Kinship models

An introduction to the analysis of population-level kinship structures

Diego Alburez-Gutiérrez†

†Kinship Inequalities Research Group,
Max Planck Institute for Demographic Research

International Institute for Population Sciences,
India, November 24-27, 2025

MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH
MAX-PLANCK-INSTITUT FÜR DEMOGRAFISCHE FORSCHUNG

Agenda

1. The Goodman-Keyfitz-Pullum kinship equations
2. Matrix kinship models
3. Implementations

What are kinship models?

1. Kinship is an *emergent property* of demographic systems
2. Simplified representation of interaction between reproduction and death
3. Not restricted to humans[1]

→ What are emergent properties?
→ Can you think of other emergent properties in nature, society, or demography?

[1] Coste, C. F. D., Bienvenu, F., Ronget, V., Ramirez-Loza, J.-P., Cubaynes, S., & Pavard, S. (2021). The kinship matrix: Inferring the kinship structure of a population from its demography (T. Coulson, Ed.). *Ecology Letters*, *24*(12), 2750–2762. https://doi.org/10.1111/ele.13854

Formal models of kinship

Given a set of:

- age-specific fertility rates
- survival probabilities
- simplifying assumptions

The models produce:

1. Number of (living/dead) kin
2. Age distribution of relatives
3. From the point of view of an average member of the population ('Focal')

Focal: an average member of the population

Typology of kinship models

No	time	sex	state	reference
1	**invariant**	**female**	**age**	[2]
2	variant	female	age	[3]
3	invariant	two	age	[4]
4	invariant	female	multiple	[5]
5	variant	two	multiple	[6]

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

[3] Caswell, H., & Song, X. (2021). The formal demography of kinship. III. kinship dynamics with time-varying demographic rates. *Demographic Research*, *45*, 517–546

[4] Caswell, H. (2022). The formal demography of kinship IV: Two-sex models and their approximations. *Demographic Research*, *47*, 359–396

[5] Caswell, H. (2020). The formal demography of kinship II: Multistate models, parity, and sibship. *Demographic Research*, *42*, 1097–1146

[6] Williams, I., Alburez-Gutierrez, D., & DemoKin Team. (2023). *DemoKin: 1.0.3*. https://CRAN.R-project.org/package=DemoKin

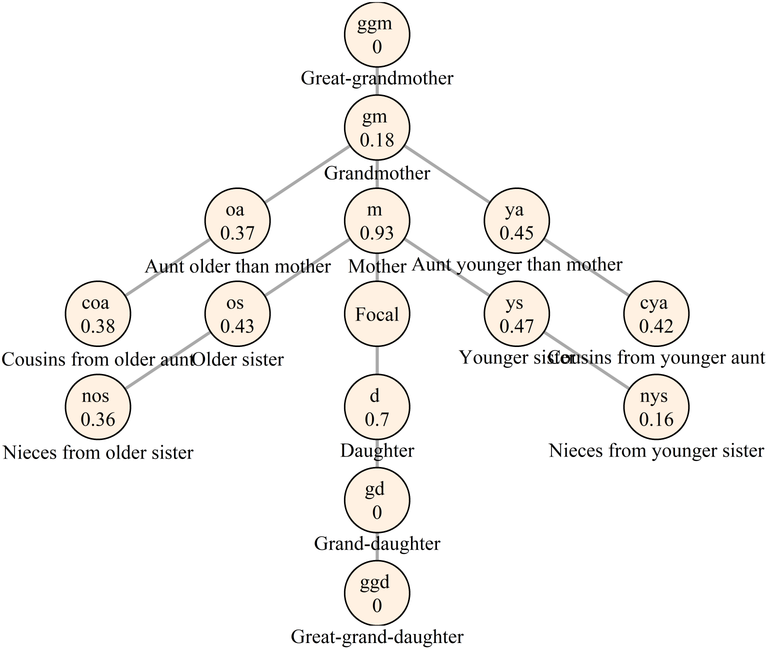
Model characteristics

Define the following model characteristics:

