ADA: Demographic Methods

Monica Alexander

Week 1: Introduction

Overview

- Course structure and goals
- What is demography?
- Some important introductory demographic concepts
- Getting started with R / git

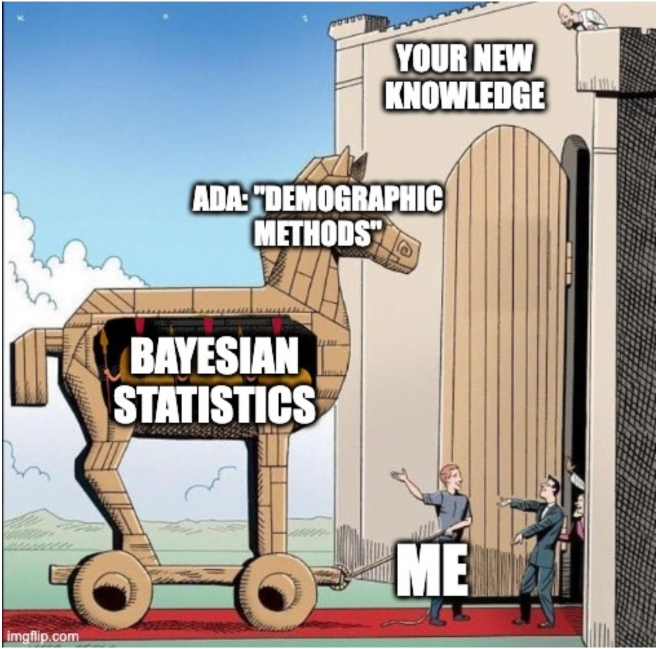
Palaver

Course structure and goals

- Introduce you to demographic methods
  - focus is on methods rather than substantive areas
- Get comfortable with fitting these methods in R
- Learn about different sorts of datasets
- Illustrate real-world issues that you can help solve

Course materials: Quercus, but mostly:
https://github.com/MJAlexander/soc6708

Demography and also



Expectations

- Most readings are recommended, not compulsory
- The starred reading will be discussed in a bit more detail
- Project should be something of interest to you (use your own data, etc)
  - improve data manipulation and graphing in R
  - research reproducibility (Quarto)
  - apply something demographic-y that we've covered in the course
  - paper, and presentation in week 12

Roadmap

1. Intro
2. Mortality
3. Mortality
4. Fertility
5. Population projections
6. Migration
7. Kinship
8. Bayes
9. Bayes regression
10. Bayesian demography
11. The future
12. Project presentations

What is demography?

What is demography?

Demography is the scientific study of population dynamics. We are interested the size, composition and distribution of populations over time, and study these changes with respect to the three main population processes:

- Births (Fertility)
- Deaths (Mortality)
- Migration

Note that these fundamental aspects of population change lead explicitly to other important compositional and distribution components of population, e.g. kin structures.

What is demography?

Demography links the **individual** to the **aggregate**.

- Individual level behavior affected by prevailing social, cultural, environmental, economic conditions
- When aggregating individuals to populations (where a ‘population’ can be any collection of interest), we see clear patterns
- These patterns in turn tell us something about the likely demographic conditions faced by an individual in a population (e.g. life expectancy, fertility rates)

What is distinct about demography?

- focuses on how individual behaviors affect aggregate structures
- obsessed with measurement
- compositional effects

Different types of demography

- Economic demography
  - economic consequences of demographic change
- Family / Social demography
  - social / cultural effects on demographic change, how demography affects family and kinship structures
- Biodemography
  - biological mechanisms for ageing, genetics
- Mathematical demography
  - formal demographic relationships and models
- Statistical demography
  - statistical models for demographic processes, agent-based studies

Why is demography important: developed countries

CANADA

Canada’s 85-and-over population set to triple over next 40 years: StatCan

By Staff • The Canadian Press

Posted June 24, 2024 5:51 pm · 4 min read

Why is demography important: developing countries

The Economist explains

Why nobody knows how many Nigerians there are

No census has yet arrived at an accurate figure

BILL GATES AND ALIKO DANGOTE SUPPORT POLIO ERADICATION EFFORTS IN NIGERIA

Gates and Dangote emphasized the need to eradicate polio, strengthen routine immunization, and improve primary health care.

