

# A national population cohort model incorporating age structure and population growth in 15 countries in R/HEEMOD

David Tordrup, Triangulate Health Ltd

[dtordrup@gmail.com](mailto:dtordrup@gmail.com)

R Seminar – London 9<sup>th</sup> July 2019



TRIANGULATE  
**HEALTH** LTD  
SCIENCE | POLICY | ECONOMICS

# The analytical issue

- Acute health-related event early in life (0-14 years of age)
- Very high prevalence in some settings (>80%)
- Immediate complications (same year)
- Follow-on complications over the life course (until old age)
- Quantifying the economic burden from the healthcare payer perspective (health care resource utilization)
- Quantifying the value of prevention, i.e. avoidable health burden from primary prevention



# Markov model in R

- Package: HEEMOD (<https://cran.r-project.org/web/packages/heemod/index.html>)
- Allows simple models to be built relatively easily
- Allows flexibility for more complex modelling scenarios
  - E.g. vector of transition probabilities to implement changes over time

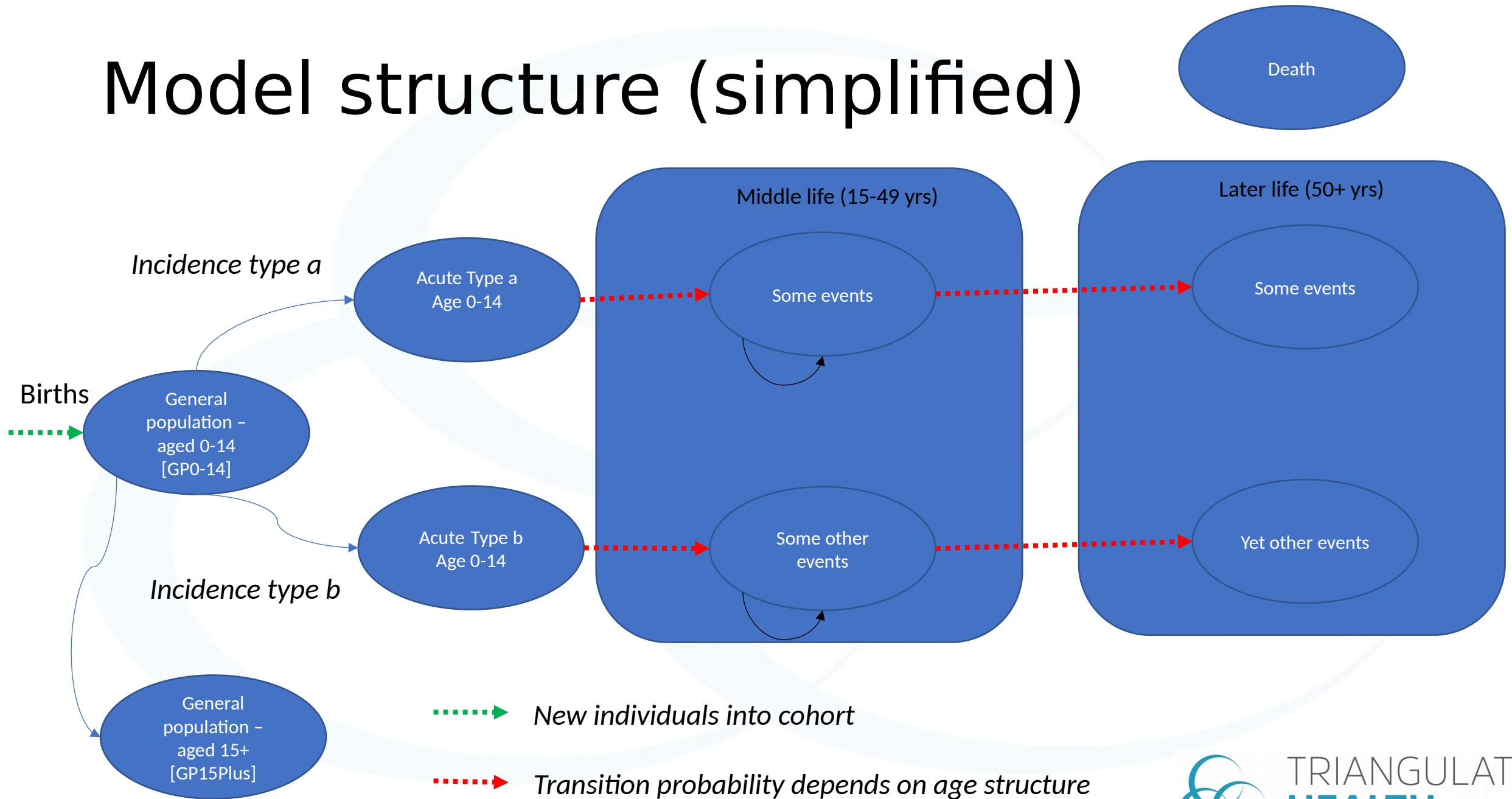


# Hypothetical vs national cohort

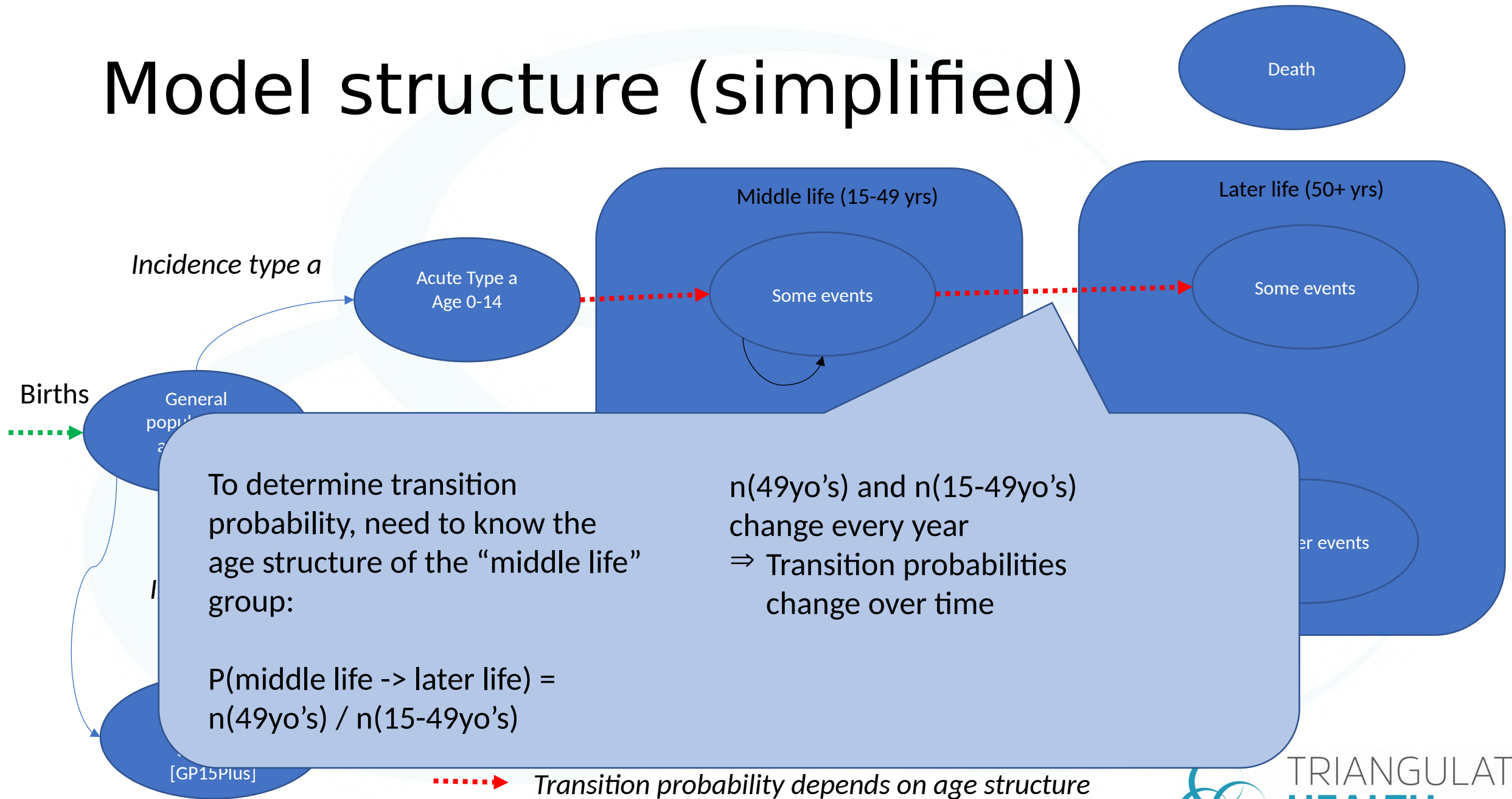
Hypothetical cohort	National cohort
Fixed size	Population growth (+births)
Average age (maybe with distribution)	Age structure of the population
Life-time horizon = life expectancy	Life-time horizon = indefinite



