Modelling Kinship Networks using Analytical and Microsimulation Approaches

A Research Note

Liliana P. Calderón-Bernal[1,2][0009−0002−7629−9718] and Sha Jiang[3][0000−0002−8230−4264]

[1] Department of Digital and Computational Demography, Max Planck Institute for Demographic Research, Rostock, Germany
calderonbernal@demogr.mpg.de
[2] Demography Unit, Department of Sociology, Stockholm University, Sweden
[3] Department of Digital and Computational Demography and Kinship Inequalities Research Group, Max Planck Institute for Demographic Research, Rostock, Germany

Abstract. Analytical and microsimulation models are the primary approaches for estimating kinship networks when detailed, long-term demographic data including kinship ties are unavailable. Although both methods use population-level demographic data as inputs, their methodological formulations, assumptions and outputs differ. Analytical models offer mathematical transparency, enabling the calculation of expected values, but are limited in capturing individual variability and interactions. Conversely, microsimulation models simulate demographic events at the individual level, enabling the estimation of the number and distribution of kin, but they are computationally intensive and rely on additional assumptions. Despite their widespread use, systematic comparisons between these approaches remain limited. This study examines the agreement between matrix kinship models and the SOCSIM demographic microsimulation programme when estimating kinship networks. Using fertility and mortality data from the World Population Prospects 2024 for Sweden between 1950 and 2100, we compared the means and distributions of the numbers and ages of different types of kin derived from both models. Our results show that both models produce closer estimates for vertical kin, such as grandparents and grandchildren, but diverge more when estimating horizontal kin, such as siblings. Microsimulation tends to underestimate the number of kin slightly, with discrepancies varying by kin type and focal age.

Keywords: Kinship · Demographic Microsimulation · Formal Demography of Kinship

1 Introduction

Kinship networks play a fundamental role throughout individuals' lives, providing support and care, and enabling intergenerational exchanges that extend

beyond the nuclear family. However, demographic changes, such as declining fertility and increasing longevity, are reshaping these networks. This raises concerns about their availability and how they change over time and with age. Understanding how many kin individuals have at different ages, and how the structure of these kin changes with age, is essential for estimating generational support, caregiving burdens, and the risk of social isolation. Nevertheless, few countries worldwide have long-term, detailed demographic data on kinship relationships at the individual level, which are essential for estimating both the number of kin and their distribution. This underscores the importance of considering alternative approaches to estimating kinship patterns using population-level rates, which are likely to be available in most countries.

In the absence of detailed individual-level demographic data, two main modelling approaches are generally used to estimate kinship networks: analytical matrix kinship models and demographic microsimulation. Both methods use population-level demographic rates as inputs but differ significantly in their methodological approaches, assumptions, advantages, and limitations. Analytical models employ matrix formulations to represent each type of kin as a population, projecting their numbers and ages based on demographic rates. This approach is generally considered more transparent, since all assumptions are explicit, and expected values can be derived numerically. However, it is limited in its ability to incorporate individual-level variability and interactions, and it does not directly provide distributions of kin numbers. In contrast, demographic microsimulation generates demographic events at the individual level using a competing-risk framework. This involves exposing individuals to the risk of experiencing demographic events, such as birth, death, and partnership, and produces fully recorded synthetic populations alongside information on their kinship network. While this model enables the estimation of means and distributions of kin numbers and ages, it is computationally intensive and relies on additional assumptions. Both models have become essential for estimating kinship in historical, contemporary, and future populations, and have been widely used to study various topics, such as kin availability and kin loss. However, to the best of our knowledge, no previous study has systematically compared these methods using the same demographic data as inputs. This raises questions about their level of agreement and disagreement, and their preferred areas of application.

This study addresses this gap by comparing the two-sex time-invariant and time-varying matrix kinship models [2,7,4] and the SOCSIM demographic microsimulation [9]. Specifically, we aim to understand how closely these models agree in estimating the means and distributions of the numbers and ages of various types of kin for an average individual, hereafter called the focal individual, and the potential causes and implications of any discrepancies. Our focus is on outcomes that are most relevant to family and kinship demography, such as the analytical and synthetic means and the distributions of various kin numbers and ages, the differences in the estimation between ascendant, collateral, and descendant kin, and the implications of model choice.

Our work builds on previous methodological advances in the formal demography of kinship [2,7,4], and microsimulation models [9], as well as their applications to issues such as kin availability and kin loss. Although the formal demographic analysis of kinship has already been expanded to include stages [3], kin loss [6], stochasticity [5], and lifetime kin overlap [8], which could also be included in the microsimulation models, these extensions are not considered here. Our analysis focuses on the characteristics, assumptions, data requirements, and outputs of the two models in their two-sex stable and dynamic settings, particularly the means and distributions of the numbers and ages of different types of kin. As previous studies have focused on methodological development or single-model application, ours is, to our knowledge, the first to systematically compare the agreements and discrepancies between the two modelling approaches.

2 Data and Methods

We ran demographic microsimulation and formal demographic models using age-specific fertility rates for women and age-specific probabilities of dying (qx) or surviving (px) for both sexes from the World Population Prospects (WPP) 2024 Revision [11] for Sweden over the period 1950–2100. We implemented the demographic microsimulation using the 'rsocsim' R-package [10] and the formal demographic models using the 'Demokin' R-package [12]. In both cases, we assumed a stable population before 1950 and did not use any population data to initialise the models. We ran the microsimulation for 200 years using the WPP rates from 1950 to obtain a stable population by 1st January 1950. This estimate corresponds to the output of running the two-sex time-invariant kinship matrix model [4] with the 1950 WPP rates, or obtaining the period estimates for (1st January) 1950 from the two-sex time-varying kinship matrix model [7,4] using the WPP annual rates from 1950 to 2100.

To minimise variability in kin numbers and distribution arising from microsimulation stochasticity and to produce more reliable estimates for comparison with the analytical model, we ran a single large simulation with an initial population of 200,000 individuals, instead of multiple small simulations. The accuracy of the microsimulation output was assessed by comparing the age-specific fertility and mortality rates and the summary measures (total fertility rate and life expectancy at birth) derived from the simulation inputs (WPP 2024) and outputs. Fig. A1 in the Appendix shows that the output estimates are broadly consistent with the inputs. The accuracy of output mortality rates for infants, children and young people, which are already very low in the input data for the most recent periods, could be improved by increasing the size of the simulation.