Demographic methods

Useful to model demographic processes, for:

- **Understanding**
- **Generalization**
- **Projection**

Demographic methods

Demographic methods fall into one of two categories:

- **Mathematical models** based on simplifying assumptions, distributional assumptions
  - Can get closed-form estimates of demographic quantities, what is likely to happen in future
- Models based on **empirical regularities**, from observing patterns in real data
  - Can make assumptions and generalizations about demographic processes

We will look at both sorts of models in this course.

Fundamental demographic concepts

The balancing equation of population change

- tracking population size (P) over time (t)
- enter a population: births (B), in-migration (I)
- exit a population: deaths (D), out-migration (O)

So we have the balancing equation, or demographic identity:

$$P(t+1) = P(t) + B[t,t+1] - D[t,t+1] + I[t,t+1] - O[t,t+1]$$

By definition, this must always hold. But issues:

- Data may not exist
- Data come from different sources
- Measurement, coverage error

Demographic rates

- Useful to standardize size of flows (births, deaths, migrants) based on the size of the population that is producing the flows
- Compare to 'population at risk'
- Exposure has two features:
  - number of people in the population
  - length of time they were exposed to be counted
  - = 'Person Years'

$$Rate = \frac{\text{Number of events}}{\text{Person-years of exposure to risk of event}}$$

Crude rates

The most simple demographic measures.

Crude birth rate:

$$CBR[t, t{+}1] = \frac{\text{Number of births in population from time } t \text{ to } t+1}{\text{Person-years lived in population from time } t \text{ to } t+1}$$

Crude death rate:

$$CDR[t, t{+}1] = \frac{\text{Number of deaths in population from time } t \text{ to } t+1}{\text{Person-years lived in population from time } t \text{ to } t+1}$$

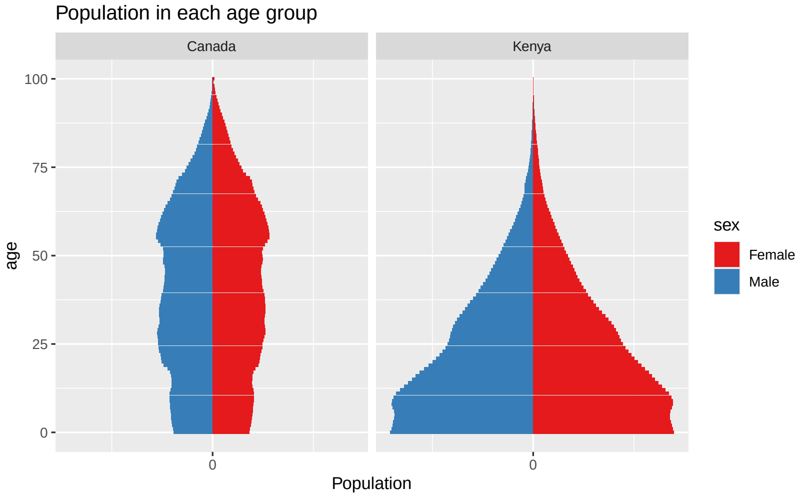
Easy to calculate: just need total counts! (don't need any info on age, sex etc)

Crude rates

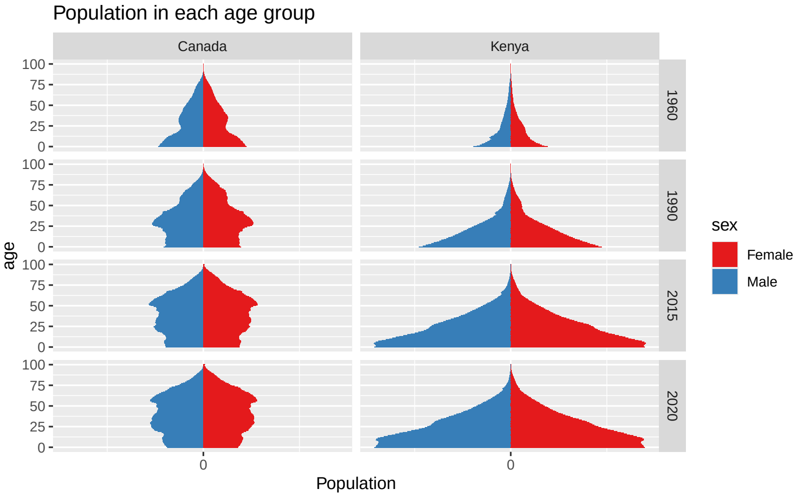
- CDR in Canada: 7.9 per 1,000 people (trending up)
- CDR in Kenya: 7.2 per 1,000 people (trending down)

Which country has worse mortality conditions?

