Age Period Cohort models: the identification problem and what to do about it

Andrew Bell

Sheffield Methods Institute, University of Sheffield, UK

@andrewjdbell

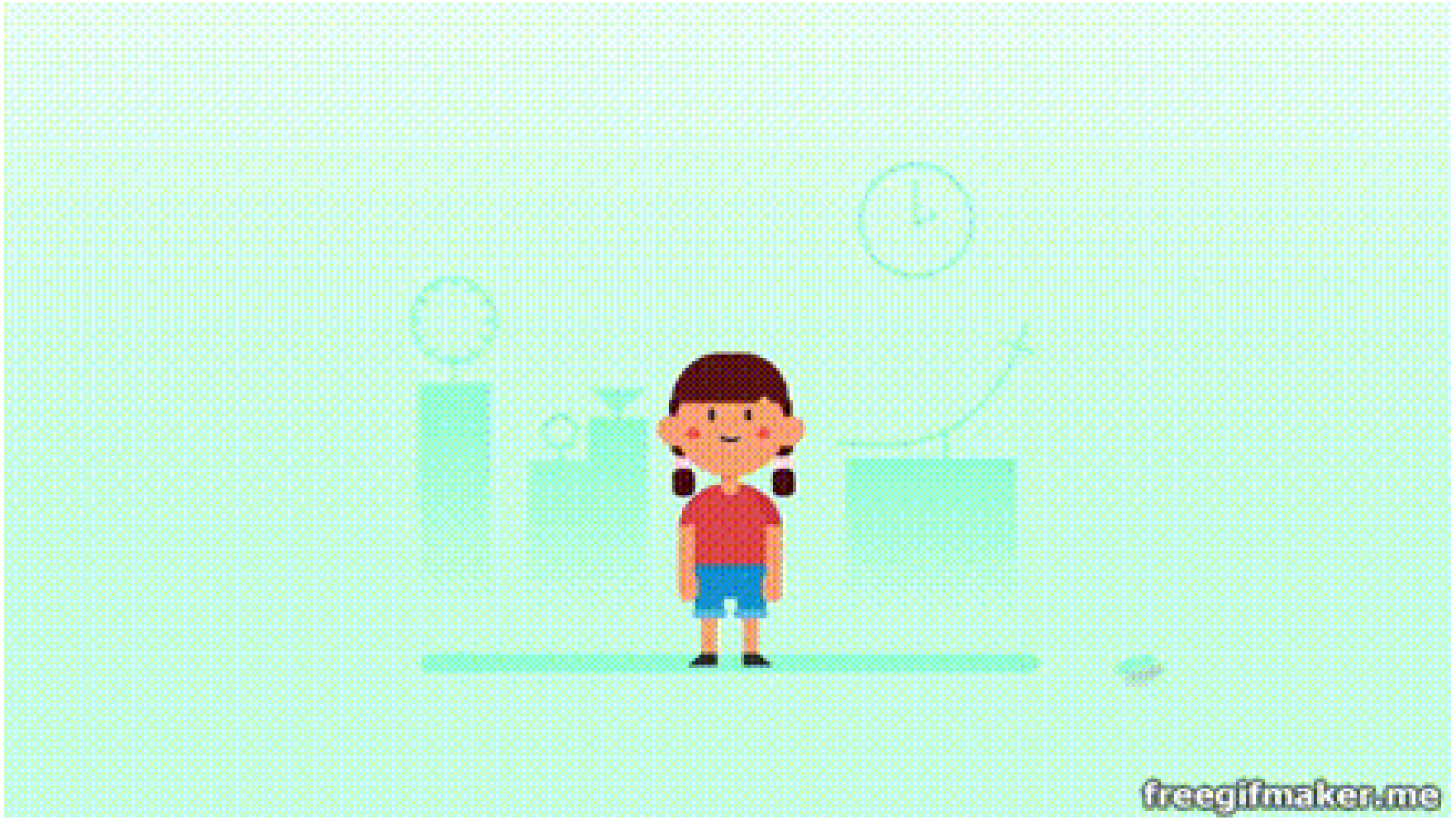
The University Of Sheffield.

Sheffield Methods Institute.

Outline

- APC effects
- The identification problem
- What is and isn’t a problem
- What *can* we do

