

Kinship assignment

Alex Sheridan, Jackson Mason-MacKay, Gilbert Habaasa

2024-11-22

You will use data on kinship structures to benchmark formal models of kinship. For this exercise, you will use the **DemoKin** R package to implement formal models of kinship. You should choose one country and run four different models according to the following specifications:

- One-sex model; approximate male kin using GKP factors
 - time-invariant rates

```
Model_1 <-
  kin(sud_surv_females[, "2023"], sud_asfr_females[, "2023"], birth_female = .5)$kin_summary %>%
  inner_join(gkp_factors, by = "kin") %>%
  mutate(count_living = count_living * factor)
```

- time-variant rates

```
Model_1v <-  
  kin(p = sud_surv_females, f = sud_asfr_females, n = sud_pop, time_invariant = FALSE, output_period = 1)  
  inner_join(gkp_factors, by = "kin") %>%  
  mutate(count_living = count_living * factor)
```

```
## Preparing output...
```

```
## Assuming stable population before 1950.
```

- Two-sex model; approximate male kin using the androgynous assumption
 - time-invariant rates

```
Model_2 <- kin2sex(
  pf = sud_surv_females["2023"],
  pm = sud_surv_males["2023"],
  ff = sud_asfr_females["2023"],
  fm = sud_asfr_females["2023"],
  time_invariant = TRUE,
  sex_focal = "f", birth_female = .5)$kin_summary
```

- time-variant rates

```
Model_2v <- kin2sex(
  pf = sud_surv_females,
  pm = sud_surv_males,
  ff = sud_asfr_females,
  fm = sud_asfr_females,
  time_invariant = FALSE,
  output_period = 2023,
  sex_focal = "f", birth_female = .5)$kin_summary
```

```
## Preparing output...
```

```
## Assuming stable population before 1950.
```

Use the output of the four models to answer the following questions:

1. Plot the **expected number of living relatives by age of focal** for each specification. For extra points (i.e., this is optional), also plot the **expected number of deceased relatives by age of focal**.

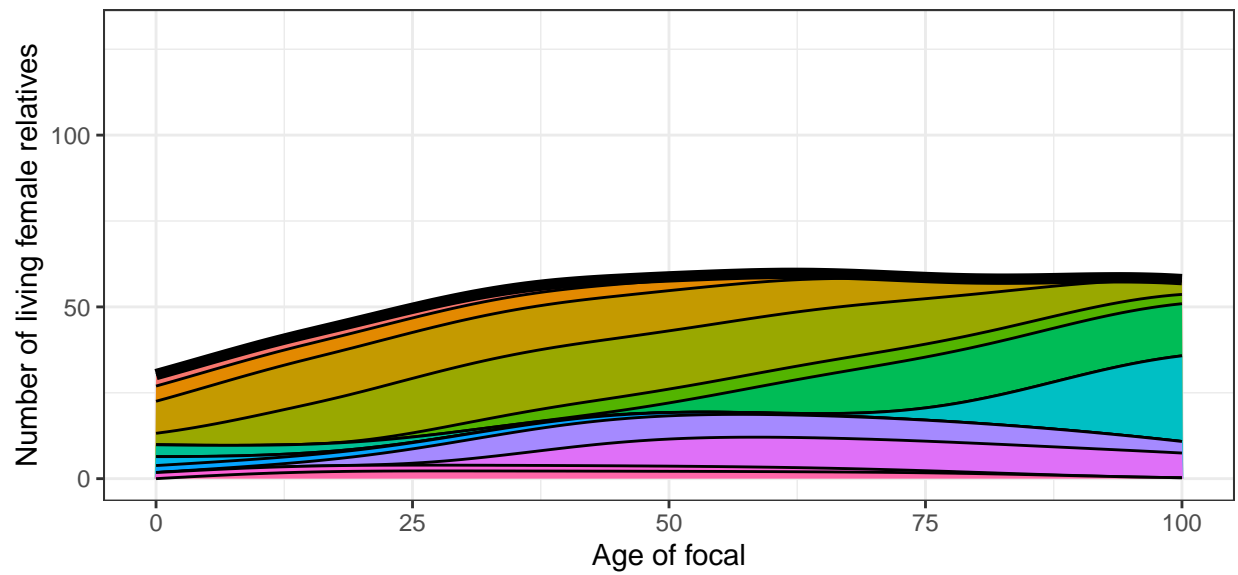
```
counts <-
  Model_1 %>%
  group_by(age_focal) %>%
  summarise(count_living = sum(count_living)) %>%
  ungroup()

Model_1 %>%
  select(age_focal, kin, count_living) %>%
  rename_kin() %>%
  ggplot(aes(x = age_focal, y = count_living)) +
  geom_area(aes(fill = kin_label), colour = "black") +
  geom_line(data = counts, size = 2) +
  labs(x = "Age of focal",
       y = "Number of living female relatives",
       fill = "Kin",
       title = "Time invariant, one sex, GKP assumption") +
  coord_cartesian(ylim = c(0, 130)) +
  theme_bw() +
  theme(legend.position = "bottom")
```

```
## Joining with 'by = join_by(kin)'
```