# Model structure (simplified)



# Model structure (simplified)



To determine transition probability, need to know the age structure of the “middle life” group:

$$P(\text{middle life} \rightarrow \text{later life}) = \frac{n(49\text{yo's})}{n(15-49\text{yo's})}$$

$n(49\text{yo's})$  and  $n(15-49\text{yo's})$  change every year  
⇒ Transition probabilities change over time

.....→ Transition probability depends on age structure

# Managing population structure

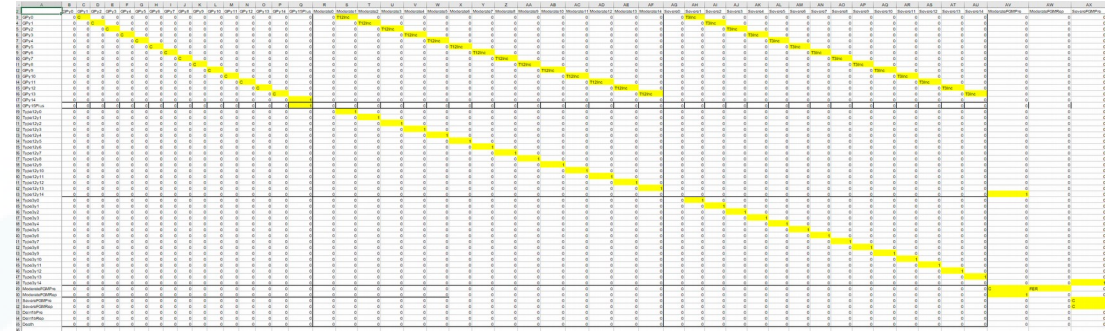
- Internally represented by a matrix with n rows (n=time horizon) and 50 columns (age 0 to 49)
- Transition probabilities calculated from partial vectors of the matrix
- Population counts calculated from
  - Initial population (5-year buckets)
  - Background mortality (flat annual rate in 0-14, 15-49 and 50+ age groups)
  - Previous year's population

Model Year	Age (number of n-year olds)							
	0	1	2	3	4	5	6	...
1	328,532	328,532	328,532	328,532	328,532	291,268	291,268	...
2	362,646	326,020	326,020	326,020	326,020	326,020	289,042	...
3	368,449	359,874	323,528	323,528	323,528	323,528	323,528	...
4	374,337	365,633	357,123	321,055	321,055	321,055	321,055	...
5	380,326	371,476	362,838	354,393	318,600	318,600	318,600	...
6	386,419	377,418	368,636	360,064	351,684	316,165	316,165	...
...	...	...	...	...	...	...	...	...
30	519,178	511,144	503,132	495,118	487,076	478,988	470,840	...



# A small comment on performance

- The age matrix is an “implicit” way of managing age structure
- I tried implementing age structure explicitly with model states in the 0-14 year age groups
- Image: Zoomed-out Excel with transition probability matrix (most are zero – “Leslie matrix”)
- Running time was unfeasibly slow on a high-spec desktop PC (4.7GHz i7/8700, 16Gb RAM, M.2 SSD)