What are APC effects



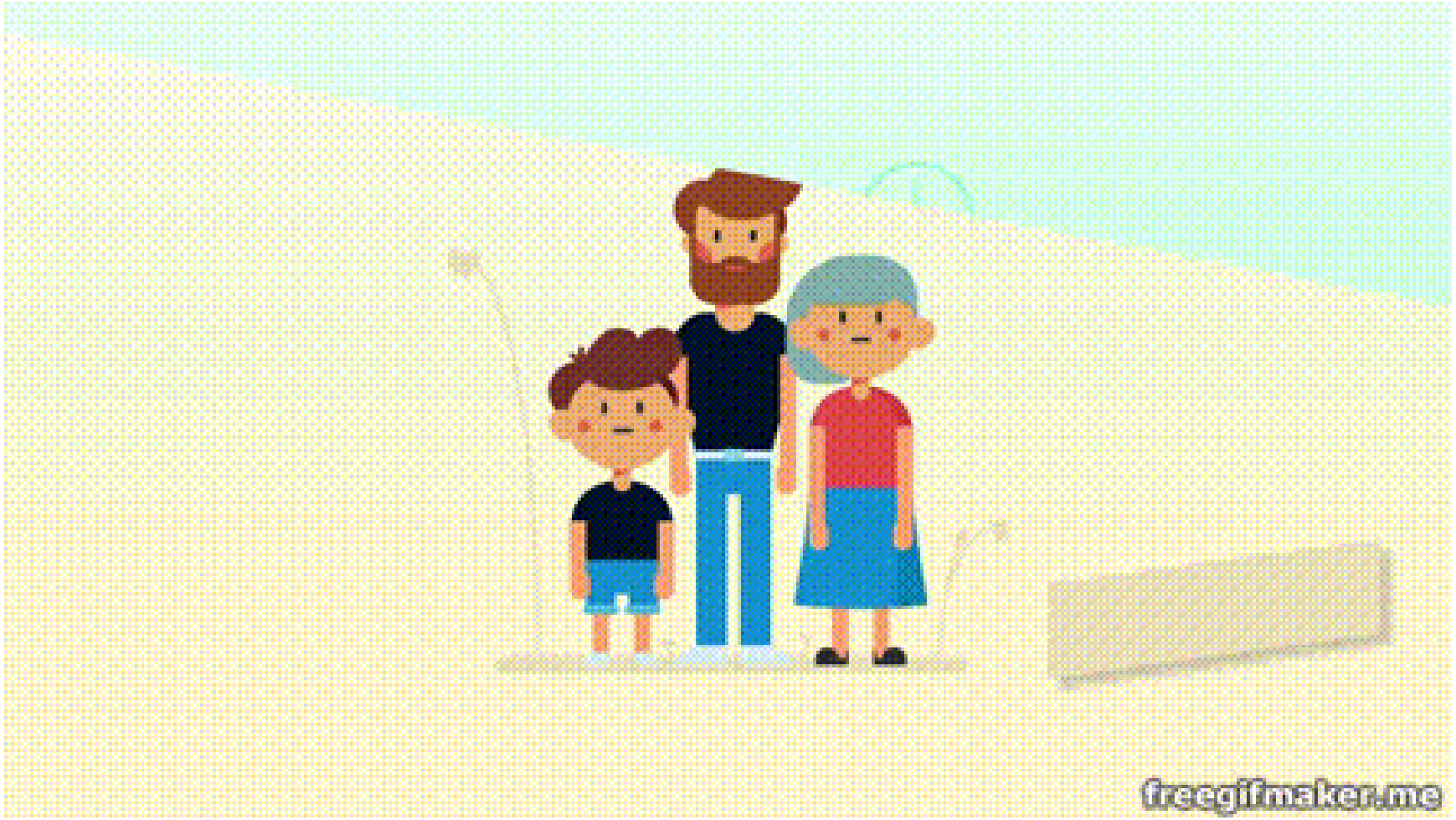
Age effects

- What happens as you get older?
- Your health declines
- Your happiness increases / decreases
- You become more religious
- You become more conservative

Or: what happens at specific ages?

