assam, West Bengal, Bihar, and Orissa, whilst male life expectancy was higher in the Northern states of Punjab and Rajasthan.<sup>55</sup> In the Southern States such as Tamil Nadu and Kerala, survival of girls was more likely. Bardhan attributes this pattern to a number of variables, including regional differences in the age of marriage, female literacy, maternal medical care, and neglect of girls.

<sup>49</sup> Nunn and Puga, 'Ruggedness'.

<sup>50</sup> Scott, *Being governed*.

<sup>51</sup> Korn, 'Family structures'.

<sup>52</sup> See e.g. Dimico et al., 'Sicilian mafia', or Dupraz, 'French and British'.

<sup>53</sup> Satya, *Cotton and famine*; Harnetty, 'Cotton exports'.

<sup>54</sup> Gupta and Swamy, 'Labor coercion'.

<sup>55</sup> Bardhan, 'Life and death'.



The literature has explored factors behind the North–South division, framing it around the Boserup thesis of women’s labour force participation.<sup>56</sup> Whilst women play a major role in paddy rice production, wheat is intensive in male labour, which may affect the economic value of men and women. Bardhan speculates that women worked in rice cultivation in the South and East, and had little involvement in agricultural lands that are suitable for wheat and millet in the North and North-West, therefore, they were more valued in the South and the East.<sup>57</sup> Rosenzweig and Schultz similarly find a positive correlation between female employment and girls’ survival.<sup>58</sup> Kishor finds a negative correlation between area cropped to rice and under-5 female mortality in census data from the 1980s.<sup>59</sup>

Other scholars have instead stressed kinship and marriage patterns. Dyson and Moore argue that differences in kinship and marriage conventions between the North and the South explain the lower status of women and greater son preference in Northern India.<sup>60</sup> In the North, women marry outside their place of birth, and therefore have less value to the parental family. Although a North–South divide did not exist in the inheritance rights of women, the regional divide was pronounced in property ownership due to the difference in marriage patterns. Under village exogamy in the North, women were not given land in dowry, whereas in the South, under endogamy and consanguinity, land could be part of dowry.

Kishor cites differences in kinship structure and female labour force participation as important determinants of child mortality in India, but shows they cannot fully explain the North–South gap.<sup>61</sup> Miller used census data from 1961 to show a North–South divide in women’s labour force participation to be correlated with restrictive social norms for women and male-biased sex ratios.<sup>62</sup> She stresses the relative prevalence of dowry in the North, rather than reciprocal marriage payments.

In the analysis that follows, we show patterns for the four regions defined in section II – East, North, South, and West. In table 1, we present the percentage of the population that is female in each of these regions in each census year, alongside the total population (male and female) within that region and year and the number of districts from which these totals are computed.<sup>63</sup> Whilst the share of the population that is female for all of India is approximately 49 per cent in each year, it is clear that there are regional differences. In the South, the percent female is above 50 per cent in all years, whilst the percentage female in the North never exceeds 48 per cent. Like the South, the East has a less male-biased sex ratio in all years. Sex ratios are clearly male biased in the West, albeit less so than in the North.

Because these broad regional tables can mask sub-regional differences, in figure 1 we show a map of the percentage of the population that is female in each district in each census year. These numbers capture the population across all religions. North of a line that roughly connects Surat on the western coast of India to Almora in the Himalayas, the percentage of the population that is

<sup>56</sup> Boserup, *Woman’s role*.

<sup>57</sup> Bardhan, ‘Life and death’.

<sup>58</sup> Rosenzweig and Schultz, ‘Market opportunities’.

<sup>59</sup> Kishor, ‘Child mortality’.

<sup>60</sup> Dyson and Moore, ‘Kinship structure’.

<sup>61</sup> Kishor, ‘Child mortality’.

<sup>62</sup> Miller, ‘Endangered Sex’; Miller, ‘Female labor’.

<sup>63</sup> Because we aggregate these numbers from the data on districts, and not all districts are covered in the age, sex, and civil condition sections of the census, there will be differences from comparable totals in the census reports.

**TABLE 1** Percentage female by region.

Year	Region	Age group	Religion	Total population	Female population	Percentage female	Districts
1881	All	ALL	All	218 513 488	107 036 368	0.490	313
1881	East	All	All	77 879 456	38 901 016	0.500	85
1881	North	All	All	68 014 832	32 153 816	0.473	122
1881	South	All	All	34 585 484	17 469 608	0.505	34
1881	West	All	All	38 033 720	18 511 924	0.487	72
1891	All	All	All	261 527 392	128 323 552	0.491	327
1891	East	All	All	87 442 432	43 659 020	0.499	99
1891	North	All	All	73 386 912	34 794 292	0.474	109
1891	South	All	All	55 436 664	27 776 536	0.501	55
1891	West	All	All	45 261 376	22 093 702	0.488	64
1901	All	All	All	288 797 920	141 924 672	0.491	366
1901	East	All	All	95 183 440	47 371 448	0.498	108
1901	North	All	All	83 109 440	39 568 864	0.476	129
1901	South	All	All	59 158 796	29 698 648	0.502	44
1901	West	All	All	51 346 248	25 285 704	0.492	85
1911	All	All	All	310 586 272	151 784 928	0.489	380
1911	East	All	All	103 952 560	51 486 344	0.495	109
1911	North	All	All	89 319 328	41 892 736	0.469	144
1911	South	All	All	65 072 468	32 707 190	0.503	64
1911	West	All	All	52 241 904	25 698 652	0.492	63
1921	All	All	All	313 544 544	152 601 360	0.487	408
1921	East	All	All	106 821 520	52 536 576	0.492	115
1921	North	All	All	87 600 008	40 971 624	0.468	139
1921	South	All	All	66 393 796	33 296 438	0.501	65
1921	West	All	All	52 729 204	25 796 720	0.489	89
1931	All	All	All	348 890 624	169 314 256	0.485	506
1931	East	All	All	118 339 632	57 857 312	0.489	114
1931	North	All	All	96 779 920	45 213 448	0.467	142
1931	South	All	All	74 184 472	37 161 520	0.501	66
1931	West	All	All	59 586 596	29 081 970	0.488	184

*Note:* These figures sum populations over all districts reporting civil condition by age in a given census year.

female is markedly lower than elsewhere. Higher female percentages of the population are visible particularly in southern peninsular India and in contemporary Burma.<sup>64</sup>

In figure 2, we present similar maps of the population aged 0–10. Here, the regional patterns are roughly similar. The general gradient separating the northwest from the rest of South Asia is

<sup>64</sup> Whilst the stability of this geographic pattern is obscured by the fact that Baluchistan, much of Burma, and many of the princely states are not censused in 1881 (Hyderabad, the Rajputana Agency, and the Orissa Tributary States all stand out for their absence), this general geographic pattern is visible across census waves.

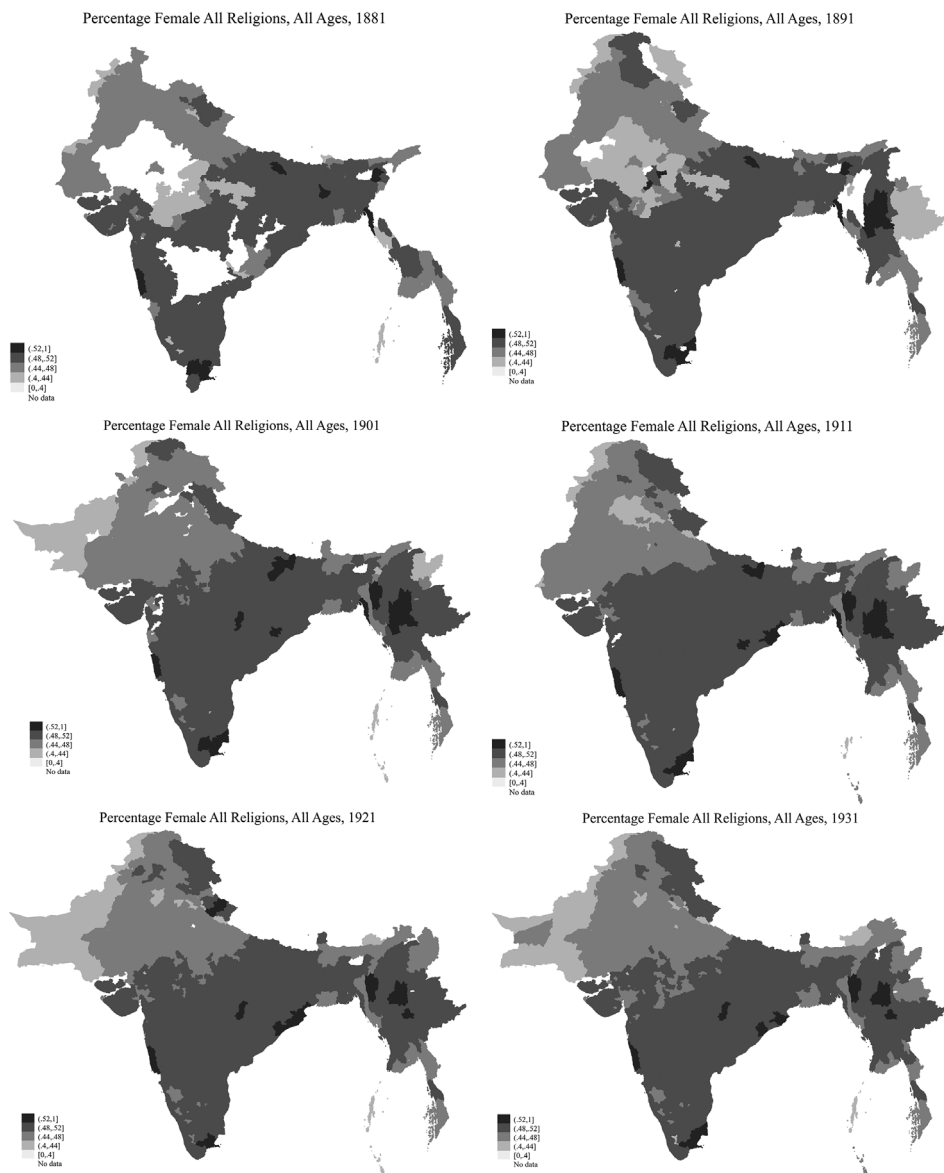
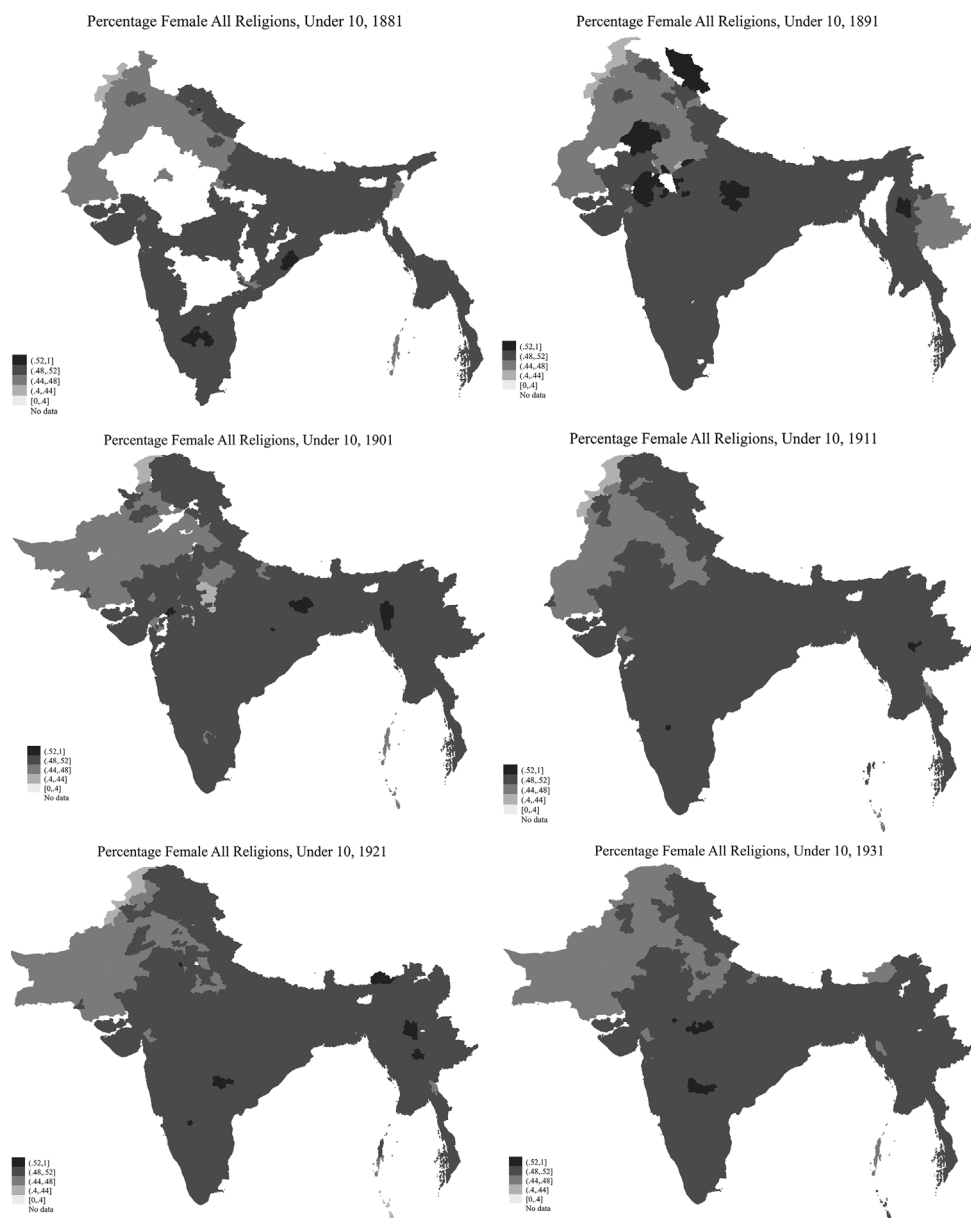


FIGURE 1 Percentage female: all ages.

again clear. Some of the more extreme outliers now have more balanced sex ratios, reflecting the prevalence of male migrants of working age in some of these frontier areas.