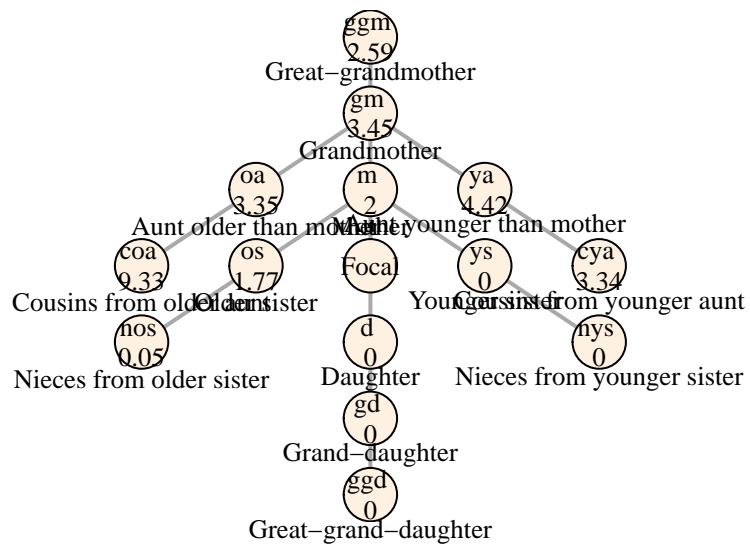
```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

Time invariant, one sex, GKP assumption



Older than mother	Cousins from younger aunts	Grandmothers	Mother
Younger than mother	Daughters	Great-grand-daughters	Nieces from older sister
Sisters from older aunts	Grand-daughters	Great-grandmothers	Nieces from younger sis

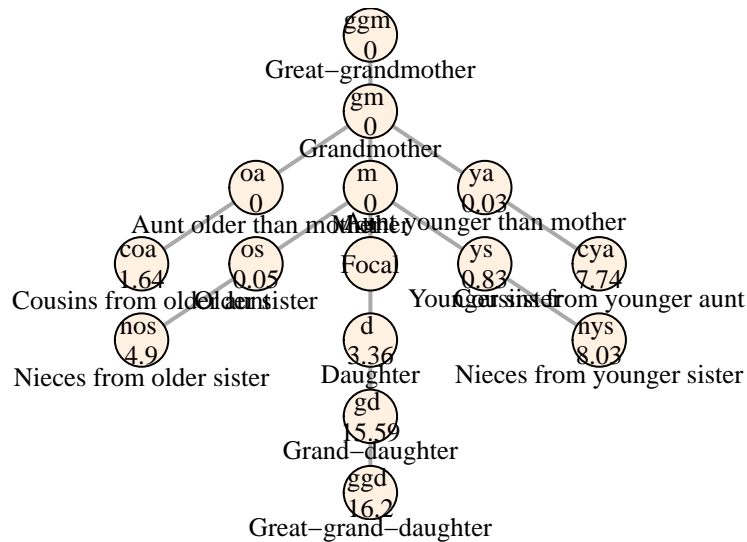
```
Model_1 %>%
  filter(age_focal == 0) %>%
  select(kin, count = count_living) %>%
  plot_diagram(rounding = 2)
```