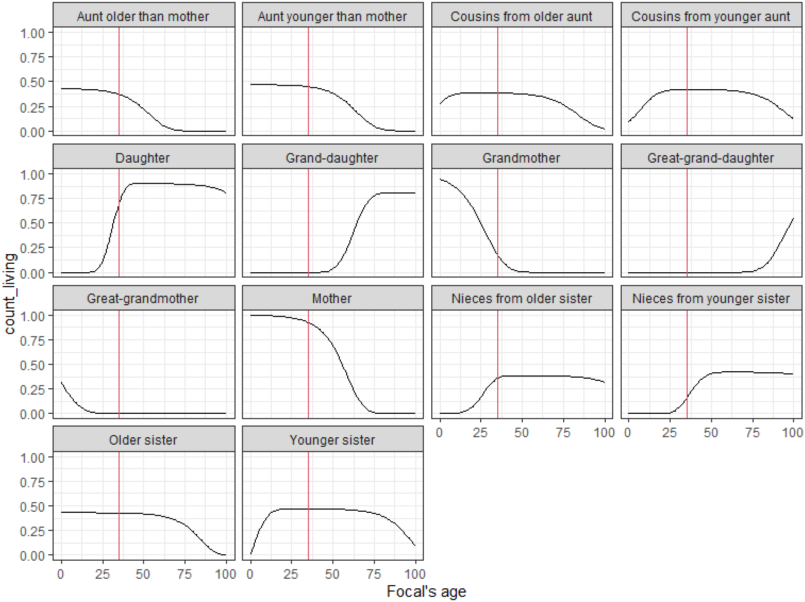
1. Time-(in)variance
2. One/two-sex models
3. (Multi)state models

The Goodman-Keyfitz-Pullum kinship equations

The tree of life



Expected number of kin



Daughters

$B_1(a)$ is the expected number of living daughters in a time-invariant female-only population[7]:

$$B_1(a) = \int_{\alpha}^{a} m(x) l(a-x)\, dx \tag{1}$$

where:

- $m(x)$ are fertility rates of mothers
- $l(a-x)$ are survival probabilities of daughters

[7] Goodman, L. A. (1974). Family Formation and the Frequency of Various Kinship Relationships. *Theoretical Population Biology*, 27

Daughters

If $a = 20$ and $\alpha = 15$; then:

$$B_1(20) \approx \sum_{15}^{20} m(x)l(20 - x)$$

So...

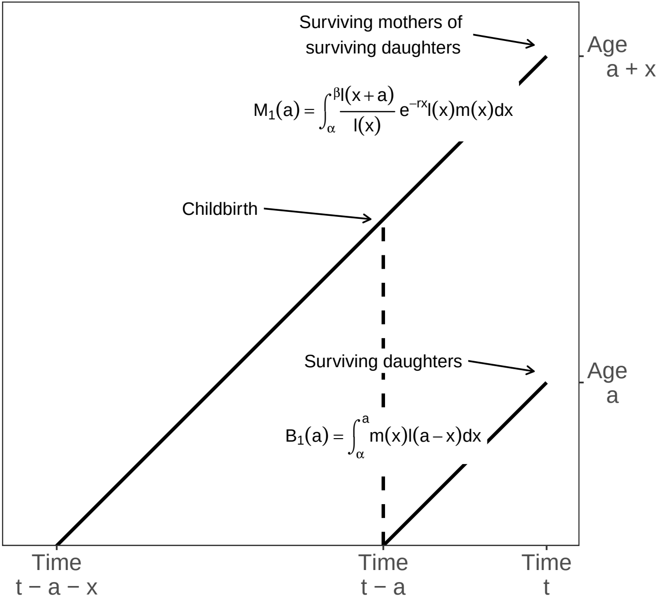
$$B_1(20) \approx m(15)l(5) + m(16)l(4) + m(17)l(3) \dots$$

Granddaughters

$B_1(a)$ is the expected number of living granddaughters in a time-invariant female-only population[8]:

$$B_2(a) = \int_{\alpha}^{a} m(x) \int_{\alpha}^{a-x} l(y)m(y)l(a-x-y)\, dy\, dx \qquad (2)$$

[8] Goodman, L. A. (1974). Family Formation and the Frequency of Various Kinship Relationships. *Theoretical Population Biology*, 27


Surviving mothers of
surviving daughters
Age
a + x
$M_1(a) = \int_\alpha^\beta \frac{l(x+a)}{l(x)} e^{-rx} l(x) m(x) dx$
Childbirth
Surviving daughters
Age
a
$B_1(a) = \int_\alpha^a m(x) l(a-x) dx$
Time
t − a − x
Time
t − a
Time
t

