Probabilistic projection of UK kinship

Joe William Barrass Butterick[1], Joanne Ellison[1], Peter W. F. Smith[1], Erengul Dodd[1], Jakub Bijak[1], Jonathan J. Forster[2], and Jason Hilton[1]

[1]University of Southampton
[2]University of Warwick

1 Introduction

The ages at which people die and have children are changing. Rising living standards are extending life expectancy, for instance, by 2070 there are expected to be double the number of people aged 80 and above in the UK (ONS, 2024). At the same time, improved access to labour market opportunities is delaying the average age of childbearing. These socio-demographic factors are re-shaping kinship networks. An emergent so-called "double sandwich" generation (due to 4 generations co-existing at one time) of 40-50 year-old individuals must simultaneously care for grand-children – thus enabling their children to work, and parents – who on average being aged 80 plus, also require support (Butterick et al., 2024). Family networks play a major role in the process of ageing and are expected to continue as the main source of late life care. Because demographic trends are consistently altering the inter-generational structures of family, the types, or nature, of kin available to care is changing. It is thus of great policy interest to understand in 50 years time, which kin are available to offer an individual financial, emotional, and other forms of support.

A simple and computationally efficient way to estimate kinship is through a matrix population model (Caswell, 2019). Matrix models take as inputs, demographic rate data, and give as outputs expectations of (or potentially variances in) the age-distributions of kin relative to a reference individual, "Focal". The theoretical synthesis has been extended to account for kinship within time-varying demographic rates (Caswell & Song, 2021), multi-state (Caswell, 2020) and two-sex populations (Caswell, 2022). Recently stochastic population dynamics (Caswell, 2024) have also been considered. Accompanying these formal models is the R package "DemoKin" implementation; recently used by Alburez-Gutierrez et al. (2023) to produce probabilistic global kinship forecasts. In this research we apply a similar approach, forecasting UK kinship networks, structured by parity level, up to the year 2070. We achieve this by combining the models proposed by Caswell (2020), Caswell & Song (2021) and Caswell (2022) into an extended framework. We additionally develop an R function to model kin in two sex multi-state time-varying populations: now a part of the DemoKin package (Alburez-Gutierrez et al., 2021). Our proposed model is informed with historic (1938-2022) demographic rates taken from the Office for National Statistics (ONS). The model is informed with projected (2023-2070) demographic rates, which are obtained from Bayesian statistical models. In what follows, for the first time, estimates of kin networks defined through sex-specific mortality and sex- and parity-specific fertility rates are procured. This level of detail regarding model inputs will to provide more accurate inferences of kin. Uncertainty in our kin estimates naturally emerges through the uncertainty in the statistical forecasts.

2 Data and Methods

Our methods involve informing a matrix projection model with realistic demographic data. We require as inputs: sex, age, and parity specific fertility rates, and sex and age specific mortality rates. With these data, we can estimate the historic unfolding of kin-networks over an observed time period. Additionally we can estimate the size and composition of kin-networks, with a quantification of uncertainty, going into the future.

2.1 Formal matrix models

The formal models of kinship proposed by Caswell & Song (2021), Caswell & Song (2021) and Caswell (2022) are combined, allowing for one to project kin structured by age, stage and sex, within a time-variant demography. To be clear: we do not present a framework, but instead utilise multiple established frameworks. For the reader unfamiliar with these formal models, below we recapitulate their methods. A more detailed summary of the formal models of kin can be found in the above-mentioned papers; a guided tutorial of their used can be found on the DemoKin website https://github.com/IvanWilli/DemoKin.

Consider a population defined at time t. Suppose that the population is composed ages $\mathcal{A} = \{1, \dots, n\}$, and sexes $\mathcal{S} = \{\mathrm{f}, \mathrm{m}\}$. Suppose that within each sex, individuals are further defined through parity level $\mathcal{P} = \{p_0, p_1, p_2, p_{3+}\}$. Let Focal be aged x at time t, and let $\tilde{\mathbf{k}}_t(x)$ represent a vector of the expected numbers of some particular kin. Since we are dealing with a population structured by age, sex, and parity, this vector will be of dimension $n \times 2 \times 4$. Write, for each sex $i \in \mathcal{S}$, the vector $\tilde{\mathbf{k}}_t^i(x)$ to represent a block-structured vector of kin with parity level embedded within age. Then create the vector $\tilde{\mathbf{k}}_t(x)$ by stacking one sex on the other:

$$\tilde{\mathbf{k}}_t(x) = \begin{pmatrix} \tilde{\mathbf{k}}_t^{\mathrm{f}}(x) \\ \tilde{\mathbf{k}}_t^{\mathrm{m}}(x) \end{pmatrix}. \tag{1}$$

We project $\tilde{\mathbf{k}}_t(x)$ using the dynamic system

$$\tilde{\mathbf{k}}_{t+1}(x+1) = \begin{cases} \tilde{\mathbf{U}}_t \tilde{\mathbf{k}}_t(x), & \text{(no reproduction)} \\ \tilde{\mathbf{U}}_t \tilde{\mathbf{k}}_t(x) + \tilde{\mathbf{F}}_t^* \tilde{\mathbf{k}}_t^{(\mathrm{subs})}(x) & \text{(dependent reproduction)} \\ \tilde{\mathbf{U}}_t \tilde{\mathbf{k}}_t(x) + \tilde{\mathbf{F}}_t \tilde{\mathbf{k}}_t^{(\mathrm{subs})}(x) & \text{(independent reproduction)} \end{cases} \tag{2}$$

where the matrices

$$\tilde{\mathbf{U}}_t = \tilde{\mathbf{U}}_t^{\mathrm{f}} \oplus \tilde{\mathbf{U}}_t^{\mathrm{m}}, \qquad \tilde{\mathbf{F}}_t^* = \begin{pmatrix} (1-\alpha)\tilde{\mathbf{F}}_t^{\mathrm{f}} & 0 \\ \alpha \tilde{\mathbf{F}}_t^{\mathrm{f}} & 0 \end{pmatrix}, \qquad \tilde{\mathbf{F}}_t = \begin{pmatrix} (1-\alpha)\tilde{\mathbf{F}}_t^{\mathrm{f}} & (1-\alpha)\tilde{\mathbf{F}}_t^{\mathrm{m}} \\ \alpha \tilde{\mathbf{F}}_t^{\mathrm{f}} & \alpha \tilde{\mathbf{F}}_t^{\mathrm{m}} \end{pmatrix}, \tag{3}$$

account for (i) the independent survival of female and male kin, structured by parity level, (ii) reproduction of only female kin, structured by parity level, and (iii) the reproduction of both female and male kin, again structured by parity level. The scalar α reflects the proportion of male newborns relative to females. Details of the projection system in a two-sex population is explained by Caswell & Song (2021).