In the absence of accurate age-specific marriage rates by sex for the entire period, we employed the 'marriage after childbirth' directive in 'rsocsim' to establish lifelong partnerships. Under this directive, when a previously single woman is scheduled to have a child, she is paired with a living single male partner. Despite its name, this approach generates lifelong partnerships associated with childbirth rather than formal marriages. Following the approach of Alburez-Gutierrez and

colleagues [1], partners for each woman were selected from all living single men to minimise the squared difference between the observed distribution of 'man's age - woman's age' and a normal distribution with a mean of two and a standard deviation of three. No attributes other than age and singleness are considered during partner selection. Once formed, the partnership persists for subsequent births, assigning the same father for all births while both partners are alive. If a partner dies, a new partner is selected for any subsequent births according to the same age difference. Fertility and mortality are otherwise independent: women experience fertility according to the age-specific rates, and mortality operates separately for men and women. There is no further joint demographic behaviour in the model. Further details on this partnership option can be found in the supplemental materials of Alburez-Gutierrez et. al. [1].

To ensure comparability with the microsimulation setup when running the two-sex analytical model, we created a schedule of male fertility rates approximating the mean age difference implemented in SOCSIM. Specifically, we defined male fertility rates as equivalent to female fertility rates, but shifted upwards by two years. This means that, for each age, a man's fertility rate reflects that of a woman who is two years younger. This approximation of the target mean age difference between partners under the 'marriage after childbirth' directive ensures that the paternity schedule in the analytical models follows a similar pattern to that in the microsimulation. Additionally, we set the proportion of female births in the analytical models to 0.488 to align with the sex ratio at birth in the microsimulation.

We also harmonised the kin types, population exposure, grouping variables and output measures across models. All estimates refer to the mean and distribution of the number and ages of each type of kin of an average 'focal' individual at the beginning of each calendar year. The analysis considered the following types of kin: great-grandparents, grandparents, parents, aunts and uncles, siblings, cousins, nieces and nephews (so-called 'niblings'), children, grandchildren and great-grandchildren. Results were averaged across both sexes. To illustrate the differences in the estimates for ascendant, collateral and descendant kin, we use grandparents, siblings and grandchildren as examples in the main text. The full results for all kin types are included in the appendix.

The two models produce different types of output. The microsimulation provides a synthetic population that is 'fully registered' and includes an identification number 'pid' and the dates of birth and death of every individual who has ever been alive during the simulated period, as well as the 'pid' of some of their close kin. This information allows to reconstruct the entire kinship network of each individual at each age and year, and to aggregate numbers and ages afterwards. Conversely, the formal model provides the expected numbers and distribution of ages of each kin of a focal individual, at a given age and year, but it does not directly provide the distributions of numbers. Therefore, some procedures were necessary to make the estimates comparable.

Using the microsimulation output, we traced the kinship networks of all synthetic individuals, including great-grandparents, grandparents, parents, aunts

and uncles, siblings, cousins, nieces and nephews, children, grandchildren and great-grandchildren, based on the 'pid' of parents and children. We then estimated the number and ages of each type of kin for each individual at each age and year, considering the dates of birth and death of both the individuals and their kin. Next, we aggregated the numbers and ages of each type of living kin at each age and calendar year across all individuals in the synthetic population. Next, we calculated the mean number and mean age of each type of kin for an average 'focal' individual at each age and year. These results are similar to those calculated with the analytical models. Finally, we estimated the distribution of numbers and ages of each type of kin across all individuals at each age and year within the entire synthetic population.

Using the analytical kinship models, we estimated the mean and variance of the numbers and ages of each type of kin at each age of a focal individual. Besides providing the mean age of kin, the analytical model provides the age distribution of each type of kin at each age of a focal individual, yielding results comparable to those obtained using microsimulation. However, these models do not directly provide distributions of the numbers of each type of kin. To address this issue, we adopted the approach outlined by Hal Caswell [5], which shows that the variances of most kin numbers align with a Poisson distribution, while those of parents and ancestors align with a binomial distribution. Therefore, for each age, year, and kin type, we approximated the distribution of kin numbers by applying a binomial distribution to the expected numbers of kin that are a priori bounded (e.g. parents, grandparents, and great-grandparents) and a Poisson distribution to all others.

By examining the models' outputs with similar input data, we aim to assess their agreement and identify where and why their results align or diverge. These comparisons provide insight into the strengths of each modelling approach and the implications of their assumptions.

3 Results

Using analytical and microsimulation models, we estimated the size and structure of various types of living kin by the age of a focal individual over the period 1950–2100. The figures show the mean number or mean age of living kin and their proportional distributions (on the y-axis) and the age of a focal individual (on the x-axis) at each selected year (1950, 2000, 2050 and 2100), as estimated from both models. The results are shown for three types of kin: grandparents, siblings and grandchildren, which represent ascendant, collateral and descendant kin. The figures for other types of kin (great-grandparents, parents, aunts and uncles, cousins, nieces and nephews, children and great-grandchildren) can be found in the Appendix of the paper. All estimates correspond to the 1st of January of each year. The estimates for 1950 correspond to a 'stable population', obtained by either running a pre-simulation using 1950 rates for 200 years, or assuming an internally time-invariant population based on 1950 rates before running the time-varying matrix model. The estimates for 2000, 2050, and 2100

correspond to the dynamic setting of both models using rates from 1950 to 2100.

Fig. 1 shows the mean number of living grandparents, siblings, and grandchildren by the age of a focal individual in 1950, 2000, 2050, and 2100, as estimated by analytical (dark red lines) and microsimulation (dark teal lines) models. The comparison reveals a consistent pattern of agreement in stable (1950) and dynamic (2000, 2050 and 2100) settings. The two models produce estimates that are closer for vertical kin (e.g., grandparents and grandchildren) than for horizontal kin (e.g., siblings). The mean numbers are slightly lower for all types of kin in the microsimulation model, but the magnitude of the difference varies depending on the type of kin and the age of a focal individual. This difference is slightly larger for grandparents (ascendant kin) when the focal individual is young, for siblings (collateral kin) when the focal individual is middle-aged and for grandchildren (descendant kin) when the focal individual is old. For other types of kin, the difference in means follows a similar pattern, depending on whether they are ascendant, collateral, or descendant kin, see Fig. A2 in the Appendix.