To what extent do the spatial patterns in colonial sex ratios persist into the present? To answer this question, we turn to the 2011 Census of India,<sup>65</sup> and compute the percentage female for each district that existed in 2011, for age bins that mimic those we can construct for the colonial period: 0–9, 10–19, 20–39, and 40+. We show maps of the percentage of female in each of these age bins across districts in figure 3. In the 0–9 and 10–19 age bins, the familiar pattern of male bias appears in the north-west, particularly in Punjab and Haryana, as well as in inland Maharashtra.

<sup>65</sup> <https://censusindia.gov.in/2011census/C-15.html>



**FIGURE 2** Percentage female: under 10.

In table 2, we measure the correlation between past and present sex ratios. We regress the percentage female in a district in 2011 on its percentage female in 1931 by age bin. We report specifications both with and without fixed effects for the states that existed in 2011.

Three conclusions are apparent from this table. First, colonial sex ratios predict current sex ratios, even at the district level. In the columns without fixed effects, the  $R^2$  values range from 0.116 in the 40+ age bin to 0.293 in the 10–19 age bin. These are equivalent (taking the square root) to correlation coefficients between 0.341 and 0.541. Second, the coefficients and  $R^2$  values

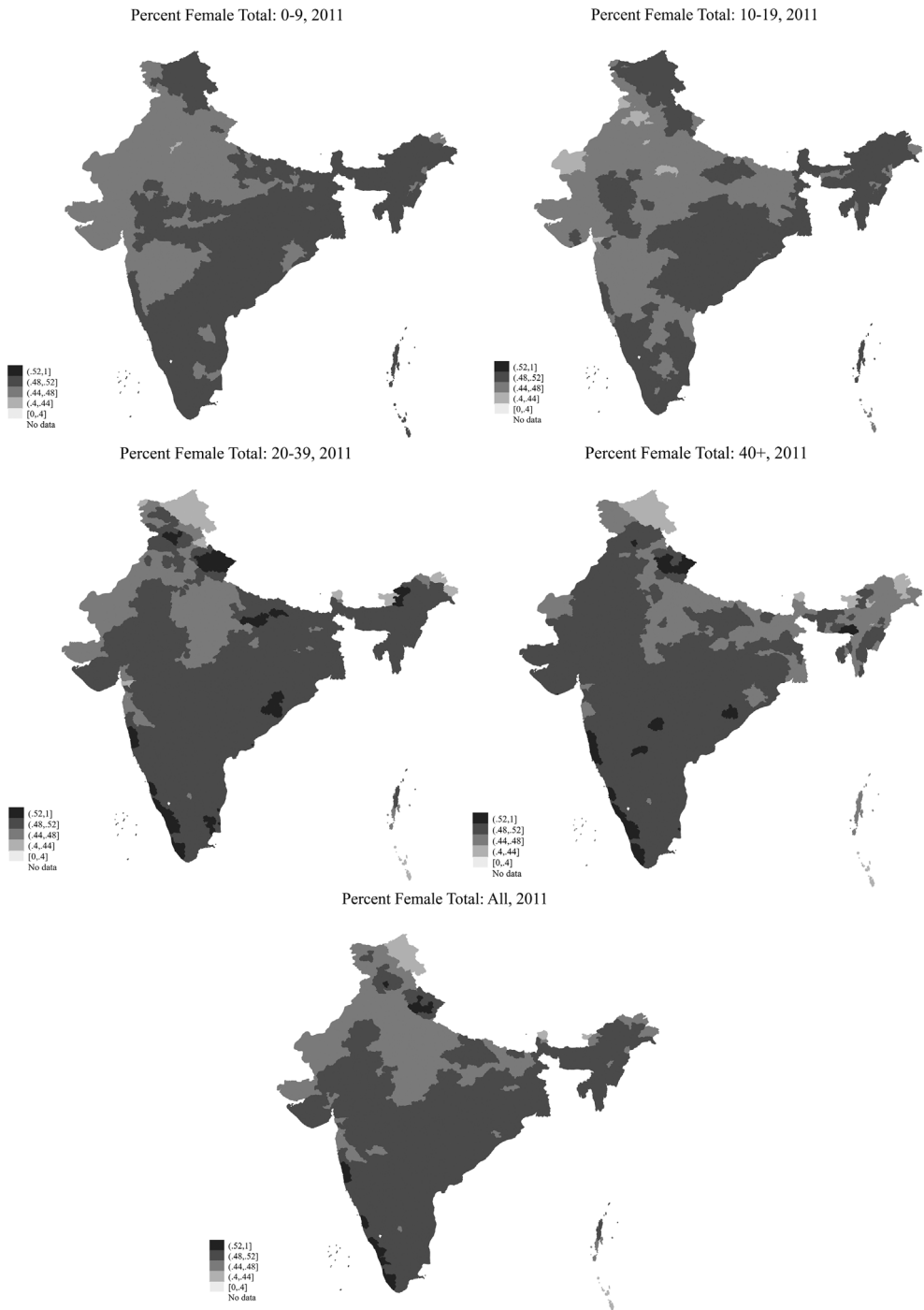


FIGURE 3 Percentage female: 2011.

**TABLE 2** Correlations between percentage female in 1931 and 2011.

	Percent female 0–9		Percent female 10–19	
Percentage female 1931: 0–10	0.565*** (0.057)	0.365*** (0.056)		
Percentage female 1931: 10–20			0.404*** (0.037)	0.234*** (0.052)
Observations	625	625	625	625
$R^2$	0.269	0.708	0.302	0.625
State FE	No	Yes	No	Yes
	Percent female 20–39		Percent female 40+	
Percentage female 1931: 20–40	0.215*** (0.041)	0.168*** (0.048)		
Percentage female 1931: 40+			0.166*** (0.018)	0.120*** (0.028)
Observations	625	625	625	625
$R^2$	0.146	0.410	0.131	0.640
State FE	No	Yes	No	Yes

Note: Observations are at the 2011 district level. \*\*\* Significant at 1%, \*\* significant at 5%, \* significant at 10%. Robust standard errors in parentheses.

are greater at younger ages – persistence in the sex ratio is most pronounced among children and adolescents. Third, the correlations between the colonial and present-day percentages female survive controlling for state fixed effects. It is not only the regions of colonial India that had male-biased sex ratios that continue to have them today – this persistence is apparent across districts within a given state.

In figure 4, we show the scatterplots of the raw data underlying these correlations. Two conclusions emerge from these figures. The first is that outliers, rather than driving the correlations in table 2, undermine them. The points that lie far from the central cloud of points are often frontier districts in regions such as Arunachal Pradesh that have lower-quality data and male-biased sex ratios in either the colonial or modern periods, or large cities such as Kolkata in which adult populations were once overwhelmingly male but have since become more balanced. The second conclusion is that it is in the 0–10 and 10–20 age bins that past and present sex ratios are most clearly correlated, and the relationship flattens at later ages.

In figure 5, we extend the analysis from table 2. For each census year between 1961 and 2011, we regress the percentage female in a district on its percentage female in 1931. For the years 1961–91, we use data from Vanneman and Barnes, whilst for 2001 and 2011 we obtain data directly from the census.<sup>66</sup> We use the same procedure as we did for table 2 to merge post-colonial districts to those that existed in 1931. In figure 5, we show the  $R^2$  from a bivariate regression of the percentage female in a given year on the 1931 percentage female. We also run this same regression with state-fixed effects and show the  $R^2$  net of these fixed effects. It is clear from both figures that the importance of colonial sex ratios diminishes over time. Whilst correlation between the 1931 percentage female and the 2011 percent female that constitutes the result in table 2 is statistically significant, it is much weaker than the correlations for earlier years.

<sup>66</sup> Vanneman and Barnes, *Indian District Data*.

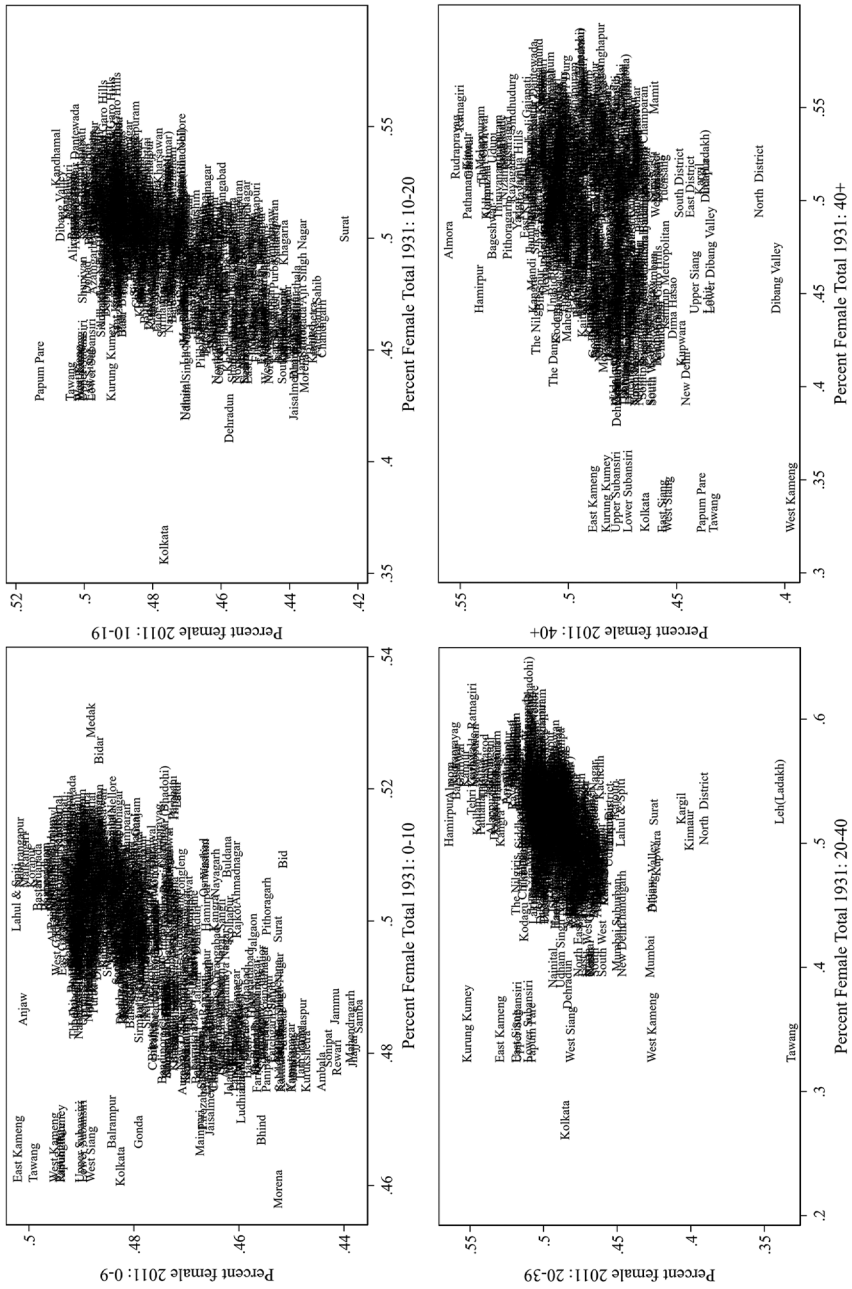


FIGURE 4 Percentage female: 2011 versus 1931.

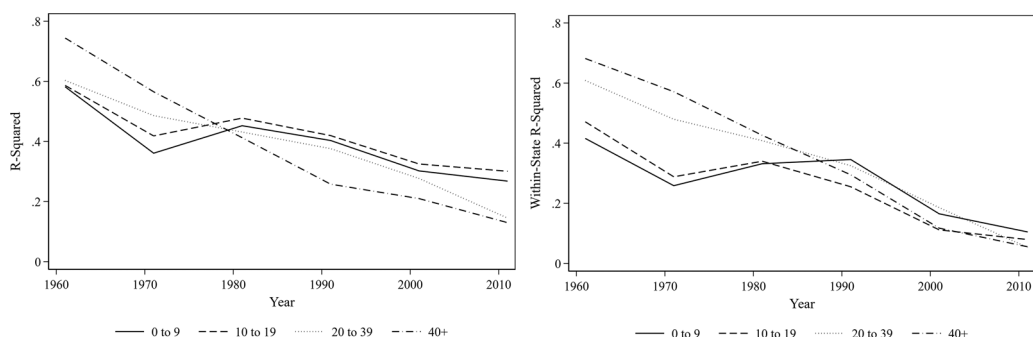


FIGURE 5 R<sup>2</sup> due to 1931 percentage female over time.

In sum, our district-level data have allowed us to document the stability of spatial patterns in the sex ratio over more than a century both across and within regions, and to identify colonial-era variation in the sex ratio at a more granular geographic level than has been done previously.

#### IV | RELIGION AND SEX RATIOS IN COLONIAL INDIA

Where do women go missing? In this section, we describe how sex ratios differ by religion, and how these differences correlate over space.

Religious norms, especially the importance of sons and male relatives in performing the last rites, have been cited as an explanation of son preference in Hindu communities.<sup>67</sup> The female deficit was higher among Muslims compared with Hindus in North India, but not in the East.<sup>68</sup> Male-biased sex ratios were more prevalent in the higher castes.<sup>69</sup> The census reports documented meticulously the sex ratio by age cohort by region and religion, but rarely commented on Hindu–Muslim differences in sex ratios. Province-level data from the census of 1901 do not show systematic differences in sex ratios across religions, although they show differences across provinces. The sex ratio for Muslims was less biased against women and girls than the sex ratio for Hindus and Sikhs in the North. In Bengal, a province with a large Muslim population, the share of women and girls in the population declined faster among Muslims after age 30.<sup>70</sup>

Recent evidence on sex ratios shows a different picture. Son preference is stronger among Hindus compared with Muslims.<sup>71</sup> Muslim women show lower son preference in surveys.<sup>72</sup> The sex ratio at birth is more balanced among Muslims and the Muslim population has better child survival rates compared with Hindus.<sup>73</sup> Lower rates of vegetarianism, closer kinship, and better maternal health among Muslims may also contribute to girls' survival.<sup>74</sup> Bhalotra et al. find that

<sup>67</sup> Visaria, 'Son preference'; Jayachandran, 'Gender inequality'.