As the population is structured by parity level, these projection matrices are multi-state. For $i \in \mathcal{S}$ we define the matrices $\tilde{\mathbf{U}}_t^i$ and $\tilde{\mathbf{F}}_t^i$ to be block-structured with form:

$$\tilde{\mathbf{U}}_t^i = \mathbf{\Psi}_{s,n}^{\dagger} \mathbb{U}_t^i \mathbf{\Psi}_{s,n} \mathbb{T}_t^i \quad \text{and} \quad \tilde{\mathbf{F}}_t^i = \mathbf{\Psi}_{s,n}^{\dagger} \mathbb{H}_t^i \mathbf{\Psi}_{s,n} \mathbb{F}_t^i. \tag{4}$$

Here, $\mathbb{U}_t^i = \mathbf{U}_{p_0}^i(t) \oplus \cdots \oplus \mathbf{U}_{p_{3+}}(t)$ and $\mathbb{F}_t^i = \mathbf{F}_{p_0}^i(t) \oplus \cdots \oplus \mathbf{F}_{p_{3+}}(t)$ are block-diagonal matrices with blocks respectively representing the survival of, and reproduction by, kin of sex i and parity $p_i \in \mathcal{P}$. The block-diagonal matrices $\mathbb{T}_t^i = \mathbf{T}_{p_0}^i(t) \oplus \cdots \oplus \mathbf{T}_{p_{3+}}^i(t)$ and $\mathbb{H}_t^i = \mathbf{H}_{p_0}^i(t) \oplus \cdots \oplus \mathbf{H}_{p_{3+}}^i(t)$ respectively have blocks yielding the probabilities of transitioning from parity p_i to a higher level and the redistribution of newborns into a specific age-class. The multi-state approach is explained in Caswell et al. (2018) and we refer the reader there for further details.

With the above-mentioned conceptual groundwork, the RHS of Eq (2), from the top line to bottom describes the time-change in kin under three distinct cases: (i) kin which Focal cannot acquire any more of during its life (e.g., older siblings); (ii) kin which Focal can acquire during its life, but which cannot be procured independently by female and male subsidisers (e.g., younger siblings with parents); (iii) kin which Focal can acquire during its life, which are independently procured by both female and male subsidisers (e.g., nieces and nephews). Notice that case (ii) assumes a female dominant population, only counting descendants of the female direct ancestors of Focal. This assumption circumvents the possibility that Focal's parents or grand-parents reproduce independently. We use Eq (2) to produce all results in this research.

2.2 Matrix model inputs: Fertility

The above framework requires as inputs, parity-specific fertility rates, structured by sex. At each time-period t, the matrix $\mathbb{F}_t^i$ has block-diagonals with entries: the expected age, sex, and parity specific fertility rate. The upper-block of $\mathbb{F}_{1990}^{\mathrm{f}}$, for example gives the expected age-specific fertility rate for females in parity 0 in 1990. We source historic rates from data (or estimate them if lack of data), and predict future rates.

2.2.1 Female

Female single-year age-specific fertility rates by parity (ASFRP) for England & Wales are sourced from the ONS, up to the year 2022. Data include estimates of age- and parity-specific births (B_x^p) and exposures (E_x^p), as well as and mid-year population estimates, revised following the 2021 Census. The female reproductive age range is assumed to be $x = 15 - 44$. Time period t is considered alongside cohort c and age x through the relation $t = c + x$. Parity-specific data is available only from 1965, and by cohort. To obtain a longer time span for kinship to run, we are currently estimating period-based fertility going back to 1938.

Fertility projections are obtained using the Bayesian parity-specific generalised additive model (GAM) proposed by Ellison et al. (2023). This approach models fertility rates as smooth functions of age and cohort, simultaneously for each parity level. In particular, the method involves a two-dimensional smoothing of age-cohort surfaces of conditional rates for each parity. Cumulated fertility across cohorts is then calculated thus allowing for one to reduce dimensionality and average over parity. Full details of the statistical model can be found in Ellison et al. (2023). Visuals of its predictive fit and forecast are shown in Appendix (5.1).

2.2.2 Male

Due to the lack of parity-specific data, we cannot fit a statistical model to male fertility. Instead, for each time-period we apply parity-specific information from female births, and develop a method which distributes the numbers of births from males into birth orders. We then project the exposures of specific a cohort of males entering fertility, and use the time-dependent estimates of males' births by order to calculate parity specific rates: through the usual division of numbers over exposed.

The model of Ellison et al. (2023) is used to estimate male age-specific birth numbers for time-period t: $B(t) = B_{15}(t) + \cdots + B_{51}(t)$ where we consider the male age range $x_{\mathrm{m}} = 15\ldots, 51$. We assume that the male birth orders comprising each $B_{x_{\mathrm{m}}}(t)$ can be inferred from the female parity levels at that time-period. That is, for each age x_{m} of father, we find the distribution of ages of mothers, and the corresponding probabilities that mother was that age;

$$\boldsymbol{\phi}_{x_{\mathrm{m}}}(t) = (\phi_{x_{\mathrm{m}}}^{x_{\mathrm{f}},15}(t), \phi_{x_{\mathrm{m}}}^{x_{\mathrm{f}},16}(t), \ldots, \phi_{x_{\mathrm{m}}}^{x_{\mathrm{f}},44}(t))^{\dagger}, \quad ||\boldsymbol{\phi}_{x_{\mathrm{m}}}|| = 1 \tag{5}$$

where, e.g., $\phi_{15}^{18}(2020)$ gives the probability that given a 15 year old father in 2020, the age of mother to child is 18. We weight the respective parity distribution $\mathbf{p}_{x_{\mathrm{f}}}(t)$ (where $||\mathbf{p}_{x_{\mathrm{f}}}(t)|| = 1$) of mother of age x_{f} using the probability that mother is that given age. Thus, for each age x_{m} of father and each time-period t, we obtain a set of probable parity levels (summing to one):

$$\mathbf{p}_{x_{\mathrm{m}}}(t) \sim \sum_{x_{\mathrm{f}}} \phi_{x_{\mathrm{m}}}^{x_{\mathrm{f}}}(t)\mathbf{p}_{x_{\mathrm{f}}}(t). \tag{6}$$