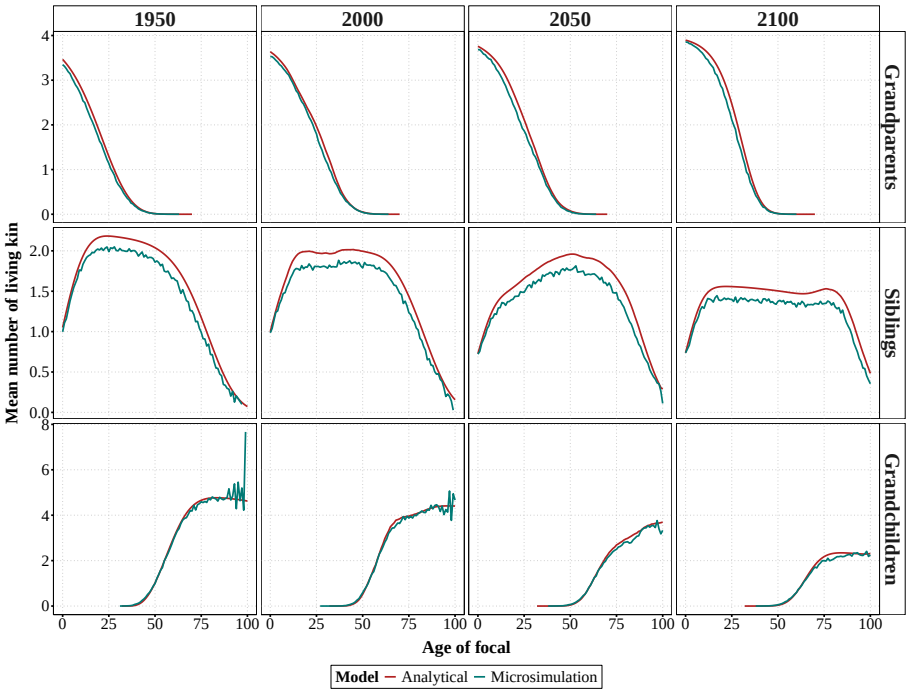

Fig. 1. Mean number of living grandparents, siblings, and grandchildren by age of a focal individual in 1950, 2000, 2050, and 2100, estimated from analytical and microsimulation models using WPP 2024 fertility and mortality data.

Fig. 2 shows the mean age of living grandparents, siblings and grandchildren by age of a focal individual in 1950, 2000, 2050 and 2100, as estimated from analytical and microsimulation models. As with the mean numbers, the difference in the mean ages of kin between the two models shows a consistent pattern in stable settings (1950) and dynamic settings (2000, 2050 and 2100), but varies depending on the type of kin. The mean ages of kin are slightly overestimated in the microsimulation when the focal individual is younger, and fluctuate as the focal individual becomes older. For grandparents (as well as other types of ascendant kin), the microsimulation does not estimate the mean age of kin at advanced ages, likely because no individuals survive to these ages in the synthetic population (see Fig. A3 in the Appendix for the mean age of all types of living kin).

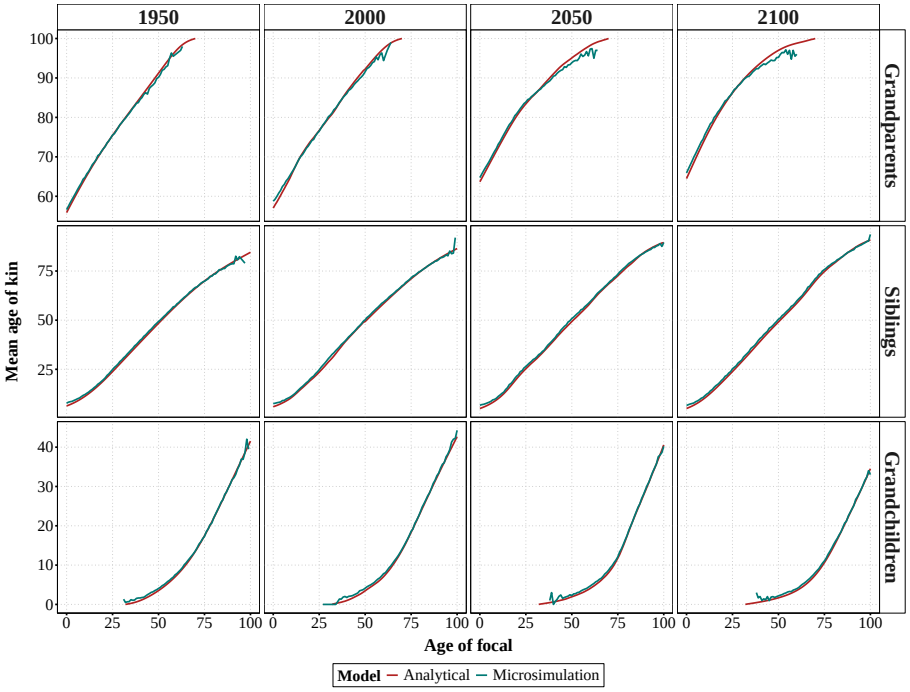

Fig. 2. Mean age of living grandparents, siblings and grandchildren by age of a focal individual in 1950, 2000, 2050 and 2100, estimated from analytical and microsimulation models using WPP 2024 fertility and mortality data.

Fig. 3 shows the distribution of the numbers of living grandparents, siblings and grandchildren by the age of a focal individual in 1950, 2000, 2050 and 2100, as estimated from analytical and microsimulation models. The figure shows a similar pattern in stable (1950) and dynamic (2000, 2050 and 2100) settings, as well as a similar shape of the age distribution of the number of kin in the an-

alytical and microsimulation models. The overall discrepancies between models seem larger for grandchildren (descendant kin) than for siblings (collateral) or grandparents (ascendant kin), as it is also the case for other types of descendant kin, such as nieces and nephews, children and grandchildren (See Fig. A4 in the Appendix for the distribution of the numbers of all types of living kin). However, when looking at the extremes of the distribution, i.e., the proportion of individuals with zero or large ($\geq$ 75th percentile) numbers of kin, the discrepancies appear for all types (see Fig. A5 in the Appendix). While the analytical model produces larger numbers of kin surviving until older ages of the focal individual, the microsimulation model estimates larger proportions of individuals with no kin at earlier ages.

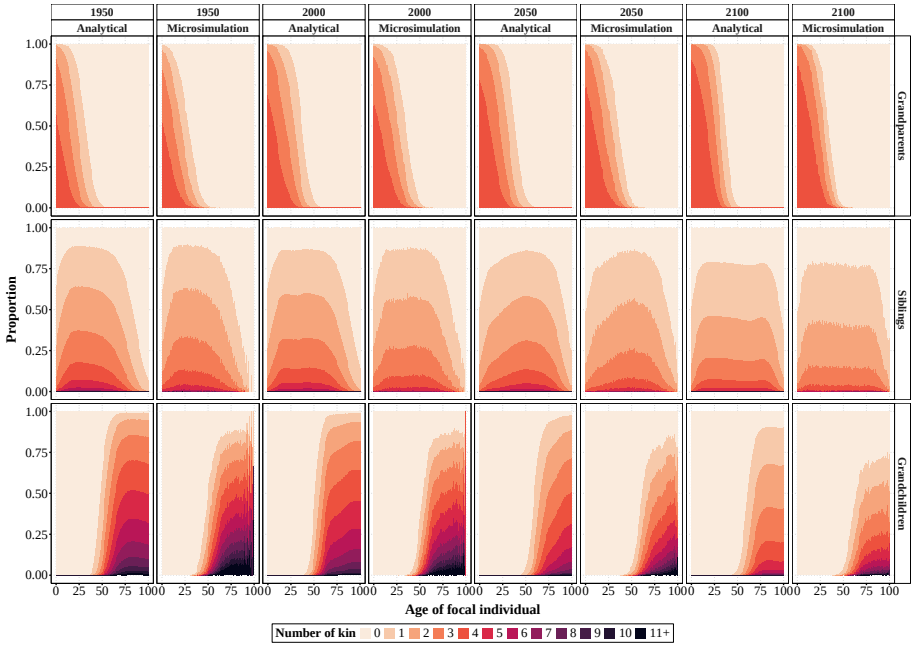

Fig. 3. Distribution of the numbers of living grandparents, siblings and grandchildren by age of a focal individual in 1950, 2000, 2050 and 2100, estimated from analytical and microsimulation models using WPP 2024 fertility and mortality data.