Population structures



Coffins and pyramids



Age-specific rates

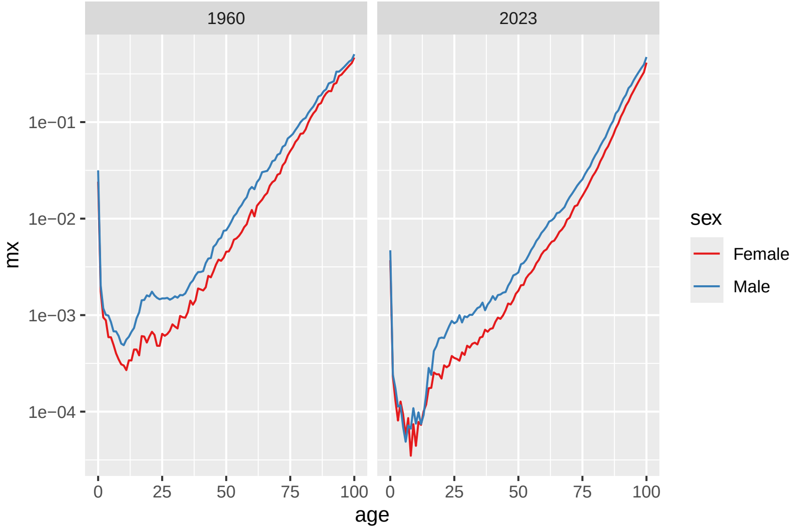
Same deal but by age:

Age-specific fertility rate for people aged x to $x+n$ (for a particular time period):

$$ASFR[x, x+n] = \frac{\text{Births to people aged } x \text{ to } x+n \text{ to people aged}}{\text{Person-years lived for population aged } x \text{ to } x+n}$$

etc

Characteristic shapes of age-specific mortality rates



Standardization

- Given age-specific rates and the population structure, there are two things that can affect crude rates.
- We are (usually) interested in the effect of the outcome, not the effect of population age structure
- Pick a population to 'standardize' the population age structure. This has populations $P[x, x+n]$

Then

$$\text{Age-standardized rate} = \sum_{\text{all ages}} \text{rate}[x, x+n] \cdot \frac{P[x, x+n]}{\sum P[x, x+n]}$$

Standardization

For example,

$$\text{Age-standardized mortality rate} = \frac{\sum_{\text{all ages}} \text{ASMR}[x, x+n] \cdot P[x, x+n]}{\sum P[x, x+n]}$$

- CDR in Canada: 7.9 per 1,000 people
- CDR in Kenya: 7.2 per 1,000 people

Using Canada's population:

- Age-standardized mortality rate in Canada: 7.9 per 1,000 people
- Age-standardized mortality rate in Kenya: 22 per 1,000 people

Age-standardized rates

$$\text{Age-standardized mortality rate} = \frac{\sum_{\text{all ages}} \text{ASMR}[x, x+n] \cdot P[x, x+n]}{\sum P[x, x+n]}$$