Multiplying $\mathbf{p}_{x_{\mathrm{m}}}(t)$ through B_x gives the number of births by males disaggregated into birth orders $B_{x_{\mathrm{m}}}^{o}(t), o = 0, 1, 2, 3+$, in that year. The method ensures for consistency, that the sum of birth-order specific divided through exposures recovers the estimates ASFR for men. The algorithm is applied for each posterior sample in the female forecast separately as well as the historic time-series. Once again, fitted medians and forecasts used to inform the model presented below, see Appendix (5.1).

2.3 Matrix model inputs: Mortality

Unlike fertility, we assume that the force of mortality is independent with respect to parity level. That is, age-specific survivorship is equal amongst parity classes. Thus, although mortality differs between sex, within sex, there is no difference in the age-specific survivorship of an individual with zero children compared to one with two children. As such, we are able to use the same statistical model to fit and forecast ASSPs for males and females.

Single-year age-specific survival-probabilities (ASSP) data are sourced from the ONS. Female data range from 1938-2022 and male data 1944-2022. Forecasts are performed using the Bayesian model proposed by Hilton et al. (2018). The model assumes a Bayesian Generalised Additive Model (GAM) is used to fit mortality rates for the majority of the age range considered, while assumes parametric fits for infant and old age ranges. Death counts are assumed Negative Binomial. Age x, period t and cohort c terms are included as smooth functions and a linear interaction between time and age accounts for temporally "linear" mortality improvements with age. Again writing cohort by $c = t - x$, the cohort smooth is constrained such that the first and last cohort terms are set to zero. For a full description of this model see Hilton et al. (2018). To see the fitted medians and forecasts used to inform the model presented below, see Appendix (5.2).

3 Results

The preliminary results here are based on a historic time-scale (2000-2022)[†] and forecast time-scale (2023-2070). We inform the model with 500 independent draws from the posterior distributions of the fertility and mortality forecasts to quantify uncertainty caused by rates in future kinship.

3.1 By cohort of birth

Figure 1 illustrates the age-specific parity progressions of a typical Focal, born in the year 2000. The uncertainty associated with Focal's parity increases with the forecast horizon, and up to the age of 44 when Focal ceases reproduction. By the end of reproduction, we predict that Focal is slightly more likely to have two children (i.e., be in parity 2) compared to none, one or more than three.

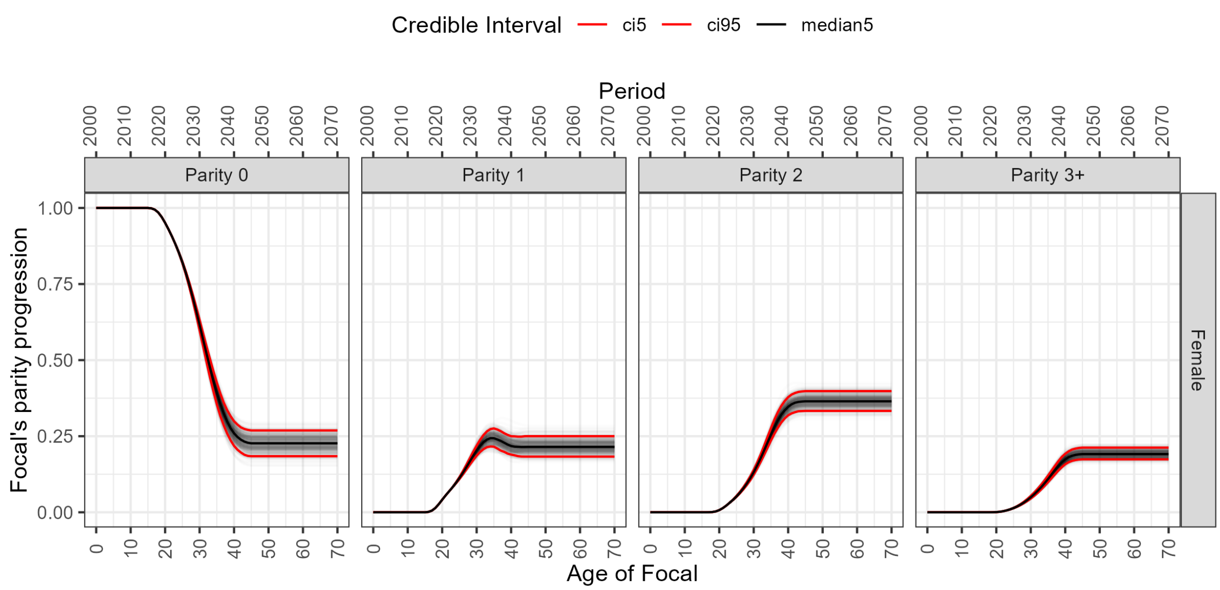

Figure 1: Expected (black) and 95% CI (red) parity level of Focal born in 2000, as a function of Focal's age. Note that we assume Focal is female. Top axis shows time period for reference. We use 500 independent samples from the statistical forecasts to generate the CI. Each sample results in a probability Focal is in parity 0, 1, 2, or 3+, and sums to one.

The expected number of offspring a typical Focal individual will have and their expected parity levels, are depicted in Figure 2. As a function of Focal's age, we see that the number of offspring is uni-modal. Decreases in the number of offspring in parity 0 for Focal older than 50 could be due to death of kin or parity progression; or both. We see that as Focal reaches ages 45+, there are increases in the number of offspring in parity 1, and when Focal reaches 50+, there are increases in parity 2. When Focal is of age 60+, a small number of her offspring are predicted to have moved to parity 3+. Not that each grey line in the figure corresponds to a unique combinations of posterior samples for ASFR and ASSP. Hence each grey line gives summing the grey lines over the rows gives the total expected number of offspring of Focal (marginalising over parity).

