ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 061/26

Sex, Lies and Birth Statistics: The Mysterious Case of the Spanish Missing Women

Manuel Bagues, Carmen Villa

www.rfberlin.com MARCH 2026

Sex, Lies and Birth Statistics: The Mysterious Case of the Spanish Missing Women

Authors

Manuel Bagues, Carmen Villa

Reference

JEL Codes: J16, J13, C18

Keywords: Sex ratio at birth, birth registry, coding errors, missing women in Spain

Recommended Citation: Manuel Bagues, Carmen Villa (2026): Sex, Lies and Birth Statistics: The Mysterious Case of the Spanish Missing Women. RFBerlin Discussion Paper No. 061/26

Access

Papers can be downloaded free of charge from the RFBerlin website: https://www.rfberlin.com/discussion-papers

Discussion Papers of RFBerlin are indexed on RePEc: https://ideas.repec.org/s/crm/wpaper.html

Disclaimer

Opinions and views expressed in this paper are those of the author(s) and not those of RFBerlin. Research disseminated in this discussion paper series may include views on policy, but RFBerlin takes no institutional policy positions. RFBerlin is an independent research institute.

RFBerlin Discussion Papers often represent preliminary or incomplete work and have not been peer-reviewed. Citation and use of research disseminated in this series should take into account the provisional nature of the work. Discussion papers are shared to encourage feedback and foster academic discussion.

All materials were provided by the authors, who are responsible for proper attribution and rights clearance. While every effort has been made to ensure proper attribution and accuracy, should any issues arise regarding authorship, citation, or rights, please contact RFBerlin to request a correction.

These materials may not be used for the development or training of artificial intelligence systems.

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

Sex, Lies and Birth Statistics: The Mysterious Case of the Spanish Missing Women

Manuel Bagues & Carmen Villa*

February 22, 2026

Abstract

Official Spanish birth registry data report sex ratios well above expected levels between 1975 and 2000, peaking at 109 boys per 100 girls in the early 1980s, the highest in the world at that time. Prior research has attributed these elevated ratios to factors such as maternal age, birth order, and differential prenatal care. We show that they instead reflect systematic coding errors by the Spanish Statistical Office. Census data reveal normal sex ratios for the same cohorts. The birth registry also exhibits implausible monthly volatility and asymmetrically distributed outliers, consistent with one-directional miscoding of females as males. Additional corroborating evidence comes from provisional birth statistics, which show significantly lower sex ratios than the finalised records, and from anomalous patterns in adjacent fields on the birth registration form. Our findings underscore the responsibility of statistical agencies to validate administrative records and cross-check them against alternative sources.

JEL Codes: J16, J13, C18

Keywords: Sex ratio at birth, birth registry, coding errors, missing women in Spain.

*Manuel Bagues, Department of Economics, University of Warwick. Carmen Villa, Department of Economics and Jacobs Center for Productive Youth Development at the University of Zurich, and Institute for Fiscal Studies.

1 Introduction

The sex ratio at birth typically falls within a narrow range of 105–106 boys per 100 girls across human societies. Larger ratios are generally considered indicative of sex-selective practices, as documented in parts of Asia and Eastern Europe, as well as among Asian immigrants in Western countries (Dubuc & Coleman 2007, Almond & Edlund 2008, Almond et al. 2013, González, Libertad 2018, Chao et al. 2019). Such imbalances have important implications for marriage and labour markets, drawing considerable research and policy attention (e.g. Angrist 2002, Edlund et al. 2013, Wang et al. 2017, Grosjean & Khattar 2019).

Spain presents a puzzling case. According to official birth statistics, beginning in 1975 it experienced a dramatic spike in sex ratios, reaching 109 boys per 100 girls in the early 1980s. This was the highest ratio in the world at that time, as reported in UN data, which draw on national official records (United Nations 2024). This elevated ratio then declined gradually through the 1980s and 1990s, returning to normal levels by the end of the century. Since sex-selective abortion was unheard of in Spain during this period (Echavarri & Beltrán Tapia 2025), researchers have sought other explanations for the anomaly. Early work argued that increases in maternal age favouring male births may have played a role (Gutiérrez-Adán et al. 2000), while others pointed at factors such as race, birth order, and solar radiation intensity (Sánchez Barricarte 2023). However, none of these forces can explain such a large gap as the one observed in Spain in the late 20th century. An extensive literature has shown that natural variation in the sex ratio due to biological and demographic factors is modest and rarely pushes ratios beyond 107 (Hesketh & Xing 2006, Chao et al. 2019, Jacobsen et al. 1999, Chahnazarian 1988).

More recently, Echavarri & Beltrán Tapia (2025) argue that the Spanish anomaly reflects son preference coupled with the introduction of prenatal sex determination. Once ultrasound technology became widely available through the expanding health system, they contend, mothers who learned they were carrying a boy took better care of themselves during pregnancy, increasing male fetal survival and pushing the sex ratio upward. If true, this interpretation would imply that tens of thousands of female fetuses died in Spain in the final decades of the 20th century due to differential maternal care. Yet the biological plausibility of this mechanism is limited. Fetal sex cannot be reliably determined by ultrasound until at least 12 weeks of gestation (Efrat et al. 1999), by which point only 1–5 percent of pregnancies end in miscarriage (Michels & Tiu 2007). Even under extreme assumptions about differential care, such low baseline rates of fetal loss cannot plausibly generate a 2–3 percentage point deviation from biologically expected sex ratios.

We provide evidence for a simpler explanation: the anomalies observed in Spain from 1975 to roughly 2000 are due to data quality issues in the birth registry rather than real demographic patterns. First, census data from the same cohorts show sex ratios consistently within the normal biological range, with no unusual spike during 1975–2000. Second, the official birth registry data exhibit extreme monthly volatility within provinces that is biologically implausible but consistent with sporadic data entry errors. Third, the distribution of outliers is asymmetric, with far more observations showing implausibly high sex ratios than implausibly low ones, consistent with systematic miscoding of females as males.

We verify that the divergence between the birth registry and census sex ratios cannot be explained by sex differences in child mortality or sex-biased migration. Adjusting census sex ratios for differential male-female child mortality closes only a small fraction of the gap, and the number of native-born individuals enumerated in the census closely matches birth registry totals once child deaths are accounted for, ruling out large-scale differential out-migration.

The nature and origin of the errors suggest they arose during the processing of birth records at the Spanish Statistical Office (*Instituto Nacional de Estadística*, INE). First, for some years (1975 to 1982), preliminary statistics published before the final figures were released show sex ratios closer to normal levels, suggesting that the errors were introduced during the finalisation of the records rather than their initial recording. Second, the errors are not confined to sex data: our analysis of outlier observations shows they are much less likely to contain information on variables that should be routinely recorded, such as birth weight, and previous work has documented large inconsistencies in the birth order variable in the same database (Devolder et al. 2010), consistent with broader data processing failures. Third, the timing of the errors, which begin appearing around 1975, coincides with the introduction of computer systems for data processing at INE, suggesting a mismanaged digital transition.

Our findings have important implications. First, we show that the "missing women" in Spanish birth statistics are a data artifact rather than evidence of biological or behavioural patterns, offering a new perspective on a long-standing demographic puzzle (Gutiérrez-Adán et al. 2000, Sánchez Barricarte 2023, Echavarri & Beltrán Tapia 2025). Second, since these are the figures reported to international aggregators such as the UN Demographic Yearbook, the World Health Organisation, and Eurostat, cross-country studies exploring determinants of sex ratios may be contaminated by erroneous Spanish data (e.g. Grech et al. 2003, Bongaarts & Guilmoto 2015, Alesina et al. 2018, Chao et al. 2019). Last, our work provides a cautionary tale on the reliability of administrative data and recommends that statistical agencies validate and, where necessary, flag or correct records.