<sup>68</sup> *ibid.*, p. 499.

<sup>69</sup> Chakraborty and Kim, 'Kinship institutions'.

<sup>70</sup> Census Report of 1901, pp. 238–249

<sup>71</sup> Jayachandran, 'Gender inequality'; Borooah et al., 'Missing women'; borooah2005religion.

<sup>72</sup> Visaria, 'Son preference'.

<sup>73</sup> Borooah and Iyer, 'Vidya, Veda, and Varna'.

<sup>74</sup> Bhalotra et al., 'Muslim advantage'.

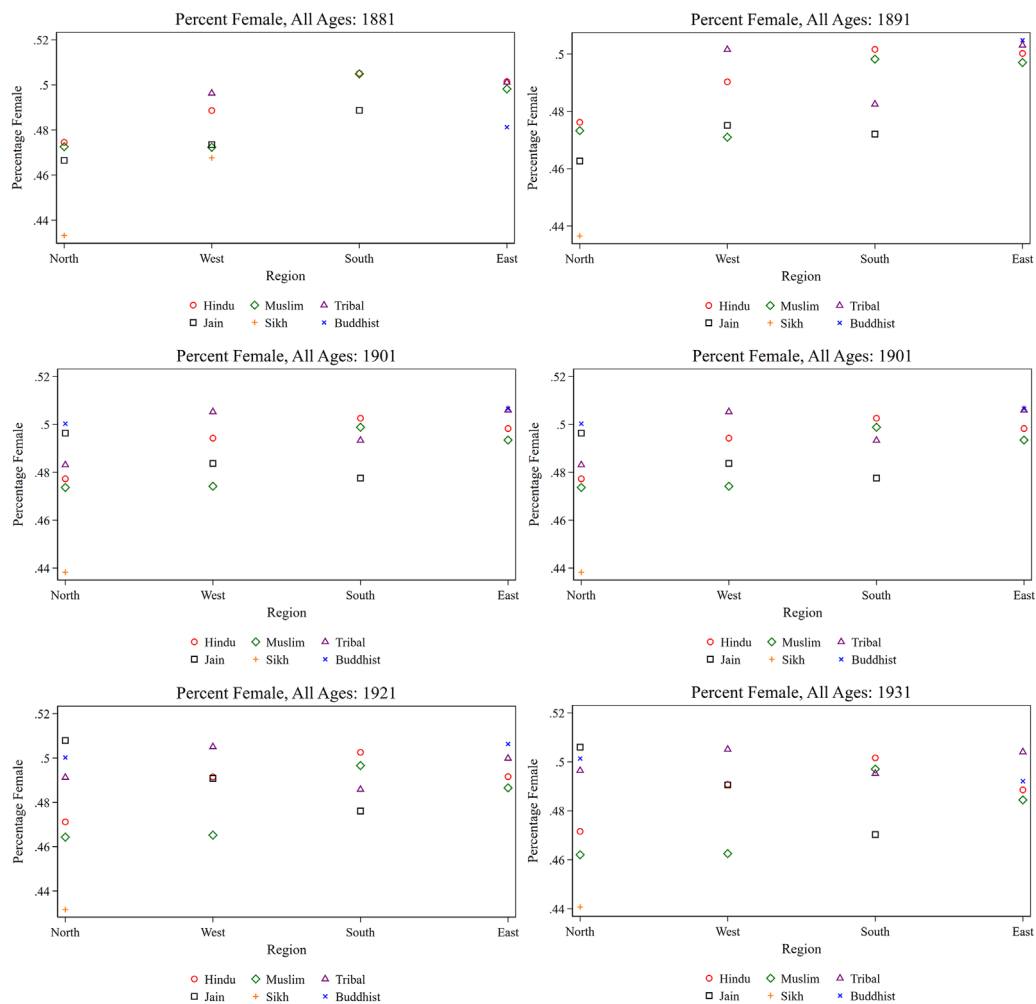


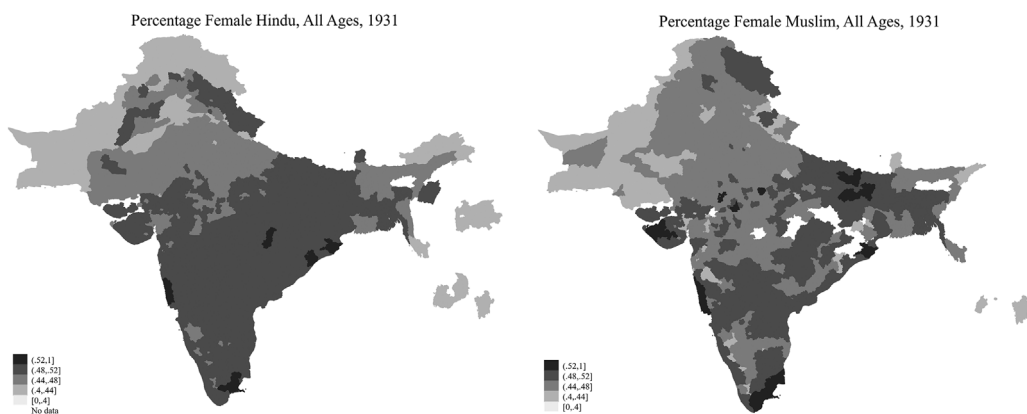
FIGURE 6 Percentage female by region and religion.

abortion is not widely practised among Muslims, and therefore in recent data, Muslims have more equal sex ratios at birth compared with Hindus.<sup>75</sup>

Our data show that, whilst there are differences in sex ratios across religions, these differences vary by region. Whilst Sikhs have the most male-biased sex ratios and Buddhists have the most balanced sex ratios, these groups are regionally concentrated. Consider figure 6, in which we show the percentage of the population that is female by religion and region.<sup>76</sup> Whilst it is clear that there are differences in the share female by religion, the share female for any one religious group often differs markedly by region. In many cases, a group that is more predominantly female than another religious group in one region is instead less female in another region. Within regions, we will show below that religions often exhibited correlations in sex ratios across districts. This is especially true of Hindus and Muslims, the two largest groups. Figure 7 shows the sex ratio by

<sup>75</sup> Bhalotra et al. 'Religion and abortion'.

<sup>76</sup> We exclude religion-region combinations with fewer than 10 000 persons in these figures.



**FIGURE 7** Percentage female: Hindus and Muslims.

religion for Hindus and Muslims for 1931. Each of these maps show regional variation similar to what we find for the overall population.

In our data, we can compute sex ratios by religion at the level of districts. In table 3, we assess the degree to which female percentages in the population are correlated across religious communities in the same district. In particular, using the sample of districts in 1901, we compute correlation coefficients for the percentage female across religions. The top panel (A) reports the correlations for all ages, whilst the middle panel (B) reports the correlations for the 0–10 age bin. Note that the number of observations will vary considerably across comparisons. Where a religion is a small minority in any given province, data on the sex composition of that minority are not reported by age, and thus we cannot include districts from that province in the comparison.

Despite this limitation, a clear pattern emerges: in districts where the population composition is male biased for one religion, the population composition tends to be male biased for others. The correlation in sex ratios between Hindus and Muslims is high. It is weaker between Hindus and other religious groups. In many cases, this correlation is statistically significant at conventional levels. Consider Hindus and Muslims, India's two largest religious groups. In the under-10 age bin, the correlation in percentage female is just shy of 0.3, and is significant at the 5 per cent level.

In appendix table A1 we add province fixed effects. We do this for two reasons. First, fixed effects allow us to show that the correlations across religions are not driven solely by broad regional comparisons, but also hold within regions. Second, fixed effects allow us to show that differences in data quality across provinces do not drive our findings. In particular, for each census year  $t$  and for each non-Hindu religion  $r$ , we estimate the following specification:

$$\text{PercentFemale}_{rdt} = \alpha + \beta \text{HinduPercentFemale}_{dt} + \delta_p + \epsilon_{rd}$$

Here,  $\text{PercentFemale}_{rdt}$  is the percentage of the population of religion  $r$  in district  $d$  in census year  $t$  that is female, and  $\alpha$  is a constant.  $\text{HinduPercentFemale}_{dt}$  is the percentage of the Hindu population that is female in district  $d$  in census year  $t$ ,  $\delta_p$  is a set of province fixed effects, and  $\epsilon_{rd}$  is the error term. We report robust standard errors. Whilst the strength of the correlations is smaller within regions than across regions as measured by the within  $R^2$ , the general finding of correlations across religions – particularly for Muslims, Hindus, Sikhs, and Jains – remains present.

Whilst the 'Age, Sex and Civil Condition' section of the census does not report data by caste, we have used two alternative sources to shed light on the role of caste in explaining sex ratios in

**TABLE 3** Correlations across religions in 1901.**Panel A. All ages**

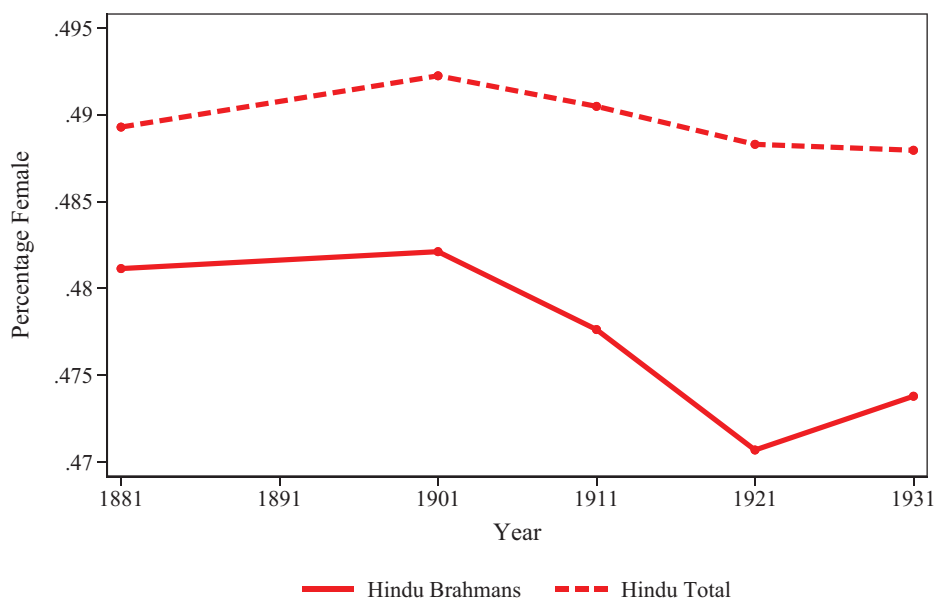
	<b>Perc. fem. Buddhist</b>	<b>Perc. fem. Hindu</b>	<b>Perc. fem. Jain</b>	<b>Perc. fem. Muslim</b>	<b>Perc. fem. Sikh</b>
<u>Perc. fem. Hindu</u>	-0.1975				
<i>p</i> -Value	0.1693				
No. of observations	50				
<u>Perc. fem. Jain</u>	0.0400	0.8679			
<i>p</i> -Value	0.9745	0.0000			
No. of observations	3	88			
<u>Perc. fem. Muslim</u>	0.1732	0.7309	0.3554		
<i>p</i> -Value	0.2339	0.0000	0.0007		
No. of observations	49	338	88		
<u>Perc. fem. Sikh</u>	0.7016	0.2761	0.1144	0.2495	
<i>p</i> -Value	0.2984	0.0158	0.4821	0.0298	
No. of observations	4	76	40	76	
<u>Perc. fem. Tribal</u>	0.0182	0.5001	0.6942	0.4536	0.2955
<i>p</i> -Value	0.9102	0.0000	0.0000	0.0000	0.1516
No. of observations	41	143	45	135	25

**Panel B. Population aged 0–10**

	<b>Perc. fem. Buddhist</b>	<b>Perc. fem. Hindu</b>	<b>Perc. fem. Jain</b>	<b>Perc. fem. Muslim</b>	<b>Perc. fem. Sikh</b>
<u>Perc. fem. Hindu</u>	0.0365				
<i>p</i> -Value	0.8031				
No. of observations	49				
<u>Perc. fem. Jain</u>	-1.0000	0.1824			
<i>p</i> -Value	1.0000	0.0948			
No. of observations	2	85			
<u>Perc. fem. Muslim</u>	0.6689	0.2872	0.1587		
<i>p</i> -Value	0.0000	0.0000	0.1468		
No. of observations	48	338	85		
<u>Perc. Fem. Sikh</u>	0.9185	0.0938	-0.1602	0.2241	
<i>p</i> -Value	0.0815	0.4332	0.3654	0.0584	
No. of observations	4	72	34	72	
<u>Perc. fem. Tribal</u>	-0.1583	0.1944	0.5146	0.1286	0.1258
<i>p</i> -Value	0.3359	0.0205	0.0004	0.1386	0.5972
No. of observations	39	142	43	134	20

Note: Observations are at the district level. *p*-Values of 0.0000 are less than 0.0001.

the Hindu population. First, using the aggregate totals in the first volume of each year's census, we report the percent female for both Hindu Brahmans and for all Hindus. We exclude 1891 due to lack of data. In figure 8, we show that the percentage of female among Hindus is lower among Brahmans, widening over time to more than a percentage point by 1911.



**FIGURE 8** Percentage female: Brahmans and all Hindus.

Second, using the 1901 census data on caste, we estimate the following regression:

$$\text{PercentFemale}_{gd} = \alpha + \beta \text{Brahman}_{gd} + \delta_d + \epsilon_{dt}$$

Here,  $\text{PercentFemale}_{gd}$  is the percent female, ranging from 0 to 1, of group  $g$  in district  $d$ . We restrict the sample so that group  $g$  can take two values – Brahman Hindus and non-Brahman Hindus – whilst  $\alpha$  is a constant,  $\delta_d$  is a district fixed effect, and  $\epsilon_{dt}$  is the error term. We cluster standard errors by district. We estimate  $\beta$  to be  $-0.00962$ , with a standard error of  $0.00336$ . That is: the magnitude of the gap between Brahmans and non-Brahmans looking within districts is roughly one percentage point, almost exactly the same as the aggregate gap in the population captured in figure 8. This intra-district gap is consistent with the gap Chakraborty and Kim find in 1901 for Punjab, Bengal, and Madras.