- Midlife crisis around 45
- Low self esteem between ages 11 and 15

What are APC effects



Period effects

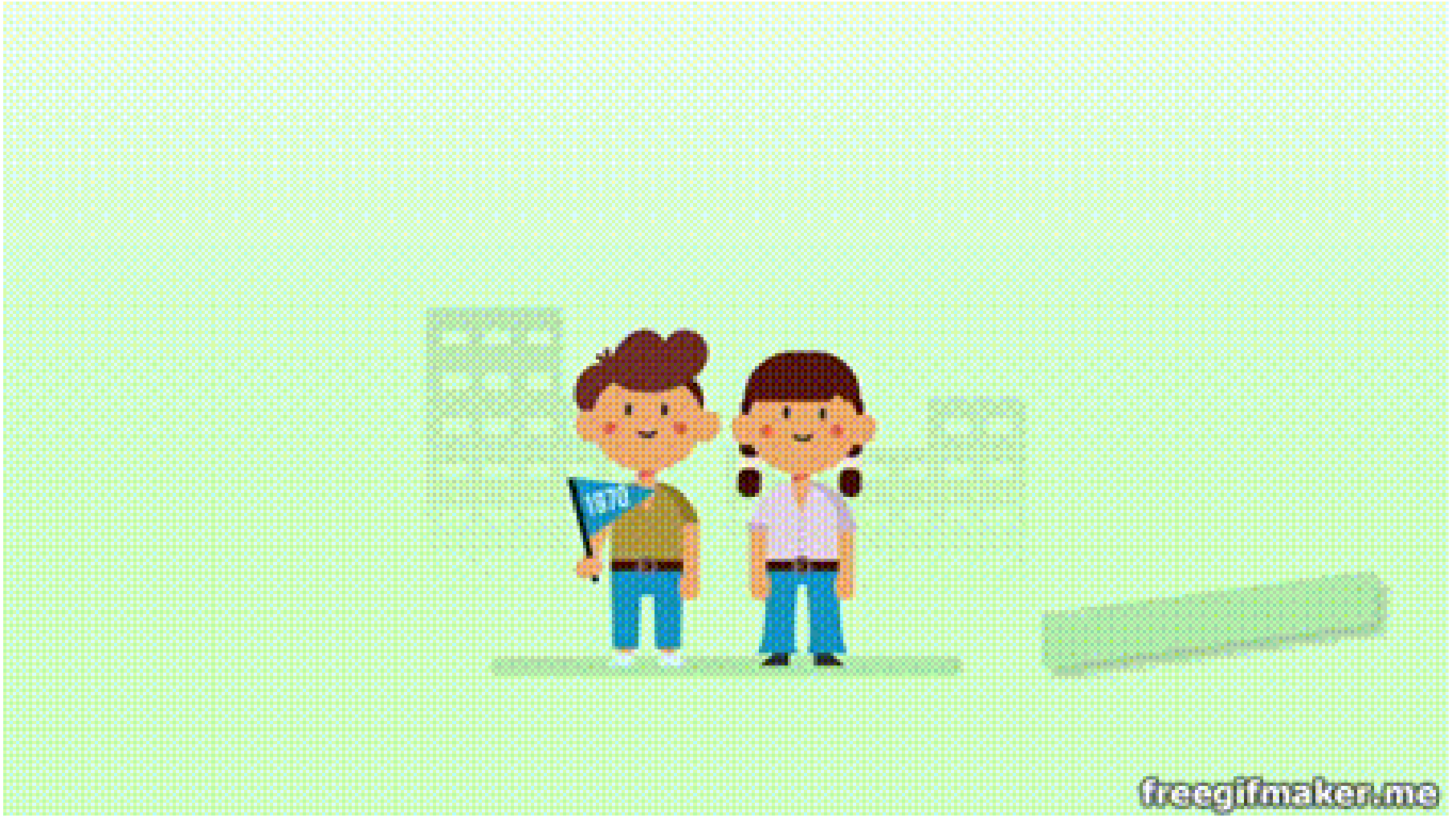
Things gradually improve… somehow...

- Health services improve gradually and incrementally, for all age groups
- Everyone becomes less likely to go to church
- Living conditions improve for everyone

A response to a specific event that effects everyone

- A recession makes everyone poorer
- A particular government policy makes people less healthy
- A cold winter, or war, increases mortality

What are APC effects



Cohort effects

Things gradually improve… by generation replacement...

- Health services for infants improve, and generations take those benefits on for the rest of their lives
- The new generation are less likely to go to church
- Living conditions improve for everyone

A response to a specific event that effects everyone

- A recession makes everyone *born in that recession* poorer for the rest of their lives
- People coming of age, politically, react negatively to a particular government
- A cold winter, or war, harms childrens' development

What are APC effects (Suzuki 2012, 452)

- A: I can’t seem to shake off this tired feeling. Guess I’m just getting old. [Age effect]
- B: Do you think it’s stress? Business is down this year, and you’ve let your fatigue build up. [Period effect]
- A: Maybe. What about you?
- B: Actually, I’m exhausted too! My body feels really heavy.
- A: You’re kidding. You’re still young. I could work all day long when I was your age.
- B: Oh, really?
- A: Yeah, young people these days are quick to whine. We were not like that. [Cohort effect]

At least 5 a day

Are you getting enough?