Fig. 4 shows the distribution of the ages of living grandparents, siblings, and grandchildren at the selected ages of the focal individual (0, 20, 40, 60, and 80) in 1950, 2000, 2050, and 2100, estimated from analytical and microsimulation models. The shape of the age distribution of kin is relatively similar for the stable and dynamic settings of the two models for most kin types and ages of focal and kin, although the microsimulation produces more variation across ages, likely due to its stochasticity. However, the main differences appear at the

youngest and oldest ages of the focal individual and their kin. The distribution of ages of kin across synthetic individuals is more irregular, and in some cases non-existent for grandchildren (as well as other descendant kin) when the focal individual is younger, and for grandparents (and other ancestors) when the focal individual is older (see Fig. A6 in the Appendix for the distribution of ages of all types of living kin.). This is likely due to the small number of individuals in a synthetic population at early and older ages, while the analytical model derives the estimates from proportions. Otherwise, the proportion of siblings with the same age as the focal is always 0 in the microsimulation, but corresponds to the mean of the distribution in the analytical model.

Discussion

Measuring kinship dynamics can be challenging due to the limited availability of demographic data including information on vital events and kinship ties that extend beyond the nuclear family. In the absence of such data, two main approaches are used to estimate kinship networks: analytical models and microsimulation. Although both models use population-level data as inputs, they differ in their methodological foundations, assumptions and outputs. Analytical models use a matrix formulation to represent each type of kin as a population, projecting their numbers and ages as the focal individual ages. Conversely, microsimulation models simulate demographic events at the individual level using a competing risk framework and exposing synthetic individuals to the risk of experiencing such events. In this study, we compared two-sex time-invariant and time-varying matrix kinship models with the SOCSIM demographic microsimulation programme to understand how closely these two approaches agree in estimating the size and structure of kinship networks. The analysis focused on the means and distributions of the numbers and ages of various types of kin estimated from both models. To our knowledge, this is the first study to compare the two modelling approaches using the same demographic data.

The comparison between the analytical and microsimulation models reveals agreements and discrepancies in the estimation of the size and structure of kinship networks. While the two models produce similar results for the mean number and age of vertical kin, such as grandparents and grandchildren, they diverge more in their estimation of horizontal kin, such as siblings. The microsimulation model tends to underestimate the mean number of kin, with variations depending on the type of kin and the age of the focal individual. These discrepancies are most evident for grandparents when the focal is young, for siblings when the focal is middle-aged, and for grandchildren when the focal is older. Similarly, the mean ages of kin differ, with the microsimulation model producing slightly higher estimates for a younger focal and showing greater variation as the focal ages. The distribution of kin numbers and ages also varies, with larger discrepancies for descendant kin, such as grandchildren.

Both modelling approaches have limitations that should be acknowledged. Although analytical models are more transparent and based on explicit assump-

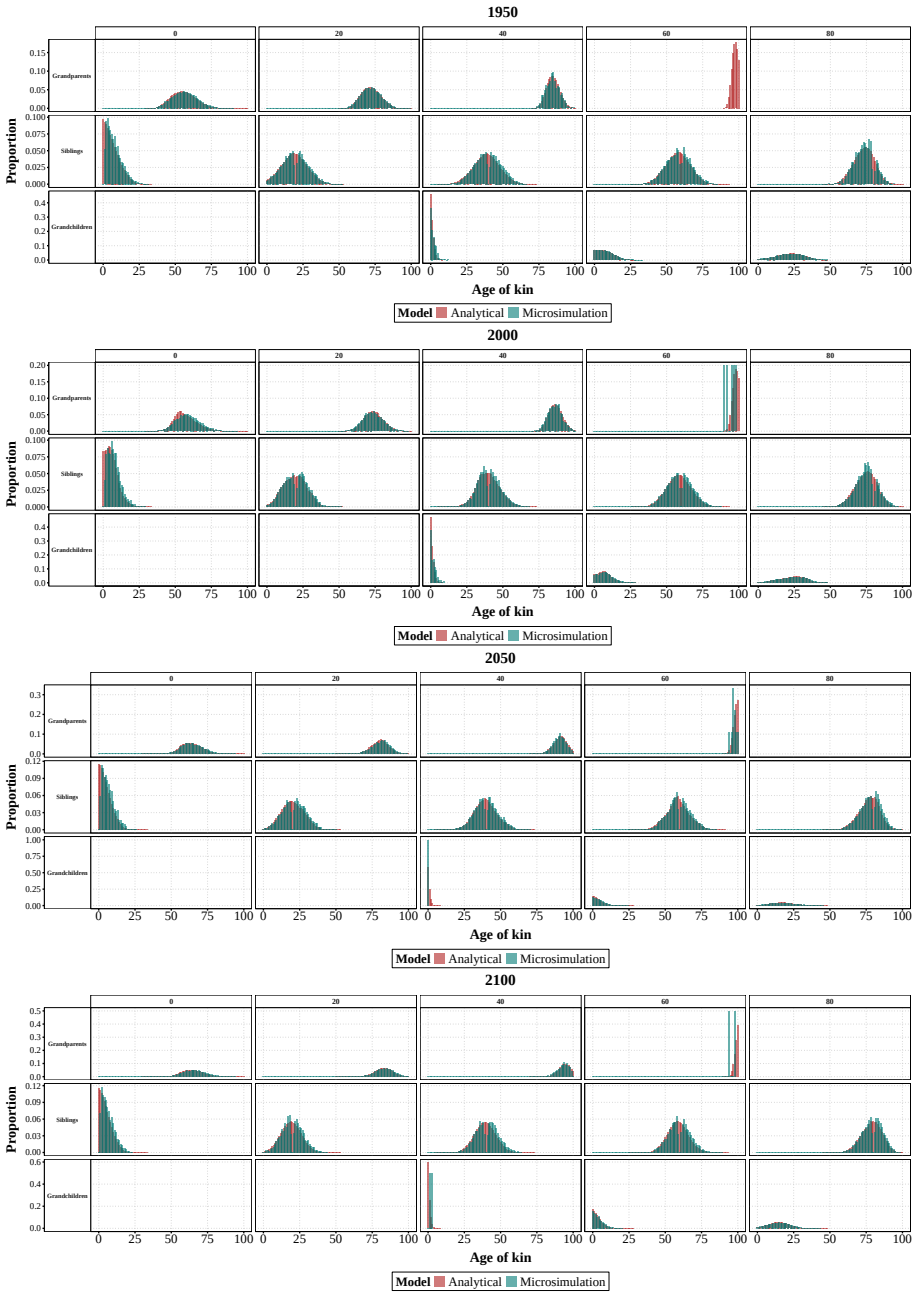

Fig. 4. Distribution of ages of living grandparents, siblings and grandchildren at selected ages of focal individual (0, 20, 40, 60, 80) in 1950, 2000, 2050 and 2100, estimated from analytical and microsimulation models using WPP 2024 fertility and mortality data.