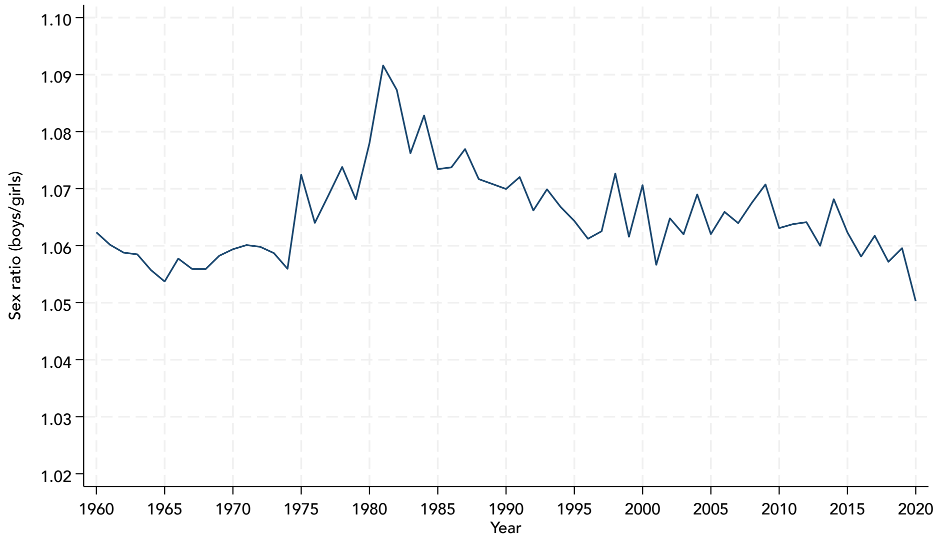
2 Institutional Context

In Spain, all births must be registered at the Civil Registry (*Registro Civil*) within a few days of the date of birth. Registration occurs at the office corresponding to the place of birth, though since 2015 it can be completed directly within hospitals. At the time of registration, parents complete a statistical bulletin recording demographic information about the newborn — including sex — as well as parental characteristics such as age, province of residence, and marital status. This bulletin is separate from the legal registration document and is designed specifically for statistical purposes. We reproduce the form used during 1981–1994 in Appendix A1. Statistical bulletins flow from local Civil Registry offices to regional statistical institutes, which handle initial processing, coding and data cleaning before transmitting records to INE's central offices (Juárez 2014).

INE compiles these bulletins into official birth statistics.[1] Until 1975, data were collected and tabulated manually; computerised systems were introduced that year (Devolder et al. 2010). Individual birth records have been publicly available since 1975, with aggregate statistics compiled in the *Movimiento Natural de la Población* dataset and published annually in INE's *Anuario Estadístico de España*.

[1]Detailed information on these statistics is available at `https://www.ine.es/dynt3/metadatos/en/RespuestaPrint.htm?oper=307`.

Figure 1: Sex ratio in Spain, 1960-2020



Notes: Male to female sex ratio in Spain, as measured by birth registry data from INE (2024*d*).

3 Empirical Analysis

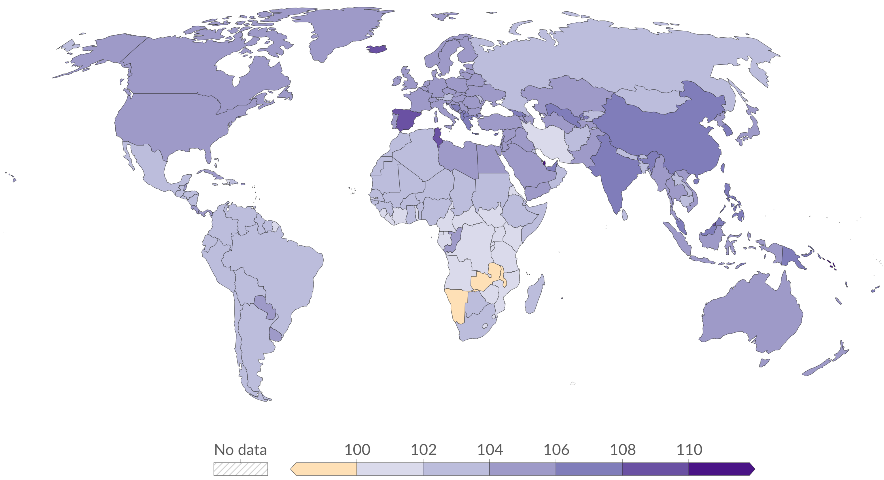
We begin by documenting the anomalous sex ratios in the birth registry and comparing them against census data. We then examine whether the divergence between the two sources can be explained by differential child mortality or sex-biased migration, before turning to direct evidence of coding errors in the microdata. All analyses draw on publicly available data from INE and full details on sources and construction are provided in the Data Appendix (Appendix B).

3.1 Sex ratio in the birth registry

We start by analysing the temporal evolution of the sex ratio in Spain between 1960 and 2020, relying on the aggregate annual figures reported by the INE on the number of male and female births. These data constitute Spain's official birth statistics, are used in demographic research, and are reported to international organisations. As shown in Figure 1, the Spanish sex ratio, as reported in the birth registry, shows a sudden spike after 1975, reaching 109 boys per 100 girls in 1981. These unusually high ratios persist through the 1990s, before gradually declining through the 1980s and 1990s to levels closer to the expected biological range of 105–106, though still occasionally above it.

To place this anomaly in international context, Figure 2 compares Spain's sex ratio with those of other countries using UN data. According to these records, by 1982, Spain had the highest sex ratio at birth among major countries.[2]

Figure 2: Sex ratio at birth, 1982



Source: United Nations Statistics Division (2024).

3.2 Sex ratio in Spain: birth registry vs. census data

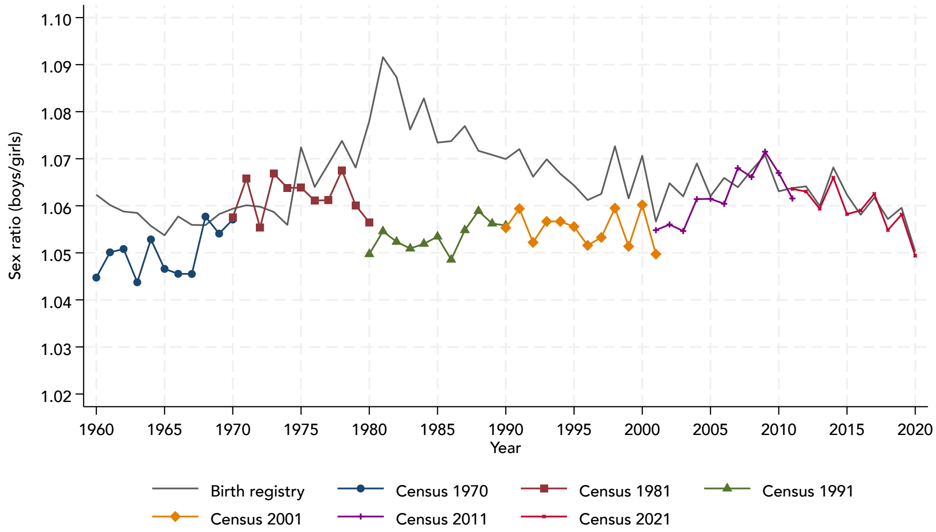
To investigate whether the elevated sex ratios in the birth registry reflect a real demographic phenomenon, we compare them against census data, an independent source that counts the resident population directly rather than relying on birth record processing.

Spanish censuses are conducted every ten years and provide comprehensive information on the resident population, including age and place of birth. We use publicly available census data covering the period 1970–2021. For each census year, we calculate the sex ratio of children at each age, allowing us to construct cohort-specific sex ratios for those born in the preceding decade. The census captures the realised sex ratio in the population. Any divergence between the two series should, in principle, be fully explained by sex differences in mortality or migration in the intervening years, provided both sources are measured without error.

Figure 3 compares sex ratios as measured in the birth registry with those observed in census data for cohorts born between 1960 and 2020. A large divergence between the two

[2]Spain had the highest sex ratio among countries with populations above 250,000. Qatar, San Marino, Solomon Islands, Palau, Brunei, Andorra, and Iceland — all with smaller populations — reported higher ratios. For very small populations the variance of the sex ratio is high and extreme values are more likely to reflect sampling variation than a true demographic pattern.