Mothers

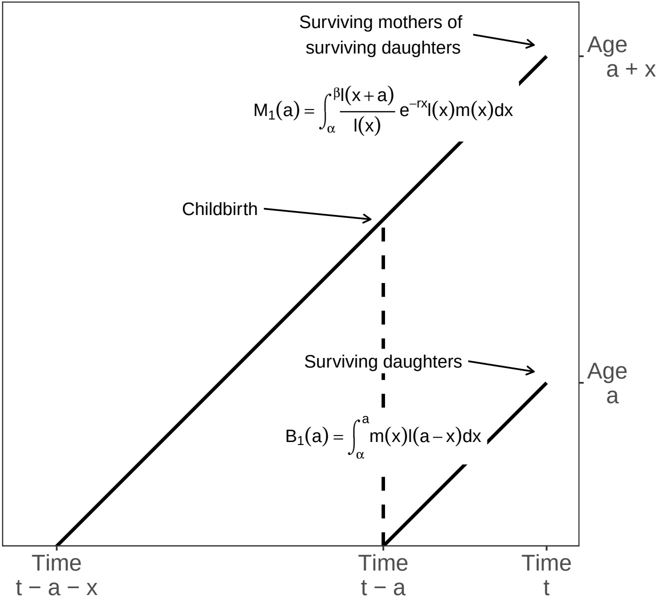
$M_1(a)$ is the probability of having a living mother in a time-invariant female-only population[9]:

$$M_1(a) = \int_{\alpha}^{\beta} \underbrace{\frac{l(x+a)}{l(x)}}_{\substack{\text{prob of surviving} \\ \text{from } x \text{ to } a+x}} \times \underbrace{W(x)}_{\substack{\text{age distribution of} \\ \text{mothers}}} dx. \tag{3}$$

where:

- $W(x) = e^{-rx} l(x) m(x)$ is the age distribution of mothers
- $l(x)$ are survival probabilities
- $m(x)$ are fertility rates
- r is the population growth rate
- α-β is the reproductive period

[9] Goodman, L. A. (1974). Family Formation and the Frequency of Various Kinship Relationships. *Theoretical Population Biology*, 27


Surviving mothers of
surviving daughters
Age
a + x
$M_1(a) = \int_\alpha^\beta \frac{l(x+a)}{l(x)} e^{-rx} l(x) m(x) dx$
Childbirth
Surviving daughters
Age
a
$B_1(a) = \int_\alpha^a m(x) l(a-x) dx$
Time
t − a − x
Time
t − a
Time
t

Grandmothers

$M_2(a)$ is the expected number of living grandmothers in a time-invariant female-only population[10]:

$$M_2(a) = \int_{\alpha}^{\beta} \underbrace{M_1(a)}_{\substack{\text{prob of having} \\ \text{living mother}}} \times \underbrace{W(x)}_{\substack{\text{age distribution of} \\ \text{mothers}}} dx. \tag{4}$$

[10] Goodman, L. A. (1974). Family Formation and the Frequency of Various Kinship Relationships. *Theoretical Population Biology*, 27

Great-grandmothers

$M_3(a)$ is the expected number of living great-grandmothers in a time-invariant female-only population[11]:

$$M_3(a) = \int_{\alpha}^{\beta} \underbrace{M_2(a)}_{\substack{\text{number of}\\ \text{grandmother}}} \times \underbrace{W(x)}_{\substack{\text{age distribution of}\\ \text{mothers}}} dx. \tag{5}$$

[11] Goodman, L. A. (1974). Family Formation and the Frequency of Various Kinship Relationships. *Theoretical Population Biology*, 27

Sisters

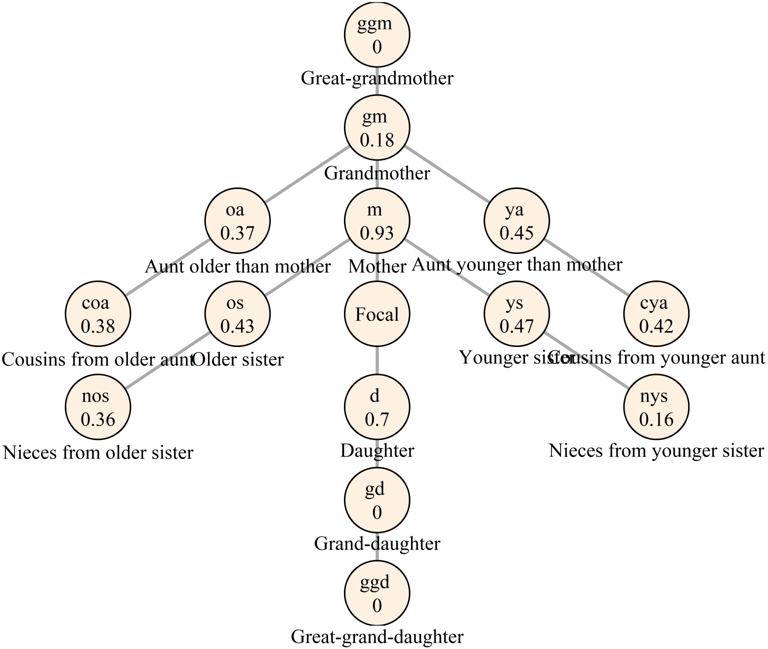
$B_1(a)$ is the expected number of living older sisters in a time-invariant female-only population[12]:

$$S^{old}(a) = \int_{\alpha}^{\beta} \int_{\alpha}^{x} m(y) l(a + x - y) W(x) \, dy \, dx \tag{6}$$

$$S^{young}(a) = \int_{\alpha}^{\beta} \int_{0}^{a} \left[\frac{l(x+u)}{l(x)} \right] m(x+u) l(a-u) \, du \, W(x) \, dx \tag{7}$$

[12] Goodman, L. A. (1974). Family Formation and the Frequency of Various Kinship Relationships. *Theoretical Population Biology*, 27

Why do we model younger and older sisters/kin separately?



Demographic subsidy

"New members of the population arise not from reproduction of current members, but from elsewhere"[13]

- Can you think of other instances of 'subsidy' in demography?

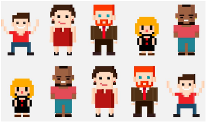
[13] Caswell, H. (2019). The formal demography of kinship: A matrix formulation. *Demographic Research*, *41*, 679–712

Break

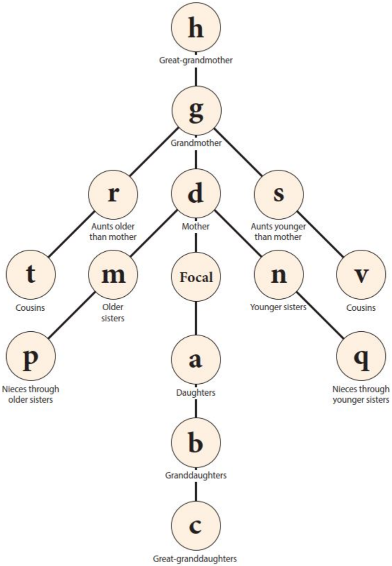
Matrix kinship models

From recursive equations to matrix operations

1. The relatives of Focal constitute a population
2. They can be modelled using traditional projection methods
3. Matrix operations provide an efficient implementation



The tree of life (2)



Implementation: time-invariant, one-sex models[14]

The models are of the general form:

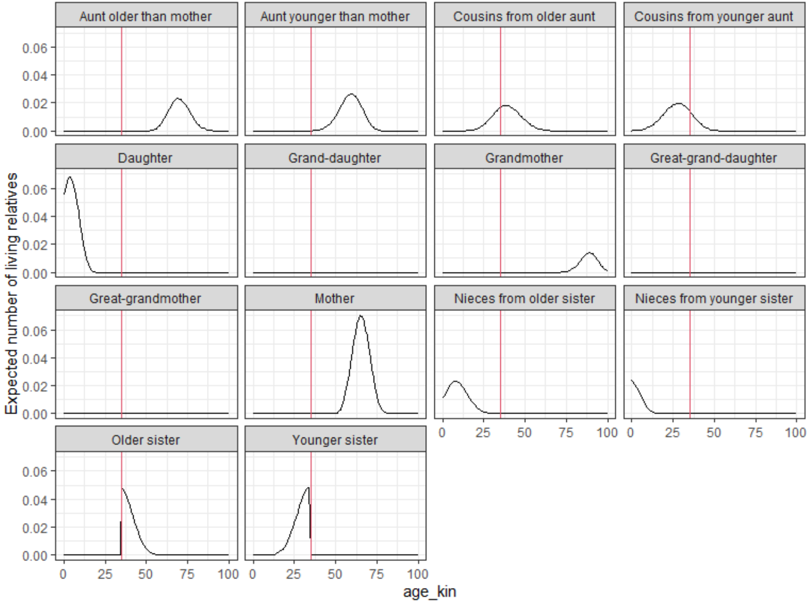
$$\underbrace{\mathbf{k}(x+1)}_{\substack{\text{age structure of kin} \\ \text{at Focal's age } x+1}} = \underbrace{\mathbf{U}\,\mathbf{k}(x)}_{\substack{\text{ageing and survival} \\ \text{of existing kin}}} + \underbrace{\begin{cases} \mathbf{0} \\ \mathbf{F}\,\mathbf{k}^{*}(x) \end{cases}}_{\substack{\text{new kin members} \\ \text{added to the population}}} .$$

where:

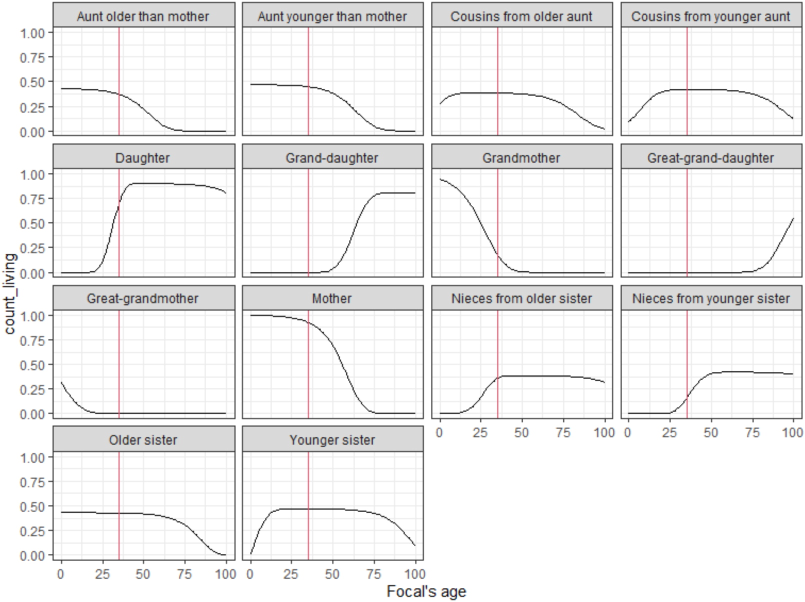
- $\mathbf{U}$ a matrix with survival probabilities in the subdiagonal
- $\mathbf{F}$ a matrix with fertility rates in the first row

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

Age distributions of kin



Expected number of kin



Daughters

Daughters ($\mathbf{a}$) are the result of the reproduction of Focal:

$$\underbrace{\mathbf{a}(x+1)}_{\substack{\text{age structure of daughters}\\ \text{at Focal's age } x+1}} = \underbrace{\mathbf{U}\,\mathbf{a}(x)}_{\substack{\text{ageing and survival}\\ \text{of existing daughters}}} + \underbrace{\mathbf{F}\,\mathbf{e}_x}_{\substack{\text{new daughters}\\ \text{(subsidy)}}}. \tag{8}$$

$$a(0) = \mathbf{0}.$$

where:

- $\mathbf{U}$ is a matrix with survival probabilities in the subdiagonal
- $\mathbf{F}$ is a matrix with fertility rates in the first row
- $\mathbf{F}\,\mathbf{e}_x$ is the subsidy vector
- e_x is the unit vector for age x
- $a(0)$ is the distribution of daughters at Focal's birth

Mothers

The population of mothers ($\mathbf{d}$) of Focal consists of at most a single individual:

$$\underbrace{\mathbf{d}(x+1)}_{\substack{\text{age structure of mothers} \\ \text{at Focal's age } x+1}} = \underbrace{\mathbf{U}\,\mathbf{d}(x)}_{\substack{\text{ageing and survival} \\ \text{of existing mothers}}} + \underbrace{\mathbf{0}.}_{\substack{\text{new mothers} \\ \text{(subsidy)}}} \tag{9}$$

$$d(0) = \pi.$$

where:

- $b(0)$ is the distribution of mothers at Focal's birth
- π is the distribution of ages of mothers in the population

All models[15]

Table 1: Summary of the components of the kin model given in equations (4) and (5)