tions, they rely on numerical derivations that may not fully capture the heterogeneity of real-world dynamics as they present estimates for an average individual who survives to old age. Conversely, microsimulation models produce detailed synthetic populations, which are constrained by their reliance on random exposure to the risk of experiencing demographic events. This stochastic nature means that if no individuals in the synthetic population survive to certain ages, there will be no kin to account for. This could lead to an underestimation of the number of kin compared to the expected numbers from analytical models. Furthermore, both models are limited in their modelling of partnership dynamics and in their incorporation of resemblance in demographic behaviours within families. This could result in kinship distributions that are more homogeneous than those observed in real populations. Despite their advantages and limitations, analytical and microsimulation models remain the main approaches for modelling kinship networks using population-level demographic data, particularly in the absence of more detailed data.

Acknowledgments. We would like to thank Diego Alburez-Gutierrez, Martin Kolk, and Emilio Zagheni for their valuable comments and feedback throughout the research process. We would also like to thank Tom Theile for his support with `rsocsim` and Ivan Williams for his help with `DemoKin`. Liliana P. Calderón-Bernal gratefully acknowledges the resources provided by the International Max Planck Research School for Population, Health and Data Science (IMPRS-PHDS).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Alburez-Gutierrez, D., Mason, C., Zagheni, E.: The "Sandwich Generation" Revisited: Global Demographic Drivers of Care Time Demands. Population and Development Review **47**(4), 997–1023 (2021). https://doi.org/10.1111/padr.12436, https://onlinelibrary.wiley.com/doi/abs/10.1111/padr.12436, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/padr.12436
2. Caswell, H.: The formal demography of kinship: A matrix formulation. Demographic Research **41**, 679–712 (Sep 2019). https://doi.org/10.4054/DemRes.2019.41.24, https://www.demographic-research.org/volumes/vol41/24/
3. Caswell, H.: The formal demography of kinship II: Multistate models, parity, and sibship. Demographic Research **42**, 1097–1146 (Jun 2020). https://doi.org/10.4054/DemRes.2020.42.38, https://www.demographic-research.org/volumes/vol42/38/
4. Caswell, H.: The formal demography of kinship IV: Two-sex models and their approximations. Demographic Research **47**(13), 359–396 (Sep 2022). https://doi.org/10.4054/DemRes.2022.47.13, https://www.demographic-research.org/volumes/vol47/13/
5. Caswell, H.: The formal demography of kinship VI: Demographic stochasticity and variance in the kinship network. Demographic Research **51**, 1201–1256 (Nov 2024). https://doi.org/10.4054/DemRes.2024.51.39, https://www.demographic-research.org/articles/volume/51/39

6. Caswell, H., Margolis, R., Verdery, A.: The formal demography of kinship V: Kin loss, bereavement, and causes of death. Demographic Research **49**, 1163–1200 (Dec 2023). https://doi.org/10.4054/DemRes.2023.49.41, https://www.demographic-research.org/articles/volume/49/41
7. Caswell, H., Song, X.: The formal demography of kinship III: Kinship dynamics with time-varying demographic rates. Demographic Research **45**, 517–546 (Aug 2021). https://doi.org/10.4054/DemRes.2021.45.16, https://www.demographic-research.org/volumes/vol45/16/
8. Caswell, H., Vries, L.d.: The formal demography of kinship VII: Lifetime kin overlap within and across generations. Demographic Research **53**, 123–174 (Jul 2025). https://doi.org/10.4054/DemRes.2025.53.5, https://www.demographic-research.org/articles/volume/53/5
9. Hammel, E.A., Hutchinson, D.W., Wachter, K.W., Lundy, R.T., Deuel, R.Z.: The SOCSIM demographic-sociological microsimulation program: operating manual. No. 27 in Research series, Institute of International Studies. University of California, Berkeley (1976), oCLC: 2704303
10. Theile, T., Alburez-Gutierrez, D., Calderón-Bernal, L.P., Snyder, M., Zagheni, E.: rsocsim: SOCSIM with R (2023), https://github.com/MPIDR/rsocsim, https://mpidr.github.io/rsocsim/
11. United Nations, Department of Economic and Social Affairs, P.D.: World Population Prospects 2024 (2024), https://population.un.org/wpp/downloadsfolder=Standard%20Projections&group=Most%20used
12. Williams, I., Alburez-Gutierrez, D., Song, X., Caswell, H.: DemoKin: An R package to implement demographic matrix kinship models (2021), https://github.com/IvanWilli/DemoKin