Figure 3: Sex ratio in Spain, birth registry vs Census, 1960-2020



Notes: Sex ratio in Spain, as measured by birth registry data (2024*d*) and Census (2025).

sources is evident from approximately 1975 to 2000. While the birth registry reports sex ratios well above the expected range of 105–106 boys per 100 girls, peaking at 109 boys per 100 girls in 1981, census data from the same cohorts show sex ratios much closer to the expected biological range – 105 to 106. The two data sources align closely both before 1975 and after 2000, suggesting the problem is not systematic undercounting of female births across all periods, but rather a specific issue affecting the 1975–2000 window.

Next, we examine whether this divergence could reflect real demographic differences rather than data errors. We consider two possibilities: sex differences in child mortality and sex-biased migration.

3.2.1 Differential child mortality

Since male child mortality typically slightly exceeds female mortality at every age, the census sex ratio among survivors is likely to understate the sex ratio at birth.[3] We adjust for this by multiplying the census sex ratio by the ratio of female to male cumulative survival probabilities: $\widehat{\text{SR}}^{\text{adj}} = \text{SR}^{\text{census}} \times (\hat{S}_{\text{female}}/\hat{S}_{\text{male}})$, where $\hat{S}_s$ is estimated from INE mortality tables.[4] As shown in Figure A2, the adjustment substantially narrows the gap

[3]Throughout this section, "child mortality" refers to mortality at ages 0–10. Age-0 mortality (infant mortality in the strict demographic sense) accounts for the largest share of the correction, with the remaining ages contributing progressively smaller increments.

[4]For the cohort born in 1981 and observed in the 1991 census at age 10, the estimated survival probabilities are 0.987 for girls and 0.982 for boys, yielding an adjustment factor of 1.005 — raising the

for cohorts born in the late 1970s, but leaves it largely unchanged for cohorts born in the 1980s and 1990s. In either case, the implied correction is far too small to explain a 2–3 percentage point divergence between the two sources.

3.2.2 Sex-biased migration

To address potential sex-biased immigration, we restrict the census to individuals born in Spain. This yields sex ratios virtually identical to those of the full census sample, well within the normal biological range, leaving the gap with the birth registry essentially unchanged (Figure A3).[5]

To assess the potential impact of sex-biased emigration, we compare the number of individuals born in Spain who appear in the census with total births recorded in the birth registry for the same cohort. While we cannot directly observe the number or sex composition of emigrants, this comparison provides an indirect estimate of the overall scale of out-migration. A retention ratio close to 1 would indicate that out-migration was too limited in scale to explain the divergence. As shown in Figure A4, once estimated child deaths are added back, the adjusted ratio averages around 0.98, suggesting that at most 2% of those born in Spain may have emigrated before the census. For sex-biased emigration to explain a 2–3 percentage point divergence in sex ratios, virtually all of these emigrants would need to have been male — a degree of sex bias that is implausible and unsupported by any evidence.[6]

3.3 Evidence of coding errors

The census comparison suggests that the birth registry sex ratios are implausible, but it does not identify the source of the error. We now examine the microdata directly to characterise the nature of the anomalies. We present three pieces of evidence. First, sex

census sex ratio by roughly 0.46 percentage points. For the cohort born in 1990 and observed in the same census at age 1, the factor is only 1.002 (survival probabilities of 0.993 for girls and 0.991 for boys), an uplift of 0.16 percentage points. The difference reflects both the shorter exposure window and the steep decline in child mortality over the decade. Averaged across all cohorts, the adjustment raises the census sex ratio by 0.43 points for the 1981 census, 0.28 for 1991, 0.14 for 2001, 0.08 for 2011, and 0.05 for 2021 — an order of magnitude smaller than the observed divergence between the two sources.

[5] The native-born sample is available from the 1991 census onwards, but country of birth is not provided in publicly available tables in earlier censuses.

[6] The adjusted ratio slightly exceeds 1 for some birth years in the early 1980s. This is consistent with evidence of under-registration of births in the province of Barcelona during that period (Devolder et al. 2010): if some births went unrecorded in the registry but the children were nonetheless enumerated in the census, the denominator would be understated and the ratio would exceed 1.

ratios exhibit implausibly large variation across months within the same province. Second, outlier province-month cells with anomalously high male shares also show unusually high rates of missing birth weight data, consistent with a mechanical coding error that corrupted multiple consecutive fields on the registration form. Third, provisional statistics published shortly after the reference year show systematically lower sex ratios than the final series, and the extreme provincial outliers present in the official microdata are absent from the provisional data, suggesting part of the errors were introduced during the finalisation of the records rather than their initial recording.

3.3.1 Implausible monthly variation in sex ratios within provinces

We use the birth registry microdata to examine the nature and distribution of the anomalies more closely. Importantly, the aggregate statistics are fully consistent with the microdata, ruling out the possibility that the errors were introduced during aggregation and confirming that they are present at the level of individual records.

If the elevated sex ratios in the birth registry reflected real biological or behavioural patterns, we would expect relatively smooth variation across months within provinces. However, as we show below, we find extensive evidence showing that the variation across months within the same province is implausibly large.

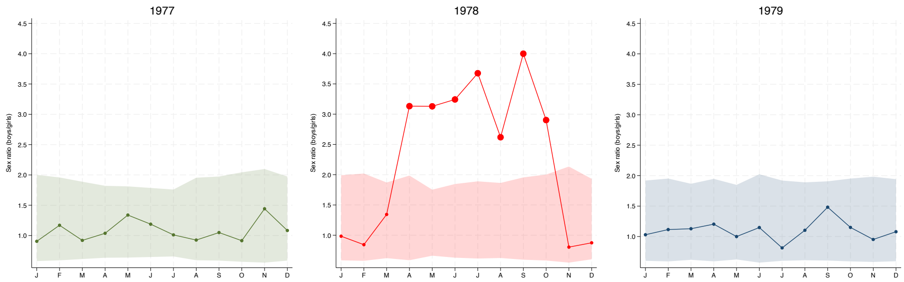
Figure 4 illustrates this pattern for three provinces. The shaded areas represent 99% confidence intervals based on the mean in all other Spanish provinces for that year. Note that this approach is conservative: the national average itself is likely inflated by the very errors we seek to detect, making it harder to identify outliers. The most extreme cases are Teruel and Valladolid. In Teruel, the sex ratio is within normal ranges in 1977 and the beginning of 1978, around 1.07 boys per girl, and then suddenly increases to over 3 boys per girl between April and October, before returning to normal levels. The probability of observing a ratio so extreme in any month just by chance is below 10^{-8} under a standard binomial model.[7] Valladolid exhibits even more dramatic swings: in 1982, the sex ratio rises to over 8 boys per girl in March and April, before returning to normal levels in surrounding months. The probability of observing a ratio so extreme just by chance is essentially zero (p-value$< 10^{-70}$).[8] Other provinces show anomalies that,

[7]In Teruel in 1978, the average number of births per month was 137. Given the sex ratio observed in the rest of Spain that year (1.072), the probability of observing a sex ratio exceeding 3 boys per girl in a sample of this size has a two-tailed p-value of 3.1×10^{-8}.

