Week 3 Methods: Disease

ECON 125: The Science of Population

Setup

Today, we data on the HIV epidemic in South Africa. For each year, we have:

- Midyear number of people living with HIV (0-14 and 15+)
- Number of new HIV infections (0-14 and 15+)
- Number of new AIDS deaths (0-14 and 15+)
- Midyear population (0-14 and 15+)

Start by setting up R and loading the dataset

```
# Load tidyverse and clear the R environment
library(tidyverse)
rm(list=ls())

# Load dataset and assign it the name country_year_df
df <- read_csv(url("https://github.com/tomvogl/econ125/raw/main/data/UN
```

Variables

Here are the variables in the dataset

```
names(df)
```

```
## [1] "year"                "aids_deaths_0_14"   "aids_deaths_15plus"
## [4] "hiv_living_0_14"     "hiv_living_15plus"  "hiv_new_0_14"
## [7] "hiv_new_15plus"      "pop_0_14"           "pop_15plus"
```

Prevalence

Recall that:

$$\text{Prevalence} = \frac{\text{People with disease}}{\text{Population}}$$

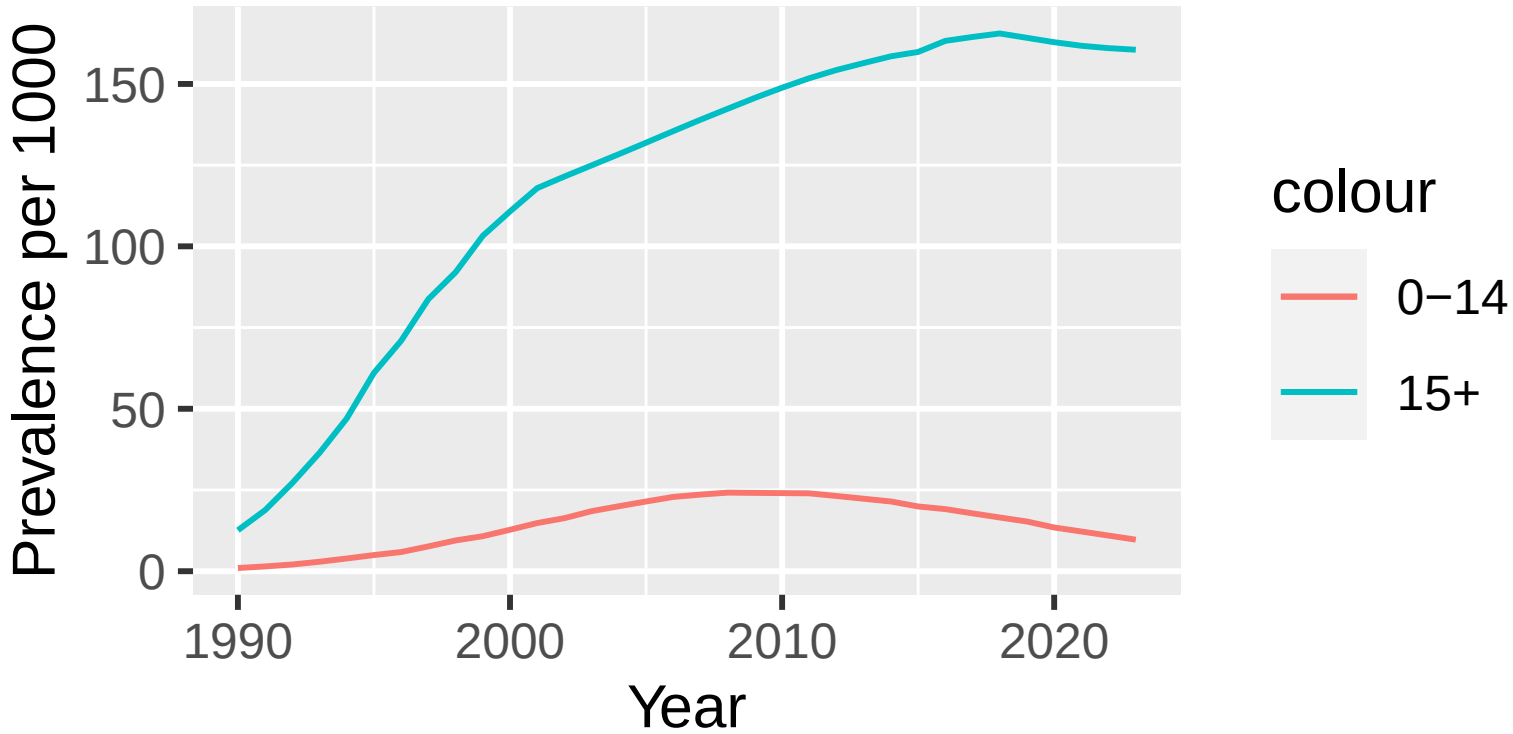
We will measure it per 1000 population

Let's implement it in R

```
df <-
  df %>%
  mutate(prev_0_14 = 1000*hiv_living_0_14/pop_0_14,
         prev_15plus = 1000*hiv_living_15plus/pop_15plus)
```