In table 4, we consider whether these interreligious correlations across districts are a feature only of the colonial period. Using data from the 2011 Census of India,<sup>77</sup> we report correlation coefficients analogous to those in table 3 for the most recent Indian census. The top panel (A) reports these correlations for all ages, whilst the bottom panel (B) reports correlations for the 0–9 age group. Here we also find significant correlations between the percentages of the population that is female coming from different religious groups in the same district. For Hindus and Muslims, the correlation is slightly above 0.40 for all ages, and just below 0.33 for the 0–10 age bin. Several other correlations are significant and positive at the 5 per cent level, including the Hindu–Buddhist, Hindu–Sikh, Muslim–Christian, Muslim–Sikh, and Sikh–Jain correlations. As in table 3, we present in appendix table A1 that even within states there are correlations in the share female between Muslims and Hindus and between Sikhs and Hindus.

<sup>77</sup> <https://censusindia.gov.in/2011census/C-15.html>

**TABLE 4** Correlations across religions in 2011.

<b>Panel A. All ages</b>						
	<b>Pct. Female: All</b>	<b>Pct. female: Buddhist</b>	<b>Pct. female: Christian</b>	<b>Pct. female: Hindu</b>	<b>Pct. female: Jain</b>	<b>Pct. female: Muslim</b>
<u>Pct. female: Buddhist</u>	0.1221					
<i>p</i> -Value	0.0020					
No. of observations	640					
<u>Pct. female: Christian</u>	0.3894	0.3664				
<i>p</i> -Value	0.0000	0.0000				
No. of observations	640	640				
<u>Pct. female: Hindu</u>	0.3945	0.2919	0.5727			
<i>p</i> -Value	0.0000	0.0000	0.0000			
No. of observations	640	640	640			
<u>Pct. female: Jain</u>	0.0659	0.1104	0.0923	0.1884		
<i>p</i> -Value	0.0961	0.0052	0.0197	0.0000		
No. of observations	638	638	638	638		
<u>Pct. female: Muslim</u>	0.3597	-0.0609	0.2080	0.4074	0.0550	
<i>p</i> -Value	0.0000	0.1240	0.0000	0.0000	0.1654	
No. of observations	640	640	640	640	638	
<u>Pct. female: Sikh</u>	0.1439	0.0924	0.3237	0.4913	0.1597	0.4880
<i>p</i> -Value	0.0003	0.0194	0.0000	0.0000	0.0001	0.0000
No. of observations	640	640	640	640	638	640
<b>Panel B. Population aged 0–9</b>						
	<b>Pct. female: All</b>	<b>Pct. female: Buddhist</b>	<b>Pct. female: Christian</b>	<b>Pct. female: Hindu</b>	<b>Pct. female: Jain</b>	<b>Pct. female: Muslim</b>
<u>Pct. female: Buddhist</u>	0.0516					
<i>p</i> -Value	0.1940					
No. of observations	636					
<u>Pct. female: Christian</u>	0.3713	-0.0186				
<i>p</i> -Value	0.0000	0.6401				
No. of observations	640	636				
<u>Pct. female: Hindu</u>	0.8357	-0.0646	0.2853			
<i>p</i> -Value	0.0000	0.1034	0.0000			
No. of observations	640	636	640			
<u>Pct. female: Jain</u>	0.0038	0.0228	0.0659	0.0013		
<i>p</i> -Value	0.9235	0.5690	0.0982	0.9740		
No. of observations	631	627	631	631		
<u>Pct. female: Muslim</u>	0.4425	0.0411	0.2667	0.3293	0.1197	
<i>p</i> -Value	0.0000	0.3012	0.0000	0.0000	0.0026	
No. of observations	640	636	640	640	631	
<u>Pct. female: Sikh</u>	0.0903	0.0141	0.0945	0.1070	0.0863	0.0374
<i>p</i> -Value	0.0231	0.7230	0.0174	0.0071	0.0311	0.3481
No. of observations	633	630	633	633	625	633

Note: Observations are at the 2011 district level. *p*-Values of zero correspond to  $p < 0.00005$ .



In sum, our results suggest that differences across regions are shared across religions. The correlation in sex ratios across religions over space has, we believe, gone largely overlooked in the literature. The striking similarity of Hindu and Muslim sex ratios in the colonial period, by contrast, differs from the present day.

## V | GEOGRAPHY AND SEX RATIOS IN COLONIAL INDIA

Where do women go missing? In this section, we describe how sex ratios correlate with geographic variables, in particular with those that shape agricultural practices that may influence the status of women.

The literature has largely focused on possible links between agricultural patterns and the sex ratio. Crops have been divided into those that use female labour intensively and those that do not. Measures of geographic suitability for various types of agriculture have the advantage of being largely exogenous to human action, and thus allow the researcher to avoid confounding effects of the possible co-determination of agricultural practices and gender outcomes. Prior empirical work has shown that, across countries and across districts in the present day, areas more suited to plough agriculture have more male populations.<sup>78</sup> The rice–wheat distinction highlighted by the literature on India<sup>79</sup> does not map cleanly into the use of the plough: [Alesina et al.](#) identify both crops as ‘plough-positive’, and in the 1921 Agricultural Statistics of India, the share of cultivated area planted to either crop correlates positively with the number of ploughs per capita.<sup>80</sup> However, rice cultivation is more likely to involve women compared with wheat cultivation.<sup>81</sup> We also include other geographic variables that have been relatively neglected in the literature on India, including ruggedness, presence of a coastline, malaria suitability, and suitability for cash crops such as tea and cotton.

*Correlations:* In figure 9, we demonstrate that these patterns are apparent in the raw data. We present scatterplots of the under-10 sex ratio against three of our key geographic variables in the sample of districts from 1901. These variables are dryland rice suitability, wetland rice suitability, and wheat suitability. The positive correlation of the share female with suitability for both types of rice and the negative correlation with suitability for wheat are visible, and clearly not driven by outliers.

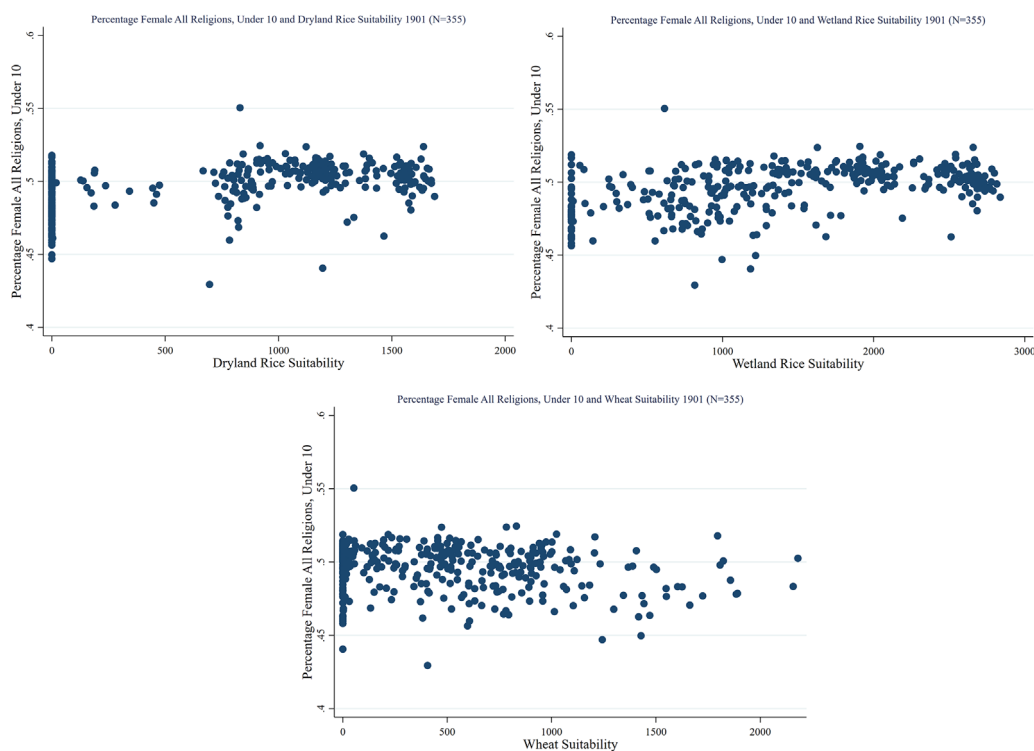
In table 5, we present pairwise correlations between our key geographic variables and the percentage female across districts in the 1901 sample. We do this separately for each of the age bins in our data. Several, but not all, of these correlations are significant at the 5 per cent level. Greater suitability for dryland and wetland rice predicts a greater female share across all age bins. Wheat suitability, by contrast, predicts a more male population in all age bins. Latitude correlates negatively, and longitude positively, with the share female in several age bins and census years. Coastal districts have a greater share of younger women in several years. Cotton suitability predicts a greater share female across age bins, whilst tea positively predicts the presence of younger women.

<sup>78</sup> [Alesina et al.](#), ‘Traditional agricultural practices’; [Carranza](#), ‘Soil endowments’.

<sup>79</sup> [Bardhan](#), ‘Life and death’.

<sup>80</sup> [Alesina et al.](#), ‘Gender roles’. By contrast, the share planted to cotton or tea correlates negatively with this measure of plough prevalence. In the 1910 Agricultural Statistics of India, the share of cultivated land that is irrigated correlates negatively across districts with the share of cultivated land planted to rice and positively with the share planted to wheat. The correlation of cotton or tea cultivation with irrigation is insignificant in these data.

<sup>81</sup> [Boserup](#), *Woman’s role*.



**FIGURE 9** Correlations with crop suitability: under 10.

*Regressions:* Because these pairwise correlations do not account for the likely correlations between the geographic variables we consider, we use table 6 to report corresponding regression results. For 1901, we regress the percentage of the population that is female on the geographic controls reported in table 5. We do this separately for each of our major age bins. For ease of interpretation, we standardize the percentage female to have a normal distribution with mean zero and standard deviation one, and we do the same for all continuous geographic controls. Indicators (coastal, river) are unchanged, thus coefficients can be interpreted as standardized betas, that is, how a one standard deviation change in a geographic control predicts changes in the outcome variable, again in standard deviations.

Whilst some of the correlations in table 5 retain their signs and significance, this is not universally the case. Latitude negatively predicts the share female in all but the oldest age bin, but this is significant only for the youngest bin. Longitude predicts a higher share female in the two youngest age bins, that is, the East–West gradient in South Asia emerges as a predictor of sex ratios as robust as the North–South gradient when controlling for other observable geographic characteristics of districts.

Contrary to what the literature suggests, wheat suitability is no longer a significant predictor of sex ratios. Dryland rice suitability, but not wetland rice suitability, significantly predicts a greater share female in all but one age bin. Cotton suitability correlates with a greater share female from adolescence onwards. The correlation between the coastal dummy and the percentage female in the 10–20 age bin is no longer significant. Correlations with ruggedness are no longer significant. Malaria suitability now becomes a negative predictor of the female share in several age bins.

**TABLE 5** Correlations of percentage female with geographic variables in 1901.

	<b>Under 10</b>	<b>10–20</b>	<b>20–40</b>	<b>Over 40</b>	<b>All ages</b>
Latitude	−0.4181	−0.3880	−0.1903	−0.1824	−0.2391
<i>p</i> -Value	0.0000	0.0000	0.0003	0.0005	0.0000
No. of observations	357	357	357	357	357
Longitude	0.3182	0.4567	0.0506	0.0073	0.1411
<i>p</i> -Value	0.0000	0.0000	0.3406	0.8906	0.0076
No. of observations	357	357	357	357	357
Costal	0.1022	0.1177	−0.0737	−0.1120	−0.0536
<i>p</i> -Value	0.0537	0.0262	0.1649	0.0344	0.3122
No. of observations	357	357	357	357	357
Ruggedness	−0.0538	−0.1028	−0.1924	−0.2348	−0.1946
<i>p</i> -Value	0.3107	0.0522	0.0003	0.0000	0.0002
No. of observations	357	357	357	357	357
Major river	0.0094	0.0679	0.0089	−0.0143	0.0347
<i>p</i> -Value	0.8590	0.2007	0.8676	0.7871	0.5135
No. of observations	357	357	357	357	357
Malaria suitability	0.1330	0.2275	−0.1066	−0.1339	−0.0220
<i>p</i> -Value	0.0119	0.0000	0.0441	0.0113	0.6789
No. of observations	357	357	357	357	357
Cotton suitability	0.1204	0.2109	0.2693	0.2619	0.2722
<i>p</i> -Value	0.0229	0.0001	0.0000	0.0000	0.0000
No. of observations	357	357	357	357	357
Wheat suitability	−0.1941	−0.2053	−0.1228	−0.1817	−0.1547
<i>p</i> -Value	0.0002	0.0001	0.0203	0.0006	0.0034
No. of observations	357	357	357	357	357
Dryland rice suitability	0.4266	0.4682	0.1459	0.1699	0.2322
<i>p</i> -Value	0.0000	0.0000	0.0057	0.0013	0.0000
No. of observations	357	357	357	357	357
Wetland rice suitability	0.3758	0.4617	0.1070	0.0241	0.1794
<i>p</i> -Value	0.0000	0.0000	0.0434	0.6499	0.0007
No. of observations	357	357	357	357	357
Tea suitability	0.1513	0.2547	−0.0581	−0.1645	0.0027
<i>p</i> -Value	0.0042	0.0000	0.2735	0.0018	0.9595
No. of observations	357	357	357	357	357

Note: Observations are at the district level. *p*-Values of 0.0000 are less than 0.0001.

The conditional correlations reported in table 6 are based off comparisons of all districts in the data, and thus may be confounded by unobserved determinants of sex ratios that vary across broad regions of South Asia. In table 7 then, we add province fixed effects to our regressions, that is, the regression coefficients reported in this table now capture comparisons across districts within the provinces of colonial India, rather than comparisons across provinces. Because fixed effects can exacerbate attenuation bias due to measurement error, we do expect some coefficients to become insignificant due to this loss of signal.