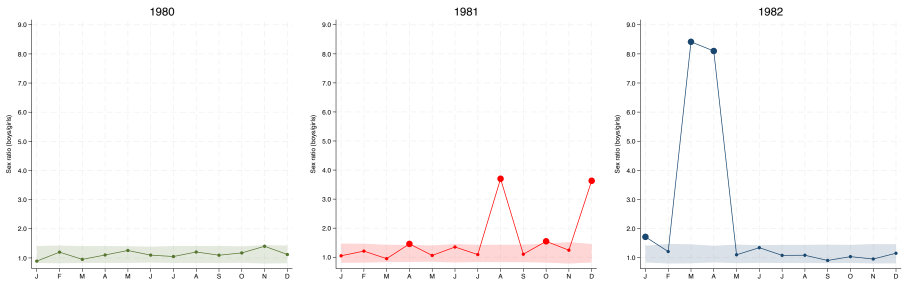
[8]In March 1982, 446 males were registered against only 53 females in Valladolid (499 births total).

while less extreme, remain biologically implausible. In Barcelona — a large province where sampling variation should be minimal — the sex ratio rises to 1.33 boys per girl in January 1984 and 1.29 in March 1984, before returning to normal levels in surrounding months, with p-values below 10^{-10} and 10^{-8} respectively.[9]

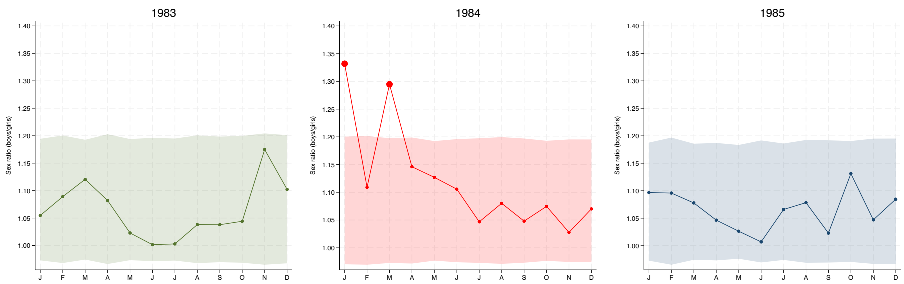
Figure 4: Implausible monthly variation in sex ratios: selected cases



(a) Teruel, 1977-1979



(b) Valladolid, 1980-1982



(c) Barcelona, 1983-1985

Notes: Sex ratios by month for selected provinces and years based on birth registry microdata from INE (2024*c*). The shaded area represents the 99% confidence interval based on the mean in all other Spanish provinces for that year.

The examples above are not isolated cases. To assess the prevalence of anomalies

Given the sex ratio observed in the rest of Spain that year (1.088), the two-tailed p-value of observing a sex ratio exceeding 8 boys per girl in a sample of this size is below 10^{-70}.

[9] In January 1984, 2,191 males were registered against 1,645 females in Barcelona (3,836 births total). Given the sex ratio observed in the rest of Spain that year (1.079), the two-tailed p-value of observing a sex ratio this extreme in a sample of this size is 1.0×10^{-10}. In March 1984, 2,266 males were registered against 1,750 females (4,016 births total), with a two-tailed p-value of 9.9×10^{-9}.

Table 1: Expected and observed number of outliers in birth registry data, 1975–2000

	Left tail (more girls)			Right tail (more boys)		
	Expected	Observed	Ratio	Expected	Observed	Ratio
p<0.01	81	126	1.56	81	139	1.72
p<0.001	8	19	2.38	8	45	5.63
p<0.0001	1	3	3.00	1	24	24.00

Notes: Expected vs. observed frequency of extreme sex ratio values in province-month data based on birth registry microdata from INE, 1975–2000 (2024*c*). For each province-month, the reference mean is the population-weighted national average computed excluding the focal province. Two-tailed test; each tail contains half the expected proportion under the null. "Right tail" refers to observations with unusually high sex ratios (more boys than expected); "left tail" refers to unusually low sex ratios (more girls than expected).

systematically, we compute a z-score for each province-month cell and we test how many province-month observations deviate significantly from the year-specific national average between 1975 and 2000.[10] Under the null hypothesis of no systematic error, these z-scores should be approximately normally distributed.

Panel A in Table 1 reports the expected and observed frequencies of extreme observations at various significance thresholds. Our sample comprises 16224 province-month observations. At the most extreme threshold (p<0.0001), we would expect approximately 1 observation in each tail by chance alone. Instead, we observe 27 – over an order of magnitude more than expected – indicating systematic data quality problems.

The outliers are not randomly distributed: they are heavily skewed toward implausibly high sex ratios. At p<0.0001, we observe 24 observations in the right tail versus only 3 in the left. At p<0.001, we observe 45 right-tail outliers versus 19 in the left tail – more than twice as many. If the excess outliers reflected random measurement error, we would expect roughly equal numbers in both directions. The one-directional pattern points instead to systematic miscoding of females as males, which would bias aggregate sex ratios upward.

If the anomalies in the birth registry reflected real demographic patterns, we would expect similar irregularities in the census. Table A1 in the Appendix shows this is not the case: the distribution of outliers in census data is symmetric and consistent with

[10] For each province-month, we compute a z-score as $z = \frac{r_{pm} - \bar{r}_y}{\sigma_{pm}}$, where r_{pm} is the observed sex ratio in province p and month m, $\bar{r}_y$ is the national average for year y, and σ_{pm} is the standard error based on the number of births in that province-month, assuming a binomial distribution.

random variation, both for 1981–2000 (Panel A) and for 2001–2020 (Panel B).[11] This reinforces our interpretation that the unusually high sex ratios are data artifacts rather than genuine demographic phenomena.

Finally, the pattern of outliers in the birth registry changes substantially after 2000: the distribution becomes symmetric and the observed frequencies align closely with what would be expected under random variation alone (Table A1, Panel C). Whatever caused the coding errors seems to have eventually been corrected.

We assess whether removing the obvious anomalies would suffice to reconcile the birth registry with census data by excluding the 63 province-month observations with p-values below 0.001: 45 in the right tail (excessively high sex ratios) and 19 in the left tail (excessively low sex ratios). Figure A6 compares the raw national sex ratio series against the cleaned series. Removing the extreme outliers reduces the magnitude of the anomaly, particularly during the peak years of 1981–1982, but the cleaned series remains substantially above the expected range of 105–106 throughout 1975–1995. Thus, the problem extends beyond the most severe cases. Many smaller errors – individually undetectable but still systematically biased toward recording females as males – likely pervade the data.

3.3.2 Missing information on birth weight

Further evidence of a systematic coding error comes from the birth weight field, which immediately follows the sex field on the statistical bulletin (see Figure A1). We define as outlier cells those province-month-year combinations that, as documented in Table 1, exhibit implausibly high sex ratios with a p-value below 0.001. Table A2 regresses an indicator for missing birth weight on outlier status and a male indicator, controlling for province $\times$ year and month fixed effects — so we are effectively comparing anomalous months against other months in the same province and year, accounting for seasonal patterns. On average, around 16% of observations have missing weight information, with outlier cells showing a significantly higher rate, 2.8 percentage points more (column 1). Column 2 shows that within outlier cells, observations coded as male are 97% (11.4

[11] To conduct this analysis at the province-month-year level, we use census microdata samples rather than aggregate tabulations. The microdata are available for censuses from 1991 onwards, with sampling rates of 2% (1991), 5% (2001), 10% (2011), and 10% (2021). Our sample comprises 11.6 million individual records, which we collapse into approximately 24,000 province-by-month-by-year-of-birth cells covering births from 1981 to 2020.

percentage points) more likely to have missing weight than those coded as female ($p < 0.001$). The same underlying data entry failure that miscoded the sex field might have also disrupted the recording of the weight field — consistent with a mechanical error that corrupted multiple consecutive fields on the statistical bulletin simultaneously, rather than an isolated error confined to the sex variable alone.

3.3.3 Provisional vs. final birth registry data

Additional support for the hypothesis that the high sex ratio is largely due to data processing comes from comparing provisional and final INE statistics. Between 1975 and 1982, INE published provisional birth counts in the Annual Statistical Yearbook (*Anuario Estadístico de España*) approximately one year after the reference year.[12] As shown in Figure A5, the sex ratio in the provisional series is consistently below the official series in every year, with an average difference of 0.9 percentage points (p-value $< 10^{-8}$), and the gap is particularly pronounced in 1981 and 1982, where it reaches 2.2 and 1.5 percentage points respectively.