Symbol	Kin	Initial condition	Subsidy $\beta(x)$
$\mathbf{a}$	daughters	$\mathbf{0}$	$\mathbf{F}\mathbf{e}_x$
$\mathbf{b}$	granddaughters	$\mathbf{0}$	$\mathbf{F}\mathbf{a}(x)$
$\mathbf{c}$	great-granddaughters	$\mathbf{0}$	$\mathbf{F}\mathbf{b}(x)$
$\mathbf{d}$	mothers	$\boldsymbol{\pi}$	$\mathbf{0}$
$\mathbf{g}$	grandmothers	$\sum_i \pi_i \mathbf{d}(i)$	$\mathbf{0}$
$\mathbf{h}$	great-grandmothers	$\sum_i \pi_i \mathbf{g}(i)$	$\mathbf{0}$
$\mathbf{m}$	older sisters	$\sum_i \pi_i \mathbf{a}(i)$	$\mathbf{0}$
$\mathbf{n}$	younger sisters	$\mathbf{0}$	$\mathbf{F}\mathbf{d}(x)$
$\mathbf{p}$	nieces via older sisters	$\sum_i \pi_i \mathbf{b}(i)$	$\mathbf{F}\mathbf{m}(x)$
$\mathbf{q}$	nieces via younger sisters	$\mathbf{0}$	$\mathbf{F}\mathbf{n}(x)$
$\mathbf{r}$	aunts older than mother	$\sum_i \pi_i \mathbf{m}(i)$	$\mathbf{0}$
$\mathbf{s}$	aunts younger than mother	$\sum_i \pi_i \mathbf{n}(i)$	$\mathbf{F}\mathbf{g}(x)$
$\mathbf{t}$	cousins from aunts older than mother	$\sum_i \pi_i \mathbf{p}(i)$	$\mathbf{F}\mathbf{r}(x)$
$\mathbf{v}$	cousins from aunts younger than mother	$\sum_i \pi_i \mathbf{q}(i)$	$\mathbf{F}\mathbf{s}(x)$

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

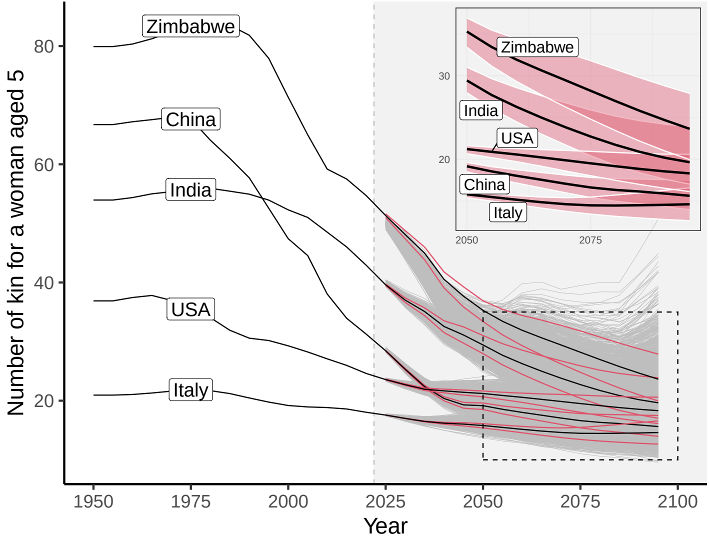
Consider a baby born in Spain in 1950...

1. How old were her grandparents when she was born, on average?
2. How many living children did she have on her 70th birthday?
3. How many grandchildren?

Break

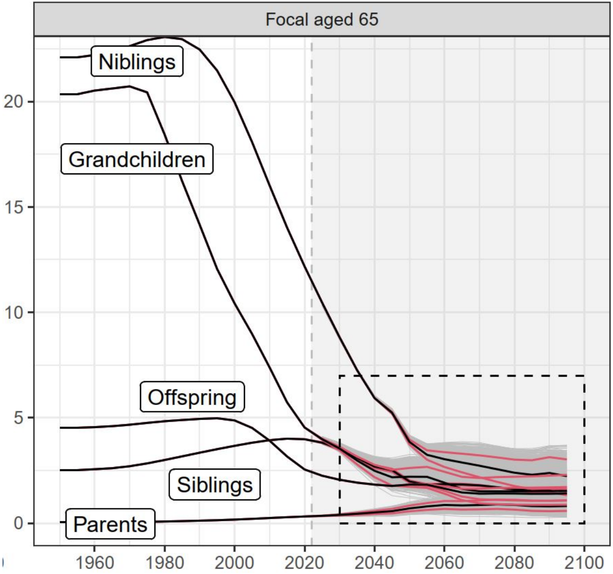
Implementations

Total number of kin (all kin combined) for a 5yo woman[16]



[16] Alburez-Gutierrez, D., Williams, I., & Caswell, H. (2023).Projections of human kinship for all countries. *Proceedings of the National Academy of Sciences, 120(52)*, e2315722120. https://doi.org/10.1073/pnas.2315722120

Number of living kin for a 65yo in China



Models vs reality

Discuss:

1. What is the relationship between demographic models and reality?
2. Would we expect kinship models to agree with 'empirical' observations of kinship?
3. Where can we find empirical data on kinship availability?