Journal of the Royal Statistical Society Series A: Statistics in Society, 2026, 00, 1–20
https://doi.org/10.1093/jrsssa/qnag045
Published: 20 March 2026
Original article

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society
OXFORD

Learning individual reproductive behaviour from aggregate fertility rates via neural posterior estimation

Daniel Ciganda[1,2], Ignacio Campón[1,2], Iñaki Permanyer[3,4], Jakob H. Macke[5,6]

[1]Max Planck Institute for Demographic Research, Rostock, Germany
[2]Statistics Institute, University of the Republic, Montevideo, Uruguay
[3]Center for Demographic Studies, Autonomous University of Barcelona, Barcelona, Spain
[4]ICREA, Catalan Institution for Research and Advanced Studies, Barcelona, Spain
[5]Machine Learning in Science, University of Tübingen & Tübingen AI Center, Tübingen, Germany
[6]Department of Empirical Inference, Max Planck Institute for Intelligent Systems, Tübingen, Germany
Address for correspondence: Daniel Ciganda, Statistics Institute, University of the Republic, Av. Gonzalo Ramírez 1926, 11200 Montevideo, Uruguay. Email: ciganda@demogr.mpg.de

Abstract

Age-specific fertility rates (ASFRs) provide the most extensive record of reproductive change, but their aggregate nature obscures the individual-level behavioural mechanisms that drive fertility trends. To bridge this micro–macro divide, we introduce a simulation-based Bayesian framework that couples a demographically interpretable, individual-level simulation model of the reproductive process with sequential neural posterior estimation (SNPE). We show that this framework successfully recovers core behavioural parameters governing contemporary fertility, including preferences for family size, reproductive timing, and contraceptive failure, using only ASFRs. The framework's effectiveness is validated on cohorts from four countries with diverse fertility regimes. Most compellingly, the model, estimated solely on aggregate data, successfully predicts out-of-sample distributions of individual-level outcomes, including age at first sex, desired family size, and birth intervals. Because our framework yields complete synthetic life histories, it significantly reduces the data requirements for building microsimulation models and enables behaviourally explicit demographic forecasts.

Keywords age-specific fertility rates, individual-level modelling, microsimulation, reproductive behaviour, simulation-based inference

1 Introduction

Globally declining fertility rates are driving a profound demographic transformation, presenting fundamental societal challenges and underscoring the need for robust models to understand and anticipate future fertility trends. Age-specific fertility rates (ASFRs), defined as the number of births to women of a given age per person-year of exposure, remain the workhorse input for these tasks because the underlying birth counts are collected continuously in every country's vital-statistics system. ASFRs series therefore provide unparalleled temporal and geographic coverage.

Received: August 22, 2025. **Revised:** February 24, 2026. **Accepted:** February 25, 2026

Yet ASFRs are only marginal aggregates. They reveal *when* births occur but not *why*. A long tradition of parametric models has been developed to describe ASFRs curves, smooth observed schedules, extrapolate fitted parameters, or convert projected total fertility rates into age patterns for cohort-component population forecasts (Brass, 1974; Chandola et al., 1999; Coale & Trussell, 1974; Mazzuco & Scarpa, 2015; Schmertmann, 2003). While existing approaches can excel at capturing aggregate ASFRs shapes, their parameters are typically macrolevel abstractions. As such, they offer limited insight into the individual behavioural mechanisms driving fertility decisions (Hoem & Hoem, 1989). Consequently, forecasts based on these models often cannot explicitly account for how evolving individual choices regarding the timing of parenthood, contraceptive practice, and birth-spacing and stopping decisions shape future fertility trends.

This disconnect is striking in light of the explosion of microlevel fertility research based on rich individual data sources. Studies using birth histories, time-use diaries, and linked administrative records have uncovered nuanced relationships between education, labour market trajectories, union stability, and childbearing. Yet virtually none of that behavioural insight is incorporated into fertility projection models. The challenge of understanding contemporary fertility lies in bridging this micro–macro divide.

This article proposes a step toward this goal. We introduce a demographically interpretable individual-level model that captures key features of modern reproductive behaviour, including stopping, spacing, and contraceptive failure. Estimating such a richly parameterized model from coarse data is a significant challenge, as many combinations of microparameters can generate similar ASFRs curves. We overcome this challenge by pairing our individual-level, mechanistic model with sequential neural posterior estimation (SNPE), a powerful simulation-based inference method. This approach allows us to demonstrate that core behavioural parameters governing the overall level and timing of fertility can be recovered from ASFRs alone. We then fit the model to US National Survey of Family Growth (NSFG) cohorts and to Demographic and Health Survey (DHS) cohorts from Colombia, the Dominican Republic, and Peru, representing four settings that span markedly different fertility regimes.

The framework's effectiveness is confirmed through rigorous validation. First, posterior predictive checks show that the observed aggregate fertility rates are highly plausible under the model, falling well within the 95% posterior predictive intervals. Most significantly, we conduct out-of-sample validation, demonstrating that the model can successfully predict distributions of individual-level characteristics, such as the empirical distributions of age at first sexual intercourse, desired family size, and birth intervals, none of which inform the estimation step.

By establishing a demographically interpretable and statistically robust link between individual reproductive behaviours and population-level fertility outcomes, our framework offers a promising path towards a more nuanced understanding of fertility dynamics. This approach lowers the barrier to building detailed microsimulation models by significantly reducing their data requirements, and could enable more accurate, behaviourally grounded population forecasts.

1.1 Previous work

Efforts to model the reproductive process mechanistically have a long and distinguished tradition in demography. Pioneered by early work on fecundability (Gini, 1924) and significantly advanced by work on 'natural fertility' (Henry, 1953), this foundation led to a first wave of mathematical and early simulation models of the reproductive process (Bongaarts, 1977; Brass, 1958; Potter, 1972; Ridley & Sheps, 1966; Singh, 1963, among others). Much of this mechanistic tradition was grounded in the proximate determinants framework, originating with Davis and Blake (1956) and later formalized and adapted by Bongaarts (1978), which articulated how behavioural and biological processes map onto observed fertility outcomes.

However, this vibrant tradition of individual-level modelling did not continue to evolve within demography at the same pace or in the same manner as in other scientific fields. While empirical fertility research using microdata flourished (Xie, 2000), the further development of these process-based simulators saw less sustained momentum. More critically, this stream of work remained largely disconnected from the concurrent revolution in statistical methodology for simulation-based inference.

Techniques like approximate Bayesian computation or other simulation-based inference (SBI) approaches that became influential in fields like ecology and genetics (Beaumont, 2010; Hartig et al., 2011), were adopted only selectively in demographic simulation work, with notable examples mainly outside fertility, including marriage and migration models analysed via Gaussian-process emulators (Bijak & Hilton, 2021; Hilton & Bijak, 2016). This slower and more uneven diffusion created a persistent gap in formally linking mechanistic understanding to empirical observations in fertility research.

Our own previous work showed that such a bridge is feasible in natural-fertility contexts (historical or religious communities without deliberate birth control) where a handful of biological parameters govern the reproductive process and can be inferred from ASFRs alone (Ciganda & Todd, 2024). The present article tackles the far more ambitious and policy-relevant task of modelling contemporary fertility, where stopping, spacing, and imperfect contraception introduce complex decision-making and heterogeneity in reproductive dynamics.

Our approach is grounded in the model-based programme in demography (Bijak, 2022; Burch, 2018; Courgeau et al., 2016), which treats formal models as carriers of substantive theory rather than purely descriptive tools. In this spirit, we understand our model as a mechanistic, process-based representation of the reproductive life course that links individual-level behaviour to macrolevel regularities. By coupling a behavioural microsimulation with a formal inference framework, we make this micro–macro connection explicit and testable. While some microsimulation models rely on existing estimates for biological parameters (see Granholm et al., 2025, for a recent example within the demographic literature), our approach seeks to infer these values directly from widely available aggregate fertility data.

1.2 Organization of the article

The remainder of this article is structured as follows: Section 2 describes our individual-level computational model of the reproductive process. Section 3 details the data sources and the construction of the main demographic quantities used in our analysis. Section 4 outlines the SNPE methodology. Section 5 presents our findings on parameter identifiability via cross-validation, model fit through posterior predictive checks, and the out-of-sample validation against microlevel data. Finally, we discuss the implications of our findings, limitations, and avenues for future research in the Section 6.

2 Model

The transition from a natural to a modern fertility regime in human populations is marked by the emergence of *stopping* and *spacing* behaviours (deciding whether and when to have children), representing a shift from a reproductive process largely determined by physiological factors to one shaped by individual preferences. As a result, fertility patterns become more diverse, with family formation increasingly tailored to individual or couple-level choices rather than biological constraints alone.

Building on these insights, our model represents the dynamics of reproductive behaviour in contemporary populations by placing fertility intentions and desires at its core. Specifically, it conceptualizes childbearing decisions as the interplay between physiological capacity, access to contraception, and individuals' subjective preferences, such as desired family size and timing.

We employ an individual-level, discrete-time simulation model. A cohort of N women is simulated from the age of potential sexual activity until the end of their reproductive capacity, with their state updated in monthly time steps.

The model begins by assigning each woman key characteristics that will shape her reproductive life. These include two critical ages: the *age at sexual initiation*, when she first becomes exposed to the risk of childbearing, and the *age at intentional reproduction*, representing the point at which she actively starts trying to conceive. Because these two ages often differ, there is typically a period during which a woman is sexually active but not yet intending to have children, an interval associated with the risk of unplanned pregnancies.