Prevalence over time

```
ggplot(df, aes(x=year)) +
  geom_line(aes(y=prev_0_14, color = "0-14")) +
  geom_line(aes(y=prev_15plus, color = "15+")) +
  scale_y_continuous("Prevalence per 1000") +
  scale_x_continuous("Year")
```



Adult prevalence rose rapidly in 1990s, slowed in 2000s as behavior changed and then treatment became widespread, peaked in 2010s

Child prevalence lower, peaked in 2010s: ART blocks mom-to-child transmission

Incidence

Recall that:

$$\text{Incidence} = \frac{\text{New cases}}{\text{Person-time at risk}}$$

Assuming infections happen halfway through the year on average, we can use:

$$\begin{aligned}\text{Incidence} &= \frac{\text{New cases}}{\text{Midyear uninfected population}} \\ &= \frac{\text{New cases}}{\text{Midyear population} - \frac{1}{2}\text{New cases}}\end{aligned}$$

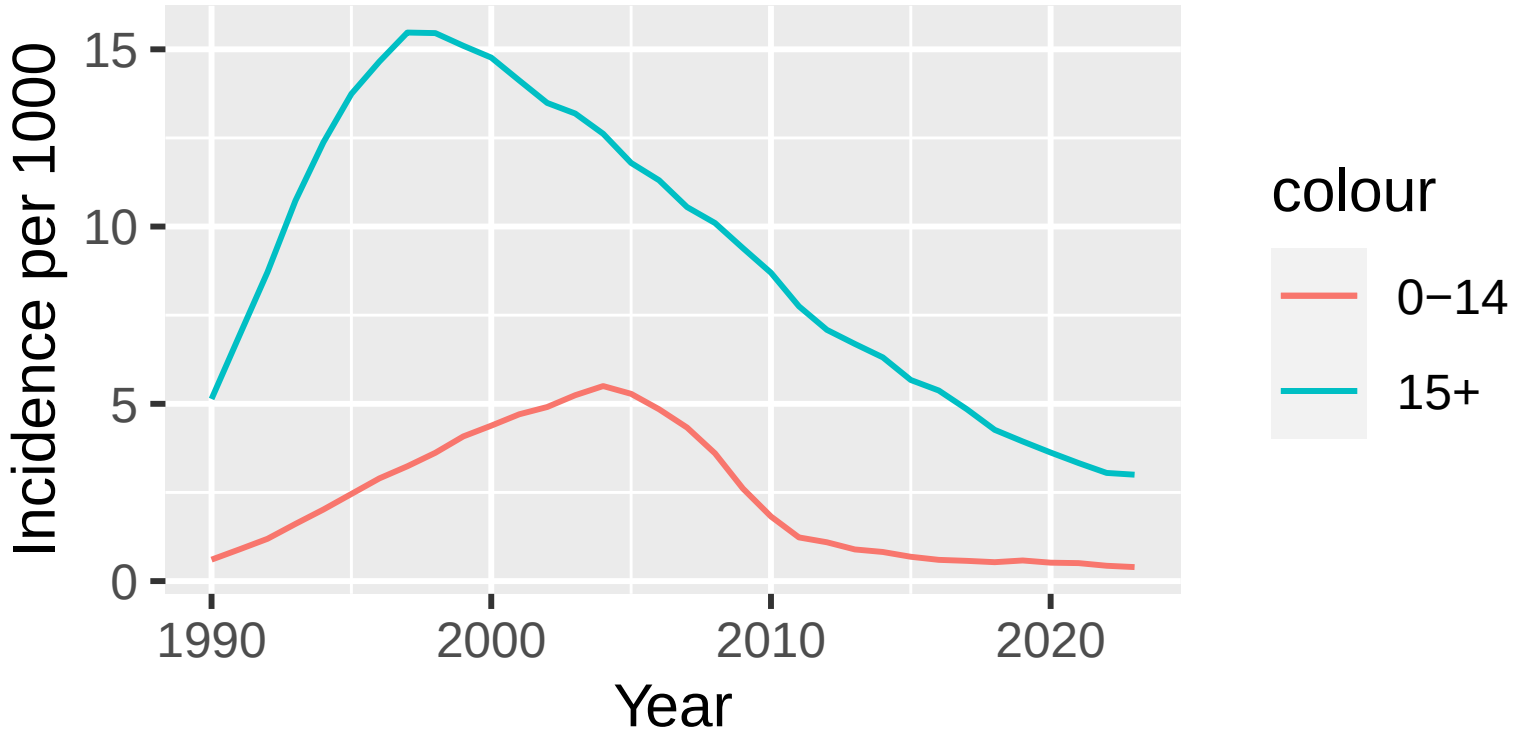
We will again measure it per 1000 population