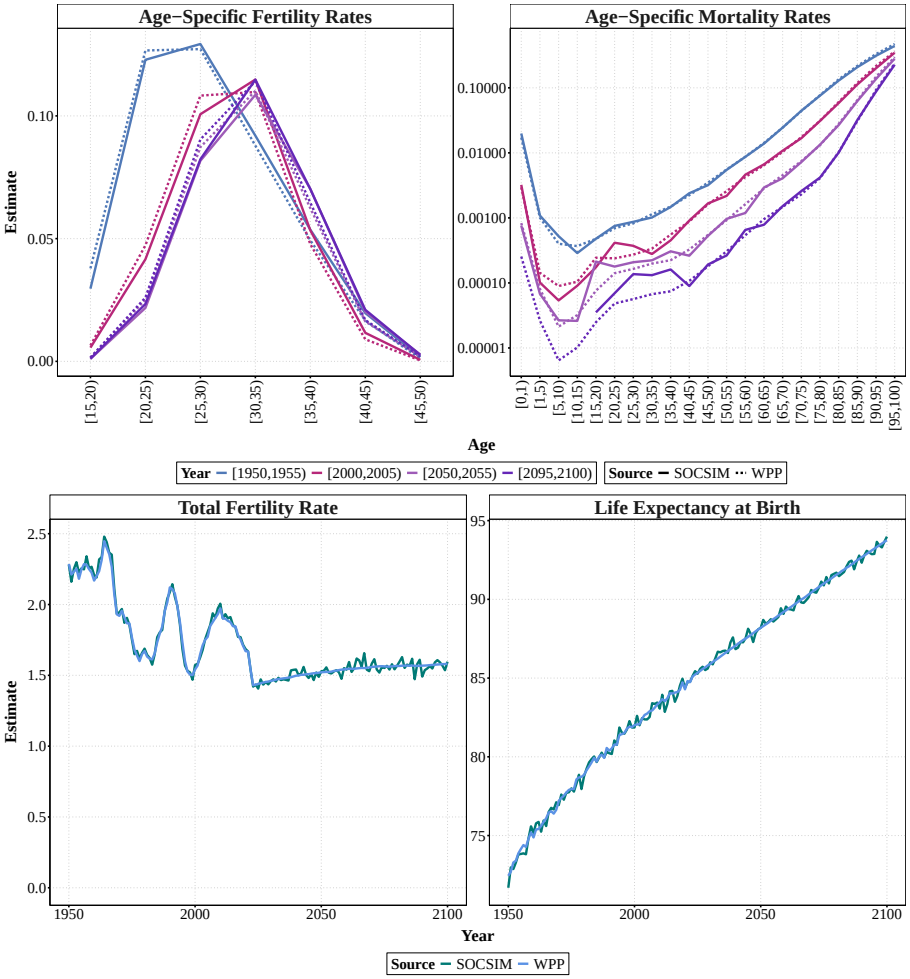

Fig. A1. Age-specific and summary demographic measures for women in Sweden, retrieved from the input data (World Population Prospects 2024 Revision) and the SOCSIM microsimulation output (ASMR in log-scale excluding NA)

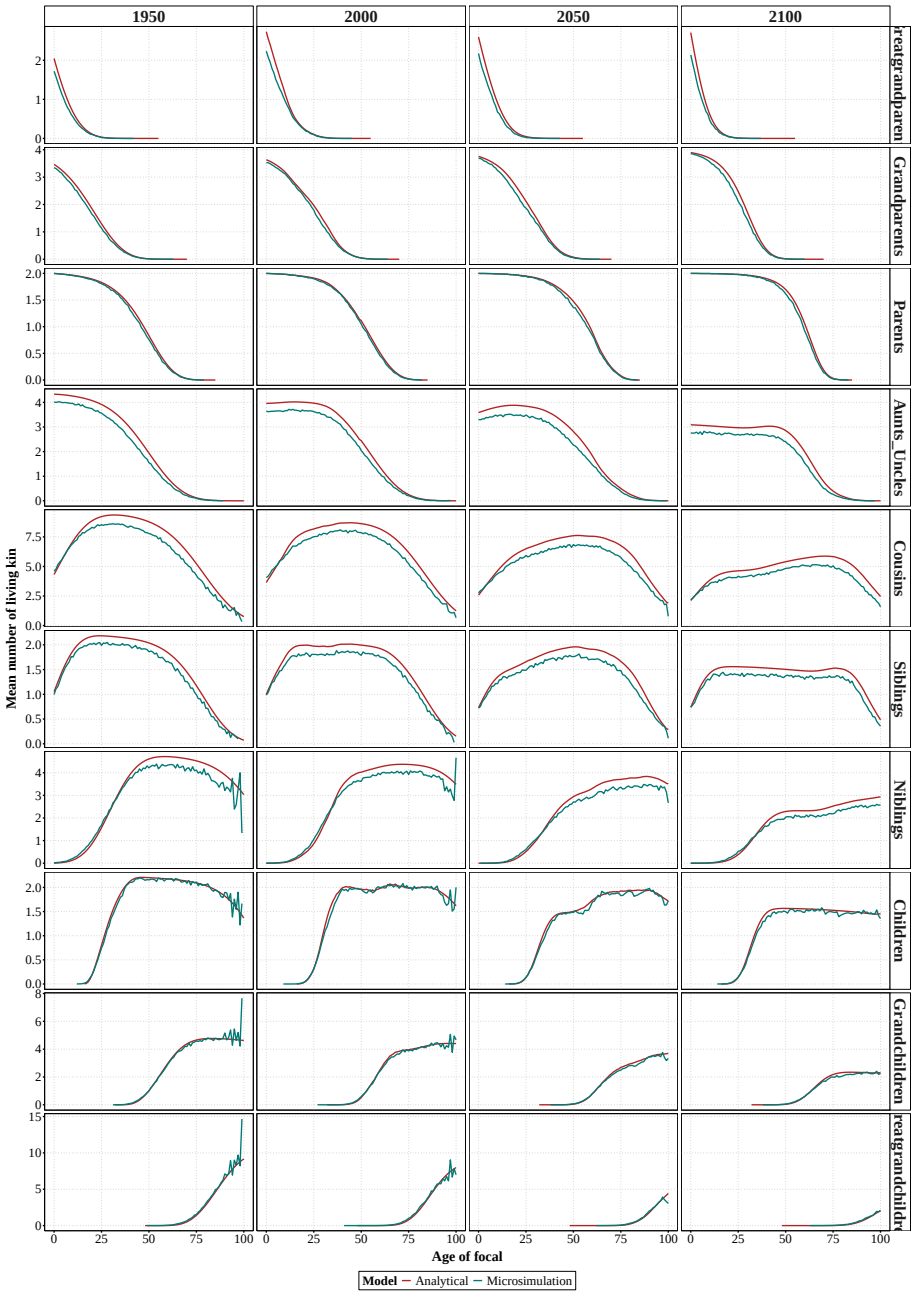

Fig. A2. Mean number of living kin by age of a focal individual in 1950, 2000, 2050 and 2100, estimated from analytical and microsimulation models using WPP 2024 fertility and mortality data.