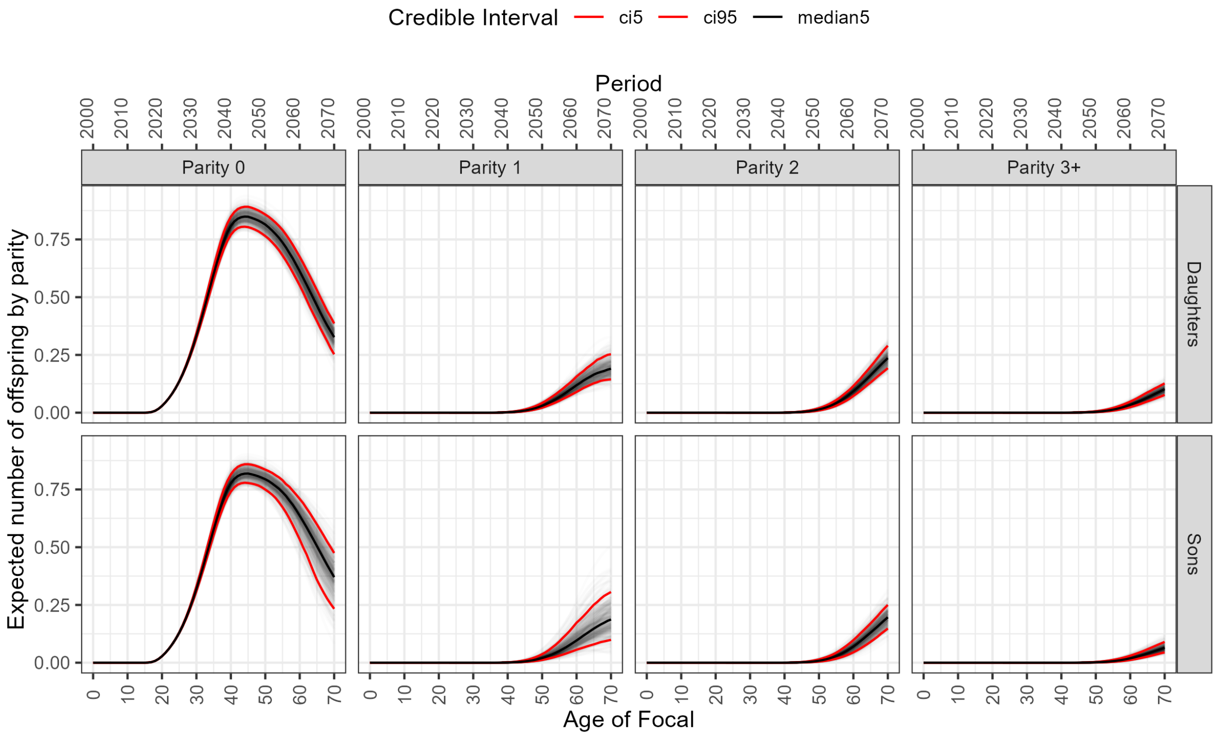

Figure 2: The expected number of (black) with 95% CI (red) of Focal's daughters (top) and sons (bottom) in each possible parity (columns), given Focal was born in 2000. We use 500 independent samples from the statistical forecasts to generate the CI.

We could also compare the estimated number of kin a typical Focal will experience, as a function of birth year. In Figure 3 we compare a 2000 cohort Focal's siblings to a 2040 cohort Focal's siblings. Here we see that a Focal born in the year 2000 is more likely to have siblings in parity 0 compared to a Focal born in 2040. In the latter cohort, the decrease in the expected number of parity 0 siblings is absorbed by an increase in the expected number of siblings in parity 1.

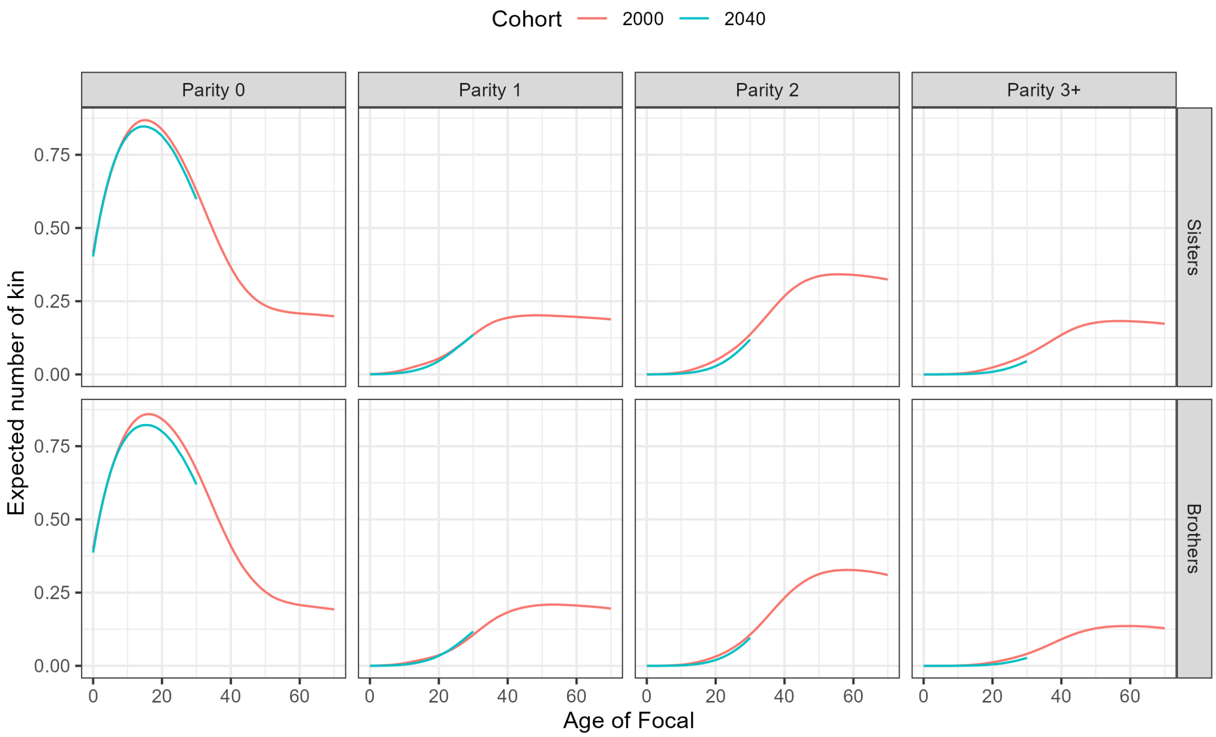

Figure 3: A comparison of the expected number of sisters (top) and brothers (bottom) for a typical Focal born in 2000 (red) and 2040 (blue). Column represents parity level of kin.