Let's implement it in R

```
df <-
  df %>%
  mutate(incid_0_14 = 1000*hiv_new_0_14/
                      (pop_0_14-0.5*hiv_new_0_14),
         incid_15plus = 1000*hiv_new_15plus/
                        (pop_15plus-0.5*hiv_new_15plus))
```

Incidence over time

```
ggplot(df, aes(x=year)) +
  geom_line(aes(y=incid_0_14, color = "0-14")) +
  geom_line(aes(y=incid_15plus, color = "15+")) +
  scale_y_continuous("Incidence per 1000") +
  scale_x_continuous("Year")
```



Among adults, behavior change reduced flow of new infections in late 90s

Among kids, incidence kept rising from mom-to-child transmission until ART

Prevalence vs Incidence

Recall from the lecture note that if incidence is low and stable, then:

$$\text{Prevalence} \approx \text{Incidence} \times \text{Average duration}$$

HIV is a lifelong chronic disease, so average duration is "forever"

If incidence is low, then:

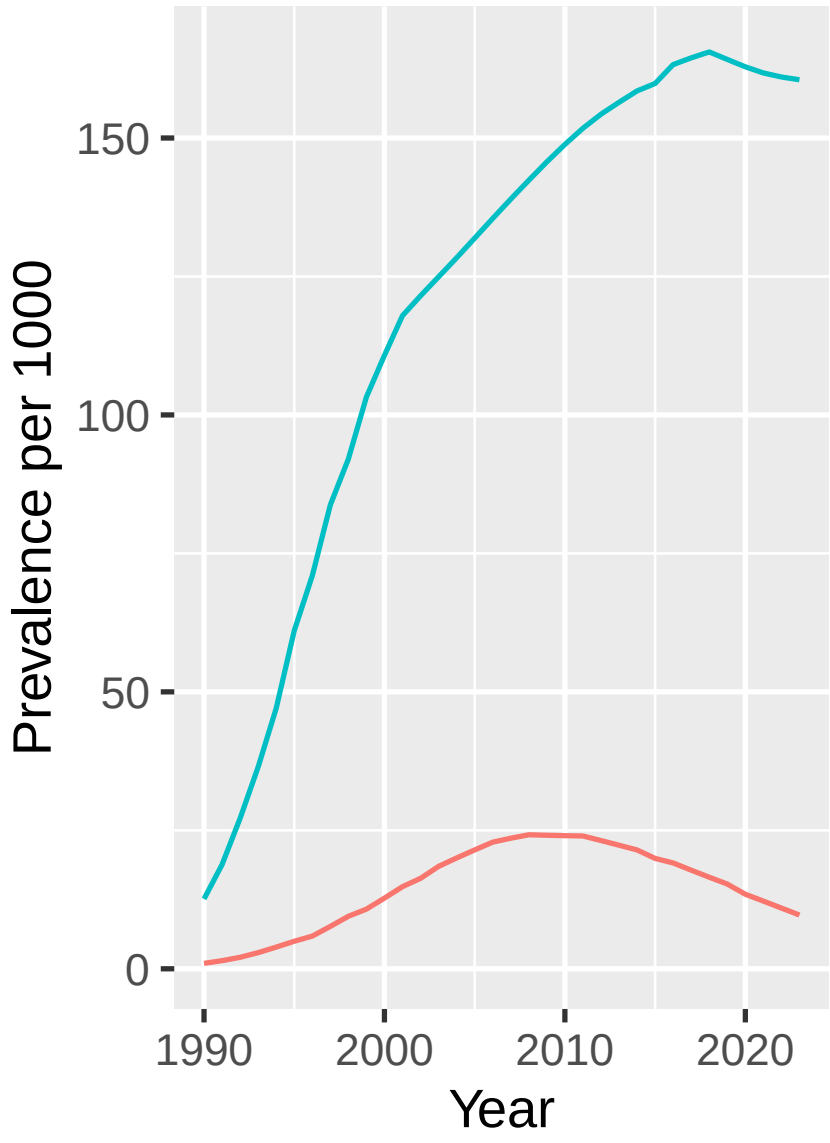
$$\text{Prevalence} \approx \text{Cumulative incidence} = \text{Area under incidence curve}$$