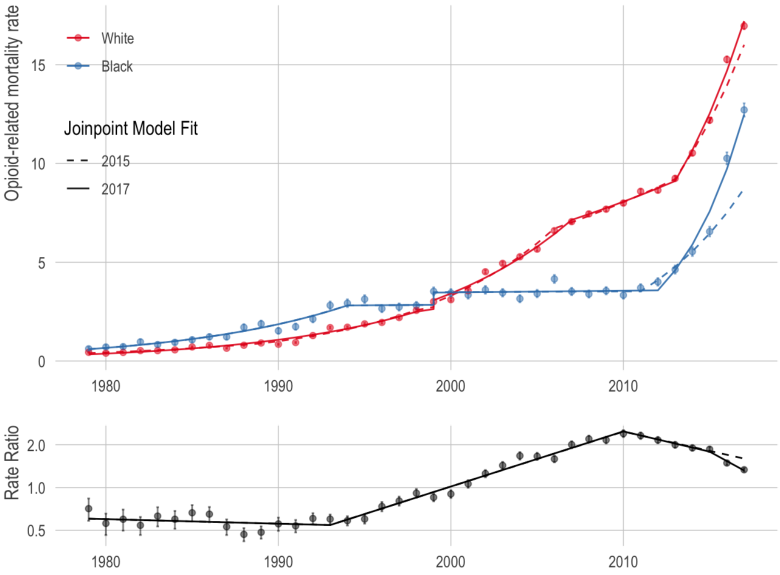
Side note I: if we moved into continuous form, call mortality rates at age x $m(x)$ and population weights $w(x)$, then

$$\text{Age-standardized mortality rate} = \frac{\int_0^{\omega} m(x) w(x) dx}{\int_0^{\omega} w(x) dx}$$

where ω is the maximum age. We will see this weighted average form a lot.

Side note II: we will switch between discrete and continuous a fair bit

Standardized mortality rates

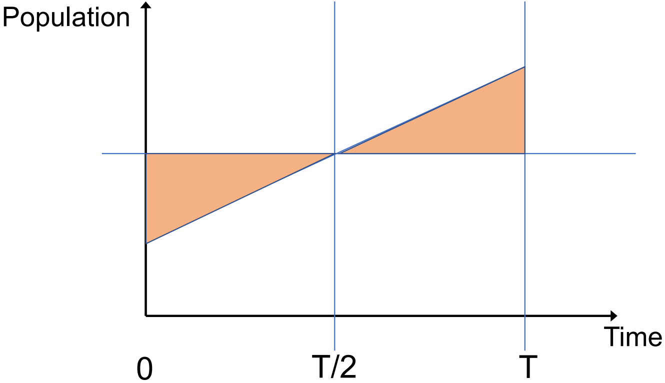


Standardization

- Standardizing age structure is called **direct standardization**
- **Indirect standardization** also exists, more common in epidemiology
  - don't know age-specific rates for a population
  - calculate 'expected deaths' based on another population's mortality
  - look at standardized mortality ratio (observed / expected deaths)

Approximation to person-years

- So far we been using person-years (PY) on the denominator
- But actually almost never have this information
- Usually just have population estimates at one point in time
- Common to use mid-year population times the period length as an **estimate** of PY



Population growth

Crude growth rate

Divide the balancing equation by population years:

$$
\begin{aligned}
\frac{P(t+1) - P(t)}{PY(t)} &= \frac{B(t) - D(t) + I(t) - O(t)}{PY(t)} \\
CGR(t) &= CBR(t) - CDR(t) + CRI(t) - CRO(t) \\
CGR(t) &= CRNI(t) + CRNM(t)
\end{aligned}
$$

Growth rate is natural increase + net migration.

Geometric growth

Let's ignore migration for now.

$$\begin{aligned} P(t+1) &= P(t) + B(t) - D(t) \\ &= P(t)\left(1 + \frac{B(t)}{P(t)} - \frac{D(t)}{P(t)}\right) \end{aligned}$$

So

$$\begin{aligned} P(1) &= P(0)\left(1 + \frac{B(0)}{P(0)} - \frac{D(0)}{P(0)}\right) \\ P(2) &= P(1)\left(1 + \frac{B(1)}{P(1)} - \frac{D(1)}{P(1)}\right) \\ &= P(0)\left(1 + \frac{B(0)}{P(0)} - \frac{D(0)}{P(0)}\right)\left(1 + \frac{B(1)}{P(1)} - \frac{D(1)}{P(1)}\right) \end{aligned}$$

etc

This is called **geometric growth**

Constant growth rate over time

If the birth and death rates (i.e. $b = \frac{B}{P}$ and $d = \frac{D}{P}$) are not changing over time

$$\begin{aligned} P(1) &= P(0)\,(1 + b - d) \\ P(2) &= P(1)\,(1 + b - d) \\ &= P(0)\,(1 + b - d)\,(1 + b - d) \\ &= P(0)\,(1 + b - d)^2 \end{aligned}$$

In general, $P(t) = A^t P(0)$ where $A = (1 + b - d)$.