Additionally, each woman is assigned a *desired family size*, reflecting the number of children she ultimately hopes to have, and a *desired number of months between births*, which influences her birth spacing behaviour.

In each monthly step, a woman's probability of conception is evaluated. This probability depends on her current age, her current parity relative to her desired family size, whether she has reached her age at intentional reproduction, and, if she has had prior births, whether her desired birth spacing interval has elapsed.

Specifically, when a woman is past her age at intentional reproduction, has not yet reached her desired family size, and is past her desired birth spacing interval (if applicable), we assume she is actively trying to conceive and uses no contraception. Conversely, if she has not yet reached her age at intentional reproduction, or has already achieved her desired family size, or is within her desired spacing interval, she is assumed to be using contraception. This results in a reduced, but non-zero, probability of conception, accounting for imperfect contraceptive effectiveness and the possibility of unplanned pregnancies throughout the reproductive lifespan.

Whenever a conception occurs, the woman is considered non-susceptible for nine months of pregnancy, followed by an additional three-month period of postpartum amenorrhoea, during which her probability of another conception is zero. This minimum postpartum insusceptibility is intended to capture the physiological delay that occurs even in the absence of prolonged breastfeeding. After this period, any additional postpartum delay associated with breastfeeding, postpartum contraception, and deliberate birth spacing is represented by the model's birth-spacing component.

2.1 Implementation

The dynamics described above are implemented by assigning the following characteristics to each woman i and defining the processes that govern her reproductive trajectory:

1. **Age at sexual initiation** (X_i)**:** Each woman's age at sexual initiation, X_i, is drawn from a lognormal distribution. This distribution is parameterized by a mean age μ_s and a standard deviation σ_s (both in years), which are converted to the distribution's underlying parameters (the mean of logarithms, $\mu_{\ln,s}$, and the standard deviation of logarithms, $\sigma_{\ln,s}$) using the standard formulas: $\mu_{\ln,s} = \ln(\mu_s^2/\sqrt{\mu_s^2+\sigma_s^2})$ and $\sigma_{\ln,s} = \sqrt{\ln(1+\sigma_s^2/\mu_s^2)}$. The value for each woman is drawn in years and then converted to months.
2. **Age at intentional reproduction** (R_i)**:** R_i (in months) is also drawn from a lognormal distribution. The mean of this distribution is set to $\mu_r = \mu_s + \delta_r$, where δ_r (in years) is an estimated model parameter. The parameters μ_r and σ_r (in years) are used to derive the underlying lognormal parameters $\mu_{\ln,r}$ and $\sigma_{\ln,r}$ using the same conversion formulas.
3. **Desired family size** (D_i)**:** D_i is drawn from a Weibull distribution, parameterized by its mean μ_d and standard deviation σ_d, which are used to approximate the Weibull shape parameter $\alpha_d = (\sigma_d/\mu_d)^{-1.086}$ and scale parameter $\lambda_d = \mu_d/\Gamma(1+1/\alpha_d)$. Samples are rounded to the nearest integer.
4. **Desired birth spacing** (B_i)**:** B_i (in months) is drawn from a lognormal distribution with a mean desired spacing μ_b and a standard deviation σ_b (both in months). These parameters are used to derive the underlying mean and standard deviation for the lognormal distribution.

The age-specific monthly probability of conception (fecundability), $\phi(x)$ at age x (in years), is a crucial component of the model. It is well established that fecundability typically increases from puberty, reaches a peak during the early adult years, and progressively declines thereafter towards menopause (Bendel & Hua, 1978; Larsen & Yan, 2000; Weinstein et al., 1990). To capture this characteristic age pattern, we model $\phi(x)$ over a reproductive window from $x_{\min} = 10$ to $x_{\max} = 50$ years. Age x is first rescaled

to $x_s = (x - x_{\min})/(x_{\max} - x_{\min})$, so $x_s \in [0, 1]$. Instead of using a standard polynomial basis, we employ Bernstein basis polynomials, which provide flexibility and ensure the curve is well-behaved at the boundaries of the reproductive window. Specifically, $\phi(x)$ is modelled as a linear combination of two Bernstein basis polynomials of degree 3:

$$\phi(x) = \beta_1[3x_s(1 - x_s)^2] + \beta_2[3x_s^2(1 - x_s)].$$

The terms $3x_s(1 - x_s)^2$ and $3x_s^2(1 - x_s)$ correspond to the two intermediate basis functions ($B_{1,3}(x_s)$ and $B_{2,3}(x_s)$) of a Bernstein basis of degree 3. The exclusion of the first and last basis functions from the model inherently forces the curve $\phi(x)$ to be zero at the start ($x_s = 0$) and end ($x_s = 1$) of the reproductive window. The parameters β_1 and β_2 are estimated and control the level and shape of the fecundability curve, allowing for a variety of empirically plausible age-fecundability profiles.

The actual probability of conception in a given month depends not only on this baseline age-specific fecundability but also on a woman's reproductive intentions and contraceptive use. As outlined earlier, if a woman i with current parity (number of previous births) k_i is *actively trying to conceive* (i.e. her current age is $\geq R_i$, her parity $k_i < D_i$, and sufficient time B_i has elapsed since her last birth), her probability of conception is $\phi(x)$.

Conversely, if a woman i with current parity k_i is sexually active but *not* actively trying to conceive, she is assumed to use contraception. Her monthly probability of conception is then $\kappa_{eff}\phi(x)$. The parameter $\kappa \in (0, 1)$ is an estimated baseline probability of contraceptive failure. To model an *increased contraceptive vigilance once desired parity (D_i) is met*, the effective failure rate κ_{eff} is defined as κ if her current parity $k_i < D_i$, and is reduced to κ^2 if $k_i \geq D_i$. This formulation reflects intensified efforts to avoid further births, substantially reducing the risk of an unplanned pregnancy (assuming $\kappa < 1$) after a woman reaches her desired family size.

Table 1 summarizes the estimated model parameters.

3 Data

Our analysis draws on two large, nationally representative sources. For countries outside the United States, we use the DHS, a leading initiative of the United States Agency for International Development (USAID) that has conducted over 300 surveys in more than 90 low- and middle-income countries. Each DHS round collects complete retrospective birth histories, allowing the reconstruction of every woman's fertility trajectory. Data for the United States are drawn from several early cycles of the NSFG. These surveys, conducted by the National Center for Health Statistics, are designed to provide detailed national data on family life, marriage, divorce, contraception, and importantly for this study, complete pregnancy and birth histories. We utilize these data as compiled and harmonized by the Integrated Fertility Survey Series (IFSS), applying IFSS sampling weights to generate population-level estimates. For both DHS and

Table 1. Model parameters and their descriptions

Parameter	Description
μ_s	Mean age at sexual initiation (in years), for a lognormal distribution.
σ_s	Standard deviation of the age at sexual initiation.
δ_r	Mean gap (in years) between sexual initiation and intentional reproduction.
σ_r	Standard deviation of the age at start of intentional reproduction.
μ_d	Mean desired family size, for a Weibull distribution.
σ_d	Standard deviation of desired family size.
μ_b	Mean desired birth spacing (in months), for a lognormal distribution.
σ_b	Standard deviation of the desired birth spacing.
κ	Probability of contraceptive failure per month.
β_1	Defines the peak of the age-specific fecundability curve.
β_2	Defines the decline of the age-specific fecundability curve.

NSFG data, to achieve adequate sample sizes while ensuring cohorts of women were exposed to broadly similar social conditions, we pool available survey rounds and organize respondents into 10-year birth cohorts. The final cohorts analysed correspond to women born between 1938 and 1948 for the United States, and between 1966 and 1975 for Colombia, the Dominican Republic, and Peru.

A key consideration in selecting empirical settings was to align them with the behavioural logic embedded in our model. Thus, we narrowed the DHS universe to countries where adolescent childbearing is primarily unintended. This was operationalized by calculating, for each potential cohort, the proportion of births to women under 18 that were declared as unplanned, retaining only those populations where this share exceeded one-half. This selection criterion identifies populations where, for a significant fraction of teenagers, sexual activity and intentional reproduction are decoupled—a precondition for our model's assumption that early fertility is largely viewed as undesirable. It also serves to exclude contexts where prevailing norms, such as marriage-centred childbearing from a young age, would imply a data-generating process markedly different from the one our model specifies.

Because DHS sample sizes and the number of survey rounds differ widely across countries, we implemented a further filter based on cohort size. We selected DHS countries where the chosen cohort included a minimum of 4,000 women, a threshold established to guarantee reasonably precise estimates of both age-specific total and unplanned fertility rates across the entire reproductive span. This resulted in the inclusion of three DHS countries: Colombia, Peru, and the Dominican Republic. Together with the United States, these four nations constitute our study sample, offering a valuable cross-section of different fertility patterns and socio-economic contexts that allows for a robust examination of our model's performance across diverse demographic and social settings.

We compute the conventional age-specific fertility rate by dividing the number of births at each age by the number of person-years of exposure contributed at that age. To test how information on birth intentionality improves our estimates, we also construct an age-specific unplanned fertility rate (ASUFRs). This rate is defined as the number of *unplanned* births at a given age divided by person-years of exposure. The classification of a birth as unplanned is based on a harmonized strategy across the DHS and NSFG that combines direct survey questions about birth timing with a parity check (i.e. whether the birth exceeded the mother's ideal family size). This hybrid approach is designed to mitigate recall bias and ex-post rationalization, which are known to affect estimates based solely on intention questions (Casterline & El-Zeini, 2007).