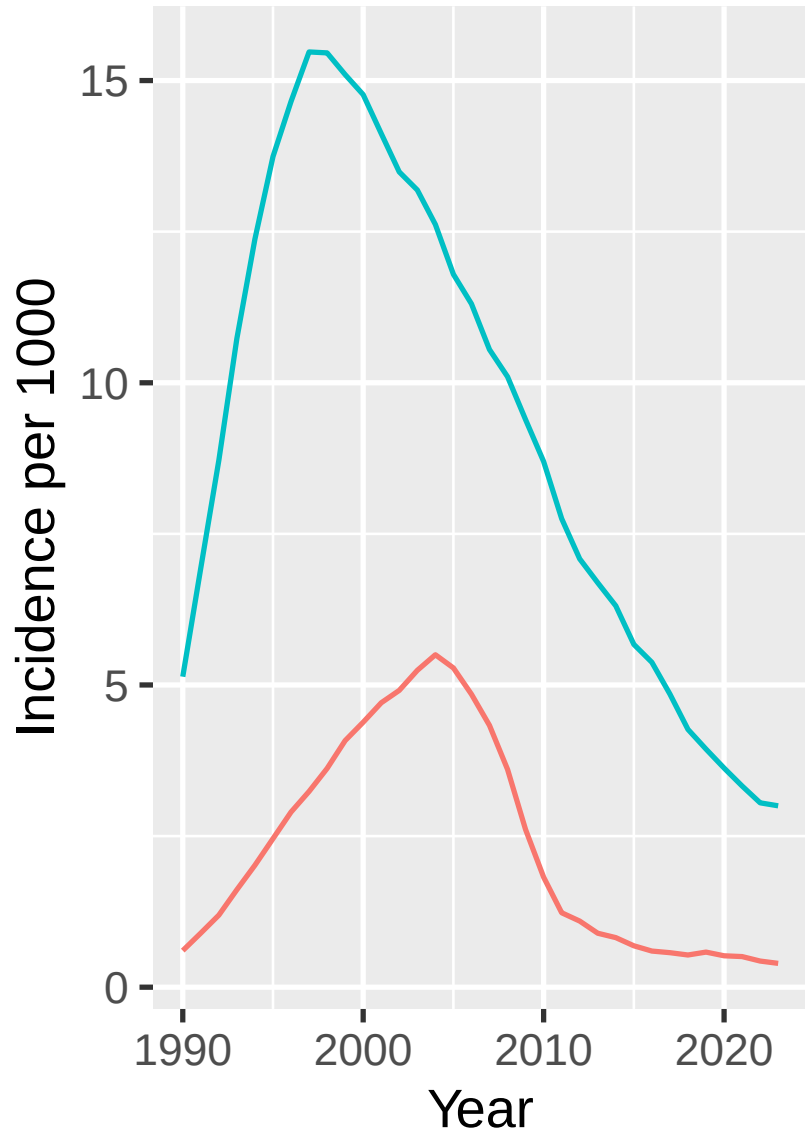
Approximation only goes so far:

- HIV incidence is high in South Africa
- AIDS kills people, and these formulas don't account for mortality

But still useful for understanding the link between the two graphs

Prevalence vs Incidence





Cause-specific mortality

Now let's come back to mortality

Recall that:

$$\text{Cause-specific mortality rate} = \frac{\text{Deaths from AIDS}}{\text{Midyear population}}$$