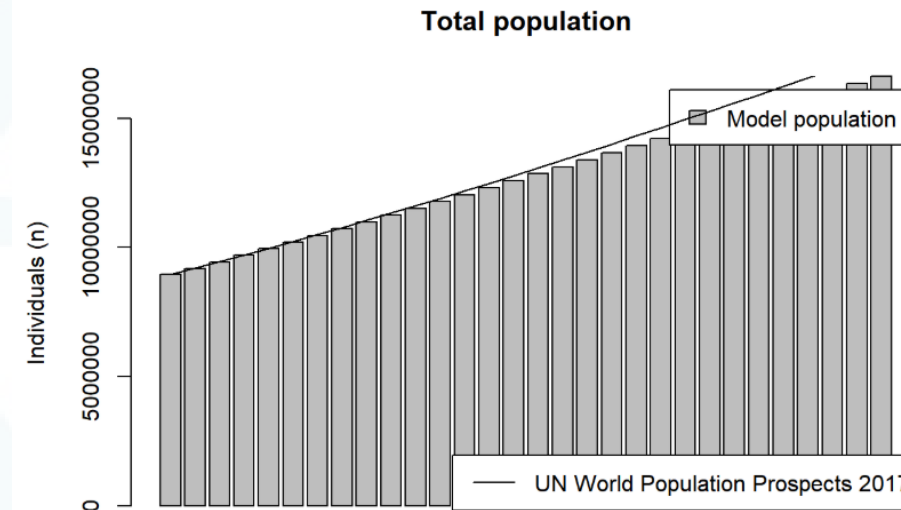


# Comparing population evolution against UN population extrapolations

- Check the validity of the population size against extrapolations made by real demographers
- As far as I know, demographic projections made in exactly the same way (“Leslie matrix”) but
  - With more granular mortality rates
  - Including migration

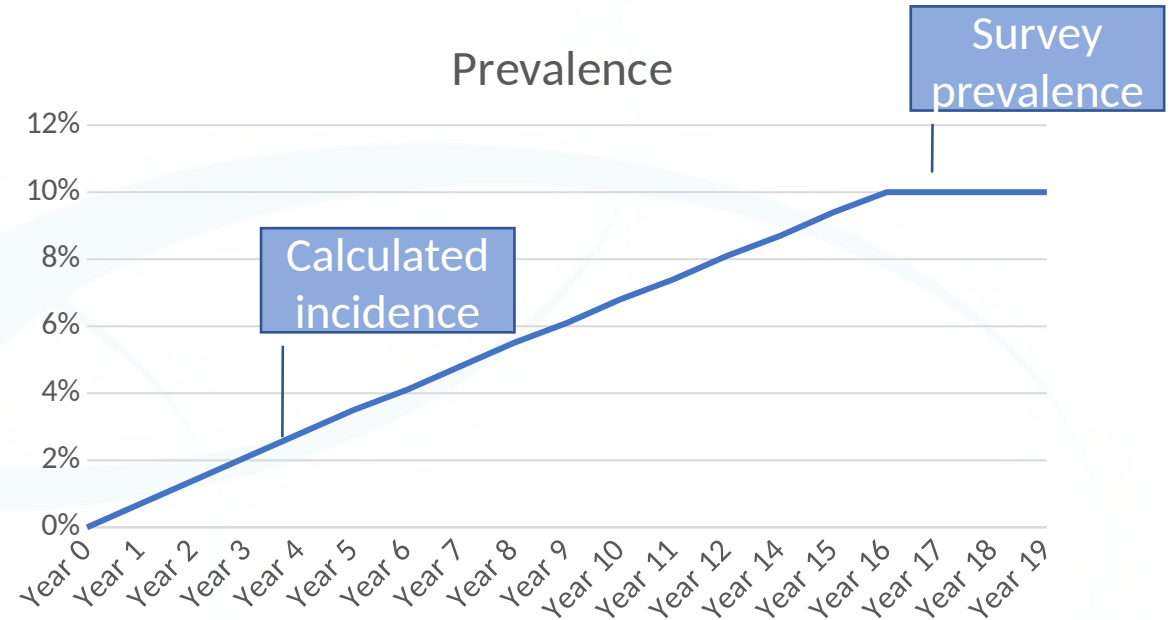
## Calculated model population

The model population is based on the UN World Population Prospects (WPP) 2017 revision. The baseline model population is taken directly from the 2018 estimate stratified by age groups [UN2017]. The model population is expanded each cycle (year) using annual births from the WPP, adjusted by the secondary sex ratio to arrive at female live births. Background mortality is modelled using the WHO Global Health Observatory (GHO) Life Table data by country [WHO2018], averaging mortality from 5-year brackets to the model age groups (0 to 14 years, 15 to 49 years, and 50 years and above).