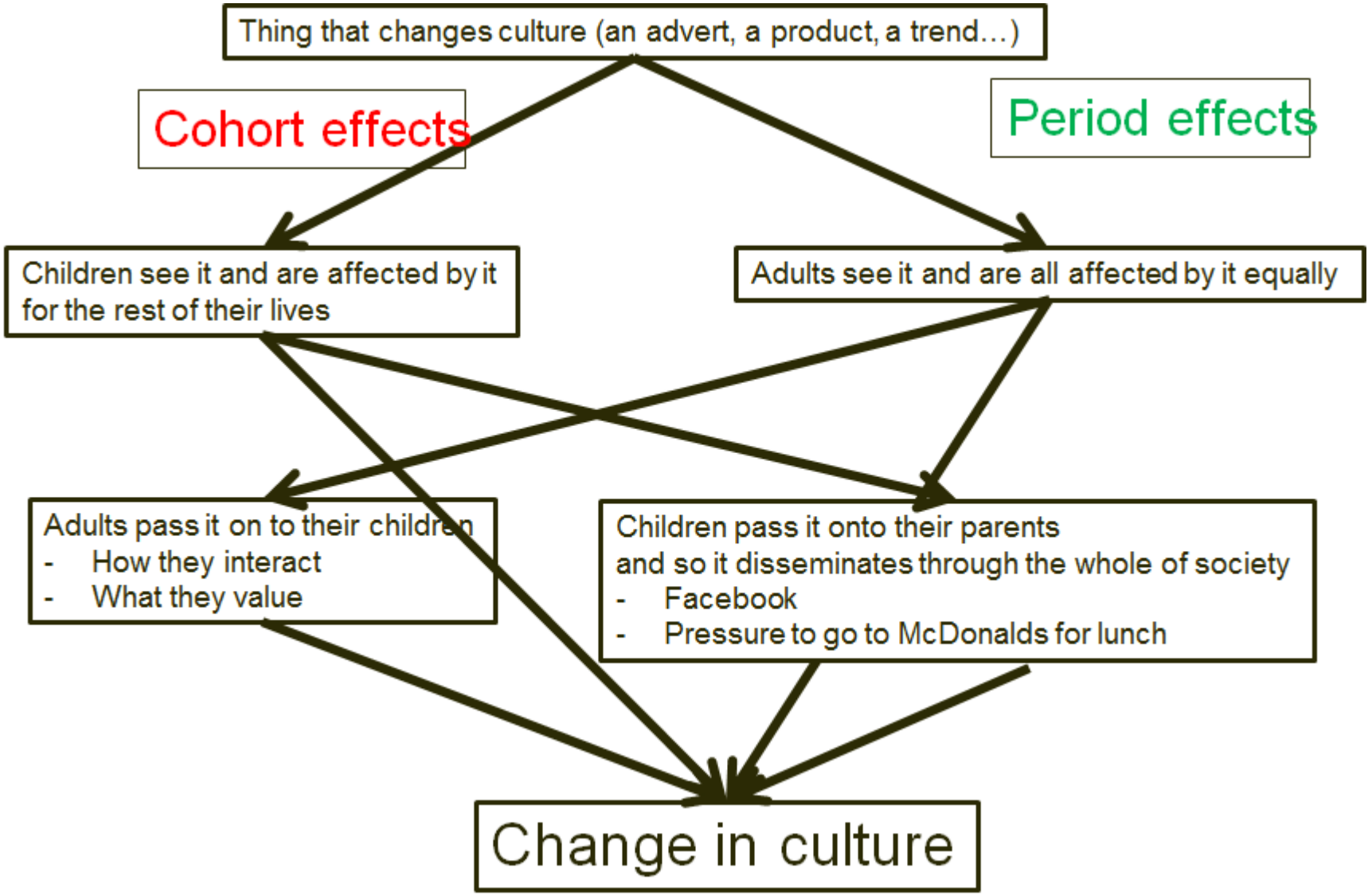
Fruits and Veggies
Good for Life

5 to 10 a day
For better health

PUBLIC HEALTH

519-376-9420 or 1-800-263-3456 www.publichealthgreybruce.on.ca

Where the lines blur



The Identification Problem

Age = Period – Cohort

“the term [confounded] is not used in the traditional design sense of experimentally confounded but in the stronger sense of logically or mathematically confounded” (Goldstein, 1979, 19)

The Identification Problem

$$Health = (1 * Age) + (1 * Period) + (1 * Cohort)$$
$$Health = (2 * Age) + (2 * Cohort)$$
$$Health = (0 * Age) + (2 * Period)$$

- All will produce *exactly* the same outcome variable
- Given that dataset, there is no logical way of telling which DGP created it
- Exact collinearity from putting all three into a regression model – model will not run.
- Grouping of one of APC breaks this collinearity, but produces arbitrary results (that depend on the chosen grouping)

What does this mean?

- Cannot hold age and cohort constant and vary period (without time travel – Suzuki 2012)
- Glenn 2005: “One of the most bizarre instances in the history of science of repeated attempts to do something that is logically impossible”
- If you have age in your model, you also have period and cohort, and vice versa (whether you like it or not)

What is and isn't a problem here

- Only applied to linear components
- Non-linear components are unaffected usually in practice
- BUT non-linear components mean very different things depending on linear components

$$Health = (1 * Age) - (0.1 * Age^2) + (1 * Period) + (1 * Cohort)$$
$$Health = (2 * Age) - (0.1 * Age^2) + (2 * Cohort)$$
$$Health = (0 * Age) - (0.1 * Age^2) + (2 * Period)$$

Various proposed solutions

- Arbitrary Constraints
- Hierarchical APC model
- Intrinsic Estimator
- Interactions
- Partial Least Squares
- …