3.2 By period in time

Figure 4 compares the average parity levels of Focal's siblings for years 2022 and 2068. First, note that increases in the x-ordinate does not reflect an individual getting older: instead it describes a cohort of birth (50/51 year old Focals in 2022 imply 1972/1973 cohorts). With this in mind, since we here use rates from 2000, our results are only time-variant up to Focal aged 22 in 2022 and Focal aged 68 in 2068. Nonetheless, we see that typical 50 year old Focal in 2068 (a 2018 cohort) expects more siblings than any other aged Focal in that period. Contrastingly, in the period 2022, a typical 20 year old Focal (a 2002 cohort) expects more siblings than any other age class.

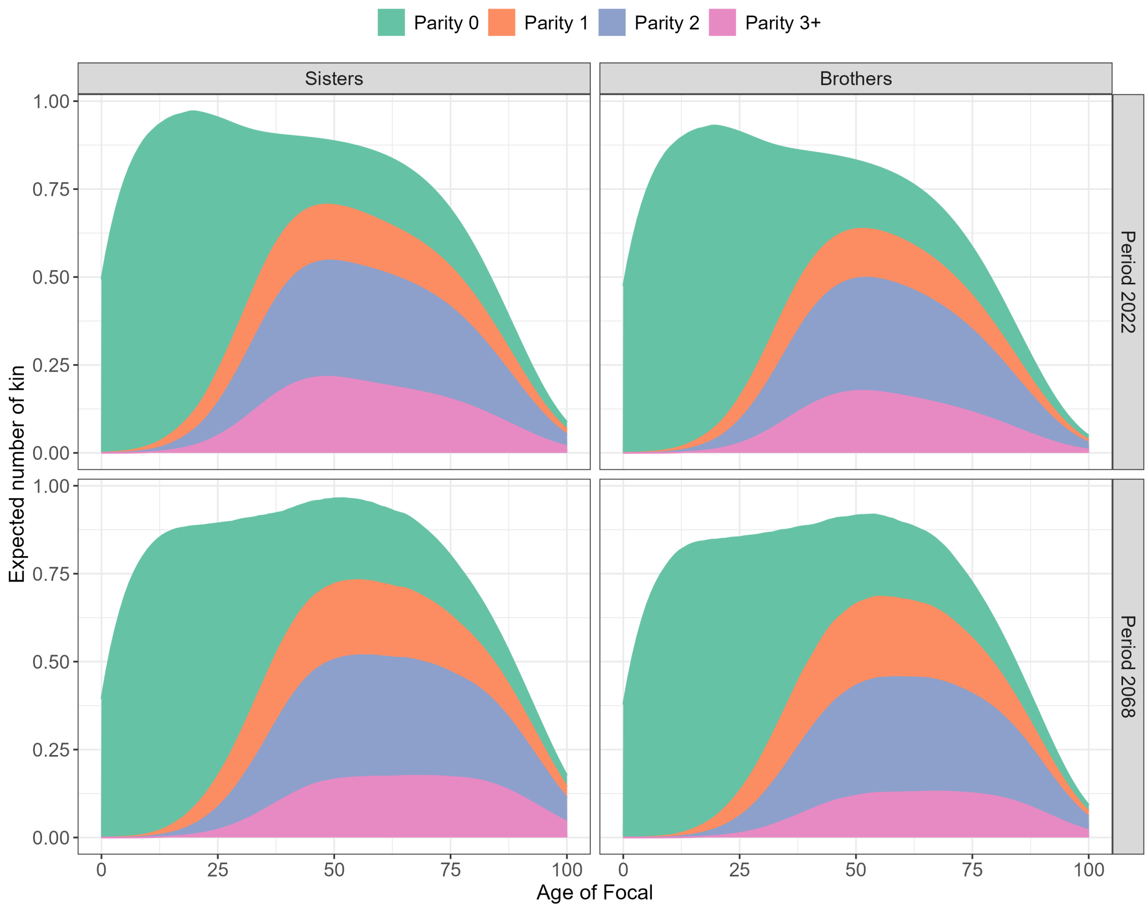

Figure 4: Expected number and parity of the brothers and sisters a typical Focal will expect in 2022 and 2068. The colours represent the proportion that each parity level contributes to the total. Left: Focal's sisters; right: Focal's brothers.

3.3 By changing time

Figure 5 demonstrates how we expect the average parity level of a parent in the UK to vary between 2000 and 2070. Illustrated are the probabilities that Focal's mother and father have had one (Focal herself), two, or more than three children. The decreases in the probabilities of parents' being of parity 1 between 2000 and 2018 is balanced out by equal increases in the probabilities that they are in parity 2 or 3+.

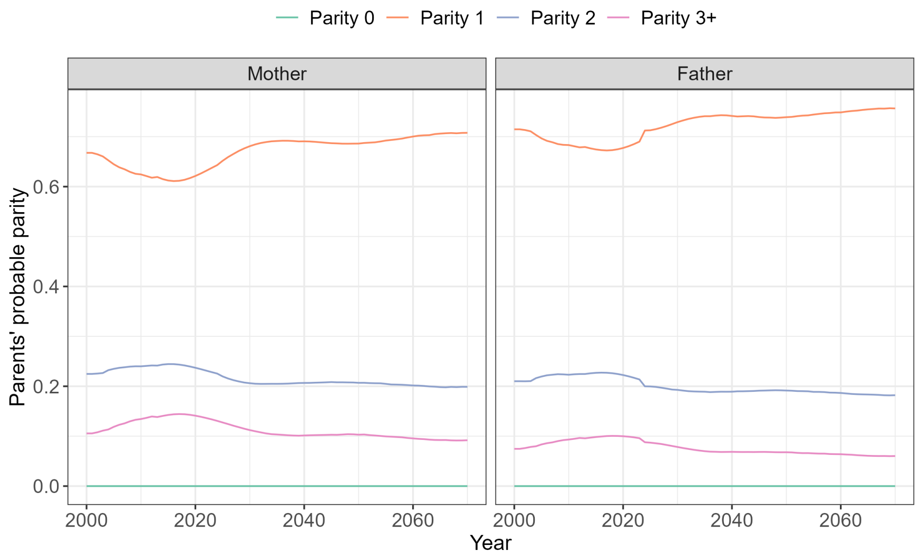

Figure 5: The probable parity levels of Focal's parents. Left: Focal's mother; right: Focal's father. Note that the probability of parity 0 is always zero since we assume Focal represents a live birth.