4 Simulation-based inference

Inferring parameters for complex process-based models is often challenging because the likelihood function is analytically intractable or computationally prohibitive to evaluate. Simulation-based inference (SBI) provides a general framework for performing Bayesian inference in such simulator-based models by replacing explicit likelihood evaluations with learned probabilistic surrogates (see Deistler et al., 2025, for an overview). The core idea is to generate synthetic data under plausible parameter settings and use this data to train an inference network that captures the relationship between simulated data and parameters. Within this family, neural posterior estimation (NPE) is a widely used and conceptually simple approach which directly learns an approximation to the posterior distribution $p(\theta \mid x)$ from simulated data pairs (θ, x) (Papamakarios & Murray, 2016). In practice, NPE proceeds by drawing parameters θ from a prior distribution $p(\theta)$, using the simulator to generate the corresponding data $x \sim p(x \mid \theta)$, and then training a flexible neural density estimator to approximate the true posterior. We denote this estimator as $q_F(x, \phi)(\theta)$, where F represents the neural network (parameterized by ϕ) that takes data x as input and defines the shape of the density.

While basic NPE is effective, our study employs a powerful sequential variant known as automatic posterior transformation (APT), sometimes referred to as SNPE-C (where the suffix 'C' follows an alphabetical naming convention for successive iterations of the method) (Greenberg et al., 2019). Like other SNPE methods, APT enhances inferential efficiency by drawing parameter samples from an iteratively updated proposal distribution, $\tilde{p}(\theta)$, rather than solely from the fixed prior. The key challenge in this approach is to correct for the use of a proposal distribution, as naively training on the

resulting samples would lead to an incorrect posterior estimate. APT solves this by analytically transforming the posterior estimate itself. For each simulated data point drawn from a proposal $\tilde{p}_r(\theta)$, it re-weights the target density using the known ratio $\tilde{p}_r(\theta)/p(\theta)$. The network is then trained by minimizing the loss with respect to this transformed 'proposal posterior', denoted as $\tilde{q}_{x,\phi}(\theta)$. A key advantage of this formulation is that because the loss is valid for any proposal, the network can be trained on the cumulative set of all simulations generated across all previous rounds. This procedure allows the network to learn the true posterior without relying on the potentially unstable importance weights used by other methods (Greenberg et al., 2019).

This adaptive refinement progressively concentrates simulation effort on regions of high posterior probability, leading to a more accurate posterior approximation with fewer overall simulations. SNPE and related NPE techniques have found successful application in diverse scientific domains, including neuroscience (Deistler et al., 2022; Gonçalves et al., 2020; Groschner et al., 2022), astrophysics (Dax et al., 2025), cognitive science (von Krause et al., 2022), and exoplanet searches (Vasist et al., 2023), among others.

A key strength of the APT framework is its flexibility, as it supports a wide range of density estimators. In this study, we employ a neural spline flow (NSF) (Durkan et al., 2019). Normalizing flows model complex probability densities by transforming a simple base distribution (typically a standard multivariate Gaussian, $u \sim \mathcal{N}(0, I)$ with density p_u) through a sequence of invertible, differentiable mappings T. The resulting approximate posterior density for the parameters $\theta = T(u)$, denoted here as $q(\theta)$ (shorthand for the estimator $q_{F(x,\phi)}(\theta)$ defined above), can be evaluated exactly using the change of variables formula:

$$q(\theta) = p_u(T^{-1}(\theta))\left|\det \frac{\partial T^{-1}}{\partial \theta}\right|. \tag{1}$$

Here, T^{-1} is the inverse transformation mapping the complex parameters θ back to the simple base space u, and the final term is the absolute value of the Jacobian determinant, which accounts for the change in volume induced by the transformation. NSFs specifically parameterize T using monotonic rational-quadratic splines. This architecture is particularly advantageous for our application because the flexibility of the spline transformations allows the network to capture highly complex, non-Gaussian, and potentially multi-modal posterior shapes. Simultaneously, the analytic invertibility of the splines ensures that the Jacobian determinant is easy to compute, facilitating stable training and efficient sampling (for a comprehensive review on normalizing flows for probabilistic inference, see Papamakarios et al., 2021).

Algorithm 1 summarizes the complete APT procedure employed in this study, which was implemented using the `sbi` Python package (Boelts et al., 2025).

Algorithm 1: Automatic posterior transformation (APT). Adapted from Greenberg et al. (2019).

Input: Simulator with implicit likelihood $p(x|\theta)$, observed data x_o, prior $p(\theta)$, conditional density family $q_{F(x,\phi)}(\theta)$, number of rounds R, simulations per round N.
Output: The final posterior approximation $q_{F(x_o,\phi)}(\theta)$.

Set initial proposal distribution $\tilde{p}_1(\theta) \leftarrow p(\theta)$;
for $r = 1$ **to** R **do**
  for $j = 1$ **to** N **do**
    Sample $\theta_{r,j} \sim \tilde{p}_r(\theta)$;
    Simulate $x_{r,j} \sim p(x|\theta_{r,j})$;
  end
  Update network weights ϕ by minimizing the negative log-probability of the parameters over all collected data $\{(\theta_{i,j}, x_{i,j})\}$ from rounds $i = 1...r$:
  $\phi \leftarrow \arg\min_\phi \sum_{i=1}^{r}\sum_{j=1}^{N} -\log \tilde{q}_{x_{i,j},\phi}(\theta_{i,j})$;
  where $\tilde{q}_{x,\phi}(\theta) := q_{F(x,\phi)}(\theta)\frac{\tilde{p}(\theta)}{p(\theta)}\frac{1}{Z(x,\phi)}$ with $Z(x,\phi) = \int_\theta q_{F(x,\phi)}(\theta)\frac{\tilde{p}(\theta)}{p(\theta)}d\theta$;
  Set the next proposal to the current posterior estimate for the observed data x_o:
  $\tilde{p}_{r+1}(\theta) \leftarrow q_{F(x_o,\phi)}(\theta)$;
end

As formally defined in Algorithm 1, the loss function depends on a normalizing constant $Z(x, \phi)$ that is required to ensure the proposal posterior is a valid probability density. However, calculating $Z(x, \phi)$ requires integrating the product of the flow-based density and the proposal ratio over the parameter space. This integral has no closed-form analytical solution for NSF, and numerical integration is computationally intractable due to the high dimensionality of our model. To circumvent this, we employ the 'atomic' version of APT (as implemented in the `sbi` package), which sums over a discrete batch of parameters (or 'atoms') sampled from the proposal distribution. By contrasting the true parameter against this batch of alternative candidates, the method avoids explicit integration while yielding a valid estimate of the posterior density.

Inference was performed over $R = 6$ rounds, with an initial batch of $N = 18{,}000$ simulations followed by $N = 6{,}000$ simulations per subsequent round. This budget was determined to be sufficient based on the observed stabilization of the mean absolute percentage error between observed and simulated ASFRs.

4.1 Inference scenarios

Simulation-based inference typically relies on summarizing potentially high-dimensional data into lower-dimensional, sufficiently informative statistics. This can be done either through explicit, predefined summaries or via learned embeddings. The ability to work with a low-dimensional yet descriptive representation of the data is often essential for making parameter inference tractable, especially for complex simulators.

When working with aggregate data, as in our study, the dimensionality reduction is an inherent feature of the observation process. For instance, vital statistics or administrative registers implicitly aggregate thousands of individual fertility trajectories into relatively low-dimensional summaries like age-specific fertility rates. We thus inherit a pre-summarized dataset, limiting our ability to engineer or learn optimal summary statistics directly from the underlying microlevel process.

A critical question then arises: do these aggregate data provide sufficient information to tightly constrain the posterior distributions of the parameters in our individual-level model? The core challenge is that different combinations of microparameters (e.g. those governing desired family size, timing of intentional reproduction, or contraceptive failure) could potentially yield very similar aggregate fertility schedules, resulting in high posterior uncertainty. We address this identifiability problem in Bayesian terms, asking to what extent different combinations of data and prior information tighten the posterior distributions of the behavioural parameters.

While ASFRs represent the baseline data available in most settings, demographic analyses can sometimes draw upon richer sources of information. We therefore test how the parameter estimation process can be strengthened by incorporating such additional information. This can be done in two primary ways: either by augmenting the data with more detailed summary statistics or by encoding external knowledge into the model through more informative priors. Given that our model explicitly simulates the planned and unplanned nature of births, information on birth intentionality is a particularly relevant source of additional data. However, since this type of data is costly to collect and often unavailable, it is valuable to assess whether similar gains in estimation accuracy can be achieved by instead using informative priors on key behavioural parameters.

We therefore evaluate parameter recovery under three increasingly informative inference scenarios:

Scenario 1: ASFRs with weak priors: This baseline scenario uses only simulated age-specific fertility rates as summary statistics. All model parameters are assigned weakly informative priors.

Scenario 2: ASFRs with informative priors. This scenario tests the impact of incorporating plausible external information. The summary statistics are still limited to ASFRs, but we use narrower, more informative priors for the mean desired family size (μ_d), the mean gap to intentional reproduction (δ_r), and the mean desired birth spacing (μ_b). All other parameters retain their weakly informative priors.