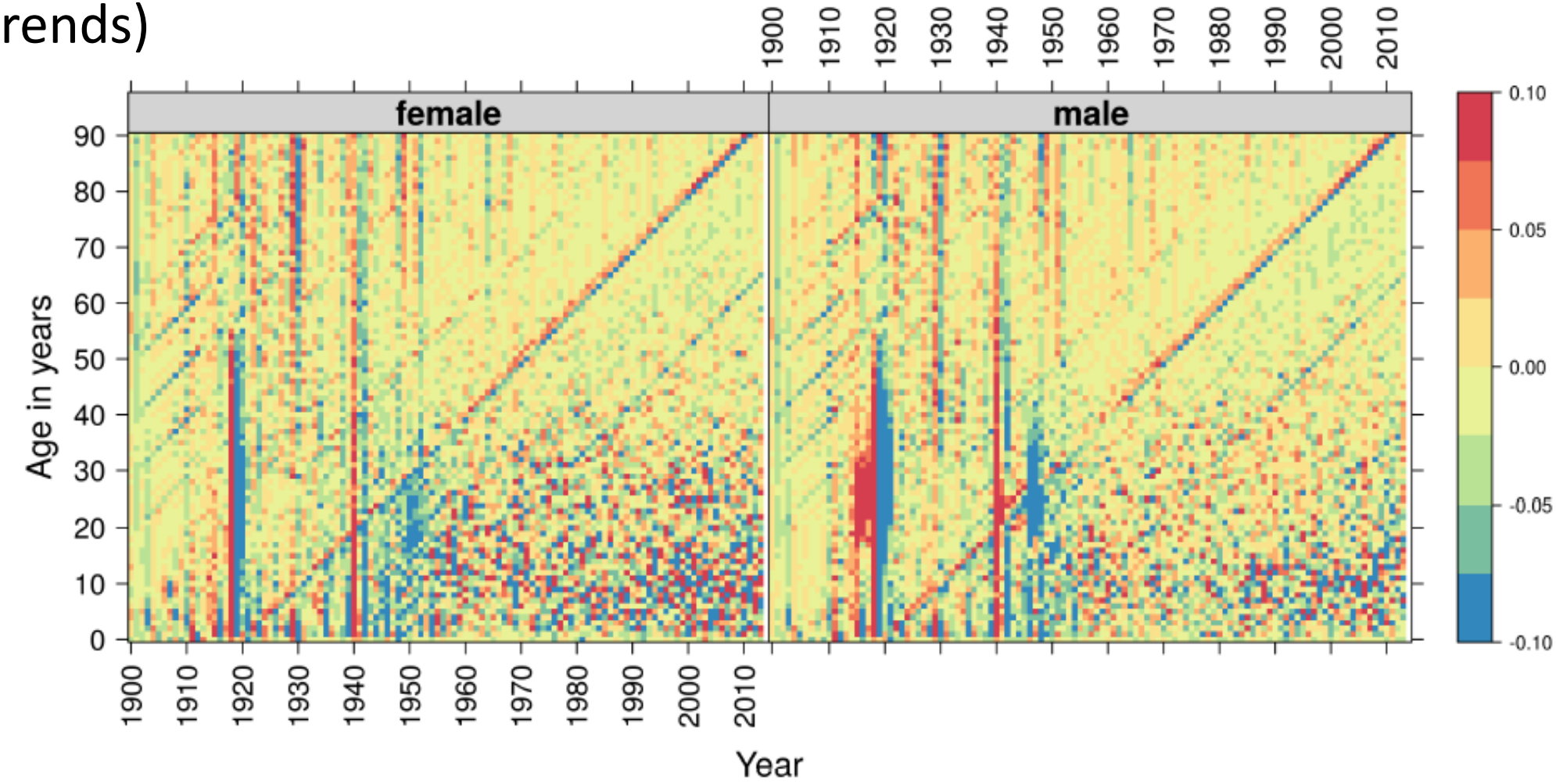
- In all cases, they have either not been tested rigorously enough, or they've been tested and shown not to work
- We know they can't work!

What can we do (without constraints) #1

- We can identify non-linearities (that is: variation around any linear trends)