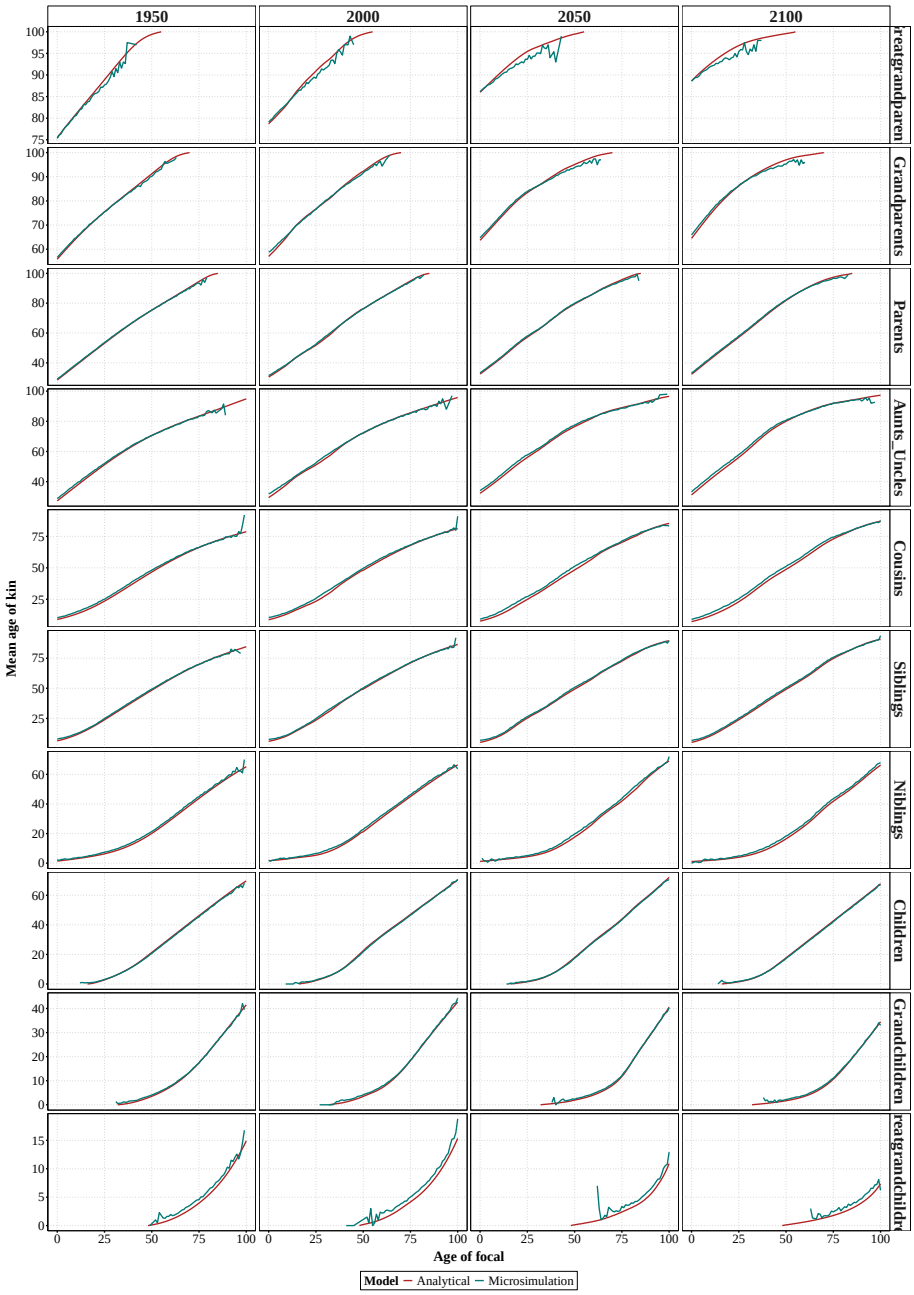

Fig. A3. Mean age of living kin by age of a focal individual in 1950, 2000, 2050 and 2100, estimated from analytical and microsimulation models using WPP 2024 fertility and mortality data.

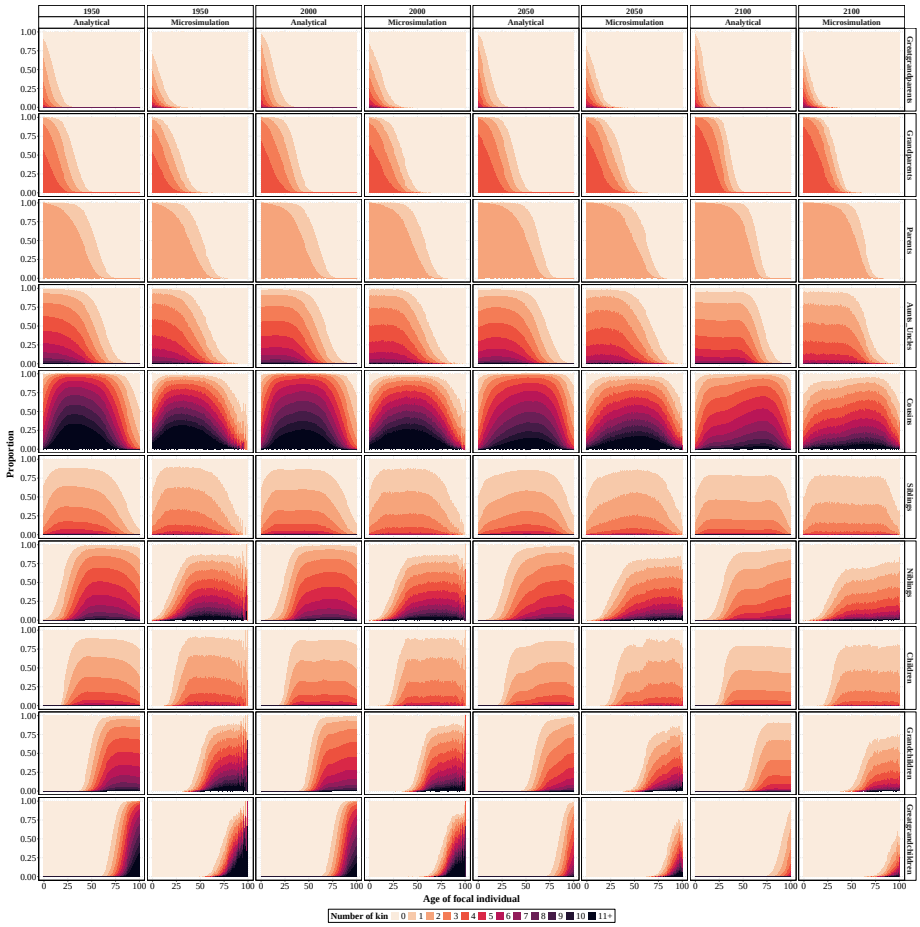

Fig. A4. Distribution of the numbers of living kin by age of a focal individual in 1950, 2000, 2050 and 2100, estimated from analytical and microsimulation models using WPP 2024 fertility and mortality data.