Scenario 3: ASFRs and ASUFRs with weak priors. This scenario tests the impact of adding more detailed data. We revert to the weakly informative priors from Scenario 1, but we augment the summary statistics to include both simulated ASFRs and age-specific unplanned fertility rates.

The informative priors used in Scenario 2 simulate a context where a researcher might leverage external information to improve estimates. For the mean desired family size (μ_d), we construct the prior directly from the empirical survey distribution for the US case, mimicking a situation where such summary data is readily available. For the remaining two parameters, we simulate the process of knowledge transfer from a data-rich to a data-poor setting by using the posteriors from our most data-intensive setup (Scenario 3) as the informative priors for Scenario 2. This mimics a situation where results from a detailed prior study on a similar cohort/country are used to inform a new analysis where only ASFRs are available.

4.2 Prior specification

For our baseline scenario (Scenario 1), we selected weakly informative priors for all eleven model parameters to allow the data to drive the inference. The specific parameterization of these priors was chosen to cover a broad yet plausible range of values based on existing demographic literature and the inherent constraints of each parameter.

For strictly positive quantities like means and standard deviations (e.g. μ_s, σ_s), we used broad Gamma distributions. For example, the prior for the mean age at sexual initiation (μ_s) contains 95% of its probability mass concentrated between 14.4 and 28.9 years. This range comfortably contains existing estimates for the mean age at onset of sexual activity across diverse Western countries and cohorts, which typically falls between 16 and 19 years (Kerry et al., 2017; Wellings et al., 2006). Likewise, the prior for mean desired family size (μ_d) is centred at 4.5 children, with a 95% probability interval of [1.9, 8.1], again comfortably encompassing the available estimates for the cohorts studied (Bongaarts & Lightbourne, 1992; Sobotka & Beaujouan, 2014; Westoff & Ryder, 2015). For parameters where reliable external estimates are lacking, such as the gap to intentional reproduction (δ_r) and mean desired birth spacing (μ_b), we employed Uniform priors to define very wide yet plausible ranges (0 to 8 years for δ_r and 10 to 100 months for μ_b). The contraceptive failure probability ($\varkappa$) was assigned a Beta distribution. Its prior reflects a wide range of plausible contraceptive efficacy levels for our study cohorts, centred around a 20% failure rate but with significant probability assigned to both lower and higher values. Finally, the priors for the Bernstein coefficients (β_1 and β_2) were selected to ensure the resulting age-fecundability curves would generously contain available estimates of conception probabilities at early, peak, and later reproductive ages (ACO, 2014; Dunson et al., 2002; Schwartz & Mayaux, 1982; Wesselink et al., 2017).

5 Results

This section presents the results from the validation experiments designed to test our framework under the three information scenarios defined above. We first assess the model's ability to recover known parameters from simulated data using a cross-validation procedure (Section 5.1). We then apply the framework to empirical data from our four study countries, evaluating both the in-sample fit through posterior predictive checks and, most critically, the out-of-sample predictive performance against microlevel data not used for estimation.

5.1 Cross-validation

Our validation procedure involves repeatedly generating datasets with known parameters and then attempting to recover these parameters using our inference procedure. Specifically, we repeat the following steps for I iterations:

1. **Generate Ground-Truth Data:** In each iteration i (for $i = 1, \ldots, I$), we draw a distinct 'true' parameter vector, denoted $\theta^{*(i)}$, from the prior distribution $p(\theta)$ defined for the given scenario. Using our individual-level simulation model, we then generate a corresponding vector of aggregate

summary statistics, denoted $x_0^{(i)}$, which serves as the pseudo-observed data for this iteration. The pair $(\theta^{*(i)}, x_0^{(i)})$ constitutes a ground-truth dataset where the data-generating parameters are known.

2. **Perform Full Posterior Inference:** For each pseudo-observed dataset $x_0^{(i)}$, we independently run our complete SNPE procedure. This involves training a neural density estimator over R rounds with N_{sim} new simulations drawn from the evolving proposal distribution in each round, to obtain an approximation of the posterior distribution $q_\phi(\theta \mid x_0^{(i)})$.
3. **Evaluate Parameter Recovery:** From each resulting posterior approximation $q_\phi(\theta \mid x_0^{(i)})$, we use the posterior mean as the point estimate for the parameters, denoted $\hat{\theta}^{(i)}$. We then assess the accuracy of recovery by comparing this estimate $\hat{\theta}^{(i)}$ with the known true generating parameters $\theta^{*(i)}$.

By performing these I independent inference runs, we obtain a distribution of estimation errors (e.g. $\hat{\theta}^{(i)} - \theta^{*(i)}$) for each model parameter under each of the three informational scenarios. We then summarize these error distributions using the normalized root-mean-squared error (RMSE). Lower RMSE values indicate better parameter recovery and allow for a direct comparison of how effectively each informational setup constrains the model parameters.

Figure 1 presents the results of our cross-validation analysis, performed over 25 folds for the seven core model parameters (Cross-validation results for the four heterogeneity parameters are presented in online supplementary material, Figure A1 in the Appendix). The first three columns display scatterplots comparing the known 'true' parameter values (horizontal axis) against the posterior mean estimates recovered by our model (vertical axis) for each of the three scenarios. Perfect parameter recovery would place all points directly on the dashed 45° reference line. The final column summarizes the performance for each parameter across scenarios by plotting the normalized RMSE.

The results demonstrate that parameter recovery is already reasonably robust in the baseline scenario. In the scatterplots for this scenario, the estimated parameter values consistently cluster around the true values on the 45-degree line, with some parameters exhibiting tighter clustering than others. This indicates that the model can successfully disambiguate the majority of its behavioural components from the aggregate fertility schedule alone. As expected, parameter recovery improves systematically when additional data is incorporated (Scenario 3). The use of informative priors (Scenario 2) yields more mixed results, reducing the RMSE for some parameters but showing slight increases for others compared to the baseline.

5.2 Model validation

The cross-validation experiments (Section 5.1) confirmed the capacity of our SBI framework to recover key microlevel parameters from aggregate ASFRs. To further scrutinize the model's performance, we conduct two additional forms of validation. First, posterior predictive checks, to assesses the model's ability to replicate the aggregate data used for inference. The second validation stage examines the model's capacity to predict out-of-sample individual-level characteristics of the observed cohorts, such as distributions of desired family size, age at first sexual intercourse, and birth interval durations, which were not used during the parameter estimation process.

We present the main validation results using the posteriors obtained under Scenario 1 (ASFRs with weak priors). We focus on this baseline scenario as it provides the most stringent test of our model's capacity to infer individual-level mechanisms solely from aggregate fertility rates. Results for the scenarios incorporating additional information (Scenarios 2 and 3) are presented in the online supplementary material, Appendix.

5.2.1 Posterior predictive checks

To perform posterior predictive checks, we draw 5,000 samples from the joint posterior distribution for each of the four countries (United States, Colombia, Dominican Republic, and Peru). For each