```

Model_1 %>%
  filter(age_focal == 90) %>%
  select(kin, count = count_living) %>%
  plot_diagram(rounding = 2)

```



```

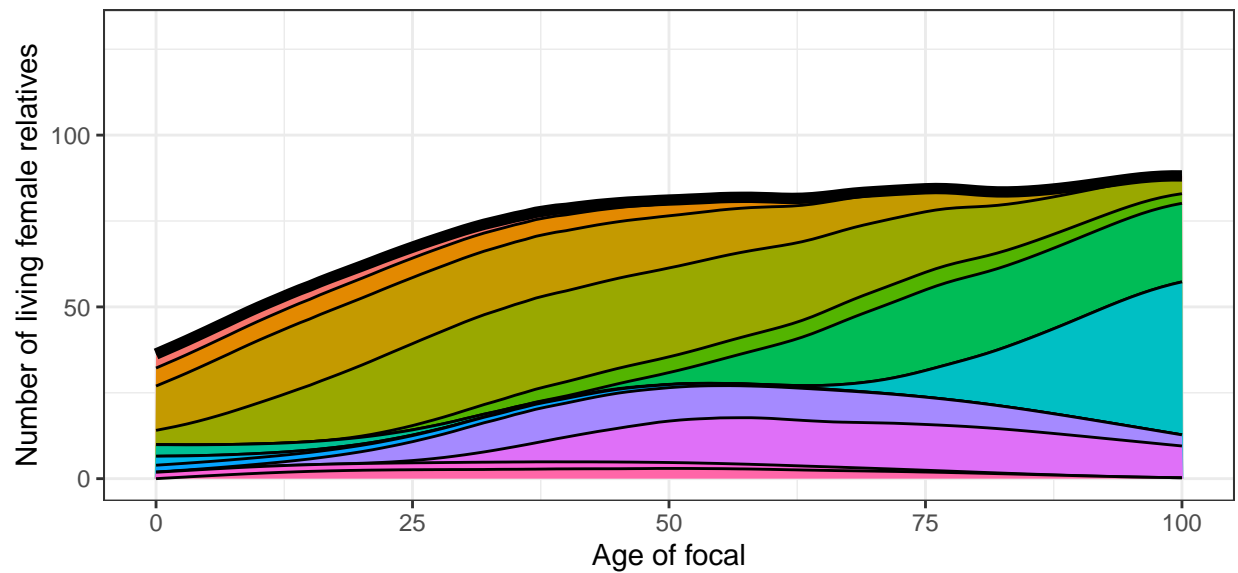
counts <-
  Model_1v %>%
  group_by(age_focal) %>%
  summarise(count_living = sum(count_living)) %>%
  ungroup()

Model_1v %>%
  select(age_focal, kin, count_living) %>%
  rename_kin() %>%
  ggplot(aes(x = age_focal, y = count_living)) +
  geom_area(aes(fill = kin_label), colour = "black") +
  geom_line(data = counts, size = 2) +
  labs(x = "Age of focal",
       y = "Number of living female relatives",
       fill = "Kin",
       title = "Time variant, one sex, GKP assumption") +
  coord_cartesian(ylim = c(0, 130)) +
  theme_bw() +
  theme(legend.position = "bottom")

```