**TABLE 6** Correlates of percentage female: regression results for 1901.

	Under 10	10–20	20–40	Over 40	All ages
Latitude	−0.346** (0.137)	−0.0890 (0.140)	−0.000879 (0.156)	0.120 (0.146)	−0.00771 (0.168)
Longitude	0.232** (0.0965)	0.388*** (0.0947)	0.101 (0.156)	0.197 (0.148)	0.159 (0.164)
Ruggedness	−0.0541 (0.174)	−0.00761 (0.191)	−0.0171 (0.163)	−0.0730 (0.157)	−0.0322 (0.192)
Malaria suitability	−0.152*** (0.0520)	−0.183** (0.0759)	−0.278** (0.116)	−0.336*** (0.0928)	−0.245** (0.114)
Major river	0.180 (0.115)	0.278** (0.127)	0.358** (0.156)	0.372** (0.144)	0.395*** (0.153)
Coastal	−0.129 (0.101)	0.0161 (0.107)	−0.207 (0.180)	−0.280* (0.169)	−0.236 (0.175)
Cotton suitability	−0.00718 (0.0688)	0.198*** (0.0739)	0.301** (0.138)	0.299** (0.131)	0.299** (0.145)
Wheat suitability	0.119 (0.136)	−0.148 (0.132)	−0.207 (0.157)	−0.221 (0.147)	−0.198 (0.166)
Wetland rice suitability	−0.00676 (0.116)	−0.0830 (0.122)	0.00261 (0.118)	−0.158 (0.106)	−0.0587 (0.120)
Dryland rice suitability	0.227*** (0.0722)	0.259*** (0.0763)	0.157 (0.0973)	0.375*** (0.0903)	0.238*** (0.0877)
Tea suitability	−0.0137 (0.0487)	0.0946* (0.0530)	−0.0487 (0.0936)	−0.135* (0.0816)	−0.0132 (0.0890)
Observations	357	357	357	357	357
R <sup>2</sup>	0.264	0.357	0.176	0.249	0.201

Note: Observations are at the district level. \*\*\* Significant at 1%, \*\* significant at 5%, \* significant at 10%. Robust standard errors in parentheses. All regressions contain a constant that is not reported.

Despite this caveat, there are meaningful patterns that emerge in table 7. First, latitude continues to predict lower under-10 female shares, even conditional on province-fixed effects and other geographic observables. The correlation between longitude and the share female, however, weakens, becoming significant only at the 10 per cent level in one column and losing significance in the others, and the East–West gradient in sex ratios is a gradient that exists across broad regions, more than within them. Malaria suitability continues to correlate negatively with the share female. Dryland rice suitability remains a positive predictor in all but one age bin.

The theories that link the sex ratio to agricultural conditions do so by linking agricultural practices to the economic role of women. We use table 8 to demonstrate that there is a strong correlation between female labour force participation and the share of the population that is female. We use data from Fenske et al., who estimate the female labour force participation rate by dividing the number of women reported as working in the occupational section of the census by the total female population in a district.<sup>82</sup> These data cover the directly ruled districts of India

<sup>82</sup> Fenske et al., ‘Demographic shocks’. Female labour force participation includes women’s participation not only in agriculture, but also industry and services.

**TABLE 7** Correlates of percentage female: regression results for 1901 with province fixed effects.

	<b>Under 10</b>	<b>10–20</b>	<b>20–40</b>	<b>Over 40</b>	<b>All ages</b>
Latitude	−0.583** (0.286)	−0.316 (0.318)	−0.201 (0.260)	−0.165 (0.242)	−0.280 (0.293)
Longitude	0.445* (0.251)	0.318 (0.259)	0.334 (0.236)	0.187 (0.233)	0.327 (0.248)
Ruggedness	−0.0750 (0.237)	−0.0770 (0.271)	−0.0805 (0.208)	−0.177 (0.193)	−0.139 (0.247)
Malaria suitability	−0.103* (0.0586)	−0.165* (0.0899)	−0.283*** (0.107)	−0.363*** (0.0805)	−0.269*** (0.102)
Major river	0.241** (0.115)	0.253* (0.136)	0.356** (0.152)	0.285** (0.135)	0.344** (0.144)
Coastal	−0.107 (0.0921)	0.0448 (0.0879)	−0.0875 (0.121)	−0.0562 (0.118)	−0.0506 (0.0967)
Cotton suitability	−0.0245 (0.0771)	0.156** (0.0752)	0.172* (0.0908)	0.0686 (0.0783)	0.117 (0.0833)
Wheat suitability	0.261 (0.225)	−0.0490 (0.248)	−0.142 (0.217)	0.0429 (0.194)	−0.0230 (0.238)
Wetland rice suitability	−0.127 (0.152)	−0.138 (0.163)	−0.0435 (0.137)	−0.199 (0.125)	−0.119 (0.149)
Dryland rice suitability	0.159* (0.0966)	0.197** (0.0984)	0.112 (0.120)	0.356*** (0.113)	0.181* (0.108)
Tea suitability	−0.00586 (0.0607)	0.126* (0.0683)	−0.0406 (0.0990)	−0.157** (0.0781)	−0.0171 (0.0824)
Observations	357	357	357	357	357
R <sup>2</sup>	0.410	0.460	0.445	0.514	0.500

Note: Observations are at the district level. \*\*\* Significant at 1%, \*\* significant at 5%, \* significant at 10%. Robust standard errors in parentheses. All regressions contain a constant that is not reported.

for the census years between 1901 and 1931. For each census year, we compute the correlation between the population that is female and this measure of women's role in the labour market. In each census year, there is a positive and significant correlation between the two variables. Female labour force participation is computed from the occupation censuses and is not narrowly focused on women's participation in rice and wheat cultivation.

Similarly, we use table 8 to present the share of the population that is female as correlating significantly with three other variables of interest – literacy, child marriage, and urbanization. We compute literacy as the share of the total population that is literate using data from Chaudhary and Fenske,<sup>83</sup> child marriage as the percentage of girls under 15 who are married using data from the censuses, and urbanization as the percentage of the population living in cities with more than 5000 persons using data from Fenske et al.<sup>84</sup> We do not have literacy or urbanization for all districts and or years. Whilst greater literacy correlates negatively with the share of the population that is

<sup>83</sup> Chaudhary and Fenske, 'Railways'.

<sup>84</sup> Fenske et al., 'Railways'.



**TABLE 8** Correlation of percentage female with female labour force participation, literacy, child marriage, and urbanization.

	Percentage female: all ages					
	1881	1891	1901	1911	1921	1931
Female labour force participation			0.4064	0.4896	0.4409	0.3129
<i>p</i>			0.0000	0.0000	0.0000	0.0000
<i>N</i>			191	198	198	197
Literacy	-0.3444	-0.4912	-0.4398	-0.4131	-0.4417	
<i>p</i>	0.0000	0.0000	0.0000	0.0000	0.0000	
<i>N</i>	188	188	195	202	200	
Child marriage	0.2054	0.1493	0.1425	0.1290	0.1466	0.1670
<i>p</i>	0.0003	0.0074	0.0071	0.0127	0.0030	0.0002
<i>N</i>	309	321	356	372	408	506
Urbanization	-0.2137	-0.1952	-0.2032	-0.2901	-0.1952	-0.2648
<i>p</i>	0.0002	0.0004	0.0001	0.0000	0.0001	0.0000
<i>N</i>	290	320	357	380	408	506

Note: Observations are at the district level;  $p = 0.0000$  reported if  $p < 0.0005$ .

female, child marriage is negatively correlated with the share of women in the population. Cities are disproportionately male.

To summarize, these correlations and regression results confirm that Northern India shows the lowest share of women in the population. Eastern and Southern India have higher shares female. Wheat suitability is an inconsistent predictor of the female share. Wetland rice suitability is a poor predictor of the female share. Dryland rice suitability is correlated positively with the share of the population that is female. Malaria consistently predicts fewer women.

The rice–wheat divide in the literature has often been made on the basis of broad regional comparisons and without adjusting for possible confounders. Once these are adjusted for, the wheat–rice dichotomy appears to be overstated, particularly for wetland rice. By contrast, the relevance of other factors, including cotton and malaria, has been relatively overlooked.

## VI | AGE AND SEX RATIOS IN COLONIAL INDIA

When do women go missing? In this section, we describe how sex ratios evolve across age bins, both by region and by religion.

*Literature:* Dyson provides estimates that life expectancy in the territory of today's India rose from 26.3 to 29.6 over the period 1891 to 1931, whilst for women the comparable increase was from 27.2 to 30.1.<sup>85</sup> Unlike the evidence from recent data, which finds the sex ratio to be unnaturally male biased at birth in some regions, the evidence from colonial censuses shows a male deficit in the age group 0–1. This became a female deficit by ages 5–10.<sup>86</sup> Life expectancy at birth moved in favour of male babies only from 1921, as mortality declined faster for male babies than female

<sup>85</sup> See tables 7.1 and 8.1 in Dyson, *Population history*.

<sup>86</sup> Bhaskar and Gupta, 'Missing girls'.



babies.<sup>87</sup> Although the regional differences in overall sex ratios have been documented,<sup>88</sup> we know little about regional variations in sex ratios by age cohorts and about how these variations change over time. To the best of our knowledge, ours is the first paper to document changes in sex ratios by age group for all of colonial India and thereby throw light on the question of missing women over the life cycle.

Bardhan cites maternal mortality as one explanation of regional differences in female survival rates in India in the 1950s and 1960s, as in states such as Uttar Pradesh and Rajasthan, female marriage ages were lower than the national average, and fertility was higher.<sup>89</sup> This would lead us to expect women to go missing in the 10–20 age bin.<sup>90</sup> Evidence on maternal mortality rates in colonial India is sparse. Most women gave birth at home, particularly in rural India, and were cared for by family members or traditional midwives.<sup>91</sup> The high risk of dying during childbirth in the Madras Presidency was described by an Indian commentator in 1866 in dramatic language: ‘mothers would eat a specially prepared death meal before going into labour, in the presence of her family and friends’.<sup>92</sup> Lang points out that British officials laid the blame on ill-trained traditional midwives, and training midwives in modern methods became a goal. The Countess of Dufferin Fund was created in 1885 to provide maternity services, and in 1914, the colonial government accepted this as a policy. However, the use of these services remained limited.<sup>93</sup>

An investigation of 16 municipalities in the Madras Presidency in the early 1930s revealed a maternal mortality rate of 15.4 per 1000, more than three times the rate in England and Wales at that time. The overall female death rate between ages 15 and 40 was 14.11 per 1000, versus 12.37 for men.<sup>94</sup> An investigation of more than 7000 hospital confinements for childbirth in Madras, Madura, Trichinopoly, and Coimbatore in 1927 and 1928 found similar rates of 17.89 maternal deaths per 1000 births.<sup>95</sup> This was comparable to 16 per 1000 in Bombay hospitals between 1925 and 1929.<sup>96</sup> Most births were at home, without access to the medical interventions available in hospitals. We would therefore expect average maternal mortality to be higher than what was reported in the urban hospitals. Balfour found the age of the mother to be an important factor in maternal mortality as well as neonatal mortality.<sup>97</sup> For every 1000 births, maternal mortality was 51 in Madras Presidency if the mother’s age was under 15 and 31–33 for mothers aged 15–25. Comparable figures for the United States in 1921 as cited in the report were 20 and 6.8, respectively. The number of mothers under the age of 15 was much higher in India, where child marriage was widespread. The study concluded that maternal mortality was highest in the case of first birth for mothers under the age of 15. Poor nutrition of the mother contributed to high maternal and infant mortality.

<sup>87</sup> Visaria and Visaria, ‘Population’; Mayer, ‘Sex ratios’.

<sup>88</sup> Visaria and Visaria, ‘Population’.

<sup>89</sup> Bardhan, ‘Life and death’.

<sup>90</sup> Jayachandran and Lleras-Muney, ‘Life expectancy’, have shown that maternal mortality, by raising life expectancy, can also increase investment in girls.

<sup>91</sup> Muraleedharan, ‘Disease and death’.

<sup>92</sup> Lang, ‘Demon dai’.

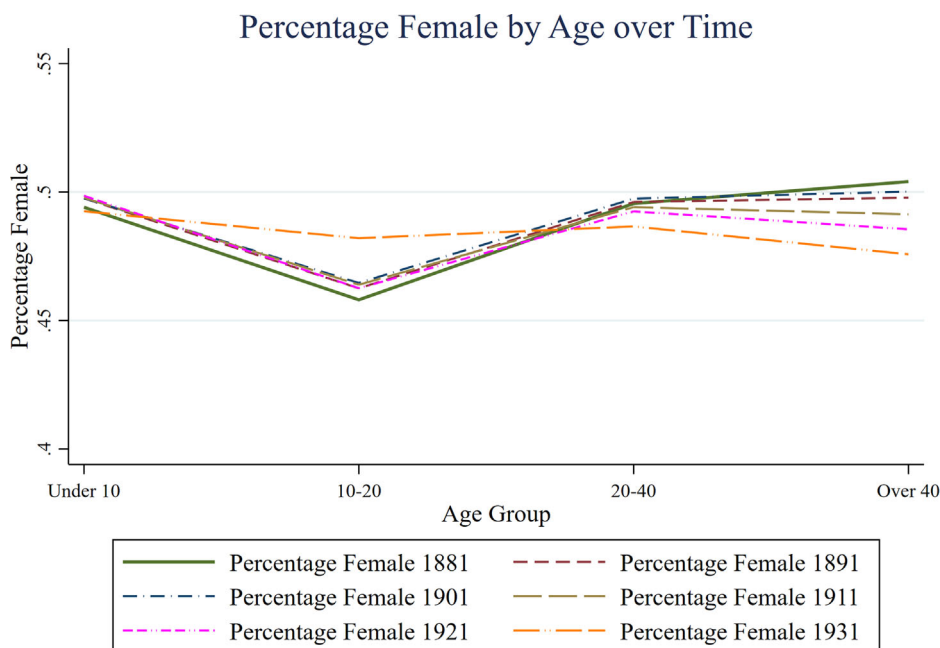
<sup>93</sup> Jat et al., ‘Maternal mortality’.