# Estimating incidence from prevalence

- The health condition is chronic and life-long, but exposure only in year 0-14
- We know prevalence from surveys, but need incidence for model probability and to model impact of interventions
- Since the condition is life-long, we can assume equalised risk over the exposure period and simply rescale the prevalence to arrive at incidence



```
HEEMOD.Rescale_prob(p=0.1, from=15,  
to=1)
```

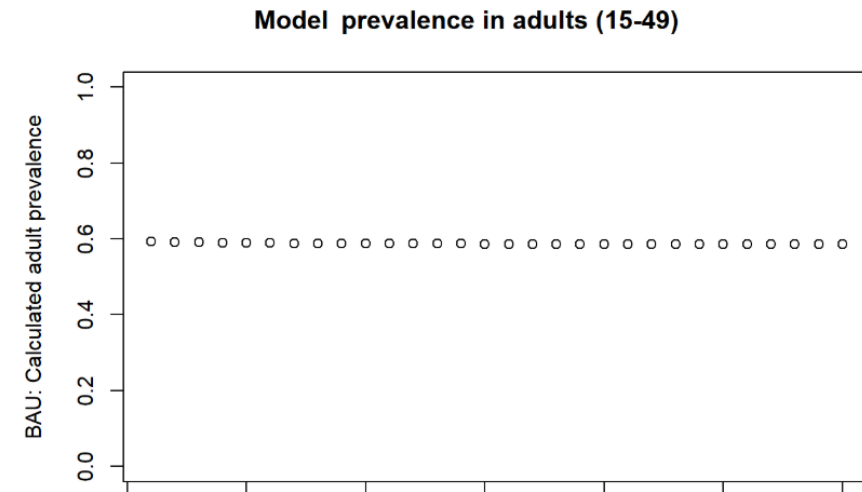


# Sense checking: Prevalence

- We expect adult prevalence to be constant over time, given the calculated incidence in 0-14 year olds

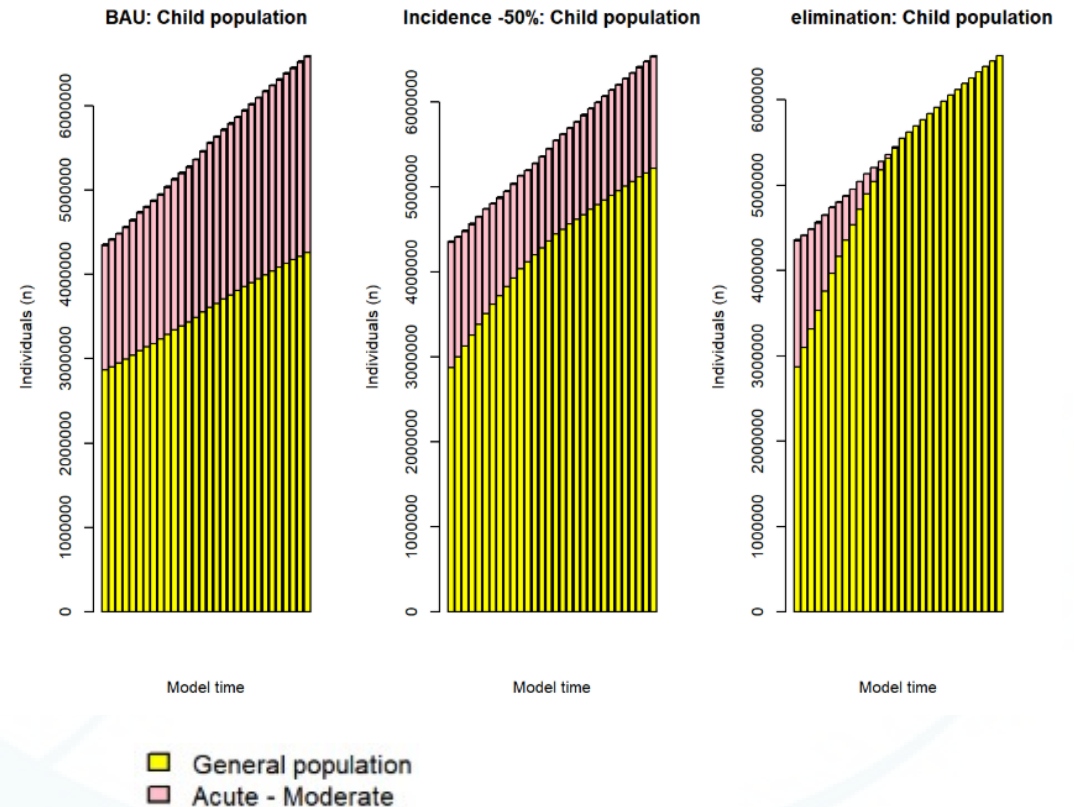
## Calculated prevalence

In the BAU scenario, we expect the model prevalence to remain relatively fixed over time, as the incidence in 0-14 year olds is calculated to result in the baseline prevalence observed among 15-49 year olds in the most recent survey data. The modelled prevalence in the adult cohort (15-49 years) over the model time horizon is shown below. Minor deviations are expected, as the incidence calculation does not take mortality rate into account.