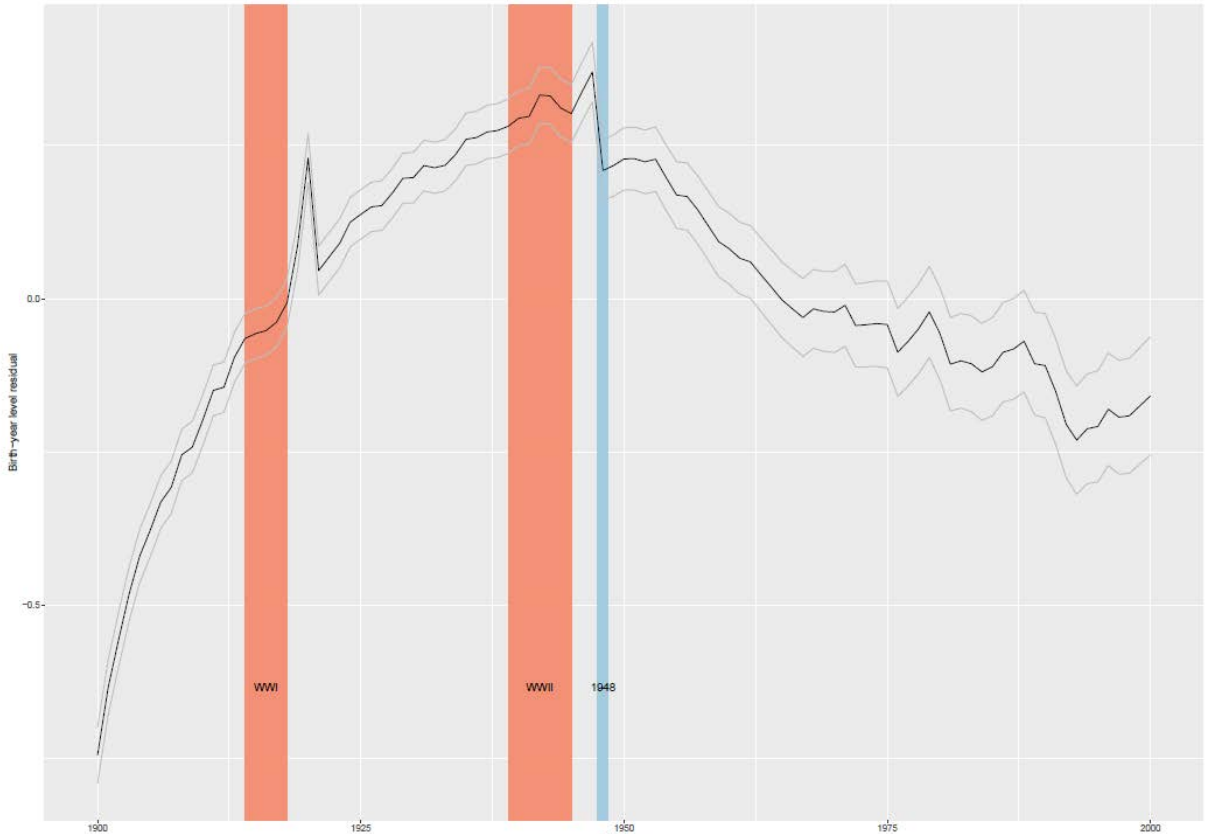
What can we do (without constraints) #1

- We can identify non-linearities (that is: variation around any linear trends)



What can we do (without constraints) #1

- We can identify non-linearities (that is: variation around any linear trends)
- But - these may mean different things depending on what the linear trend is…

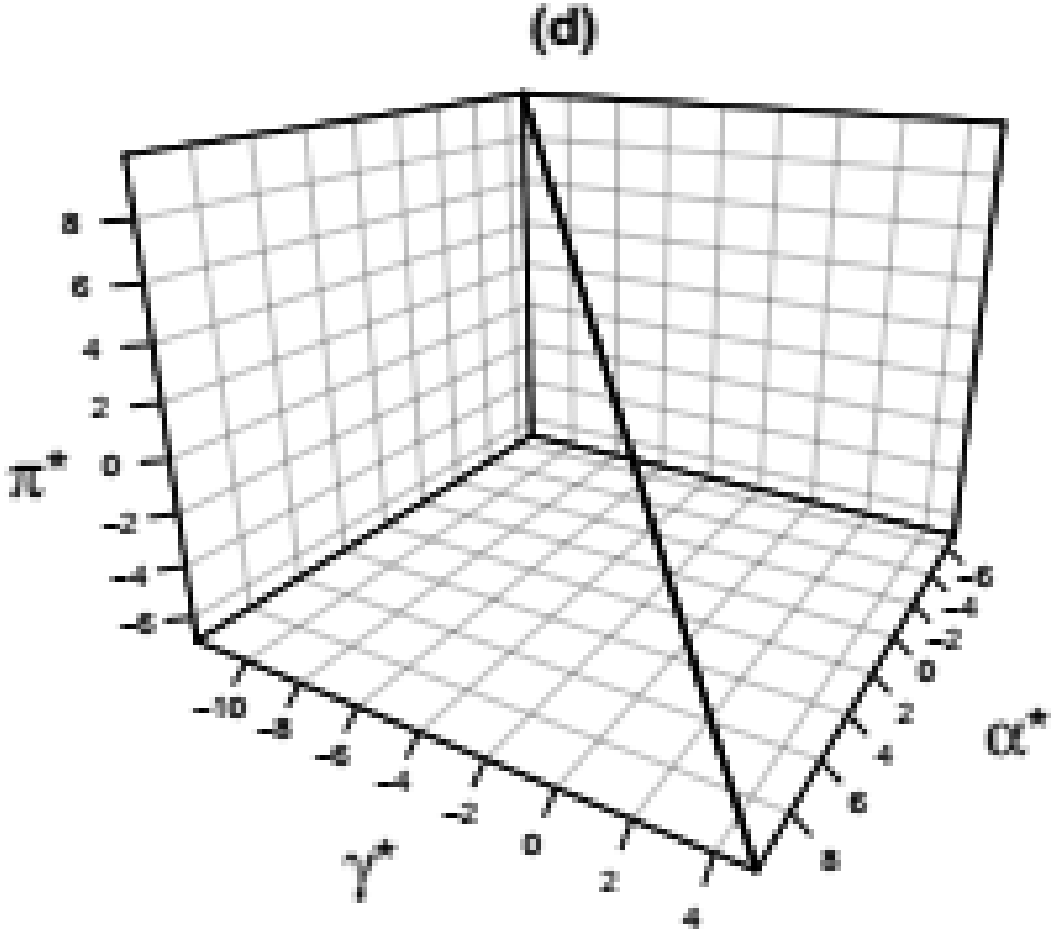


What can we do (without constraints) #2

- Identify effects that vary with APC, rather than overall APC effects themselves – Winship and Harding approach