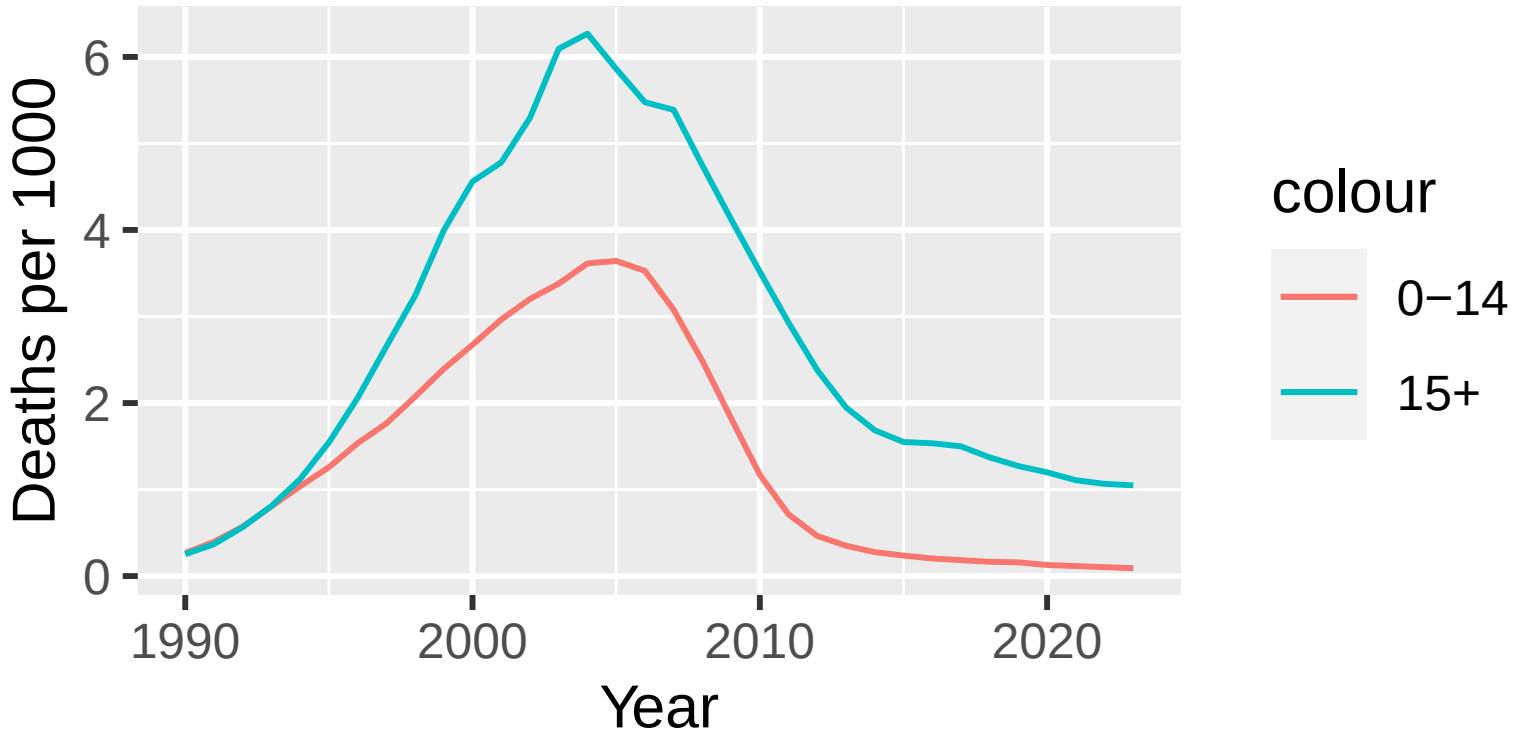
We will measure it per 1000 population

Let's implement it in R

```
df <-
  df %>%
  mutate(mort_0_14 = 1000*aids_deaths_0_14/pop_0_14,
         mort_15plus = 1000*aids_deaths_15plus/pop_15plus)
```

Cause-specific mortality from HIV over time

```
ggplot(df, aes(x=year)) +
  geom_line(aes(y=mort_0_14, color = "0-14")) +
  geom_line(aes(y=mort_15plus, color = "15+")) +
  scale_y_continuous("Deaths per 1000") +
  scale_x_continuous("Year")
```



Peak is about 6x higher than current global death rate from heart disease

How Deadly is HIV/AIDS?