Instantaneous growth rate

Consider the growth rate $r(t)$ between two times points that are very close together, Δt, and then look at the limit.

$$r(t) = \lim_{\Delta t \to 0} \frac{\Delta P(t)}{P(t)\Delta t} = \frac{\frac{dP(t)}{dt}}{P(t)} = \frac{d\ln(P(t))}{dt}$$

Taking integrals and exponents and doing a bit of rearranging we have the population at time T

$$P(T) = P(0)e^{\int_0^T r(t)dt}$$

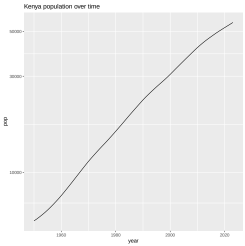
Constant growth rates

If r is constant over time then

$$P(T) = P(0)e^{Tr}$$

And so assuming constant growth and given two time points, we can calculate the implied growth rate as

$$r = \frac{log(P(t_2)) - log(P(t_1))}{t_2 - t_1}$$



Age, periods cohorts

Three dimensions of demographic time

We can express an individual's relative position in time based on three different dimensions:

- their age
- the period (year) we are in
- their cohort (e.g. birth cohort)

Demographers, sociologists, epidemiologists often interested in one, some or all of these effects on outcomes

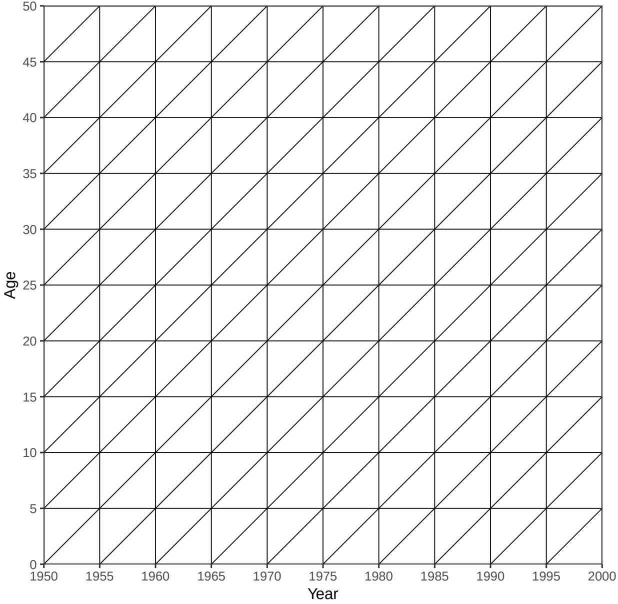
Age effect

SPADINA
1
2
System Map
ROVINCES OF CANADA
Newfoundland
HUDSON'S BAY
ONTARIO
Nova Scotia
New Brunswick
CANADA