What can we do (without constraints) #3

- We can identify the ‘line of solutions’ (not necessarily that useful)



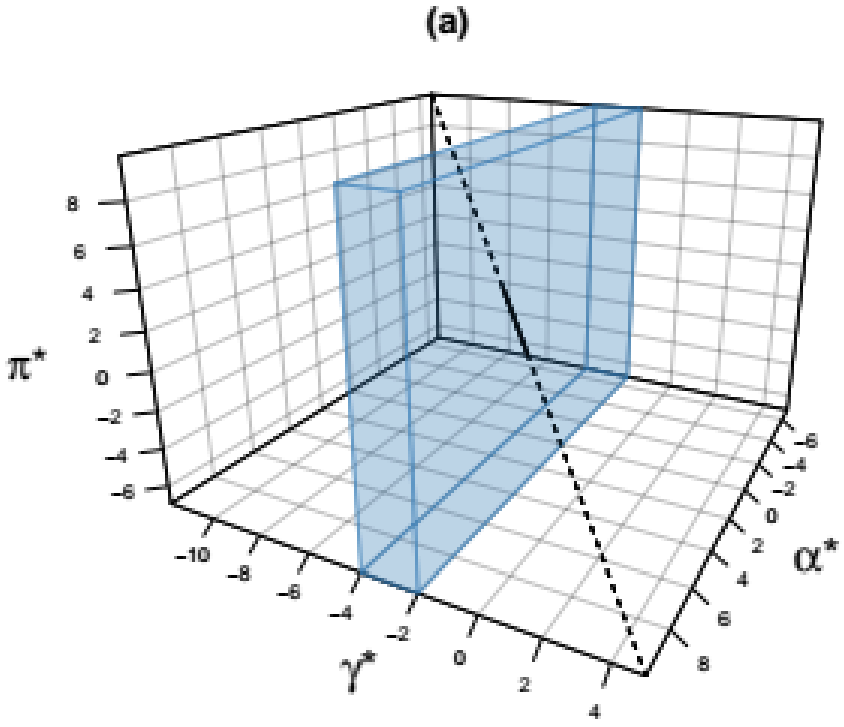
What can we do (with constraints)

- Make assumptions about linear part of age / period / cohort effects
- Effectively choosing the most plausible point on the line of solutions
- This needs to be a big assumption (eg period trend is flat)
- Smaller assumptions (eg 1990 is the same as 1991) are actually hiding a much bigger assumption – has a big effect on estimation
- Could use more than one assumption, and/or bounds

What can we do (with constraints)