<sup>94</sup> Nair, ‘Population problem’, p. 244.

<sup>95</sup> Muraleedharan, ‘Disease and death’.

<sup>96</sup> Balfour and Talpade, ‘Maternity conditions’.

<sup>97</sup> Balfour and Talpade, ‘Maternity conditions’.



**FIGURE 10** Percentage female by age and census year.

*Patterns:* In figure 10, we show the fraction of the population that is female for each of our four age bins, in each of our six census years. Note that we aggregate these numbers from our district-level data, so they will differ slightly from comparable totals in the census reports, as will the numbers by age and religion below.

Two broad patterns are apparent. The first is that the female deficit in the population emerges most strikingly after age 10. Whilst even in the 0–10 cohort, the percentage of the population that is female is universally less than 50 per cent, in the 10–20 cohort it falls below 47 per cent in all census years except for 1931. In the cohort aged 20–40, the sex ratio reverted back towards the 50 per cent mark, whilst remaining male biased. In the early census waves, the percentage female for the cohorts aged over 40 is slightly less male biased than that for the cohort aged 20–40. In the years 1911 and later, sex ratios again become more male biased in this over 40 cohort.

The explanations of female deficit in adolescence are as follows: differences in mortality across sexes by age and differential misreporting of age by sex.<sup>98</sup> The deficit of females in the population under 10 was discussed in multiple censuses. The enumerators concluded that, whilst infanticide of girls was practised in isolated communities in Punjab, Sind, Rajasthan, and Cutch, this was not of a magnitude that could impact the overall sex ratio. The mostly likely reason for the female deficit was a lack of care for girls, and therefore, higher female mortality in this age group.<sup>99</sup> The female deficit occurring in adolescence is likely to result in part from maternal mortality in a context where the median age of marriage in 1911 was 12.9 years.<sup>100</sup> The reduction in the female deficit in the working age population can also be attributed in part to famines, in which male

<sup>98</sup> Migration of men for employment within India can explain a deficit of women in adolescence in some districts, but not over the whole of India.

<sup>99</sup> Report of Census of 1901, p. 116.

<sup>100</sup> Bhat and Halli, 'Brideprice and dowry', p. 137.



mortality was higher than female mortality in the working-age population.<sup>101</sup> [Dyson](#) attributes this to greater mobility of men and to the lower likelihood of pregnancy during famines, which reduced maternal mortality.

On misreporting of age, [Dyson](#) notes that the colonial censuses had some ‘female age shifting into the reproductive span’, that is, women misstating their age towards their prime reproductive years.<sup>102</sup> The Census report of 1901 confirms that there were significant reporting issues with respect to adolescent females. There was under-reporting of unmarried girls in the age group 9–15 and of married women of 15–20.<sup>103</sup> The sharp rise in the female deficit in the age group 10–20 and its reversal in the age group 20–30 reflect in part maternal mortality and in part misreporting. In comparison with more recent censuses, this pattern is quite different. The number of missing girls at birth has risen since the 1970s, with the availability of prenatal sex selection technology. [Anderson and Ray](#) find a substantial share of missing women in older age groups.<sup>104</sup>

[Goldin and Lleras-Muney](#) find that in Massachusetts in the late nineteenth century, female mortality from infectious disease was higher than male mortality.<sup>105</sup> Such detailed mortality information is not available for India in the early twentieth century. Evidence from the sanitary report of 1914 shows mortality for disaggregated age bins at the provincial level. Female mortality across all provinces, except in the North, was lower than male mortality in the age bin 5–10. Excess female mortality emerged from the age bin 10–15 onwards. The sanitary report from Punjab in 1914 reported male and female mortality by age group for infectious diseases. However, there is no evidence of systematically higher female mortality from infectious diseases even in this province, where female mortality was generally higher than male mortality over the life cycle.<sup>106</sup>

A second pattern that is apparent is that the changes in sex ratios across cohorts becomes dampened over time. In each successive census wave, the drop in the percentage female between the 0–10 and 10–20 cohorts is less pronounced. In 1931, the age profile of the share female is visibly flatter than in other census years. These changes over time likely reflect a weakening of the processes noted above – disproportionate female mortality in adolescence, and under-reporting of female adolescents.<sup>107</sup>

It is unlikely that a reduction in child marriage in 1931 explains the flattening of the age profile in 1931. [Dyson](#) estimates that the mean age of marriage for women was 12.7 in both 1921 and 1931.<sup>108</sup> [Roy and Tam](#) have shown that the 1929 announcement of the Child Marriage Restraint Act raised rates of child marriage in anticipation of its coming into force 6 months later.<sup>109</sup> Anecdotal evidence suggests this rush to marry girls was urgent enough that some high-caste girls married lower-caste men and men with disabilities.<sup>110</sup>

<sup>101</sup> [Dyson](#), ‘Asian famines’; [Lardiniois](#), ‘Famine, epidemics and mortality’.

<sup>102</sup> [Dyson](#), ‘Demography of Berar’, p. 165.

<sup>103</sup> Census of 1901, p. 115

<sup>104</sup> [Anderson and Ray](#), ‘Missing women’; [Anderson and Ray](#), ‘Age distribution’.

<sup>105</sup> [Anderson and Ray](#), ‘Age distribution’.

<sup>106</sup> Sanitary Report India, 1914 and Punjab 1914.

<sup>107</sup> If we restrict the 1931 sample to only those districts present in 1921 (not reported), the dampening is very similar and thus is not driven by a change in sample composition. The major changes in composition are the Central India Agency and Western India States Agency being divided into more districts in 1931 than in 1921.

<sup>108</sup> See table 8.1 in [Dyson](#), *Population history*.

<sup>109</sup> [Roy and Tam](#), ‘Gender reform’.

<sup>110</sup> [Gupta](#), ‘Son preference’.

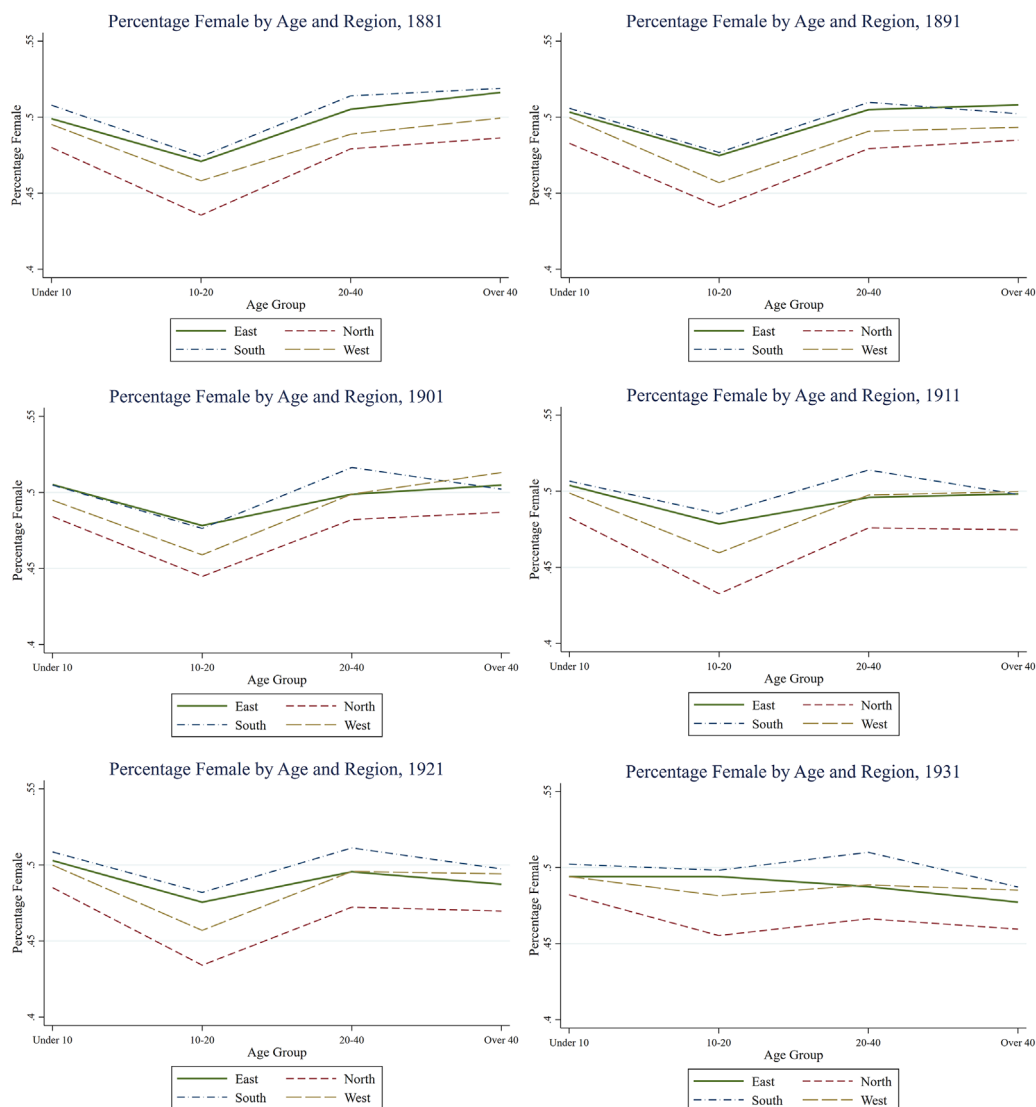


FIGURE 11 Percentage female by age and region.

In figure 11, we show figures analogous to figure 10, but we separate our sample into the four broad regions we consider – East, North, South, and West. Here, three broad patterns are apparent. First, the general trend of sex ratios becoming more male biased in the 10–20 age bin is not confined to any one region of India. Before 1931, a reduction in the share female between the 0–10 and 10–20 cohorts is visible in all four regions. It is still present in all four in 1931, though it has become less pronounced in the Eastern region.

Second, there are pronounced differences across regions. It is in the North that sex ratios are most male biased, and in the South where they are least male-biased. Indeed, in some years both the 0–10 and 20–40 cohorts are greater than 50 per cent female in the Southern region. The reduction in the share of the population that is female between the 20–40 and over 40 cohorts is apparent in several census years in both the North and South, but not in Western and Eastern India.



Third, the dampening of the differences across cohorts in 1931 is not driven by only one region alone. The reduction in the share female between the 0–10 and 10–20 cohorts, as well as the subsequent increase in the share female in the 20–40 cohort, are less pronounced in all four of the major regions.

*Persistence:* A striking finding we have noted is that the low female population share in the 10–20 age bin is common across regions, which we believe has largely gone overlooked. This pattern is consistent with a role for child marriage and maternal mortality in explaining the age-specific sex ratio even in regions where the overall sex ratio is relatively more balanced. In recent years, a significant proportion of missing women are in older age cohorts.<sup>111</sup> Anderson and Ray, for example, find that relative mortality due to cardiovascular disease in older women is higher than in developed countries.<sup>112</sup>

*Patterns:* In figure 12, we show graphs similar to figure 10, except that we now separate the sex ratio trends across age bins by religion, rather than by region. A number of key patterns stand out. The first is that there are clear differences in sex ratios across religions that are apparent throughout the life cycle. Sikhs have, in all census waves, exceptionally male populations in all age bins. In 1881 and 1911, for example, the percentage female is 45 per cent or less in all age bins. Hindus, Muslims, and Jains all have male-biased sex ratios, whilst for Tribal populations and Buddhists, sex ratios are less male biased. Indeed, for Buddhists, the share female is often greater than 50 per cent.

Second, the general pattern of a lower share female in the 10–20 age bin than in either the 0–10 or 20–40 bin is not driven by any single religion. In several years, this dip in share female in the 10–20 age bin is visible for Sikhs, Hindus, Muslims, and Jains. Third, the dampening of the differences in the share female across age bins in 1931 is driven by a similar dampening among Jains, Muslims, and Hindus that is not visible for Sikhs.

Third, although the dip in the share of females in the age group 10–20 is reversed in the age group 20–40 in all religious groups, the pattern differs across religions after age 40. For Hindus, the reduced share of women in the over 40 age bin becomes more pronounced in 1931. For Sikhs, this becomes prevalent enough in 1931 that the share female in the over 40 population is even lower than in the adolescent age bin. The share of the population that is female declined for Muslims and Tribals. Whilst male-biased sex ratios are present in early stages of the life cycle for Hindus, Muslims, and Sikhs, for Tribals this appears to be the case only in later life. Higher mortality for men from famines and from diseases such as tuberculosis may explain more favourable sex ratios for females after age 40.

*Persistence:* In figure 13, we show the equivalent of figure 12 for 2011. There are both similarities and differences with patterns in the colonial data. First, the percentage of the population that is female in the 0–9 age bin is lower today than it was nearly a century ago. Women are more likely to go missing at birth or in early childhood in the present than they were in the past. Second, whilst some religious groups experience a reduction in the share female in the 10–19 age bin, this is not apparent for all religious groups and is much less pronounced than in the colonial data. Third, whilst there has been some persistence in the relative ordering of religious groups by percentage female – note that Sikhs still have the most male-biased sex ratios today – there have been reversals. Muslims in particular have higher percentages female in the age ranges from 0–39 than the other religious groups shown, including Buddhists.

<sup>111</sup> Calvi, 'Older women'.

<sup>112</sup> Anderson and Ray, 'Missing women'; Anderson and Ray, 'Age distribution'.