Relative to the provisional series, the final one records approximately 1% more births. However, for these additional births alone to explain a 0.9 percentage point difference in sex ratios, they would need to be almost entirely male — an implausible degree of selection. While the provisional series is not available at the province-month level, the province-year totals show no trace of the extreme outliers documented in the official microdata. In Teruel in 1978, the official records register 1,085 boys against 561 girls — a sex ratio of 1.93 — whereas the provisional data report 817 boys and 799 girls, a ratio of 1.02 that falls squarely within the normal biological range. Similarly, in Valladolid in 1982, the official records show 3,861 boys against 2,600 girls (sex ratio of 1.49), while the provisional data report 3,229 boys and 3,042 girls (sex ratio of 1.06). The divergence between the two series suggests that the imbalance was largely introduced during the finalisation of the records rather than their initial recording, though the provisional figures themselves occasionally exceed 107 boys per 100 girls, hence the errors may have begun earlier but intensified at the finalisation stage.

[12] After 1982, provisional estimates were published with a lag of two to three years and converge closely to the final estimates, making a comparable analysis unavailable for later periods.

4 Discussion

We show that the anomalous sex ratios in Spanish birth registry data from 1975 to 2000 are an artifact of systematic coding errors. This has important implications for the literature exploring sex ratios in Spain during this period: earlier findings attributing them to biological or behavioural factors can largely be explained by erroneous data rather than true demographic phenomena. It also affects the validity of cross-country analyses of sex ratios that draw on Spanish data via international aggregators such as Eurostat, the WHO, and the UN. Notably, the UN Population Division classifies the Spanish birth registry as "high quality", making it unlikely that researchers would suspect underlying data issues (e.g. Chao et al. 2019).[13]

Whenever data errors are detected, the preferred solution is for the responsible statistical agency to correct the erroneous records directly. When this is not feasible, the minimum requirement is transparency: flagging the affected years in the published data and accompanying them with a methodological note would at least prevent future misinterpretation by researchers.[14]

The Spanish case provides a cautionary tale about the reliability of administrative data. We recommend that statistical agencies ensure that routine internal validation checks — such as flagging implausible variation in key variables across geographic units or time periods — are applied consistently. In the present case, a simple check on the variance of provincial sex ratios would have immediately identified the anomalous figures. More broadly, systematic validation of administrative data against alternative sources may help detect such problems before they propagate into the international databases on which researchers rely.

[13] This view was not limited to international aggregators. Sánchez Barricarte (2023), in a detailed study of the historical evolution of sex ratios in Spain, explicitly states, referring to the period 1975–1995, that "the high quality of the Civil Registry data from those years is beyond any doubt" (p. 146, translation ours), before proceeding to analyse the anomalous figures as a genuine demographic phenomenon.

[14] In our own correspondence with INE regarding specific implausible figures in the data, we were informed that the original records from that period are no longer available, hence direct correction may not be possible in this case.

References

Alesina, A., Giuliano, P. & Nunn, N. (2018), 'Traditional Agricultural Practices and the Sex Ratio Today', *PLOS One* **13**(1), e0190510.

Almond, D. & Edlund, L. (2008), 'Son-Biased Sex Ratios in the 2000 United States Census', *Proceedings of the National Academy of Sciences* **105**(15), 5681–5682.

Almond, D., Edlund, L. & Milligan, K. (2013), 'Son Preference and the Persistence of Culture: Evidence from South and East Asian Immigrants to Canada', *Population and Development Review* **39**(1), 75–95.

Angrist, J. (2002), 'How Do Sex Ratios Affect Marriage and Labor Markets? Evidence from America's Second Generation', *Quarterly Journal of Economics* **117**(3), 997–1038.

Bongaarts, J. & Guilmoto, C. Z. (2015), 'How Many More Missing Women? Excess Female Mortality and Prenatal Sex Selection, 1970–2050', *Population and Development Review* **41**(2), 241–269.

Chahnazarian, A. (1988), 'Determinants of the Sex Ratio at Birth: Review of Recent Literature', *Biodemography and Social Biology* **35**(3-4), 214–235.

Chao, F., Gerland, P., Cook, A. R. & Alkema, L. (2019), 'Systematic Assessment of the Sex Ratio at Birth for All Countries and Estimation of National Imbalances and Regional Reference Levels', *Proceedings of the National Academy of Sciences* **116**(19), 9303–9311.

Devolder, D., Ortiz, E. & Zeman, K. (2010), 'Human Fertility Database Documentation: Spain', `https://www.academia.edu/download/56874944/ESPcom.pdf`. Accessed: 21st February 2024.

Dubuc, S. & Coleman, D. (2007), 'An Increase in the Sex Ratio of Births to India-born Mothers in England and Wales: Evidence for Sex-Selective Abortion', *Population and Development Review* **33**(2), 383–400.

Echavarri, R. & Beltrán Tapia, F. (2025), 'Prenatal Care, Son Preference, and the Sex Ratio at Birth', *Demography* **62**(1), 211–236.

Edlund, L., Li, H., Yi, J. & Zhang, J. (2013), ‘Sex Ratios and Crime: Evidence from China’, *Review of Economics and Statistics* **95**(5), 1520–1534.

Efrat, Z., Akinfenwa, O. O. & Nicolaides, K. H. (1999), ‘First-Trimester Determination of Fetal Gender by Ultrasound’, *Ultrasound in Obstetrics & Gynecology* **13**, 305–307.

González, Libertad (2018), ‘Sex Selection and Health at Birth Among Indian Immigrants’, *Economics & Human Biology* **29**, 64–75.

Grech, V., Vassallo-Agius, P. & Savona-Ventura, C. (2003), ‘Secular Trends in Sex Ratios at Birth in North America and Europe over the Second Half of the 20th Century’, *Journal of Epidemiology and Community Health* **57**(8), 612–615.

Grosjean, P. & Khattar, R. (2019), ‘It’s Raining Men! Hallelujah? The Long-Run Consequences of Male-Biased Sex Ratios’, *The Review of Economic Studies* **86**(2), 723–754.

Gutiérrez-Adán, A., Pintado, B. & De La Fuente, J. (2000), ‘Demographic and Behavioral Determinants of the Reduction of Male-To-Female Birth Ratio in Spain from 1981 to 1997’, *Human Biology* **72**(5), 891–898.

Hesketh, T. & Xing, Z. W. (2006), ‘Abnormal Sex Ratios in Human Populations: Causes and Consequences’, *Proceedings of the National Academy of Sciences* **103**(36), 13271–13275.

Instituto Nacional de Estadística (2024*a*), ‘Anuario Estadístico de España, 1975–2000’, `https://www.ine.es/inebaseweb/25687.do`. Accessed: 21st February 2024.

Instituto Nacional de Estadística (2024*b*), ‘Censos de población y viviendas: Microdatos, 1991–2021’, `https://www.ine.es/prodyser/microdatos.htm`. Accessed: 21st February 2024.

Instituto Nacional de Estadística (2024*c*), ‘Estadística de nacimientos: Microdatos, 1975–2020’, `https://www.ine.es/dyngs/INEbase/operacion.htm?c=Estadistica_C&cid=1254736177007&menu=resultados&idp=1254735573002`. Accessed: 21st February 2024.

Instituto Nacional de Estadística (2024*d*), 'Estadística de nacimientos: Series anuales definitivas por sexo, provincia y año', https://www.ine.es/dyngs/INEbase/operacion.htm?c=Estadistica_C&cid=1254736177007&menu=resultados&idp=1254735573002. Accessed: 21st February 2024.

Instituto Nacional de Estadística (2024*e*), 'Movimiento natural de la población: Tasas de mortalidad por edad y sexo, 1975–2024', https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177007&menu=resultados&idp=1254735573002. Accessed: 21st February 2024.

Instituto Nacional de Estadística (2025), 'Censo de población y viviendas: Tablas de resultados', https://www.ine.es/dyngs/INEbase/operacion.htm?c=Estadistica_C&cid=1254736177108&menu=resultados&idp=1254735572981. Accessed: 21st February 2024.

Jacobsen, R., Møller, H. & Mouritsen, A. (1999), 'Natural Variation in the Human Sex Ratio', *Human Reproduction* **14**(12), 3120–3125. Publisher: Oxford Academic.

Juárez, S. P. (2014), 'Notas acerca del movimiento natural de la población para el estudio de la salud perinatal', *Gaceta Sanitaria* **28**(6), 505–507.