```
## Joining with 'by = join_by(kin)'
```

Time variant, one sex, GKP assumption



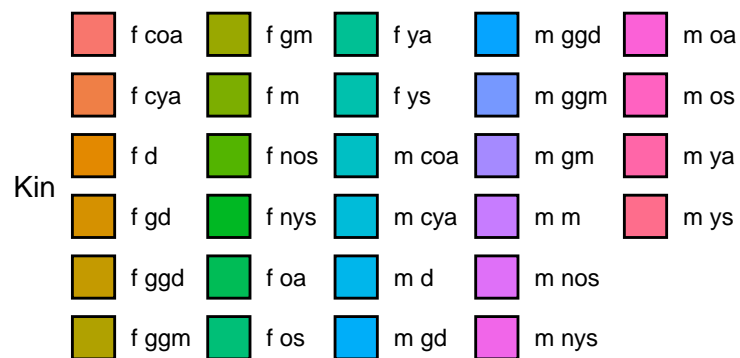
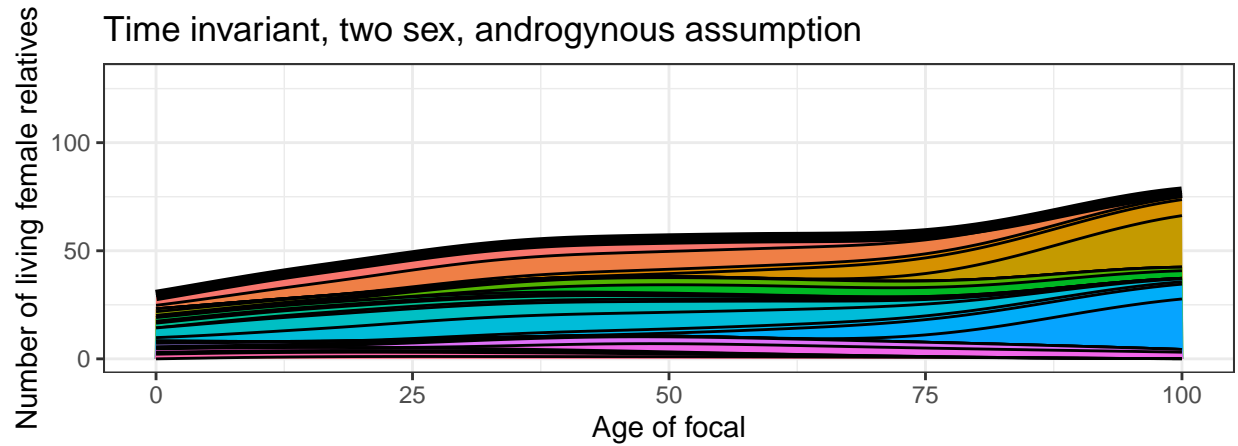
Older than mother	Cousins from younger aunts	Grandmothers	Mother
Younger than mother	Daughters	Great-grand-daughters	Nieces from older sister
Sisters from older aunts	Grand-daughters	Great-grandmothers	Nieces from younger sis

```
counts <-
  Model_2 %>%
  group_by(age_focal) %>%
  summarise(count_living = sum(count_living)) %>%
  ungroup()

Model2b <- Model_2 %>%
  mutate (kin = paste (sex_kin, kin))