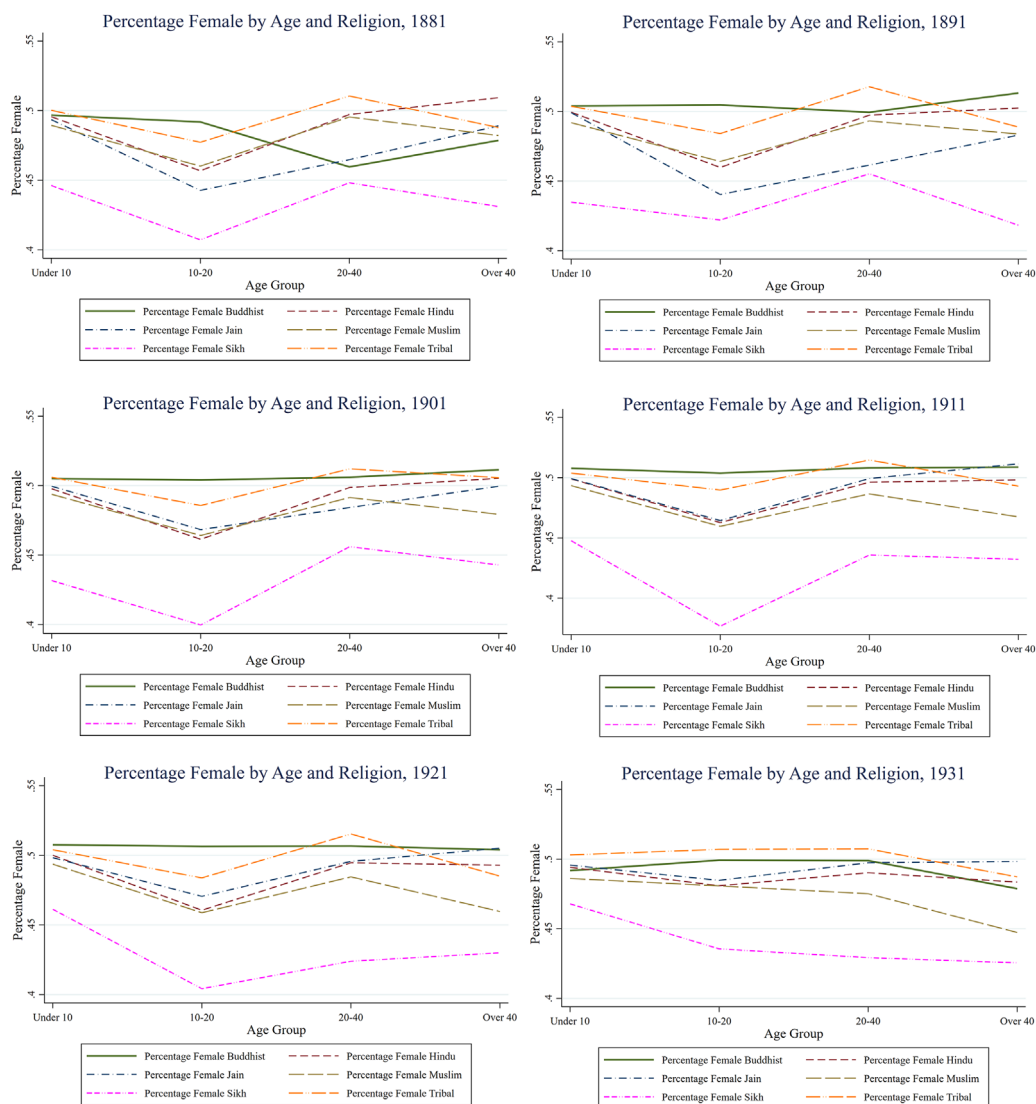
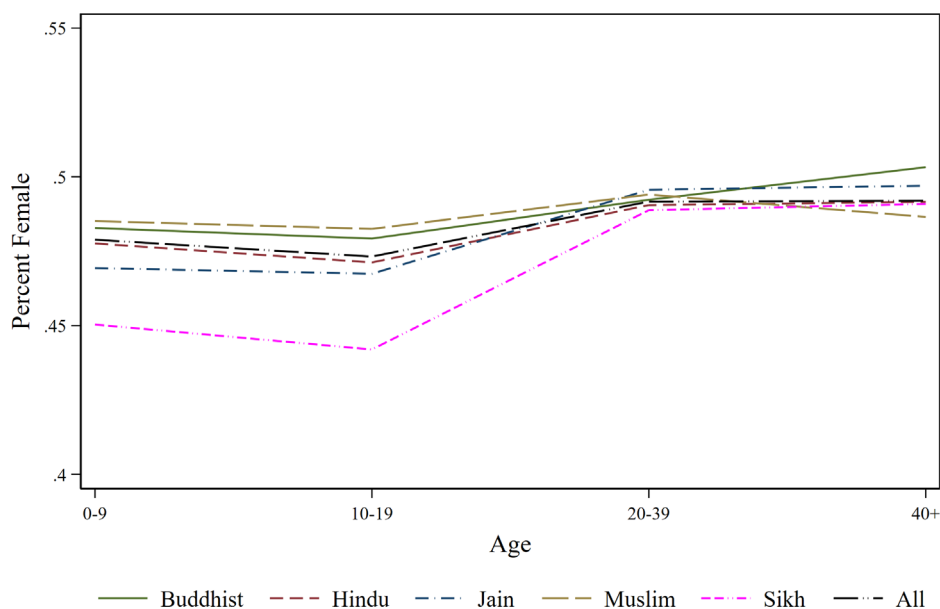


FIGURE 12 “Percentage female by age and religion”.

## VII | CONCLUSION

In this paper, we have introduced a novel dataset at the district level covering sex ratios in colonial South Asia between 1881 and 1931. We have shown that the familiar pattern of more male-biased sex ratios in Northern India and more balanced sex ratios in Southern India and modern Burma go back more than a century, and are visible in granular, district-level data. Across districts, colonial sex ratios correlate strongly with present-day sex ratios, particularly at younger ages, and even comparing districts within the same modern state. Sex ratios generally became more male-biased in adolescence, a pattern that is visible across several regions, religions, and census waves. Sex ratios are most male biased among Sikhs, but the fraction female is less than 50 per cent among Hindus, Muslims, and Jains. Female shares correlate across religious groups over space. Whilst female shares correlate with the suitability for rice and wheat cultivation, many



**FIGURE 13** “Percentage female by age and religion, 2011”.

of these correlations become weaker when adjusting for confounders and looking within, rather than across, regions of South Asia.

The existing literature has provided many explanations of male-biased sex ratios in India, and our descriptive results provide an opportunity to reevaluate these theories in light of evidence from the colonial period. Whilst many of these theories are rooted in a preference for sons, this preference is only a proximate cause of missing women and requires explanation. Some proposed causes, such as the decline in fertility since the 1960s, are unique to the post-colonial period,<sup>113</sup> and cannot account for sex ratios in India’s past. Other explanations are, however, relevant to both the past and present.

The persistence we find in sex ratios, not only across broad regions but also within states, is indicative that slowly changing, ‘deep’ determinants of sex ratios do indeed matter. This persistence is, however, strongest at younger ages and has been declining over time, suggesting that newer explanations are needed to fully account for missing women at older ages in the present. Other patterns we find are also suggestive of changes in the causes of missing women over the twentieth century, including the dampening of the female deficit in adolescence, the relatively male-skewed sex ratios for Muslims in colonial India, and the relatively more equal sex ratios for Buddhists in the past.

The first set of explanations considers religion and social norms. The literature has stressed religious reasons for Hindu son preference and the particularly skewed sex ratios among the upper castes. We do not find, however, that differences across religions are the same in all regions. Whilst the male-biased sex ratios we find among Sikhs and Hindus are consistent with these explanations, it is clear that both Muslim and Buddhist sex ratios have diverged from those in the colonial era. One explanation is that attitudes towards the modern technology of abortion are critical in understanding modern outcomes, but not colonial patterns. That we find strong correlations over

<sup>113</sup> Clark, ‘Son preference’.



space across religions, at younger age bins, in both the past and present, suggests that there are common determinants of the sex ratio across religions.

The second set of explanations considers the economic role of women. Much of this literature focuses on the cultivation of rice and wheat, and its consequences for female labour force participation, as a possible explanation for the North–South gradient in sex ratios. Related work stresses the importance of deep tillage. Our evidence here has been mixed. We do find that sex ratios correlate with female labour force participation. We do not, however, find that either cropping patterns or their deeper determinants are robust predictors of colonial sex ratios. This is particularly true for wheat and for wetland rice.

The third set of explanations is those that are absent from the literature. We have found a number of robust correlates of sex ratios, including suitability for malaria, suitability for cotton, and the presence of a major river, that have gone largely overlooked in conventional accounts. Existing explanations are thus likely to be incomplete.

There are, however, a number of limitations to our approach. Because we use the censuses of colonial India as our underlying source, we are unable to push our estimates back in time further than the second half of the nineteenth century. There will be gaps in our coverage due to the gaps in the original sources: not all districts are covered in all years, not all religions are covered in all districts, and the inconsistencies in the age ranges reported in the census compel us to use more aggregate age bins in our analysis than would be ideal. Further, any deficiencies of the original data collection process, such as the misreporting of women by age, will affect our data. It is our hope that later historians and demographers will uncover additional sources that will fill the gaps in our data, and will develop techniques permitting them to produce plausible corrections for the errors in the original sources.

## ACKNOWLEDGEMENTS

The authors are grateful to Sonia Bhalotra, Sahar Parsa, Debraj Ray, and seminar audiences at the Economic History Society Meetings, the University of Bergamo, Humboldt University, Trinity College Dublin, the University of Nottingham, the University of Oxford, the University of Warwick, and the Warwick-SUFE Workshop 2021 for their comments, and to the British Academy/Leverhulme Small Research Grants scheme for funding. The authors also thank Claudia Goldin for her insightful discussion of the paper in the virtual economic history seminar series and the editor Catherine Schenk and three anonymous referees for their valuable comments.

## ORCID

Cora Neumann  <https://orcid.org/0000-0003-1364-4975>

## REFERENCES

- Agnihotri, S., Palmer-Jones, R. and Parikh, A., 'Missing women in Indian districts: a quantitative analysis', *Structural Change and Economic Dynamics*, 13 (2002), pp. 285–314.
- Alesina, A., Giuliano, P. and Nunn, N., 'On the origins of gender roles: women and the plough', *Quarterly Journal of Economics*, 128 (2013), pp. 469–530.
- Alesina, A., Giuliano, P. and Nunn, N., 'Traditional agricultural practices and the sex ratio today', *PLoS ONE*, 13 (2018), e0190510.
- Alsan, M., 'The effect of the tsetse fly on African development', *American Economic Review*, 105 (2015), pp. 382–410.
- Anderson, S. and Ray, D., 'Missing women: age and disease', *Review of Economic Studies*, 77 (2010), pp. 1262–1300.
- Anderson, S. and Ray, D., 'The age distribution of missing women in India', *Economic and Political Weekly*, 47 (2012), pp. 87–95.



- Ashraf, N., Bau, N., Nunn, N. and Voena, A., 'Bride price and female education', *Journal of Political Economy*, 128 (2020), pp. 591–641.
- Austin, G., 'The "reversal of fortune" thesis and the compression of history: perspectives from African and comparative economic history', *Journal of International Development: The Journal of the Development Studies Association*, 20 (2008), pp. 996–1027.
- Balfour, M. I. and Talpade, S. K., 'The maternity conditions of women mill-workers in India', *Indian Medical Gazette*, 65 (1930), pp. 241–249.
- Bardhan, P. K., 'On life and death questions', *Economic and Political Weekly*, 9 (1974), pp. 1293–1304.
- Basu, A. M., 'Fertility decline and increasing gender imbalance in India, including a possible South Indian turnaround', *Development and Change*, 30 (1999), pp. 237–263.
- Behrman, J. R., Foster, A. D., Rosenzweig, M. R. and Vashishtha, P., 'Women's schooling, home teaching, and economic growth', *Journal of Political Economy*, 107 (1999), pp. 682–714.
- BenYishay, A., Jones, M., Kondylis, F., and Mobarak, A. M., 'Gender gaps in technology diffusion', *Journal of Development Economics*, 143 (2020), 102380.
- Bhalotra, S., Clots-Figueras, I. and Iyer, L., 'Religion and abortion: the role of politician identity', *Journal of Development Economics*, 153 (2021), 102746.
- Bhalotra, S., Valente, C., and Van Soest, A., 'The puzzle of Muslim advantage in child survival in India', *Journal of Health Economics*, 29 (2010), pp. 191–204.
- Bharadwaj, P. and Lakdawala, L. K., 'Discrimination begins in the womb: evidence of sex-selective prenatal investments', *Journal of Human Resources*, 48 (2013), pp. 71–113.
- Bharadwaj, P., Khwaja, A. and Mian, A., 'The big march: migratory flows after the partition of India', *Economic and Political Weekly*, 43 (2008), pp. 39–49.
- Bhaskar, V. and Gupta, B., 'India's missing girls: biology, customs, and economic development', *Oxford Review of Economic Policy*, 23 (2007), pp. 221–238.
- Bhat, P. N. M. and Halli, S. S., 'Demography of brideprice and dowry: causes and consequences of the Indian marriage squeeze', *Population Studies*, 53 (1999), pp. 129–148.
- Bordalo, P., Coffman, K., Gennaioli, N. and Shleifer, A., 'Beliefs about gender', *American Economic Review*, 109 (2019), pp. 739–73.
- Borooh, V. K. and Iyer, S., 'Religion, literacy, and the female-to-male ratio', *Economic and Political Weekly*, 40 (2005), pp. 419–427.
- Borooh, V. K. and Iyer, S., 'Vidya, Veda, and Varna: the influence of religion and caste on education in rural India', *Journal of Development Studies*, 41 (2005), pp. 1369–1404.
- Borooh, V. K., Do, Q.-T., Iyer, S. and Joshi, S., 'Missing women and India's religious demography', World Bank Policy Research Working Paper (5096) (2009).
- Boserup, E., *Woman's role in economic development* (1970).
- Calvi, R., 'Why are older women missing in India? The age profile of bargaining power and poverty', *Journal of Political Economy*, 128 (2020), pp. 2453–2501.
- Carranza, E., 'Soil endowments, female labor force participation, and the demographic deficit of women in India', *American Economic Journal: Applied Economics*, 6 (2014), pp. 197–225.
- Cervellati, M., Esposito, E. and Sunde, U., 'Epidemic shocks and civil violence: evidence from malaria outbreaks in Africa', *Review of Economics and Statistics*, 104 (2022), pp. 780–796.
- Chakraborty, T. and Kim, S., 'Kinship institutions and sex ratios in India', *Demography*, 47 (2010), pp. 989–1012.
- Chandavarkar, R., *The origins of industrial capitalism in India: business strategies and the working classes in Bombay, 1900–1940* (Cambridge, 2003).
- Chandramouli, C., Singh, A. P. and Sethi, R. C., *Census of India 2011: Administrative Atlas of India* (Office of the Registrar General & Census Commissioner, India, Ministry of Home Affairs, Government of India, 2011).
- Chaudhary, L. and Fenske, J., 'Railways, development, and literacy in India', *Journal of Economic History*, 83 (2023), pp. 1139–1174.
- Chima, R. I., Goodman, C. A. and Mills, A., 'The economic impact of malaria in Africa: a critical review of the evidence', *Health Policy*, 63 (2003), pp. 17–36.
- Clark, S., 'Son preference and sex composition of children: evidence from India', *Demography*, 37 (2000), pp. 95–108.
- Collins, W. J., 'Labor mobility, market integration, and wage convergence in late 19th century India', *Explorations in Economic History*, 36 (1999), pp. 246–277.