Michels, T. C. & Tiu, A. Y. (2007), 'Second Trimester Pregnancy Loss', *American Family Physician* **76**(9), 1341–1346.

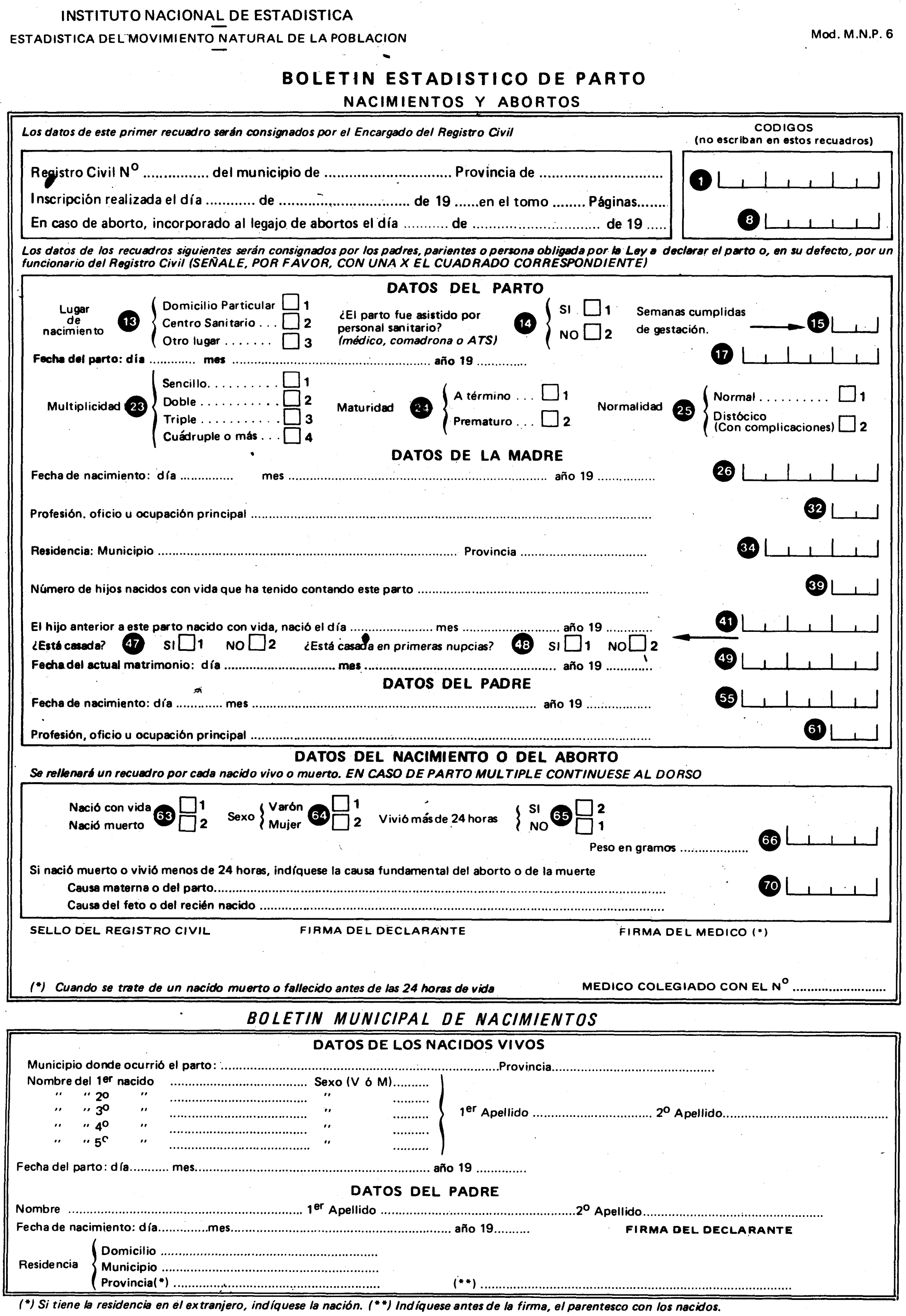
Sánchez Barricarte, J. J. (2023), 'Analysis of the Historical Evolution of Sex Ratio at Birth in Spain', *Revista Española de Investigaciones Sociológicas (REIS)* **182**(182), 139–328.

United Nations (2024), 'Sex Ratio at Birth – UN WPP', https://archive.ourworldindata.org/20260105-133053/grapher/sex-ratio-at-birth.html. Accessed: 21st February 2024.

Wang, F., Zhao, L. & Zhao, Z. (2017), 'China's Family Planning Policies and Their Labor Market Consequences', *Journal of Population Economics* **30**(1), 31–68.

A Additional Figures and Tables

Figure A1: Example of statistical bulletin form 1981-1994



INSTITUTO NACIONAL DE ESTADISTICA
ESTADISTICA DEL MOVIMIENTO NATURAL DE LA POBLACION

Mod. M.N.P. 6

BOLETIN ESTADISTICO DE PARTO
NACIMIENTOS Y ABORTOS

Los datos de este primer recuadro serán consignados por el Encargado del Registro Civil

CODIGOS
(no escriban en estos recuadros)

Registro Civil N° .... del municipio de .... Provincia de ....
Inscripción realizada el día .... de .... de 19 ......en el tomo .... Páginas.......
En caso de aborto, incorporado al legajo de abortos el día .... de .... de 19 .....

1 8

Los datos de los recuadros siguientes serán consignados por los padres, parientes o persona obligada por la Ley a declarar el parto o, en su defecto, por un funcionario del Registro Civil (SEÑALE, POR FAVOR, CON UNA X EL CUADRADO CORRESPONDIENTE)

DATOS DEL PARTO

Lugar de nacimiento 13 { Domicilio Particular 1; Centro Sanitario ... 2; Otro lugar 3

¿El parto fue asistido por personal sanitario? (médico, comadrona o ATS) 14 { SI 1; NO 2

Semanas cumplidas de gestación. 15

Fecha del parto: día .... mes .... año 19 17

Multiplicidad 23 { Sencillo.......... 1; Doble .... 2; Triple .... 3; Cuádruple o más ... 4

Maturidad 24 { A término ... 1; Prematuro ... 2

Normalidad 25 { Normal .... 1; Distócico (Con complicaciones) 2

DATOS DE LA MADRE

Fecha de nacimiento: día .... mes .... año 19 26

Profesión, oficio u ocupación principal 32

Residencia: Municipio .... Provincia 34

Número de hijos nacidos con vida que ha tenido contando este parto 39

El hijo anterior a este parto nacido con vida, nació el día .... mes .... año 19 41

¿Está casada? 47 SI 1 NO 2 ¿Está casada en primeras nupcias? 48 SI 1 NO 2

Fecha del actual matrimonio: día .... mes .... año 19 49

DATOS DEL PADRE

Fecha de nacimiento: día .... mes .... año 19 55

Profesión, oficio u ocupación principal 61

DATOS DEL NACIMIENTO O DEL ABORTO

Se rellenará un recuadro por cada nacido vivo o muerto. EN CASO DE PARTO MULTIPLE CONTINUESE AL DORSO

Nació con vida / Nació muerto 63 1 / 2 Sexo { Varón / Mujer 64 1 / 2 Vivió más de 24 horas { SI / NO 65 2 / 1

Peso en gramos 66

Si nació muerto o vivió menos de 24 horas, indíquese la causa fundamental del aborto o de la muerte
Causa materna o del parto.. 70
Causa del feto o del recién nacido ....

SELLO DEL REGISTRO CIVIL FIRMA DEL DECLARANTE FIRMA DEL MEDICO (*)

(*) Cuando se trate de un nacido muerto o fallecido antes de las 24 horas de vida MEDICO COLEGIADO CON EL N° ....

BOLETIN MUNICIPAL DE NACIMIENTOS

DATOS DE LOS NACIDOS VIVOS

Municipio donde ocurrió el parto: ....Provincia..
Nombre del 1er nacido .... Sexo (V ó M)..........
" " 2o " .... " ....
" " 3o " .... " .... 1er Apellido .... 2o Apellido...
" " 4o " .... " ....
" " 5o " .... " ....