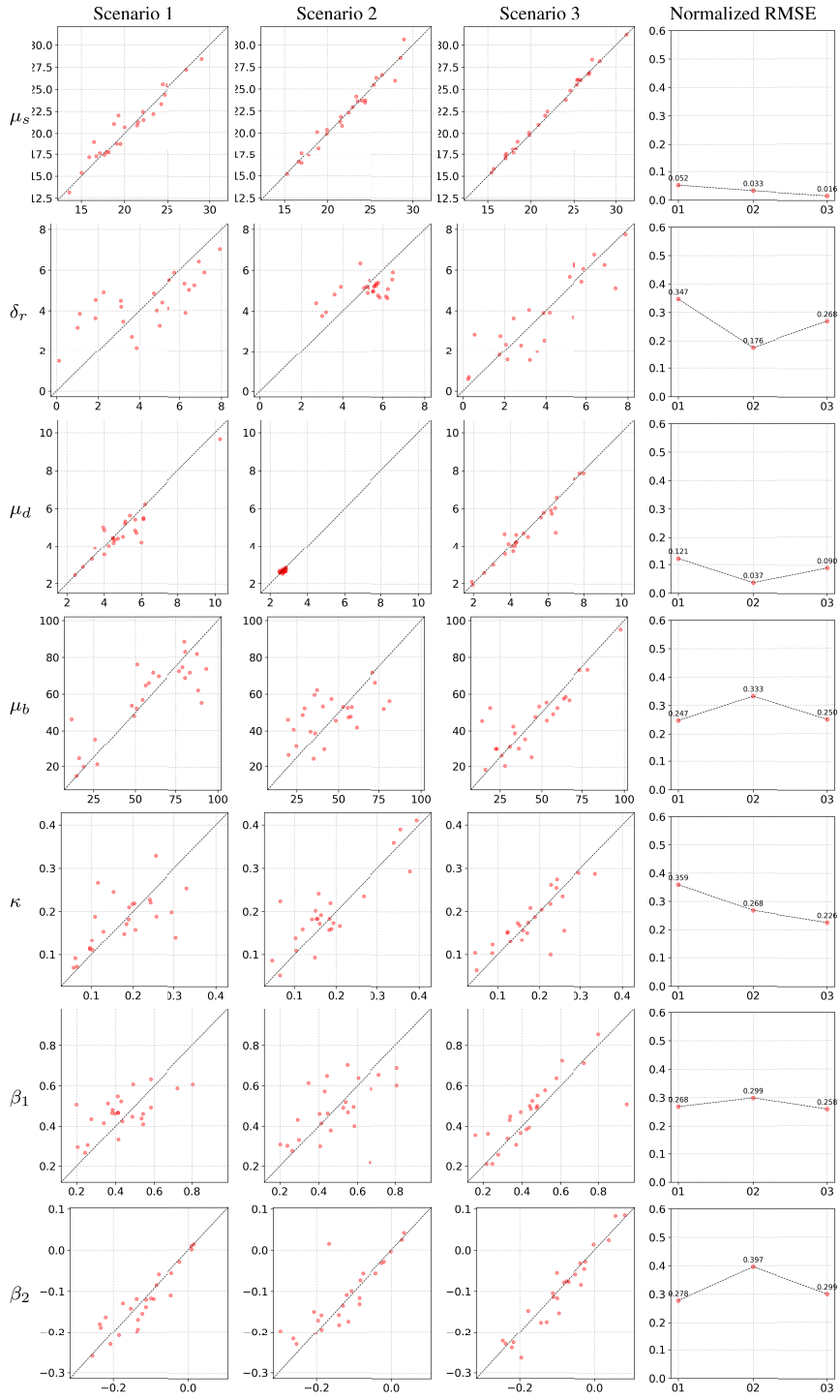

Figure 1. Cross-validation of seven core model parameters over 25 folds. Columns 1–3 display scatterplots of true parameter values (horizontal axis) versus posterior mean estimates (vertical axis) for Scenarios 1, 2, and 3, with the dashed 45° reference line indicating perfect recovery. Column 4 reports the scenario-specific root-mean-squared error. Cross-validation for the four heterogeneity parameters is shown in the online supplementary material, Appendix.

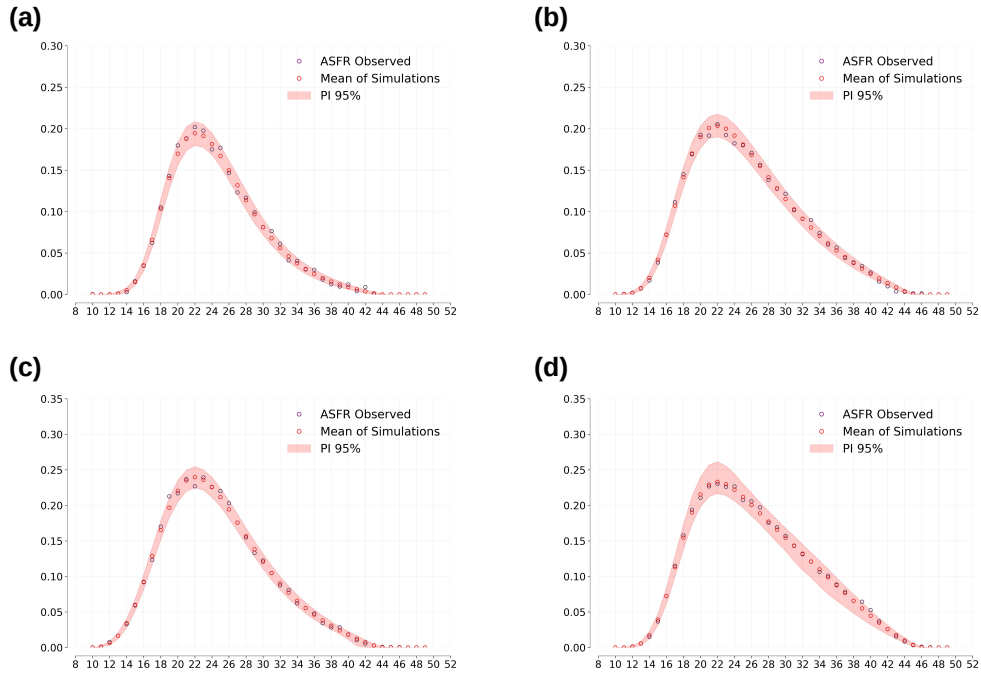

Figure 2. Posterior predictive checks for age-specific fertility rates. Comparison of observed rates (purple open circles) with posterior predictions for four countries: (a) United States, (b) Colombia, (c) Dominican Republic, and (d) Peru. Red open circles indicate the mean of 5,000 posterior predictive simulations. Red shaded areas represent the 95% posterior predictive intervals. Parameters were estimated under Scenario 1 (ASFRs with weakly informative priors). (a) United States, (b) Colombia, (c) Dominican Republic, and (d) Peru.

parameter sample, we simulate a full cohort of individual reproductive histories using our model and then recompute the ASFRs.

Figure 2 presents observed ASFRs against their corresponding posterior predictive distributions. Observed ASFRs are represented by purple open circles. Red open circles represent the mean of the 5,000 simulated ASFRs schedules, while the shaded red areas depict the 95% posterior predictive intervals.

Across all four diverse demographic contexts, the model demonstrates a strong capacity to replicate the observed data. The posterior predictive means closely track the observed ASRFs, capturing key features such as the location and height of peak fertility, and the age pattern of decline. Notably, across all countries, almost all observed points fall well within the 95% posterior predictive interval, indicating that the model, conditioned on the inferred parameters, generates aggregate outcomes highly consistent with the observed data.

These results confirm that our framework, even with minimal data inputs, can consistently and accurately reproduce the population-level fertility schedules it was intended to explain.

5.3 Estimated marginal posteriors

Figure 3 displays the marginal posterior distributions for all 11 model parameters, estimated under Scenario 1 for each of the four countries. The results clearly show that the aggregate ASFRs data provide substantial information for constraining most parameters. In nearly all cases, the posterior distributions (in red) are considerably sharper and more concentrated than their corresponding wide priors (in grey), indicating a significant reduction in uncertainty after conditioning on the observed data.

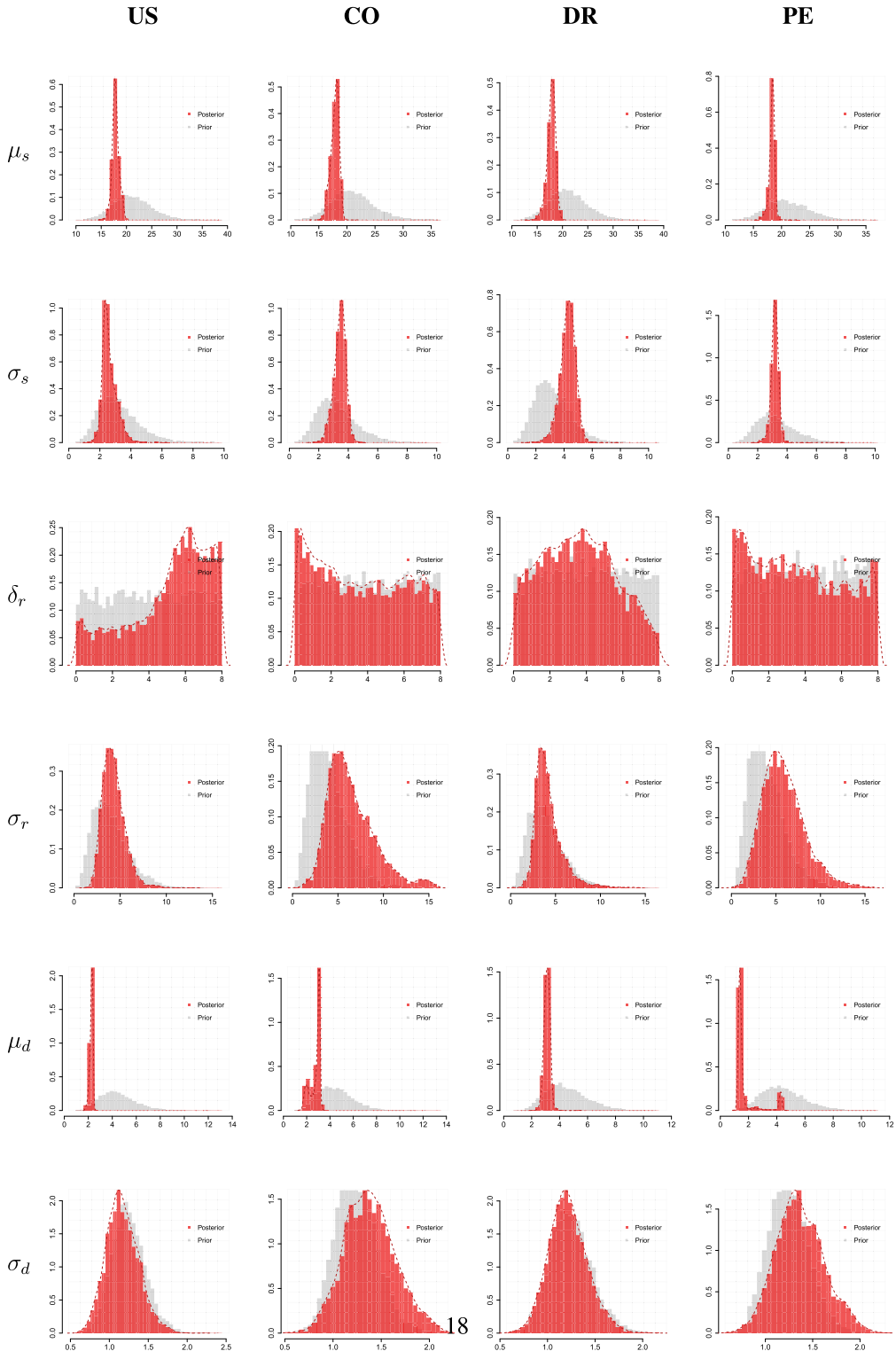

Figure 3. Marginal posterior distributions of the 11 key model parameters for the United States, Colombia, Dominican Republic, and Peru, inferred from empirical data under Scenario 1.