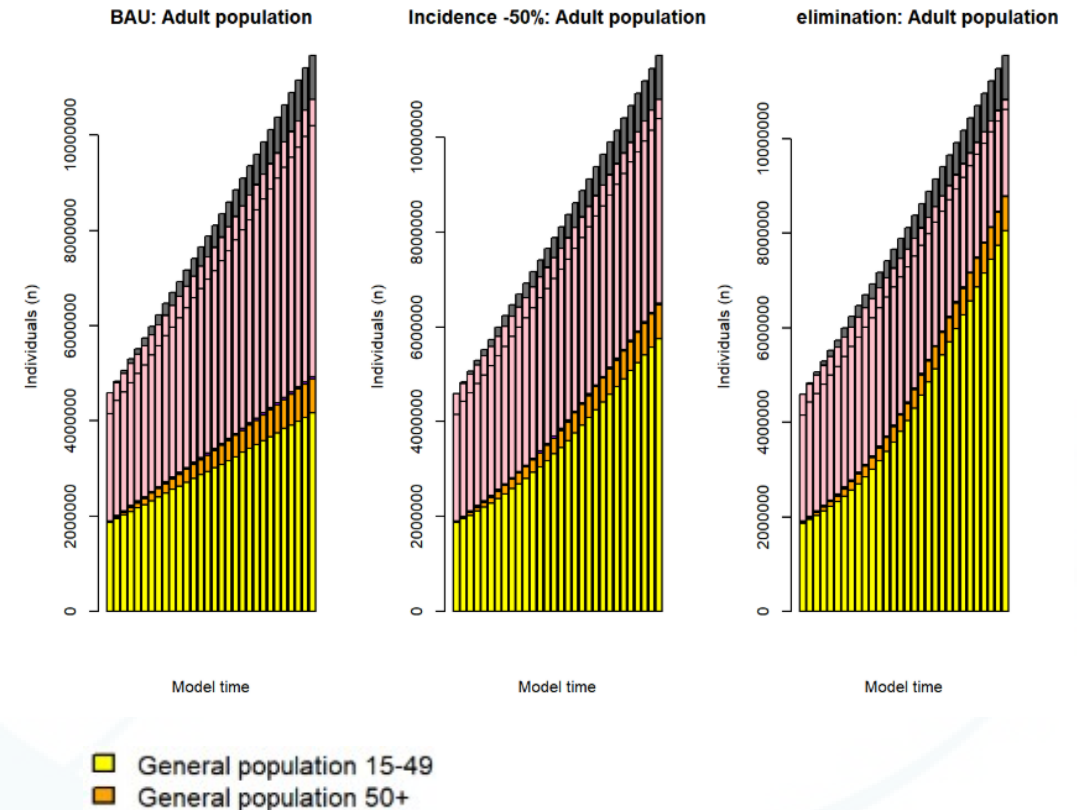
# What do interventions look like?

- We're interested in modelling interventions that reduce the incidence
- Expect to see a lag-effect, as many afflicted adult individuals are still alive



# What do interventions look like?

- We're interested in modelling interventions that reduce the incidence
- Expect to see a lag-effect, as many afflicted adult individuals are still alive



# In summary

## What we did

- While R/HEEMOD takes care of population distribution between health states, we have explicitly modelled distribution within health states
- We account for cohort (population) evolution using annual births and background mortality, with acceptable deviation from full demographic modelling
- We can model interventions on incidence with intuitive results

## Working with HEEMOD

- A lot of model states made execution too slow
- It was tricky to understand from documentation at what point in a model cycle populations were transferred between states (“life-table” seems to be half-cycle correction)
- Some small bugs in the sensitivity analysis part
- Ability to dump traces from model runs was very useful for model checking



# The end!

*I'm looking for an R/Shiny consultant Sept-Nov 2019 (relatively small LoE)*

Thanks for listening!

[dtordrup@gmail.com](mailto:dtordrup@gmail.com)



TRIANGULATE  
**HEALTH** LTD  
SCIENCE | POLICY | ECONOMICS