Fecha del parto: día.......... mes... año 19 ....

DATOS DEL PADRE

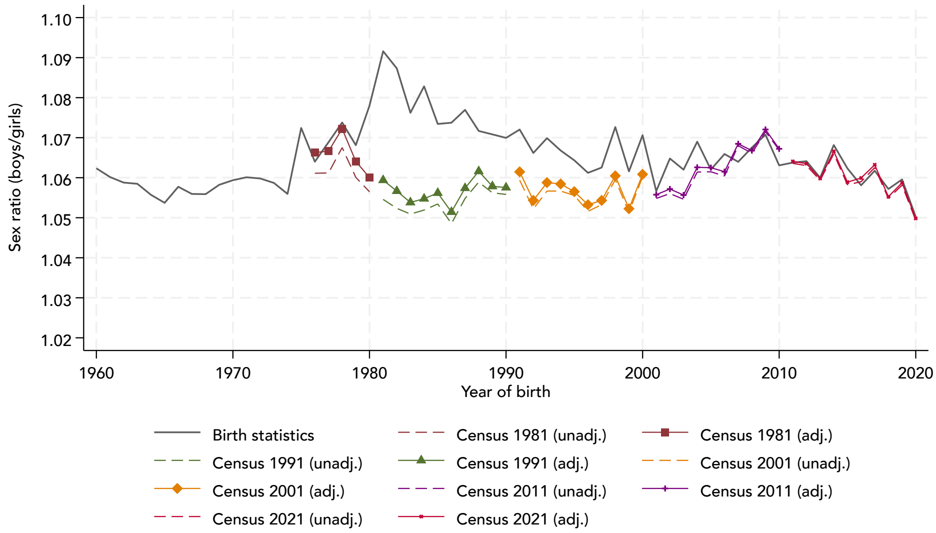
Nombre .... 1er Apellido ....2o Apellido...
Fecha de nacimiento: día.............mes... año 19......... FIRMA DEL DECLARANTE

Residencia { Domicilio ....; Municipio ....; Provincia(*) .... (**) ....

(*) Si tiene la residencia en el extranjero, indíquese la nación. (**) Indíquese antes de la firma, el parentesco con los nacidos.

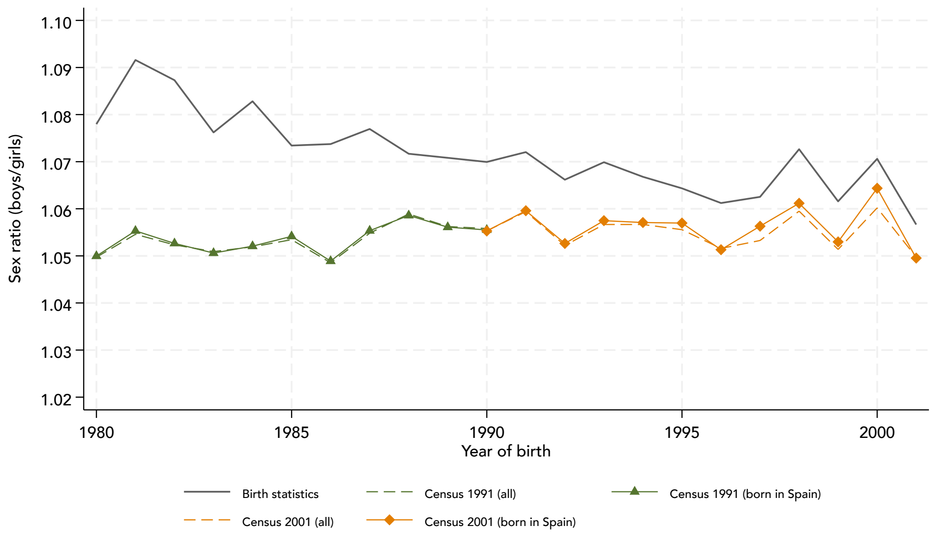
Source: https://www.ine.es/metodologia/t20/cues_nac_81-94.pdf

Figure A2: Sex ratio in Spain, birth registry vs. Census, adjusting for mortality



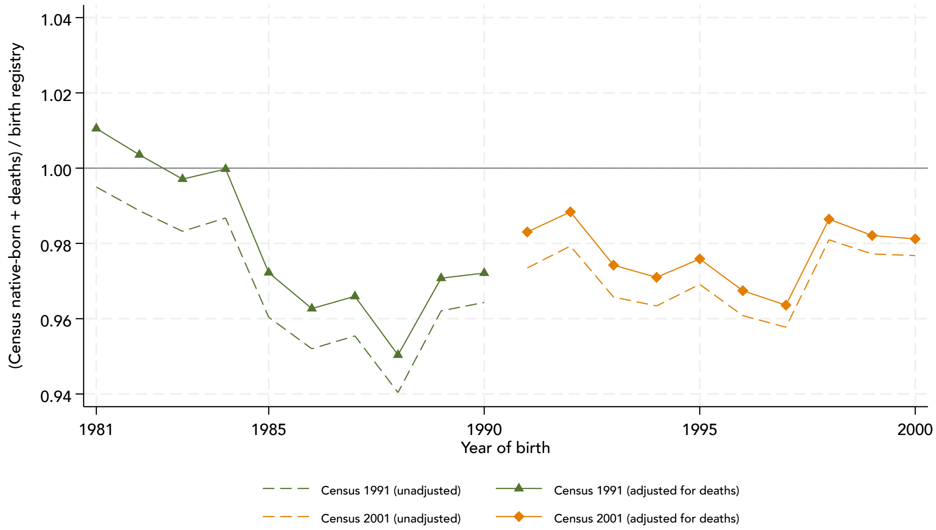
Notes: Grey line: sex ratio at birth (boys per girl) from the birth registry. Coloured series: sex ratios implied by the 1981, 1991, 2001, 2011, and 2021 censuses for birth cohorts in the ten years preceding each census (five years for 1981) from INE 2025. Dashed lines show raw census sex ratios among survivors; solid lines show ratios adjusted for differential male/female child mortality using data from INE 2024*e* as described in text.

Figure A3: Sex ratio in Spain, birth registry vs. Census for sample born in Spain



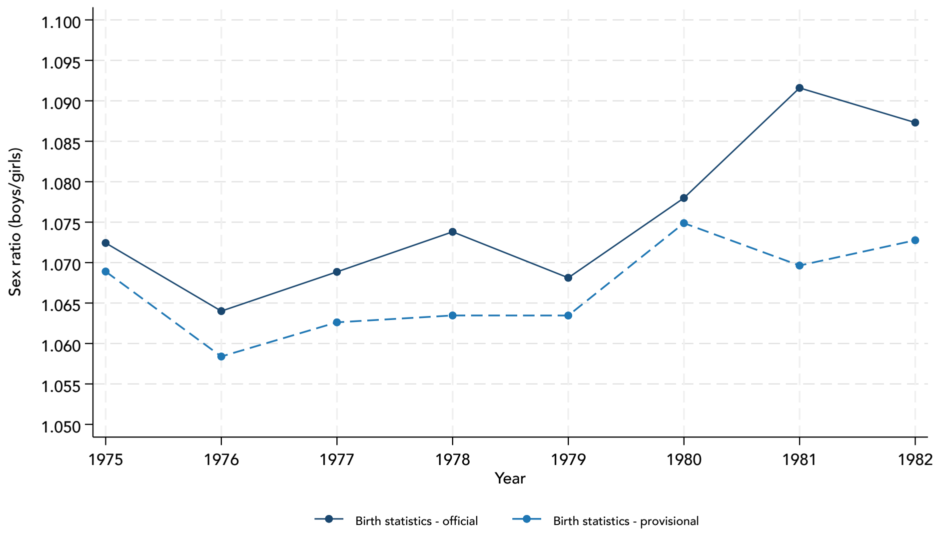
Notes: Grey line: sex ratio at birth (boys per girl) from the birth registry. Coloured series: sex ratios implied by the 1991 and 2001 censuses for birth cohorts in the ten years preceding each census from INE 2025. Dashed lines show census sex ratios for all residents; solid lines show census sex ratios restricted to individuals born in Spain.