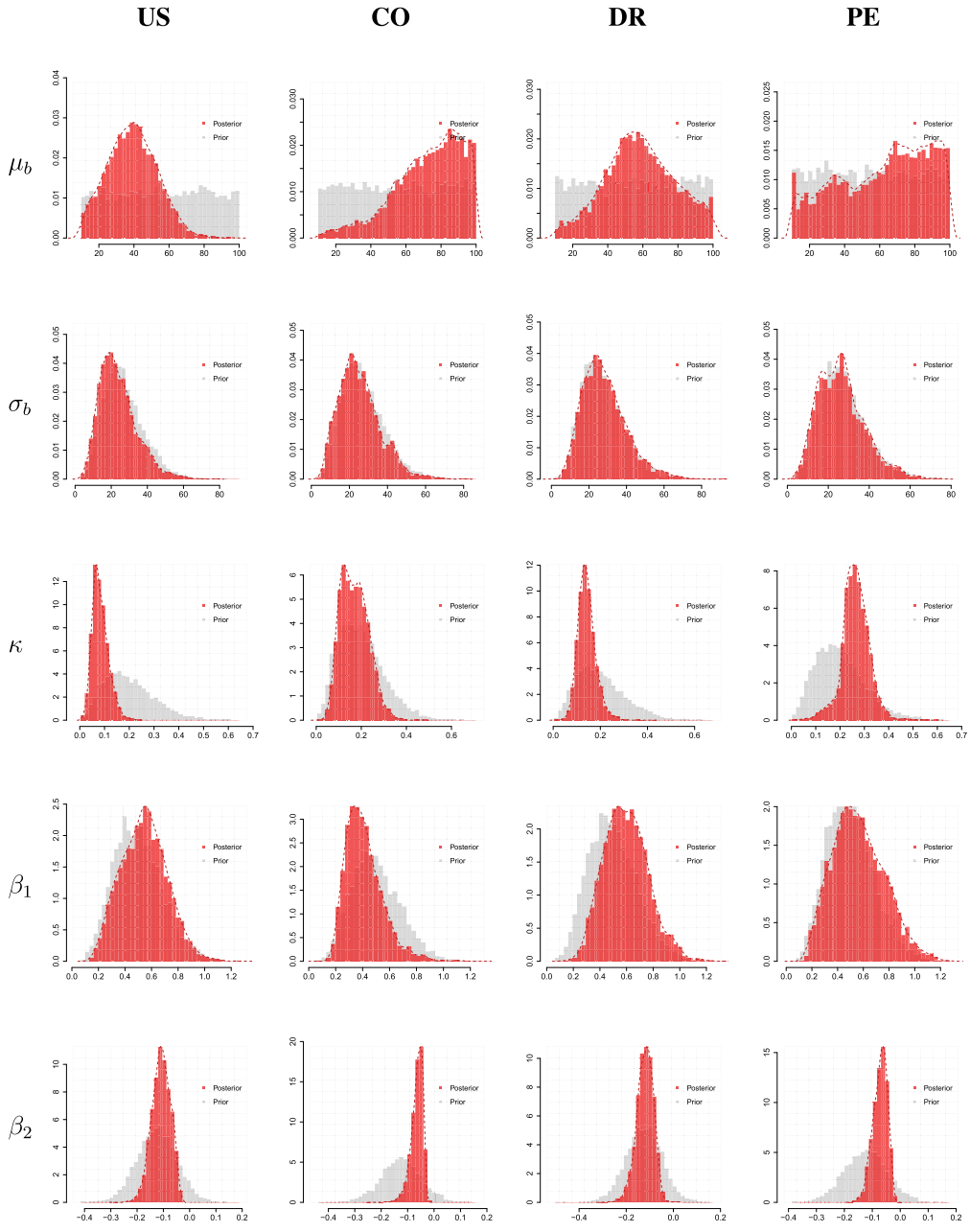

Figure 3. Continued

The degree to which parameters are constrained varies logically across the different components of the model. Parameters related to the timing of sexual initiation (μ_s, σ_s), the age-pattern of fecundability (β_1, β_2), desired family size (μ_d, σ_d), and contraceptive failure (κ) are all well-constrained, exhibiting sharp, unimodal posterior distributions. The posteriors also reveal plausible cross-national differences. For example, the mean desired family size (μ_d) for the US cohort is lower than in the other three countries, as expected. Similarly, the contraceptive failure probability (κ) is estimated to be lowest in the United States, a finding that aligns with the socio-demographic context of these cohorts. In contrast, parameters governing the timing between life-course events, like the gap to intentional

reproduction (δ_r) and the spacing between births (μ_b, σ_b), show higher posterior uncertainty. Their posteriors are wider and less distinguished from their priors, suggesting that while ASFRs alone are sufficient to identify the core components of the model, pinpointing the precise timing of *intended* fertility is more challenging without additional data on birth intentionality. However, it is important to note that this remaining uncertainty in specific parameters does not significantly degrade the model's overall predictive performance, as evidenced by the successful posterior predictive checks (Figure 2). Finally, the estimation for these timing-related parameters improves substantially when direct information on birth intentionality is included, as shown in the results for Scenario 3 (see online supplementary material, Figure A3 in the Appendix).

5.3.1 Out-of-sample validation

To assess the model's ability to generate realistic individual-level reproductive behaviours not directly used in parameter estimation, we use parameter values representative of the posterior distribution (specifically, the posterior mean obtained under Scenario 1, as described in Section 5.2) to drive new simulations. These simulations yield full synthetic reproductive histories. We then compare the distributions of key microlevel outcomes from these synthetic life courses, namely, age at first sexual activity, desired family size, and interbirth intervals, with their empirical counterparts derived from the survey data (Figure 4). This side-by-side comparison provides a direct appraisal of the model's realism in capturing individual-level reproductive patterns.

5.3.1.1 Overall fit

Across the four countries and three microlevel indicators, the distributions generated from the model's synthetic life courses generally provide a strong approximation of the empirical distributions observed in the survey data. This indicates that the parameters inferred solely from aggregate ASFRs enable the reconstruction of distinct microbehavioural dimensions without any specific tuning towards these individual-level targets. We quantify the discrepancy between the simulated and empirical distributions using the Jensen–Shannon divergence (JSD). The JSD is a symmetric measure of similarity between two probability distributions, ranging from 0 (for identical distributions) to 1 (for fully disjoint distributions) when using the base-2 logarithm. Thus, values close to 0 indicate strong agreement between the model predictions and the observed data. With the notable exception of desired family size in Peru, the JSD between simulated and observed distributions never exceeds 0.068 bits.

5.3.1.2 Age at first sex

The simulated distributions reproduce the overall shape and cross-country ordering of sexual debut: the mean age is highest in the United States and Peru, slightly lower in Colombia, and lowest in the Dominican Republic. The model naturally smooths over sharp heaping in the survey data (e.g. at age 18), and for the United States, the predicted curve is displaced slightly to the left of the observed one. Even with these minor mismatches, the alignment is striking, given the relative simplicity of the model and of the data used for its estimation.

5.3.1.3 Desired family size

Overall, the simulated preferences closely match the shape of the observed distributions, reproducing the main peak around two to three children and the gradual taper into larger family ideals. A clear deviation appears at exactly two children, where the empirical data show a sharp bump (a clear outcome of strong social norms) that the model's simple two-parameter form inevitably smooths over. A more significant discrepancy is observed for Peru, where the model's predicted distribution is shifted to the left of the empirical one (JSD: 0.144 bits). This specific mismatch is consistent with the higher posterior uncertainty observed for the Peruvian cohort and may reflect underlying issues like lower data quality or greater-than-modelled heterogeneity in reproductive behaviour in this cohort. It is important to note, however, that the alignment of these distributions improves substantially when additional information is incorporated (see online supplementary material, Figure A5 in the Appendix).