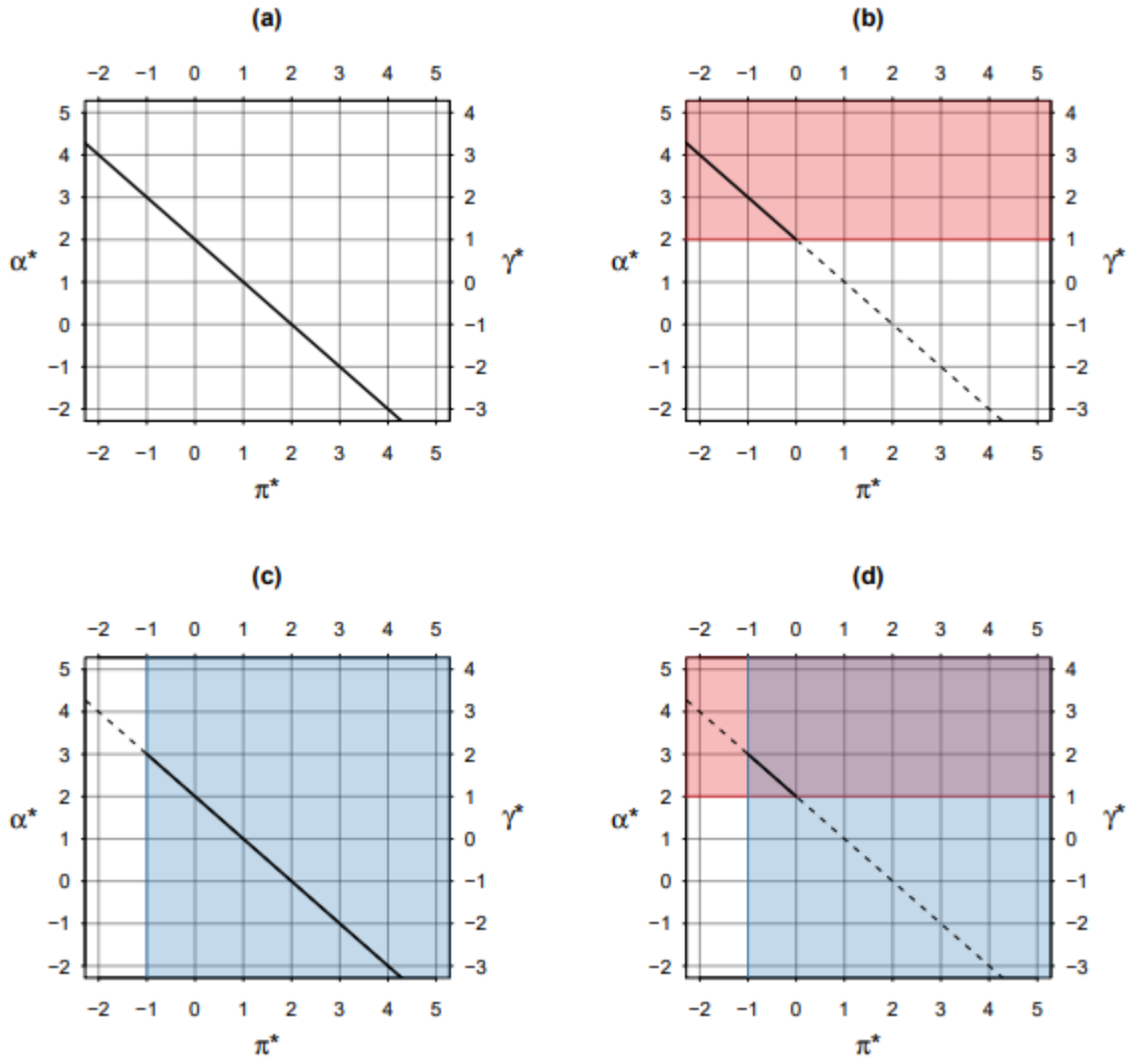
- Make assumptions about linear part of
- Effectively choosing the most plausible
- This needs to be a big assumption (eg p
- Smaller assumptions (eg 1990 is the sar a much bigger assumption – has a big e
- Could use more than one assumption

Fig. 3: Solution Line with Bounds on



What can we do (with constraints)

- Make assumptions about linear part
- Effectively choosing the most plausi
- This needs to be a big assumption (e
- Smaller assumptions (eg 1990 is the a much bigger assumption – has a b
- Could use more than one assumptic



Conclusions

- Age Period and Cohort effects can be important
- We need to take them, and the identification problem seriously, whilst acknowledging that it cannot be solved
- That involves making explicit assumptions (a wrong explicit assumption is better than a correct implicit assumption), and being honest about what our models can and cannot do
- Various visual and statistical methods – but all make assumptions or only consider non-linearities (that may be misleading

Some useful texts:

Glenn, N (2005) Cohort analysis, 2nd ed. Sage.

Suzuki, E. 2012. Time changes, so do people. *Social Science & Medicine*. **75**,pp.452–456.

On Multilevel modelling and APC

Bell, A. and Jones, K. 2014a. Another 'futile quest'? A simulation study of Yang and Land's hierarchical age–period–cohort model. *Demographic Research*. **30**,pp.333–360.

Bell, A. and Jones, K. 2014b. Don't birth cohorts matter? A commentary and simulation exercise on Reither, Hauser and Yang's (2009) age—period—cohort study of obesity. *Social Science & Medicine*. **101**,pp.176–180.

Bell, A. and Jones, K. 2018. The hierarchical age–period–cohort model: Why does it find the results that it finds. *Quality & Quantity*, 52(2), 783-799.

Bell, A and Jones, K (forthcoming) Multilevel models for age period cohort analysis. Chapter 4 in Bell (ed) Age, period cohort analysis: the identification problem and what to do about it. Routledge.

Bell, A . 2014. Life course and cohort trajectories of mental health in the UK, 1991-2008: a multilevel age-period-cohort analysis. *Social Science and Medicine*, 120, 21-30.

On constraints and the line of solution

Fosse (in review) Bounding analysis of age-period-cohort effects. https://q-aps.princeton.edu/sites/default/files/q-aps/files/apcanalysis.pdf

O'Brien (2017) Mixed models, linear dependency and identification in age period cohort models. Statistics in Medicine, 36(16), 2590-2600.

O'Brien (2011) Constrained estimators and age-period-cohort models. Sociological Methods and Research, 40(3), 419-452.

Winship and Harding (2009) A mechanism-based approach to the identification of age-period-cohort models. Sociological Methods and Research, 36(3), 362-401.

Thanks!

- Email: andrew.j.d.bell@sheffield.ac.uk
- Tweet: @andrewjdbell

- Look out for a forthcoming book (due out some time next year)
  - Bell (ed.) Age, period and cohort effects: the identification problem and what to do about it. Routledge.