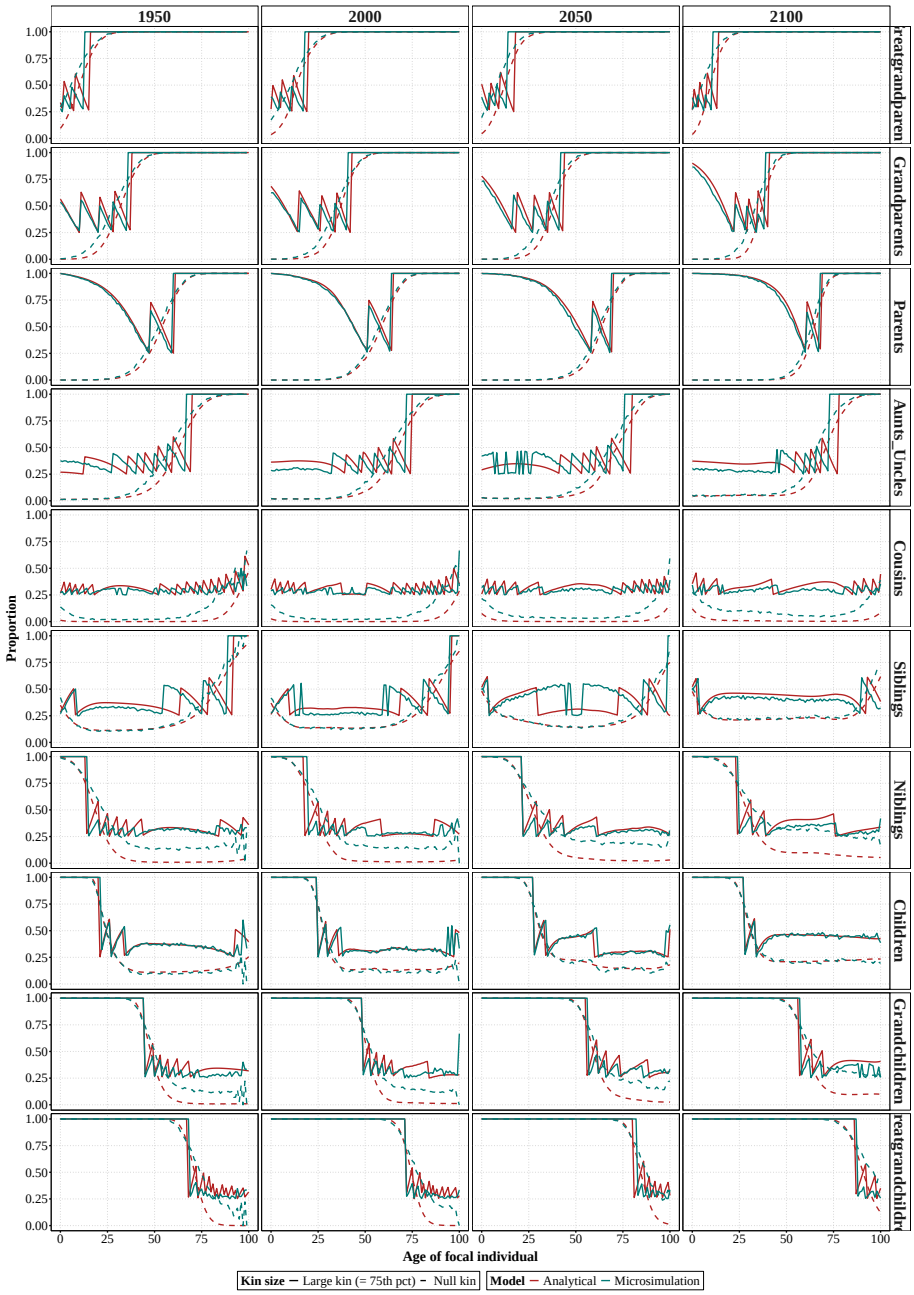

Fig. A5. Proportion of individuals with zero or large ($\geq$ 75th pct) numbers of kin by age of a focal individual in 1950, 2000, 2050 and 2100 estimated from analytical and microsimulation models using WPP 2024 fertility and mortality data.

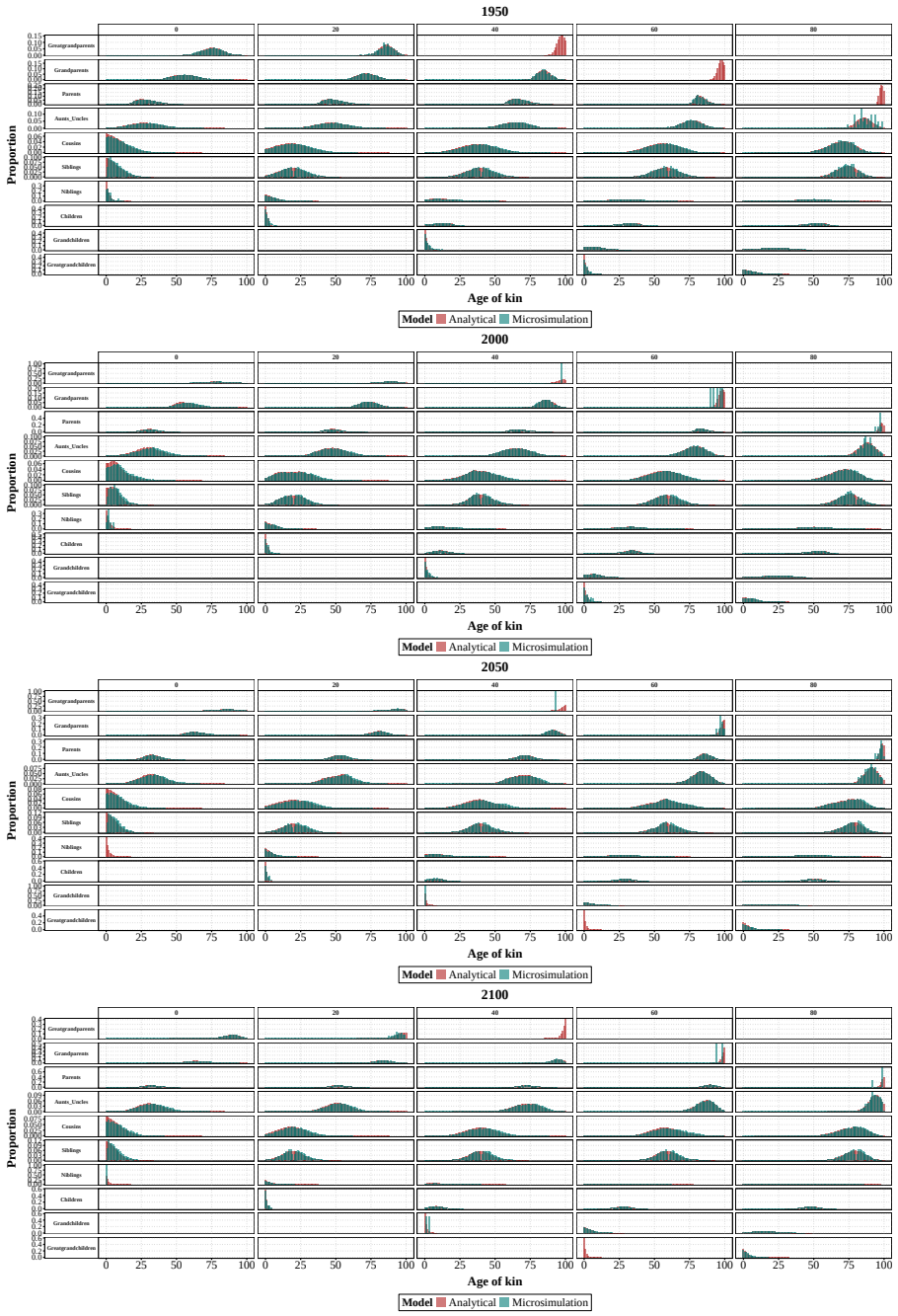

Fig. A6. Distribution of ages of kin at selected ages of a focal individual in 1950, 2000, 2050 and 2100, estimated from analytical and microsimulation models using WPP 2024 fertility and mortality data.