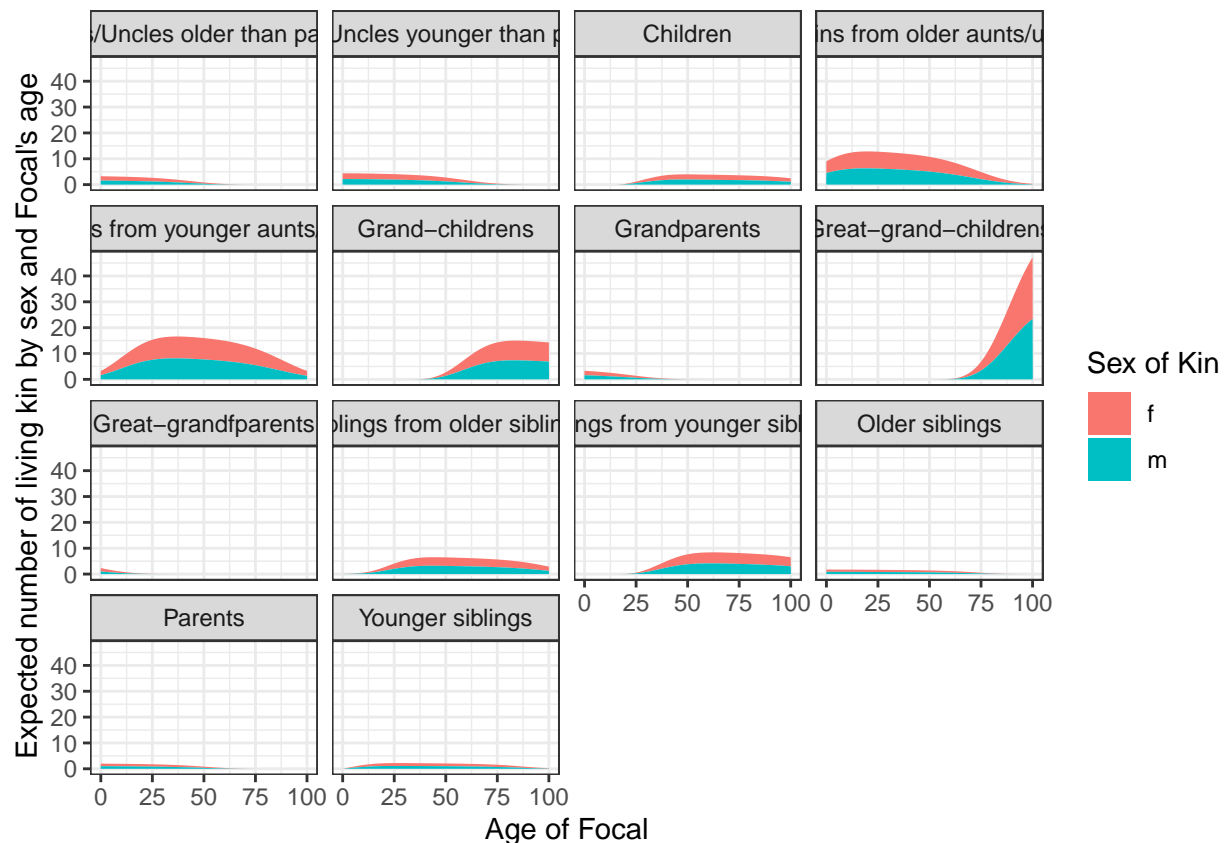
Model2b %>%
  select(age_focal, kin, count_living) %>%
  rename_kin() %>%
  ggplot(aes(x = age_focal, y = count_living)) +
  geom_area(aes(fill = kin), colour = "black") +
  geom_line(data = counts, size = 2) +
  labs(x = "Age of focal",
       y = "Number of living female relatives",
       fill = "Kin",
       title = "Time invariant, two sex, androgynous assumption") +
  coord_cartesian(ylim = c(0, 130)) +
  theme_bw() +
  theme(legend.position = "bottom")
```

```
## Joining with 'by = join_by(kin)'
```



```
Model_2 %>%
  rename_kin(sex = "2sex") %>%
  summarise(count=sum(count_living), .by = c(kin_label, age_focal, sex_kin)) %>%
  ggplot(aes(age_focal, count, fill=sex_kin))+
  geom_area()+
  theme_bw() +
  labs(y = "Expected number of living kin by sex and Focal's age",
       x = "Age of Focal",
       fill = "Sex of Kin") +
  facet_wrap(~kin_label)
```

```
## Joining with 'by = join_by(kin)'
```

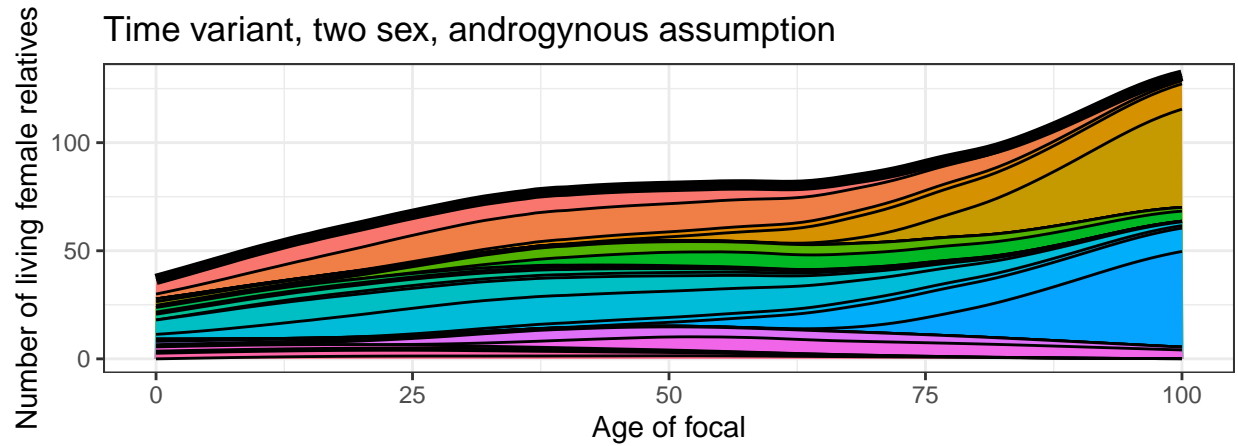