4 Conclusions

This study presents the first forecasts of kin-networks structured by age, sex, parity, including uncertainty intervals. Our main contribution is to provide a means to explore how including sex- and parity-specific data affects the estimated future unfolding of a typical individual's family. Improved understanding of how an individual's parity level can affect its kin available to care will prove important when planning social policy measures. More generally, we show the complexity of kinship structures that can be explored by combining an ensemble of the state-of-the-art formal models of kinship. Here the focus is on parity, however, our method allows for any other arbitrary characteristic affecting a population member to be considered. Examples of education or income band are of future interest. Moreover, at this preliminary stage, we merely demonstrate a framework capable of investigating complex kin-dynamics: further down the line we hope to explore them.

In the application presented we estimate the number and parity level of relatives a UK population member should expect in 50 years time, with the bonus of added uncertainty. Decreasing and postponed fertility means that future cohorts of population members will expect fewer children and grandchildren. Establishing the extent of such a drop is important; these kin prove to be a major source of social care. Many people in the latter years receive inform care from children and grandchildren. Future knowledge of someone's descendants and their parities is desirable. We also find that future cohorts will overall have fewer siblings, but, relatively have more siblings without children. One finding in this research – reflecting recent declines and age shifts in fertility – are future increases in the likelihoods that Focal's parents have no extant offspring (i.e., Focal is the first born). This translates into smaller family sizes in the future.

There are some methodological weaknesses in the above model. First, we only start projections from the year 2000 (although this should soon be resolved). The dynamic system approach requires a well defined initial condition (time-period) at which the time-variant rates are applied. Before this period, we assume a stable demography. The longer the time-variant demographic series of rates, the more accurate the results. Using 2000 as an initial means that our model does not capture demographic trends following either of the World Wars, nor the baby boom. We are currently working on estimation of rates back to 1938. As noted by Pullum (1982), there are always limitations in a mathematical approximations to kinship. In particular, we assume that the same demographic rates apply uniformally to all population members. A more realistic demographic model would account for stochasticity in the underlying birth and death processes which drive kinship. One could update the present work to account for such random events, i.e., by applying either of Caswell (2024) or Butterick et al. (2025). This would provide for kinship estimates which accommodate rate uncertainty and uncertainty in the probabilistic events population replenishment. Of course such a model would be high dimensional, and at this stage, it would be worth exploring the role of microsimulation.

To conclude, through estimating the fertility of female and male population members as a function of time, and their existing offspring number, the results here give the most detailed projections of kinship in the UK. It would be interesting to consider additional characteristics which affect both mortality and fertility schedules. Socio-economic status and educational attainment for examples are likely to considerable affect the parity of a typical population member. Moving forwards, such an extension of the analysis here should prove useful for social planning. Incorporating inter-generational transmission of values, or cultural ties, would be equally interesting. Future work could build on these ideas.

References

Alburez-Gutierrez, D., Williams, I., & Caswell, H. (2023). Projections of human kinship for all countries. *Proc. Natl. Acad. Sci.*, *120*(52).

Alburez-Gutierrez, D., Williams, I., Caswell, H., & Song, X. (2021). Demokin: An r package to implement demographic matrix kinship models.
URL https://github.com/IvanWilli/DemoKin

Butterick, J., Hilton, J., Bijak, J., Smith, P. W. F., & Dodd, E. (2025). Probabilistic projections of distributions of kin over the life course. *bioRxiv*.

Butterick, J. B., Evandrou, M., Falkingham, J., & Hilton, J. (2024). Changes in kinship: implications for the availability of kin to care. Project report.

Caswell, H. (2019). The formal demography of kinship: A matrix formulation. *Demographic Res.*, *41*, 679–712.

Caswell, H. (2020). The formal demography of kinship II: Multistate models, parity, and sibship. *Demographic Res.*, *42*(38).

Caswell, H. (2022). The formal demography of kinship IV: Two-sex models and their approximations. *Demographic Res.*, *47*.

Caswell, H. (2024). The Formal Demography of Kinship V: Demographic stochasticity, variance, and covariance in the kinship network. *BioRxiv [preprint]*.

Caswell, H., de Vries, C., Hartemink, N., Roth, G., & van Daalen, S. F. (2018). Age stage-classified demographic analysis: a comprehensive approach. *Ecol. Monogr.*, *88*(4).

Caswell, H., & Song, X. (2021). The formal demography of kinship III: Kinship dynamics with time-varying demographic rates. *Demographic Res.*, *45*.

Ellison, J., Berrington, A., Dodd, E., & Forster, J. J. (2023). 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.
URL https://doi.org/10.1093/jrsssc/qlad095

Hilton, J., Dodd, E., Forster, J. J., & Smith, P. W. F. (2018). Projecting UK Mortality by Using Bayesian Generalized Additive Models. *J. R. Stat. Soc., C: Appl. Stat.*, *68*(1), 29–49.

ONS (2024). National Population Projections.
URL https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/datasets/tablea21principalprojectionukpopulationinagegroups

Pullum, T. W. (1982). The eventual frequencies of kin in a stable population. *Demography*, *19*(4).

5 Appendix

5.1 Fertility

Figure 6 and Figure 7 respectively show median estimates for forecast parity specific fertility rates, and uncertainty intervals estimated using 500 samples of the posterior distributions of the Bayesian forecasts.