- Crosen, R. and Gneezy, U., 'Gender differences in preferences', *Journal of Economic Literature*, 47 (2009), pp. 448–74.
- Das Gupta, M., 'Explaining Asia's "missing women": a new look at the data', *Population and Development Review*, 31 (2005), pp. 529–535.
- Das Gupta, M. and Bhat, P. N. M., 'Fertility decline and increased manifestation of sex bias in India', *Population Studies*, 51 (1997), pp. 307–315.
- De Moor, T. and Van Zanden, J. L., 'Girl power: the European marriage pattern and labour markets in the North Sea region in the late medieval and early modern period', *Economic History Review*, 63 (2010), pp. 1–33.
- Desai, M., Ter Kuile, F. O., Nosten, R., McGready, R., Asamoah, K., Brabin, B. and Newman, R. D., 'Epidemiology and burden of malaria in pregnancy', *The Lancet Infectious Diseases*, 7 (2007), pp. 93–104.
- Dimico, A., Isopi, A. and Olsson, O., 'Origins of the Sicilian mafia: the market for lemons', *Journal of Economic History*, 77 (2017), pp. 1083–1115.
- Drenstedt, G. L., Crimmins, E. M., Vasunilashorn, S. and Finch, C. E., 'The rise and fall of excess male infant mortality', *Proceedings of the National Academy of Sciences of the United States of America*, 105 (2008), pp. 5016–5021.
- Dupraz, Y., 'French and British colonial legacies in education: evidence from the partition of Cameroon', *Journal of Economic History*, 79 (2019), pp. 628–668.
- Dyson, T., 'The historical demography of Berar, 1881–1980', in T. Dyson, ed., *India's historical demography* (1989), pp. 150–196.
- Dyson, T., 'On the demography of South Asian famines: part I', *Population Studies*, 45 (1991), pp. 5–25.
- Dyson, T., *A population history of India: from the first modern people to the present day* (Oxford, 2018).
- Dyson, T. and Moore, M., 'On kinship structure, female autonomy, and demographic behavior in India', *Population and Development Review*, 9 (1983), pp. 35–60.
- Fenske, J. and Kala, N., 'Linguistic distance and market integration in India', *Journal of Economic History*, 81 (2021), pp. 1–39.
- Fenske, J., Gupta, B. and Yuan, S., 'Demographic shocks and women's labor market participation: evidence from the 1918 influenza pandemic in India', *Journal of Economic History*, 82 (2022), pp. 875–912.
- Fenske, J., Kala, N. and Wei, J., 'Railways and cities in India', *Journal of Development Economics*, 161 (2023), 103038.
- Giuliano, P., 'Female labor force participation: persistence and evolution', in S. N. Durlauf and L. E. Blume, eds., *The new Palgrave dictionary of economics* (2014), pp. 1–9.
- Goldin, C., 'The U-shaped female labor force function', in T. P. Schultz, ed., *Investment in women's human capital* (Chicago and London, 1995), pp. 61–90.
- Goldin, C. and Lleras-Muney, A., 'XX>XY?: the changing female advantage in life expectancy', *Journal of Health Economics*, 67 (2019), 102224.
- Guilmoto, C. Z., Saikia, N., Tamrakar, V. and Bora, J. K., 'Excess under-5 female mortality across India: a spatial analysis using 2011 census data', *Lancet Global Health*, 6 (2018), e650–e658.
- Gupta, B., 'Where have all the brides gone? Son preference and marriage in India over the twentieth century', *Economic History Review*, 67 (2014), 1–24.
- Gupta, B., *An economic history of India: from the Mughal Empire to modern times* (Book manuscript, Under contract with Cambridge University Press, 2025).
- Gupta, B. and Swamy, A. V., 'Reputational consequences of labor coercion: evidence from Assam's tea plantations', *Journal of Development Economics*, 127 (2017), pp. 431–439.
- Harnetty, P., 'Cotton exports and Indian agriculture, 1861–1870', *Economic History Review*, 2nd ser., XXIV (1971), pp. 414–429.
- Jat, T. R., Deo, P. R., Goicolea, I., Hurtig, A.-K. and San Sebastian, M., 'Socio-cultural and service delivery dimensions of maternal mortality in rural central India: a qualitative exploration using a human rights lens', *Global Health Action*, 8 (2015), 24976.
- Jayachandran, S., 'The roots of gender inequality in developing countries', *Annual Review of Economics*, 7 (2015), pp. 63–88.
- Jayachandran, S., 'Fertility decline and missing women', *American Economic Journal: Applied Economics*, 9 (2017), pp. 118–39.
- Jayachandran, S. and Kuziemko, I., 'Why do mothers breastfeed girls less than boys? Evidence and implications for child health in India', *Quarterly Journal of Economics*, 126 (2011), pp. 1485–1538.



- Jayachandran, S. and Lleras-Muney, A., 'Life expectancy and human capital investments: evidence from maternal mortality declines', *Quarterly Journal of Economics*, 124 (2009), pp. 349–397.
- Jayachandran, S. and Pande, R., 'Why are Indian children so short? The role of birth order and son preference', *American Economic Review*, 107 (2017), pp. 2600–2629.
- Jha, S., 'Trade, institutions, and ethnic tolerance: evidence from South Asia', *American Political Science Review*, 107 (2013), pp. 806–832.
- Kishor, S., "'May God give sons to all": gender and child mortality in India', *American Sociological Review*, 58 (1993), pp. 247–265.
- Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Ehrlich Sachs, S. and Sachs, J., 'A global index representing the stability of malaria transmission', *American Journal of Tropical Medicine and Hygiene*, 70 (2004), pp. 486–498.
- Korn, E., 'On the formation of family structures', *Public Choice*, 105 (2000), pp. 357–372.
- Lang, S., 'Drop the demon dai: maternal mortality and the state in colonial Madras, 1840–1875', *Social History of Medicine*, 18 (2005), pp. 357–378.
- Lardinois, R., 'Famine, epidemics and mortality in South India: a reappraisal of the demographic crisis of 1876–1878', *Economic and Political Weekly* (1985), pp. 454–465.
- Lowes, S., 'Kinship Structure and the Family: Evidence from the Matrilineal Belt', Working Paper (University of California San Diego, 2022).
- Luke, N. and Munshi, K., 'Women as agents of change: female income and mobility in India', *Journal of Development Economics*, 94 (2011), pp. 1–17.
- Mayer, P., 'India's falling sex ratios', *Population and Development Review*, 25 (1999), pp. 323–343.
- Michalopoulos, S. and Papaioannou, E., 'Pre-colonial ethnic institutions and contemporary African development', *Econometrica*, 81 (2013), pp. 113–152.
- Miller, B. D., *The endangered sex: neglect of female children in rural north India* (Ithaca, 1981).
- Miller, B. D., 'Female labor participation and female seclusion in rural India: a regional view', *Economic Development and Cultural Change*, 30 (1982), pp. 777–794.
- Morris, D., *The emergence of an industrial labor force in India: a study of the Bombay cotton mills, 1854–1947* (Berkeley, CA, 1965).
- Muraleedharan, V. R., 'Diet, disease and death in colonial South India', *Economic and Political Weekly*, 29 (1994), pp. 55–63.
- Nair, R., 'The construction of a "population problem" in colonial India 1919–1947', *Journal of Imperial and Commonwealth History*, 39 (2011), pp. 227–247.
- Narayan, S., 'Religion and female–male ratios in India', *Indian Journal of Human Development*, 12 (2018), pp. 441–452.
- Niederle, M. and Vesterlund, L., 'Do women shy away from competition? Do men compete too much?', *Quarterly Journal of Economics*, 122 (2007), pp. 1067–1101.
- Nunn, N. and Puga, D., 'Ruggedness: the blessing of bad geography in Africa', *Review of Economics and Statistics*, 94 (2012), pp. 20–36.
- Ojha, S., 'Regulating work: decasualisation of dock labour in Colonial India', *Social Scientist*, 42 (2014), pp. 77–89.
- Padmanabha, P., 'Mortality in India: a note on trends and implications', *Economic and Political Weekly* (1982), pp. 1285–1290.
- Pande, R. P., 'Selective gender differences in childhood nutrition and immunization in rural India: the role of siblings', *Demography*, 40 (2003), pp. 395–418.
- Qian, N., 'Missing women and the price of tea in China: the effect of sex specific earnings on sex imbalance', *Quarterly Journal of Economics*, 123 (2008), pp. 1251–1285.
- Rose, E., 'Consumption smoothing and excess female mortality in rural India', *Review of Economics and Statistics*, 81 (1999), pp. 41–49.
- Rosenzweig, M. R. and Schultz, T. P., 'Market opportunities, genetic endowments, and intrafamily resource distribution: child survival in rural India', *American Economic Review*, 72 (1982), pp. 803–815.
- Roy, S. and Tam, E. H. F., 'Impact of British colonial gender reform on early female marriages and gender gap in education: evidence from Child Marriage Abolition Act, 1929', Working Paper (University of Oxford, 2021).
- Satya, L. D., *Cotton and famine in Berar, 1850–1900* (New Delhi, 1997).
- Scott, J. C., *The art of not being governed* (New Haven and London, 2009).
- Sen, A., 'More than 100 million women are missing', *New York Review of Books*, 37 (1990), pp. 61–66.



- Sudha, S. and Rajan, S. I., 'Female demographic disadvantage in India 1981–1991: sex selective abortions and female infanticide', *Development and Change*, 30 (1999), pp. 585–618.
- Vanneman, R. and Barnes, D., *Indian district data, 1961–1991: machine readable data file and codebook* (College Park, Maryland: Center on Population, Gender, and Social Inequality, 2000).
- Visaria, A., 'Religion and son preference in India and Bangladesh: Three essays on comparing Hindus and Muslims on son preference and sex differentials in child health', unpub. Ph.D, thesis, Univ. Pennsylvania (2015).
- Visaria, L. and Visaria, P., 'V—population (1757–1947)', in D. Kumar and M. Desai, eds., *The Cambridge Economic History of India*, Vol. 2: c.1757-c.1970 (Cambridge, 1983), pp. 463–532.
- Xue, M. M., 'High-value work and the rise of women: the cotton revolution and gender equality in China', Working Paper (New York University, 2018).

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Fenske, J., Gupta, B., and Neumann, C. 'Missing women in colonial India', *Economic History Review*, 78 (2025), pp. 997–1038.  
<https://doi.org/10.1111/ehr.13413>.

## APPENDIX

### Correlations of religious shares with province fixed effects

**TABLE A1** Correlations of religious shares with province fixed effects.

	(1) Percentage female Buddhist	(2) Percentage female Jain	(3) Percentage female Muslim	(4) Percentage female Sikh	(5) Percentage female Tribal
1881					
Percentage female	0.010	0.805	0.882***	0.761***	0.640**
Hindu	(0.327)	(0.689)	(0.133)	(0.145)	(0.293)
Observations	70	167	301	122	79
Within $R^2$	<0.001	0.0213	0.325	0.0557	0.237
1891					
Percentage female	0.111	1.417***	0.595***	0.547**	0.584**
Hindu	(0.262)	(0.463)	(0.138)	(0.272)	(0.266)
Observations	121	231	314	160	127
Within $R^2$	0.00109	0.0781	0.179	0.0274	0.0676
1901					
Percentage female	-0.036	1.008***	0.465***	0.807***	-0.555*
Hindu	(0.044)	(0.125)	(0.153)	(0.172)	(0.323)
Observations	48	86	333	72	140
Within $R^2$	0.00177	0.654	0.0829	0.122	0.0616

(Continues)



TABLE A1 (Continued)

	(1) Percentage female Buddhist	(2) Percentage female Jain	(3) Percentage female Muslim	(4) Percentage female Sikh	(5) Percentage female Tribal
1911					
Percentage female	-0.244	0.401	0.359**	0.741***	0.152
Hindu	(0.275)	(0.782)	(0.147)	(0.152)	(0.384)
Observations	45	115	350	89	137
Within $R^2$	0.0284	0.00530	0.107	0.126	0.00411
1921					
Percentage female	-0.138	0.864***	0.395**	0.746***	0.158
Hindu	(0.083)	(0.266)	(0.169)	(0.167)	(0.236)
Observations	50	133	385	82	174
Within $R^2$	0.0604	0.0709	0.105	0.173	0.00758
1931					
Percentage female	-0.125	1.569*	0.460***	0.494*	0.546**
Hindu	(0.204)	(0.892)	(0.168)	(0.271)	(0.215)
Observations	10	144	443	62	126
Within $R^2$	0.0382	0.103	0.115	0.0434	0.0327
2011					
Percentage female	0.122	0.187	0.166***	0.571***	
Hindu	(0.115)	(0.164)	(0.046)	(0.103)	
Observations	637	635	637	637	
Within $R^2$	0.00930	0.0154	0.0834	0.124	

Note: Observations are at the district level. \*\*\* Significant at 1%, \*\* significant at 5%, \* significant at 10%. Robust standard errors in parentheses. All regressions contain a constant and province fixed effects that are not reported.