```
counts <-
  Model_2v %>%
  group_by(age_focal) %>%
  summarise(count_living = sum(count_living)) %>%
  ungroup()

Model2bv <- Model_2v %>%
  mutate (kin = paste (sex_kin, kin))

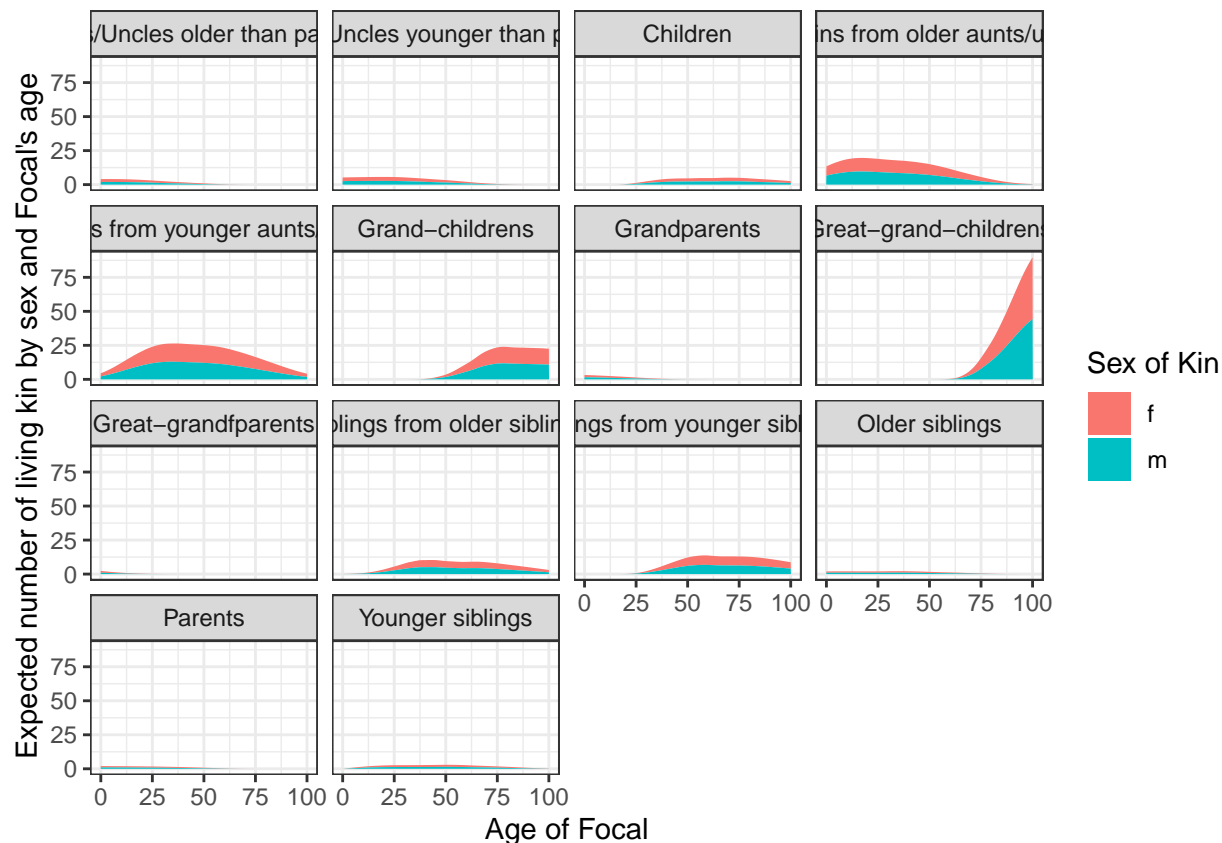
Model2bv %>%
  select(age_focal, kin, count_living) %>%
  rename_kin() %>%
  ggplot(aes(x = age_focal, y = count_living)) +
  geom_area(aes(fill = kin), colour = "black") +
  geom_line(data = counts, size = 2) +
  labs(x = "Age of focal",
       y = "Number of living female relatives",
       fill = "Kin",
       title = "Time variant, two sex, androgynous assumption") +
  coord_cartesian(ylim = c(0, 130)) +
  theme_bw() +
  theme(legend.position = "bottom")
```