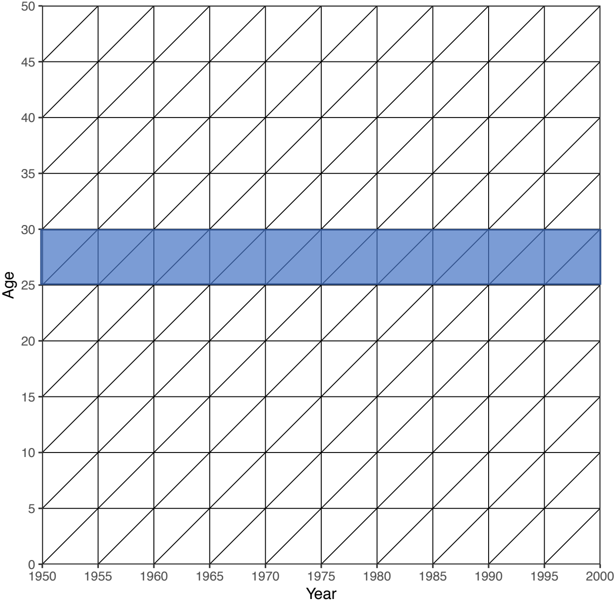
Period effect

Cohort effect

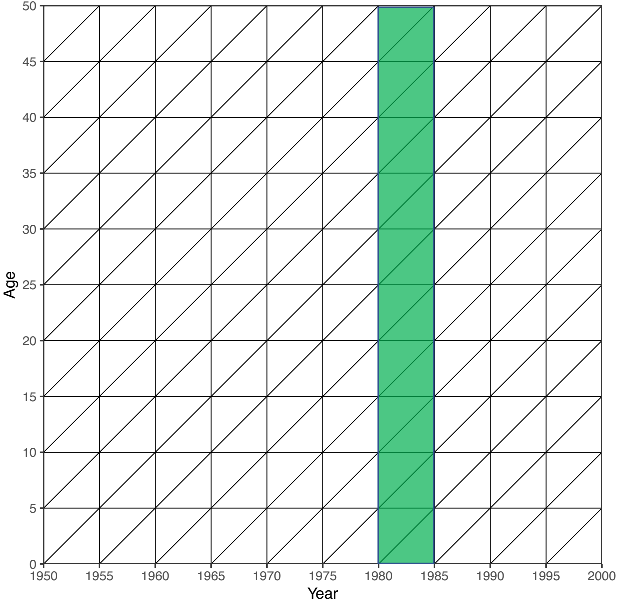
Lexis diagrams



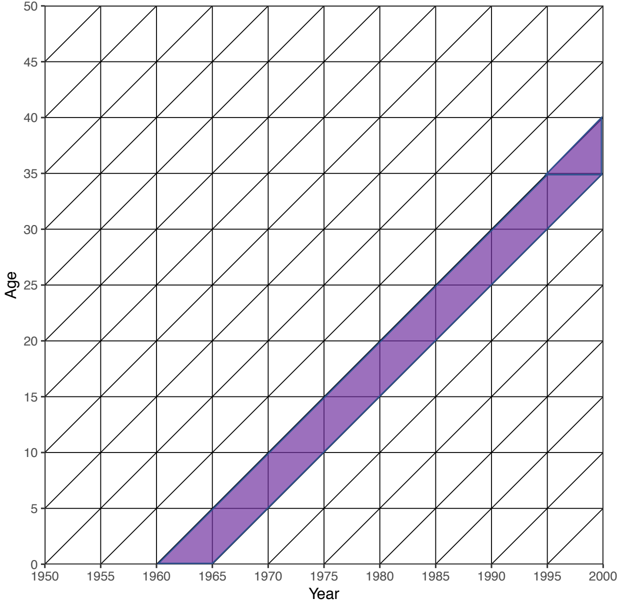
Lexis diagrams



Lexis diagrams



Lexis diagrams



Measurement issues

May want to estimate age, period and cohort effects on outcome of interest (e.g. smoking)

$$Y_i = \alpha_i + \beta_i + \gamma_i$$

But

$$Age = Period - Cohort$$

etc

so regression matrix is singular (not identifiable)

Denominators in the wild

Why do we usually care about rates?

- Usually interested in getting an idea of prevalence / incidence of an event relative to exposure to risk
- Absolute counts of an event (e.g. deaths) not usually meaningful without knowing how many people are at risk of that event
- Person-years captures the population level magnitude and time exposed

But denominators sometimes not that clear cut!

Maternal and pregnancy-associated mortality

- Usually when maternal mortality (pregnancy-related mortality up to 42 days after end of pregnancy) is reported it is in the form of a maternal mortality **ratio**: number of maternal deaths per 100,000 live births.
- Logic: number of births in a population is a good proxy for exposure to risk of maternal mortality
- Note: this is the same for pregnancy-associated mortality (deaths up to one year after end of pregancy)

But what happens when we want to compare pregnant versus non-pregnant risks?

Obstet Gynecol. 2021 Nov 1;138(5):762-769. doi: 10.1097/AOG.0000000000004567.