The case fatality rate measures a disease's deadliness: the share of people with the disease who die from it (at **any** time)

- Easy to calculate for acute conditions (e.g., COVID, malaria)
- Less so for chronic diseases (e.g., HIV/AIDS, heart disease)
- Need to follow chronic disease sufferers over many years

We cannot calculate the CFR for HIV/AIDS from these data because we only get annual snapshots

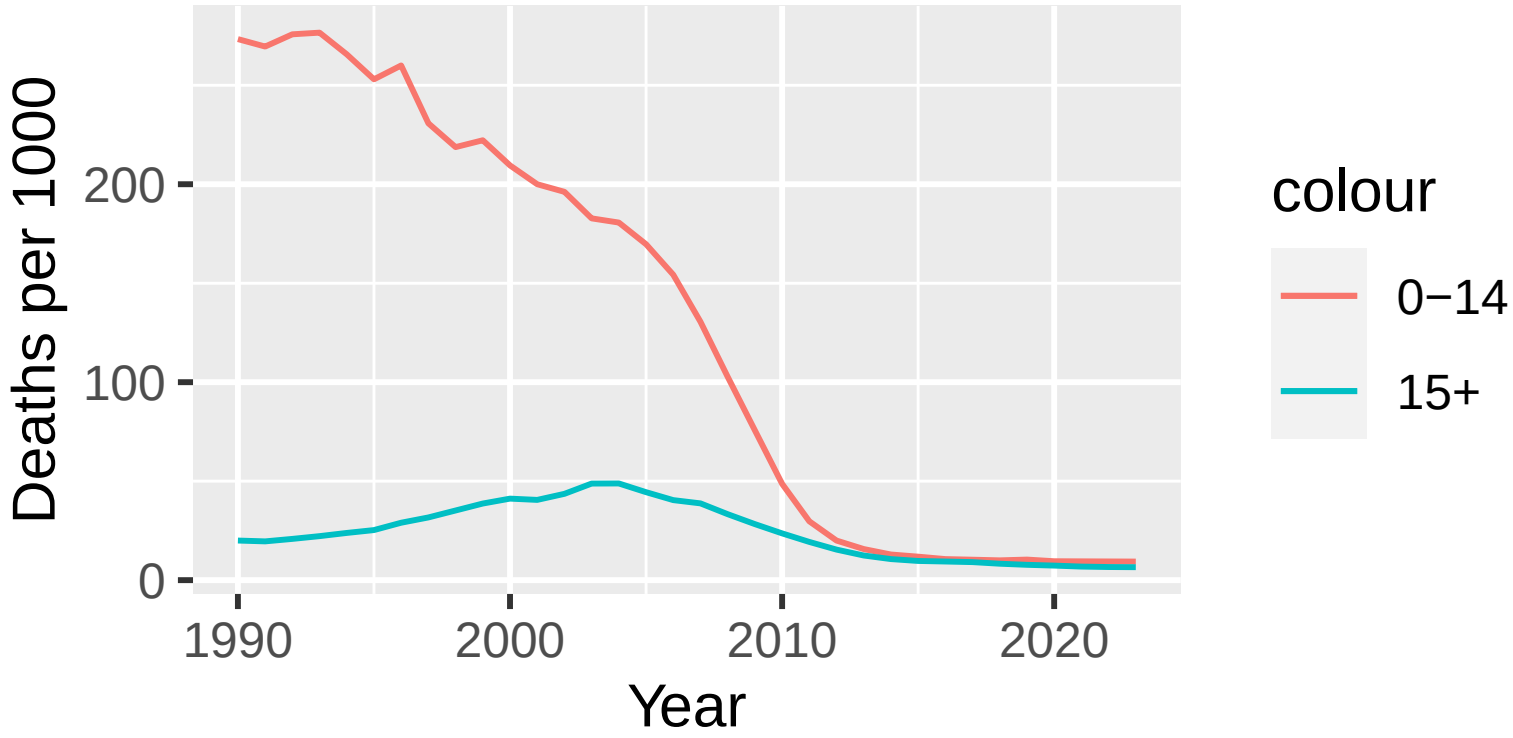
But we can still measure AIDS deaths per 1000 people living with HIV to see how much less deadly HIV/AIDS has become over time

Let's implement it in R

```
df <-
  df %>%
  mutate(ratio_0_14 = 1000*aids_deaths_0_14/hiv_living_0_14,
         ratio_15plus = 1000*aids_deaths_15plus/hiv_living_15plus)
```

AIDS Deaths per Infected over Time

```
ggplot(df, aes(x=year)) +
  geom_line(aes(y=ratio_0_14, color = "0-14")) +
  geom_line(aes(y=ratio_15plus, color = "15+")) +
  scale_y_continuous("Deaths per 1000") +
  scale_x_continuous("Year")
```



AIDS was extremely deadly for kids in the early years: ~25%