```
## Joining with 'by = join_by(kin)'
```

```
Model_2v %>%
  rename_kin(sex = "2sex") %>%
  summarise(count=sum(count_living), .by = c(kin_label, age_focal, sex_kin)) %>%
  ggplot(aes(age_focal, count, fill=sex_kin))+
  geom_area()+
  theme_bw() +
  labs(y = "Expected number of living kin by sex and Focal's age",
       x = "Age of Focal",
       fill = "Sex of Kin") +
  facet_wrap(~kin_label)
```

```
## Joining with 'by = join_by(kin)'
```



- Discuss 1-2 key insights, when would you use different specifications? Consider the specific context and the data available for the country you selected. (max 250 words)

This analysis uses data from Sudan applying a period analysis (rather than cohort) using data 1950-2023. We compared time variant and time invariant models, which apply female fertility. Males are added to the analysis by either i) applying factors which directly estimate male family members based on female family members; this assumes fertility and mortality are the same for males and females, and ii) applying the androgynous approach, where we assume fertility rates are the same for males and females, but we take the different mortality rates by sex.

For the variant vs invariant models, we were interested to see how this might affect the expected number of living grandparents a new-born would experience. For this we used the androgenous model so that we could consider grandmothers and grandfathers. We found that applying a time-invariant model results in more grandparents. This is because the time-invariant model applies 2023 mortality levels, which are lower compared to the previous years that are used by the time-varying model. The detailed results are summarised below:

For 2 sex model, without time varying At age 0 (when born) focal has: 1.014 great grandfathers 1.3 great grandmothers 1.73 grand mothers 1.59 grand fathers

For 2 sex model, time varying: 1.01 great grand fathers 1.28 great grand mothers 1.67 grand mothers 1.53 grand fathers

Regarding the androgenous vs factor approach, we assumed that the androgenous method would result in fewer kin members. This is because male mortality is higher than female, so accounting for this would result in fewer kin. However, we found the opposite to be true. In particular, applying the androgenous approach resulted in greater numbers of grandchildren for focal, after reaching ages 75+. We speculated this may be

due to Sudan's relatively high maternal mortality rates, so adjusting for the male mortality would actually lower the overall death rate and increase numbers of kin. However, looking at the data we didn't see strong evidence of this, with male mortality generally higher than female mortality. This would be interesting to investigate further!

3. Can you think of other ways of incorporating male fertility into the kinship models (beyond the options we discussed in the course)? (max 250 words)

Incorporating male fertility is tricky. Males tend to have children a little bit later than females, and have a wider period of child-bearing years that extend further into older ages. Fewer studies have looked at the distribution of male fertility, making it difficult to model.

For the case of Sudan, we could find a study or survey of male fertility in a similar country (e.g. sub-Saharan Africa with similar population composition). Such surveys have been implemented for instance from Demographic Health Surveys (DHS). We could use these results to model a distribution of male fertility (i.e. modelling the age-specific fertility rates for men). This would allow us to apply a kinship analysis along the male line.