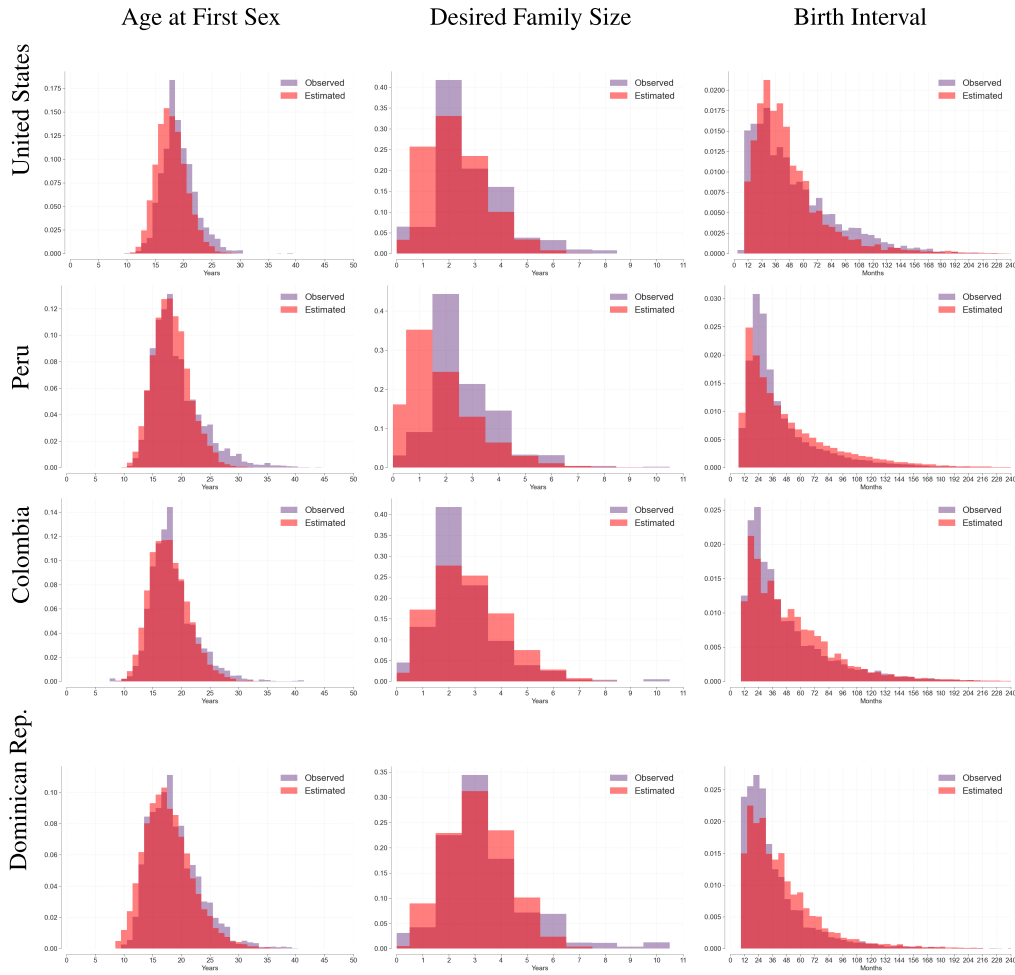

Figure 4. Validation on microlevel outcomes. Purple (dark) bars denote the observed survey distributions, while red (lighter) bars show the model-implied counterparts generated from posterior draws. Columns display key behavioural dimensions: age at first sex, desired family size, and birth interval, whereas rows correspond to each study cohort.

5.3.1.4 Birth intervals

The synthetic data recover the principal spacing pattern: a broad peak around 2–3 years followed by a gradual decline towards longer gaps. The model undershoots the cluster of very short gaps and shows a slightly heavier tail beyond 5 years. Apart from these edges of the distribution, the match is close across the main mass of birth intervals.

6 Discussion

This study has introduced a simulation-based Bayesian framework capable of inferring complex individual-level reproductive behaviours solely from aggregate population data. By integrating a demographically interpretable individual-level simulation model with SNPE, we demonstrated that the core parameters governing the reproductive process can be recovered from age-specific fertility rates alone. While incorporating additional data (Scenario 3) serves to further reduce parameter estimation error, the use of informative priors (Scenario 2) yields variable improvements. Nonetheless,

our findings show that the information contained in ASFRs alone provides a robust baseline estimation. The model's ability to accurately reproduce observed fertility schedules, as confirmed by our posterior predictive checks, offers strong support for this conclusion. However, the most compelling validation of our approach lies in the model's ability, using these inferred parameters, to accurately predict out-of-sample microlevel distributions for behaviours such as age at first sexual intercourse, desired family size, and birth intervals, none of which informed the original estimation. This establishes a new, statistically rigorous bridge between macrolevel fertility observations and their underlying microbehavioural drivers.

Methodologically, this work serves as an empirical demonstration of SNPE's utility for parameter inference in interpretable, yet complex, social science simulation models where direct likelihood calculation is prohibitive. The ability to recover these parameters from a single, widely available data series like ASFRs offers transferable insights for simulation-based inference across other disciplines facing similar inverse problems. Substantively, our framework provides a novel toolkit for demographers, enabling the estimation of behaviourally meaningful parameters (e.g. mean desired family size, age at intentional reproduction, and contraceptive failure rates) even in data-scarce contexts, thereby offering a mechanistic complement to traditional microdata analyses.

This capability to infer interpretable behavioural parameters has the potential to enhance population forecasting, moving beyond traditional methods that often rely on extrapolating aggregate trends or employ macrolevel parameters with limited behavioural grounding. One promising avenue opened by our approach is the generation of 'behaviourally explicit population forecasts' in which future scenarios are constructed from hypothesized changes in these underlying, microlevel behavioural parameters. Nevertheless, realizing this forecasting potential is likely to present a series of challenges and will require additional modelling and validation of how these behavioural parameters evolve over time. A complementary, statistical approach to incorporating individual-level information into fertility forecasting has been introduced by Ellison et al. (2024), who use a Bayesian parity-specific model combining survey birth histories with population-level fertility rates.

Crucially, by first distilling complex fertility dynamics into these more causally proximate behavioural components, our framework also facilitates a clearer understanding of how external covariates, such as education or economic conditions, shape reproductive outcomes. Machine learning models, for instance, may more effectively identify systematic relationships when predicting these interpretable behavioural parameters rather than attempting to directly predict highly aggregated and confounded ASFRs. This improved ability to model behaviour mechanisms provides more robust and theoretically sound inputs for scenario-based population projections. Furthermore, by reducing the data requirements for microsimulation, our framework makes the construction of 'demographic digital twins' a more accessible and feasible tool for policy exploration.

Despite these advances, the present framework has limitations that define important avenues for future development. First, the individual-level model, while behaviourally detailed, necessarily simplifies the complex dynamics of human reproduction. Second, the data generating process encoded by our model is suited to contemporary populations where reproduction is substantially decoupled from sexuality and where such articulable birth control decision-making is prevalent. This may limit its direct applicability to contexts with more natural fertility regimes or where reproductive agency and intentions are structured differently. Third, the framework as presented assumes relatively homogeneous cohorts. Applying the model to highly heterogeneous populations where reproductive behaviour differs systematically by socioeconomic characteristics would likely require disaggregated input data. A final consideration is that while ASFRs provide a robust foundation for estimation, applications that demand higher precision or focus on specific mechanisms may require online supplementary material, supplementary data that are not always easily available.

These limitations point toward several exciting directions for future work. First, strategies can be developed to enhance the precision and utility of the framework, particularly for analyses relying solely on ASFRs. One important avenue is the incorporation of stronger, empirically-grounded priors on the key behavioural parameters, allowing researchers to leverage existing knowledge from demographically similar populations. Another promising direction is the development of methods to generate plausible online supplementary material, supplementary data; for instance, one could use the

learned associations between covariates (e.g. education and age) and birth intentionality from available survey to produce plausible estimates in data-scarce contexts. A second major direction involves adapting the framework's core structure to broaden its applicability and complexity. A key extension would be to explicitly model population heterogeneity. In the many contexts where data such as education-specific ASFRs are available, the framework could be adapted to infer distinct behavioural parameters for different subgroups. For populations closer to natural fertility, the model could be simplified to focus on a smaller set of parameters, while for contexts where Western notions of 'unplanned' births are less salient, research could explore identifying and incorporating alternative summary statistics that better reflect local modes of fertility regulation.

Conflicts of interest

The authors declare no conflict of interest.

Funding

The authors received no specific funding for this work.

Data availability

Replication code, derived data, and instructions to reproduce the analyses are available at github.com/dciganda/asfr_snpe. The underlying individual-level survey microdata are available from the DHS Program and from the National Survey of Family Growth (NSFG) via their respective data providers.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series A*.

References

ACO. (2014). Female age-related fertility decline. *Fertility and Sterility*, *123*(3), 719–721. Committee Opinion No. 589 https://doi.org/10.1016/j.fertnstert.2013.12.032

Beaumont M. A. (2010). Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, *41*(1), 379–406. https://doi.org/10.1146/ecolsys.2010.41.issue-1

Bendel J.-P., & Hua C.-I. (1978). An estimate of the natural fecundability ratio curve. *Social Biology*, *25*(3), 210–227. https://doi.org/10.1080/19485565.1978.9988340

Bijak J. (2022). *Towards Bayesian model-based demography: Agency, complexity and uncertainty in migration studies*. Springer Nature.

Bijak J., & Hilton J. (2021). *Uncertainty quantification, model calibration and sensitivity*. In *Towards Bayesian model-based demography: Agency, complexity and uncertainty in migration studies* (pp. 71–92). Springer.

Boelts J., Deistler M., Gloeckler M., Tejero-Cantero A., Lueckmann J.-M., Moss G., Steinbach P., Moreau T., Muratore F., Linhart J., Durkan C., Vetter J., Miller B. K., Herold M., Ziaeemehr A., Pals M., Gruner T., Bischoff S., Krouglova N., … Macke J. H. (2025). sbi reloaded: A toolkit for simulation-based inference workflows. *Journal of Open Source Software*, *10*(108), 7754. https://doi.org/10.21105/joss

Bongaarts J. (1977). A dynamic model of the reproductive process. *Population Studies*, *31*(1), 59–73. https://doi.org/10.1080/00324728.1977.10412747

Bongaarts J. (1978). A framework for analyzing the proximate determinants of fertility. *Population and Development Review*, *4*, 105–132. https://doi.org/10.2307/1972149

Bongaarts J., & Lightbourne R. (1992). Fertility preferences in Latin America: Trends and differentials in seven countries. *Notas De Población*, *20*(55), 79–102.