Figure A4: Census coverage of native-born children relative to birth registry, 1981–2000



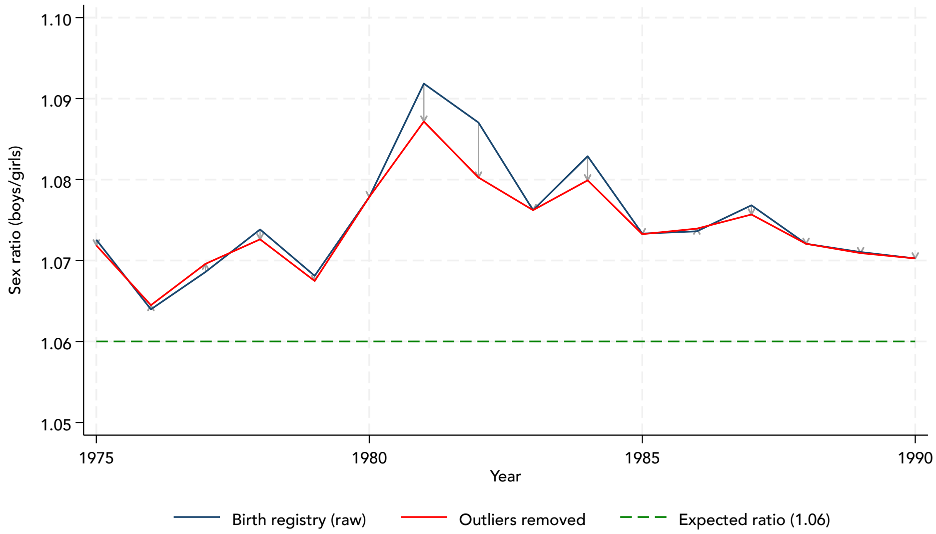
Notes: The figure plots, for each birth year, the ratio of individuals born in Spain enumerated in the census (INE, 2025) to total births recorded in the birth registry (INE, 2024*d*). Dashed lines show the unadjusted ratio; solid lines add back child deaths (ages 0–10) during the corresponding period to the census count. Birth years 1981–1990 are observed in the 1991 census; birth years 1991–2000 in the 2001 census.

Figure A5: Sex ratio in Spain, provisional vs official birth registry, 1975–1986



Notes: Sex ratio in Spain, 1975–1982, as measured by official birth registry data from INE (2024*d*) and by provisional estimates from the Annual Statistical Yearbook from INE (2024*a*).

Figure A6: Sex ratio at birth: raw data vs. outliers removed, Spain 1975–1990



Notes: Sex ratio in Spain comparing raw birth registry data from INE (2024*d*) with data after removing 64 province-month observations with p-values below 0.001 (45 above the year average, 19 below). The dashed line indicates the expected sex ratio of 1.06 (106 boys per 100 girls).

Table A1: Expected and observed outliers: census data (1981–2020) and birth registry (2001–2020)

	Left tail (more girls)			Right tail (more boys)		
Panel A: Census data, 1981–2000						
	Expected	Observed	Ratio	Expected	Observed	Ratio
p<0.01	62	43	0.69	62	38	0.61
p<0.001	6	2	0.33	6	2	0.33
p<0.0001	1	0	0.00	1	0	0.00
Panel B: Census data, 2001–2020						
	Expected	Observed	Ratio	Expected	Observed	Ratio
p<0.01	62	68	1.10	62	52	0.84
p<0.001	6	7	1.17	6	4	0.67
p<0.0001	1	1	1.00	1	0	0.00
Panel C: Birth registry data, 2001–2020						
	Expected	Observed	Ratio	Expected	Observed	Ratio
p<0.01	59	68	1.15	59	56	0.95
p<0.001	6	7	1.17	6	7	1.17
p<0.0001	1	1	1.00	1	0	0.00

Notes: Expected vs. observed frequency of extreme sex ratio values. Panels A and B show estimates from census data ((2024*b*) for 1981–2000 and 2001–2020 respectively, based on microdata samples (2% in 1991, 5% in 2001, 10% in 2011 and 2021), comprising 11.6 million individuals collapsed into approximately 25,000 province-month-year cells. Panel C shows estimates from birth registry data for 2001–2020 from INE (2024*c*). Two-tailed test using year-specific Spanish average as baseline. "Right tail" refers to observations with unusually high sex ratios (more boys than expected); "left tail" refers to unusually low sex ratios (more girls than expected).

Table A2: Correlates of missing information of birth weight

	Missing weight data = 1	
Male	0.003***	0.003***
	(0.000)	(0.000)
Outlier	0.028**	-0.041***
	(0.013)	(0.012)
Outlier × Male		0.114***
		(0.019)
Province × Year FE	Yes	Yes
Month FE	Yes	Yes
Mean	0.160	0.160
N	7,634,351	7,634,351

Notes: Sample includes births from 1980 to 2000 from official birth registry microdata (2024*c*). The outcome variable is a dummy equal to 1 if birth weight was not recorded. *Outlier* is a dummy equal to 1 if the sex ratio in a province-month-year cell falls outside the 99.9% confidence interval relative to the population-weighted national average in that year, computed excluding the focal province. All columns include province × year and month fixed effects. Standard errors are clustered at the province-month-year level.

B Data Appendix

B.1 INE birth registry – official figures

Our main focus is the publicly available official birth registry data as provided by INE. These data are extensively used in research and are the figures reported to international data aggregators.

We use information from the aggregate series reporting the number of births by sex, province of residence of the mother, and year from 1960 to 2020 to compute annual sex ratios at the province and national level (INE, 2024*d*).

We use individual-level microdata to compute monthly sex ratios at the province level for 1975–2020 and explore the attributes of outlier observations. These data are available from 1975 and record each birth with information on sex, province of inscription, month and year of birth, as well as parental characteristics including age, occupation, and birth order (INE, 2024*c*).

B.2 INE birth registry – provisional figures

We also digitise publicly available aggregate statistics from the Annual Statistical Yearbook (*Anuario Estadístico de España*), which report provisional births by sex, province, and year for the period 1975–2000 (INE, 2024*a*).[15] These provisional figures are published with a lag relative to the reference year: up to 1982, the provisional tables were published the year following the reported births; from 1983 onwards, the publication lag increases to two years.

B.3 Census

We compare the birth registry data with aggregate publicly available population census data from INE (2025) for 1970, 1981, 1991, 2001, 2011, and 2021. Spanish censuses are conducted decennially.[16] For each census, we compute the sex ratio of children by single year of birth, covering the ten birth cohorts born in the decade preceding the census. This allows us to compare the sex ratio recorded at birth in the registry with the sex ratio of survivors observed at the census date.

In addition to this aggregate data, we use microdata samples from INE (2024*b*). The samples account for 2% of the population in 1991, 5% in 2001, 10% in 2011 and 2021, comprising 11.6 million individuals, to compute province by month sex ratios used in panels A and B of Table A1.

B.4 Mortality data

To construct mortality-adjusted comparisons between the birth registry and the census, we use data on mortality rates by single year of age (ages 0–10), sex, and calendar year for the period 1975–2024, sourced from INE (2024*e*). For each birth cohort B observed at census year C, we compute the cumulative probability of surviving from birth to the

[15] Data for 1995 are not available.

[16] The reference dates are: 31 December 1970, 1 March 1981, 1 March 1991, 1 November 2001, 1 November 2011, and 1 January 2021.

census date as the product of one-year survival probabilities across ages $0, 1, \ldots, C-B-1$, where the mortality probability at each age is taken from the calendar year in which the cohort passed through that age. The mortality-adjusted sex ratio is then obtained by multiplying the census sex ratio by the ratio of female to male cumulative survival, which corrects for the fact that male mortality exceeds female mortality at every age, causing the census to understate the sex ratio at birth. We apply this adjustment for the censuses of 1981, 1991, 2001, 2011, and 2021, covering birth cohorts from 1976 to 2020.