Homicide During Pregnancy and the Postpartum Period in the United States, 2018–2019

Maeve Wallace [1], Veronica Gillispie-Bell, Kiara Cruz, Kelly Davis, Dovile Vilda

Abstract

Objective: To estimate the national pregnancy-associated homicide mortality ratio, characterize pregnancy-associated homicide victims, and compare the risk of homicide in the perinatal period (pregnancy and up to 1 year postpartum) with risk among nonpregnant, nonpostpartum females aged 10-44 years.

Methods: Data from the National Center for Health Statistics 2018 and 2019 mortality files were used to identify all female decedents aged 10-44 in the United States. These data were used to estimate 2-year pregnancy-associated homicide mortality ratios (deaths/100,000 live births) for comparison with homicide mortality among nonpregnant, nonpostpartum females (deaths/100,000 population) and to mortality ratios for direct maternal causes of death. We compared characteristics and estimated homicide mortality rate ratios and 95% CIs between pregnant or postpartum and nonpregnant, nonpostpartum victims for the total population and with stratification by race and ethnicity and age.

Results: There were 3.62 homicides per 100,000 live births among females who were pregnant or within 1 year postpartum, 16% higher than homicide prevalence among nonpregnant and nonpostpartum females of reproductive age (3.12 deaths/100,000 population, $P<.05$). Homicide during pregnancy or within 42 days of the end of pregnancy exceeded all the leading causes of maternal mortality by more than twofold. Pregnancy was associated with a significantly elevated homicide risk in the Black population and among girls and younger women (age 10-24 years) across racial and ethnic subgroups.

Conclusion: Homicide is a leading cause of death during pregnancy and the postpartum period in the United States. Pregnancy and the postpartum period are times of elevated risk for homicide among all females of reproductive age.

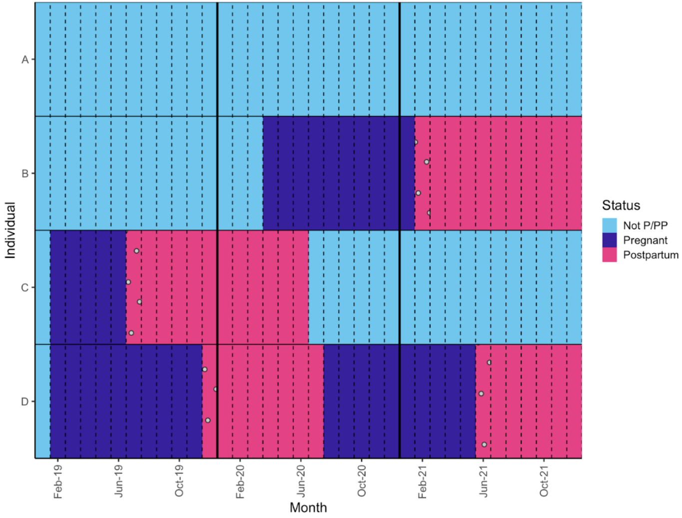
The denominators are different!!!

- And things are complicated by the fact that pregnancy-associated mortality can span a risk period of almost two years
- So just looking at population for non-pregnant population over-estimates the denominator in this group

Ideally, would do full accounting in person-time framework

- Can get pregnant/postpartum population from date of birth, -gestational age, + 1 year, as well as data on fetal deaths
- non-pregnany/nonpostpartum population is then total population-month minus the calculation above

Figure 1. Visual representation of births and relationship to pregnancy and postpartum exposure time for 4 individuals, shown across 36 months of observation spanning the years 2019-2021.



Stay tuned

Relative Risk of Homicide in Pregnant and Postpartum versus Nonpregnant and Nonpostpartum Populations: Re-estimation using a Person-time Framework

Neal Marquez[1,2], Claire R. McNellan[3], Monica Alexander[4]

[1]Population Research Center, Portland State University, Portland, Oregon

[2]Colorado State Demography Office, Colorado Department of Local Affairs, Denver, Colorado

[3]Department of Health Systems, Management & Policy, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, Colorado

[4]Department of Statistical Sciences, University of Toronto, Toronto, Ontario

[5]Department of Sociology, University of Toronto, Toronto, Ontario

Lab

Lab

- Intro to git (set up GitHub account, make repository, clone repository, push work)
- Work through quarto doc