Brass W. (1958). The distribution of births in human populations in rural Taiwan. *Population Studies*, *12*(1), 51–72. https://doi.org/10.1080/00324728.1958.10404370

Brass W. (1974). Perspectives in population prediction: Illustrated by the statistics of England and Wales. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *137*(4), 532–570. https://doi.org/10.2307/2344713

Burch T. K. (2018). *Model-based demography: Essays on integrating data, technique and theory*. Springer Nature.

Casterline J. B., & El-Zeini L. O. (2007). The estimation of unwanted fertility. *Demography*, *44*(4), 729–745. https://doi.org/10.1353/dem.2007.0043

Chandola T., Coleman D. A., & Hiorns R. W. (1999). Recent European fertility patterns: Fitting curves to 'distorted' distributions. *Population Studies*, *53(3)*, 317–329. https://doi.org/10.1080/00324720308089

Ciganda D., & Todd N. (2024). Modelling the age pattern of fertility: An individual-level approach. *Royal Society Open Science*, *11*(11), Article 240366. https://doi.org/10.1098/rsos.240366

Coale A. J., & Trussell T. J. (1974). Model fertility schedules: Variations in the age structure of childbearing in human populations. *Population Index*, *40*, 185–258. https://doi.org/10.2307/2733910

Courgeau D., Bijak J., Franck R., & Silverman E. (2016). Model-based demography: Towards a research agenda. *Agent-based Modelling in Population Studies: Concepts, Methods, and Applications*, *41*, 29–51. https://doi.org/10.1007/978-3-319-32283-4_2

Davis K., & Blake J. (1956). Social structure and fertility: An analytic framework. *Economic Development and Cultural Change*, *4*(3), 211–235. https://doi.org/10.1086/449714

Dax M., Green S. R., Gair J., Gupte N., Pürrer M., Raymond V., Wildberger J., Macke J. H., Buonanno A., & Schölkopf B. (2025). Real-time inference for binary neutron star mergers using machine learning. *Nature*, *639*(8053), 49–53. https://doi.org/10.1038/s41586-025-08593-z

Deistler M., Boelts J., Steinbach P., Moss G., Moreau T., Gloeckler M., Rodrigues P. L. C., Linhart J., Lappalainen J. K., Kurt Miller B., Gonçalves P. J., Lueckmann J.-M., Schröder C., & Macke J. H. (2025). 'Simulation-based inference: A practical guide', arXiv, arXiv: 2508.12939, preprint: not peer reviewed.

Deistler M., Macke J. H., & Gonçalves P. J. (2022). Energy-efficient network activity from disparate circuit parameters. *Proceedings of the National Academy of Sciences*, *119*(44), Article e2207632119. https://doi.org/10.1073/pnas.2207632119

Dunson D. B., Colombo B., & Baird D. D. (2002). Changes with age in the level and duration of fertility in the menstrual cycle. *Human Reproduction*, *17*(5), 1399–1403. https://doi.org/10.1093/humrep/17.5.1399

Durkan C., Bekasov A., Murray I., & Papamakarios G. (2019). Neural spline flows. *Advances in Neural Information Processing Systems*, *32*. https://doi.org/10.48550/arXiv.1906.04032

Ellison J., Berrington A., Dodd E., & Forster J. J. (2024). Combining individual-and population-level data to develop a Bayesian parity-specific fertility projection model. *Journal of the Royal Statistical Society: Series C, Applied Statistics*, *73*(2), 275–297. https://doi.org/10.1093/jrsssc/qlad095

Gini C. (1924). Premières recherches sur la fécondabilité de la femme. In *Proceedings of the International Mathematical Congress* (Vol. 2).

Gonçalves P. J., Lueckmann J.-M., Deistler M., Nonnenmacher M., Öcal K., Bassetto G., Chintaluri C., Podlaski W. F., Haddad S. A., Vogels T. P., Greenberg D. S., & Macke J. H. (2020). Training deep neural density estimators to identify mechanistic models of neural dynamics. *eLife*, *9*, Article e56261. https://doi.org/10.7554/eLife.56261

Granholm R., Gauthier A., & Stulp G. (2025). Examining the relationships between education, coresidential unions, and the fertility gap by simulating the reproductive life courses of Dutch women. *Demographic Research*, *52*(4), 797–848. https://doi.org/10.4054/DemRes.2025.52.24

Greenberg D., Nonnenmacher M., & Macke J. (2019). Automatic posterior transformation for likelihood-free inference. In *International Conference on Machine Learning* (pp. 2404–2414). PMLR.

Groschner L. N., Malis J. G., Zuidinga B., & Borst A. (2022). A biophysical account of multiplication by a single neuron. *Nature*, *603*(7899), 119–123. https://doi.org/10.1038/s41586-022-04428-3

Hartig F., Calabrese J. M., Reineking B., Wiegand T., & Huth A. (2011). Statistical inference for stochastic simulation models–theory and application. *Ecology Letters*, *14*(8), 816–827. https://doi.org/10.1111/j.1461-0248.2011.01640.x

Henry L. (1953). Fondements théoriques des mesures de la fécondité naturelle. *Revue de l'Institut International de Statistique/Review of the International Statistical Institute*, *21*(3), 135–151. https://doi.org/10.2307/1401425

Hilton J., & Bijak J. (2016). *Design and analysis of demographic simulations*. In *Agent-based modelling in population studies: Concepts, methods, and applications* (pp. 211–235). Springer.

Hoem B., & Hoem J. M. (1989). The impact of women's employment on second and third births in modern Sweden. *Population Studies*, *43*(1), 47–67. https://doi.org/10.1080/0032472031000143846

Kerry L. D. M., Mallick L., & Allen C. (2017). *Sexual and reproductive health in early and later adolescence: DHS data on youth age 10–19*. (Technical report). USAID, The DHS Program, Rockville, Maryland, USA.

Larsen U., & Yan S. (2000). The age pattern of fecundability: An analysis of French Canadian and Hutterite birth histories. *Social Biology*, *47*(1–2), 34–50. https://doi.org/10.1080/19485565.2000.9989008

Mazzuco S., & Scarpa B. (2015). Fitting age-specific fertility rates by a flexible generalized skew normal probability density function. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *178*(1), 187–203. https://doi.org/10.1111/rssa.12053

Papamakarios G., & Murray I. (2016). Fast ε-free inference of simulation models with Bayesian conditional density estimation. *Advances in Neural Information Processing Systems*, *29*.

Papamakarios G., Nalisnick E., Rezende D. J., Mohamed S., & Lakshminarayanan B. (2021). Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, *22*(57), 1–64. https://doi.org/10.48550/arXiv.1912.02762

Potter R. G. (1972). Additional births averted when abortion is added to contraception. *Studies in Family Planning*, *3*(4), 53–59. https://doi.org/10.2307/1965360

Ridley J. C., & Sheps M. C. (1966). An analytic simulation model of human reproduction with demographic and biological components. *Population Studies*, *19*(3), 297–310. https://doi.org/10.1080/00324728.1966.10406018

Schmertmann C. P. (2003). A system of model fertility schedules with graphically intuitive parameters. *Demographic Research*, *9*, 81–110. https://doi.org/10.4054/DemRes.2003.9.5

Schwartz D., & Mayaux M. J. (1982). Female fecundity as a function of age: Results of artificial insemination in 2,193 nulliparous women with azoospermic husbands. *New England Journal of Medicine*, *306*(7), 404–406. https://doi.org/10.1056/NEJM198202183060706

Singh S. (1963). Probability models for the variation in the number of births per couple. *Journal of the American Statistical Association*, *58*(303), 721–727. https://doi.org/10.1080/01621459.1963.10500882

Sobotka T., & Beaujouan É. (2014). Two is best? The persistence of a two-child family ideal in Europe. *Population and Development Review*, *40*(3), 391–419. https://doi.org/10.1111/padr.2014.40.issue-3

Vasist M., Rozet F., Absil O., Mollière P., Nasedkin E., & Louppe G. (2023). Neural posterior estimation for exoplanetary atmospheric retrieval. *Astronomy & Astrophysics*, *672*, A147. https://doi.org/10.1051/0004-6361/202245263

von Krause M., Radev S. T., & Voss A. (2022). Mental speed is high until age 60 as revealed by analysis of over a million participants. *Nature Human Behaviour*, *6*(5), 700–708. https://doi.org/10.1038/s41562-021-01282-7

Weinstein M., Wood J. W., Stoto M. A., & Greenfield D. D. (1990). Components of age-specific fecundability. *Population Studies*, *44*(3), 447–467. https://doi.org/10.1080/0032472031000144846

Wellings K., Collumbien M., Slaymaker E., Singh S., Hodges Z., Patel D., & Bajos N. (2006). Sexual behaviour in context: A global perspective. *The Lancet*, *368*(9548), 1706–1728. https://doi.org/10.1016/S0140-6736(06)69479-8

Wesselink A. K., Rothman K. J., Hatch E. E., Mikkelsen E. M., Sørensen H. T., & Wise L. A. (2017). Age and fecundability in a North American preconception cohort study. *American Journal of Obstetrics and Gynecology*, *217*(6), 667–e1. https://doi.org/10.1016/j.ajog.2017.09.002

Westoff C. F., & Ryder N. B. (2015). *The contraceptive revolution*. Princeton University Press.

Xie Y. (2000). Demography: Past, present, and future. *Journal of the American Statistical Association*, *95*(450), 670–673. https://doi.org/10.1080/01621459.2000.10474248