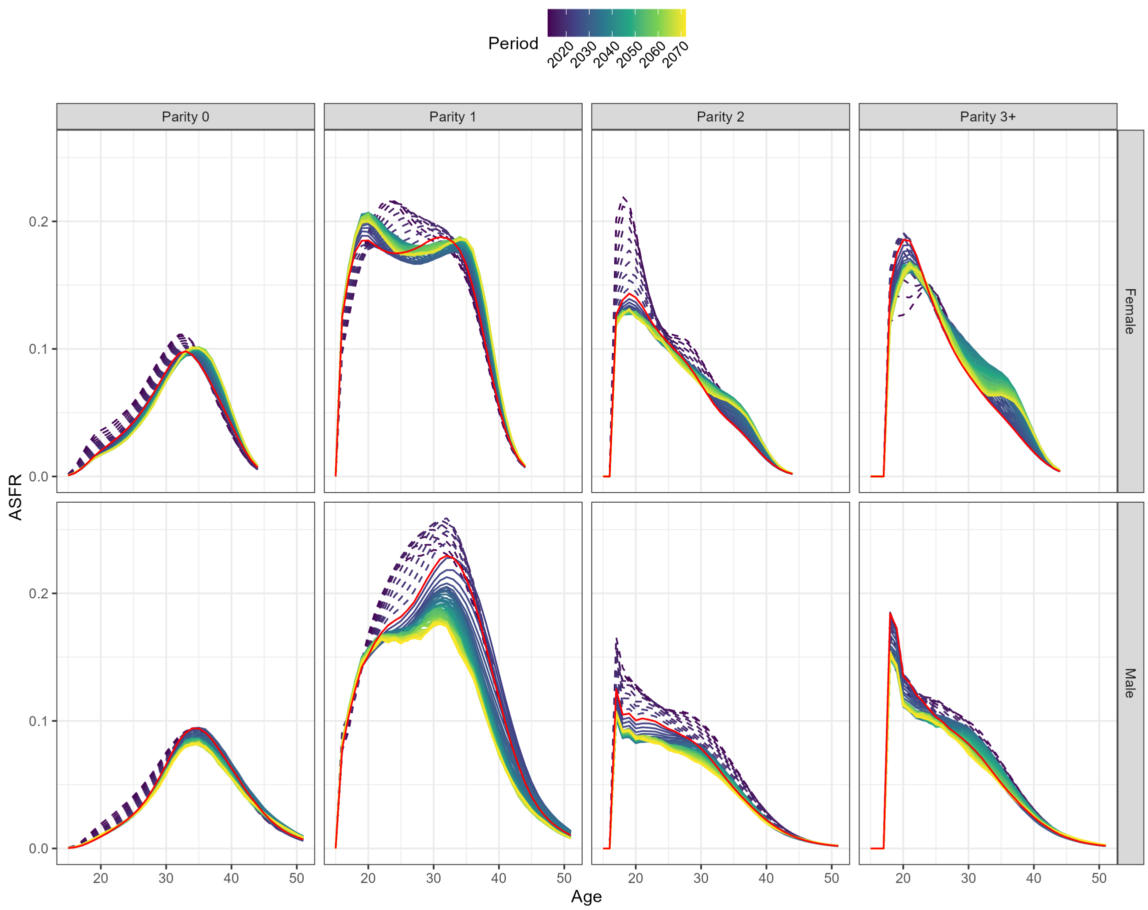

Figure 6: Fitted medians for female parity-specific fertility 2000-2070, and estimated male parity-specific fertility as explained in text. Dotted show historic fitted rates, and solid shows forecasts.

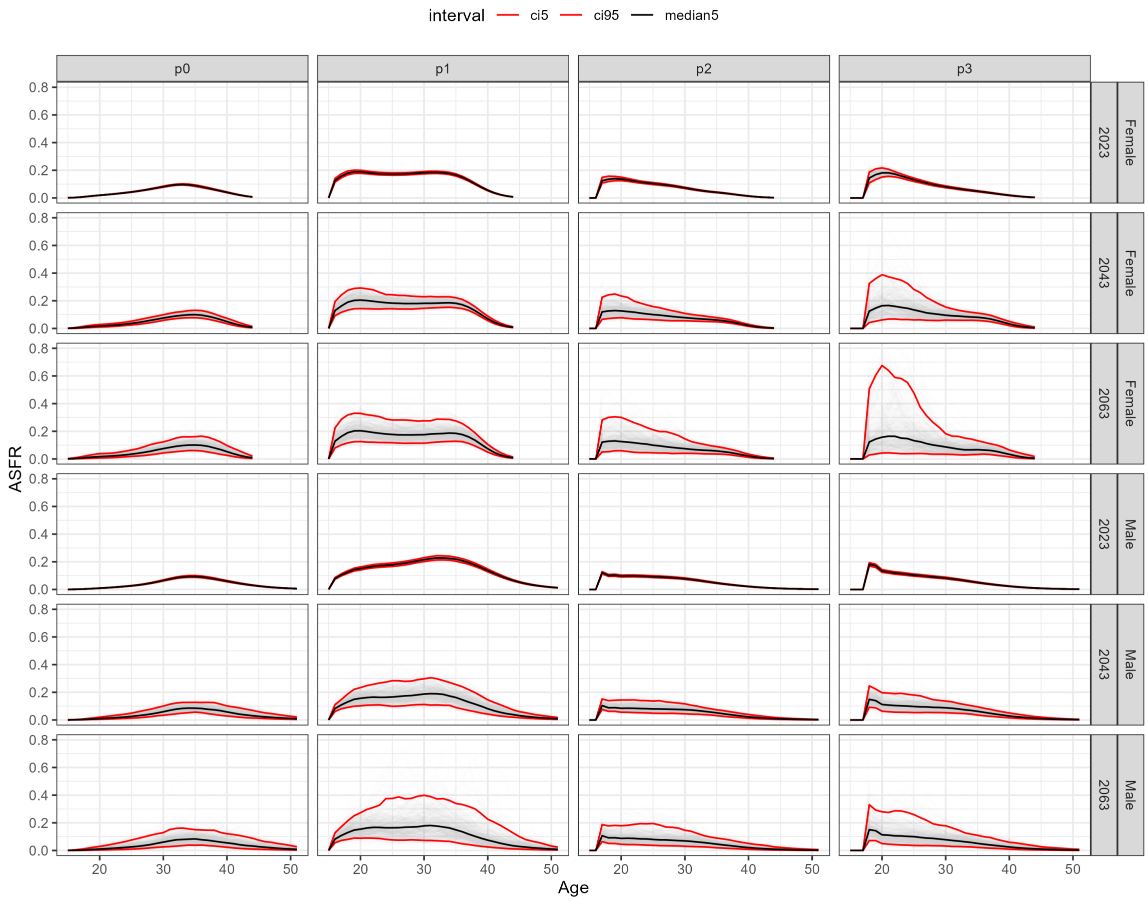

Figure 7: Forecast parity specific fertility with 95% prediction interval, using the Bayesian model for females, and the algorithm described in text for males. Forecast years selected: 2023, 2043, 2063.

5.2 Mortality

Figure 8 and Figure 9 respectively show median estimates for forecast mortality rates, and uncertainty intervals estimated using 500 samples of the posterior distributions of the Bayesian forecasts.

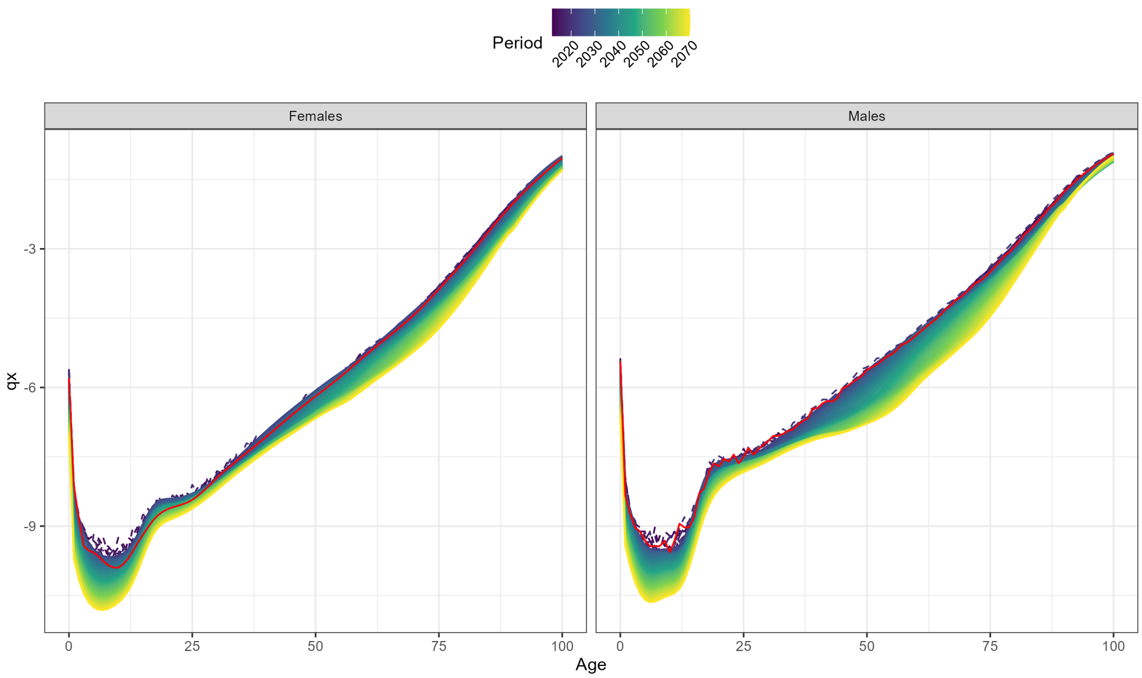

Figure 8: Fitted medians for age-specific mortality (log scale) for females and males. From 2000 to 2070. Dotted show historic fitted rates, and solid shows forecasts.

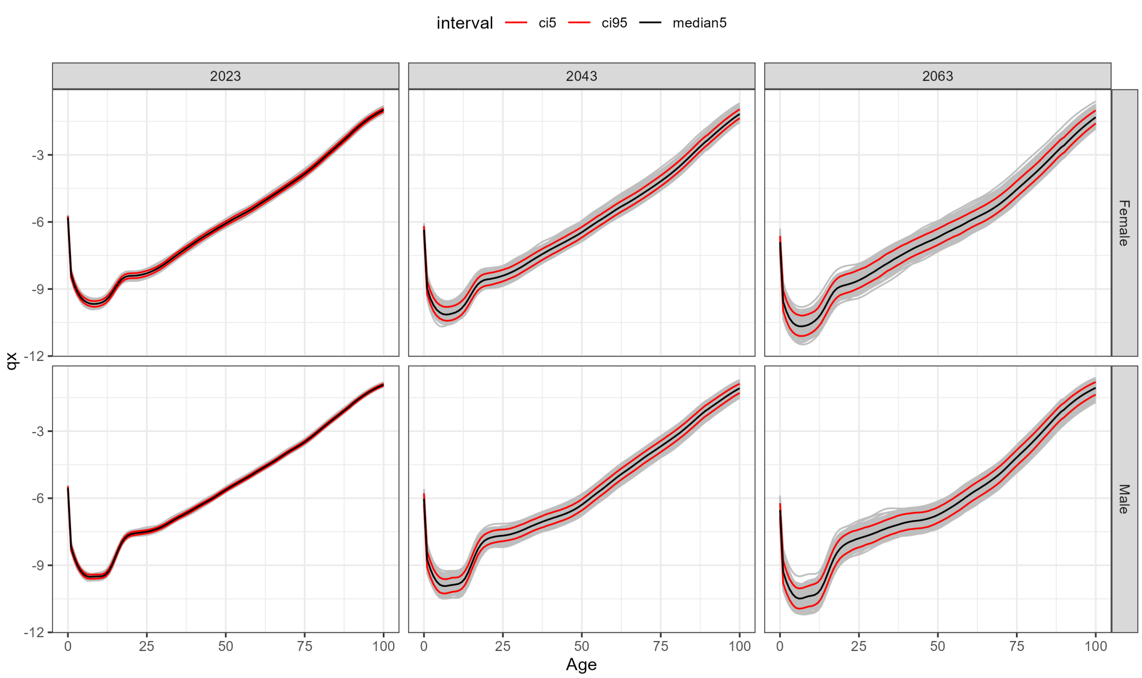

Figure 9: Forecast logarithmic death rates with 95% prediction interval, for selected future years: 2023, 2043, 2063.