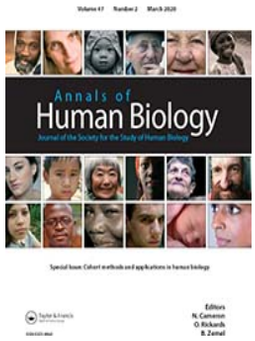

Annals of Human Biology

ISSN: 0301-4460 (Print) 1464-5033 (Online) Journal homepage: https://www.tandfonline.com/loi/iahb20

Age period cohort analysis: a review of what we should and shouldn't do

Andrew Bell

To cite this article: Andrew Bell (2020) Age period cohort analysis: a review of what we should and shouldn't do, Annals of Human Biology, 47:2, 208-217, DOI: 10.1080/03014460.2019.1707872

To link to this article: https://doi.org/10.1080/03014460.2019.1707872

Published online: 20 May 2020.

ANNALS OF HUMAN BIOLOGY
2020, VOL. 47, NO. 2, 208–217
https://doi.org/10.1080/03014460.2019.1707872

REVIEW

Age period cohort analysis: a review of what we should and shouldn't do

Andrew Bell

University of Sheffield, Sheffield, UK

ABSTRACT

Context: Age, period and birth cohort (APC) effects have been known for decades in biological, health and social sciences. However, exact collinearity between these three (Age = Year – Birth Year) leads to difficulty estimating these effects. It is thus impossible to estimate linear components of these effects without strong assumptions about at least one of these. This is problematic for anyone interested in APC patterns. Attempts to 'solve' this identification problem without strong assumptions are, in fact, making hidden unintended assumptions.

Objective: Provide an overview of what APC effects are and the nature of the identification problem, before reviewing and critiquing methodological literature across the health and social sciences. I also present an argument for what researchers should do.

Method: Non-systematic review of methodological literature across health and social sciences.

Results: Recommendations include considering non-linearities around linear APC effects and stating strong and explicit theory-based assumptions. Mechanical solutions to the identification problem do not work.

Conclusion: These recommendations acknowledge there is a 'line of solutions' of possible combinations of APC effects, and not a single answer that can be estimated empirically. None of these methods solve the identification problem – rather they acknowledge that methods are limited by assumptions.

ARTICLE HISTORY
Received 13 September 2019
Revised 12 November 2019
Accepted 25 November 2019

KEYWORDS
Age; period; cohort; identification; methodology

1. Introduction

Age, period and cohort (APC) effects are three ways that people and the societies in which they live can change over time. However, despite being fundamental to our understanding of how humans change over their life-course, and being known across epidemiology and the social sciences for 150 years, the three effects, and the processes that underlie them, are often misunderstood, and statistical attempts to uncover these effects are often fundamentally flawed to the point that results from such analysis are sometimes devoid of meaning.

At the heart of this misunderstanding is the so-called APC identification problem. In essence, this is a mathematical relationship between the three variables, such that age = period – cohort. However, the relationship is also a conceptual one. Attempts to solve this problem often take a mathematical approach that ignores the conceptual, and in the process produces results that may be mathematically appropriate, but do not necessarily bear any resemblance to the societies and people that the analysis is attempting to understand.

The conclusion of this review will, for many, be unsatisfactory – that there can be no solution to the age period cohort effect, and that any attempt to understand the effects will always rely on theoretical insights that are independent of the data being analysed. Unfortunately, there is no automatic, mechanical solution to the identification problem; researchers need to make strong assumptions in order to understand what combinations of APC are the most plausible, making those assumptions explicit at every stage of analysis.

The review will proceed as follows: First we will consider what age period and cohort are, and the substantive nature of the identification problem. This will include consideration of what different sorts of effects might mean, along with the possibility of interactions between them. The focus here will be on the processes that could drive the APC effects. We will then consider the identification problem, seeing the sorts of APC effects that are affected by it. Central to this will be the idea of the 'line of solutions', which shows that whilst we cannot know from the data alone the value of APC effects, we can know, for a given dataset, the values of two of the APC effects if we know the value of the third. Put another way, although there are an infinite number of linear combinations of APC that are 'best fitting', we can rule out the infinite number that are not best fitting.

This will be followed by suggestions of what we can do. First, there are some things that can be estimated from a dataset without making assumptions; however these are not age, period and cohort effects themselves. Second, we could make one or more assumptions – I will discuss the sorts of assumptions that could be made in order to estimate APC effects, and the fact that APC effect estimates will be

CONTACT Andrew Bell Andrew.j.d.bell@sheffield.ac.uk Sheffield Methods Institute, ICOSS Building, 219 Portobello, Sheffield, S1 4DP, UK

incorrect if those assumptions are incorrect. Finally, I will discuss some of the more problematic attempts to resolve the identification problem with the data alone and, apparently with no assumptions – it turns out that, in each case, there are in fact hidden assumptions being made that can make the results from that method likely to be incorrect.

This paper is not a systematic review of all empirical studies attempting to uncover age, period and/or cohort effects. Rather it highlights the large body of methodological work that has been done to further understand APC effects, and critiques the methods that are most regularly used, what those methods uncover, why they find the results that they do, and what their flaws are.

2. What are age, period, and cohort effects

This section outlines what each of the APC effects are substantively, considering how both continuous and discrete APC effects might occur, with reference to (among others) an example of changing obesity rates over the life course and over time.

2.1. Age effects

Age effects are perhaps the easiest and most obvious effect of interest, particularly in human biology. Age effects imply that as individuals get older, something happens: that is, they become richer, or more likely to be obese, or more likely to die.

In general, age effects are continuous effects, rather than discrete effects related to specific ages. For instance, there are biological reasons why we would expect the risk of obesity to increase gradually as people get older. However specific, discrete age effects can also happen, usually as a result of particular societal expectations or constructs. In times of war, for instance, we would expect mortality rates to be particularly high for particular discrete age groups – young men, perhaps between the ages of 18 and 40. If we believe in midlife crises, we might expect an increase in mental ill-health between the ages of 40 and 50.

Age is usually measured as the number of years since birth but might be measured with more or less precision depending on the research questions at hand and the nature of the data.

2.2. Period effects

Period effects relate to the year in which an individual is being measured – that is, the effect of contemporary time. For instance, my own, current risk of obesity might be affected by the fact that I am writing in 2019, as opposed to 2009, or 1909. We can see why we might expect to see both continuous and discrete period effects. Progress in medical treatments might lead to a reduction in mortality rates over time, helping everyone, of all ages, to live longer. Conversely, societal changes have led the environment in which we live to be more obesogenic – with more sedentary lives and higher-fat diets. We might also expect period effects that are discrete. A period effect associated with a financial crash will have effects on average levels of health and wellbeing across a society, including potentially obesity as healthier, more expensive food becomes too expensive for some. Similarly, an epidemic, or a war, is likely to increase mortality for the short period of time that that event is taking place.

Period is usually measured as year of measurement, but again, a finer or wider scale of measurement could be used depending on the data available.

2.3. Cohort effects

Finally, cohort effects are related to the generation that an individual is part of. The idea here is that being a member of a particular cohort, brought up with specific societal norms from childhood, results in particular outcomes. From a health perspective, we might expect the early years of life to be important, and that it is during this period that particular habits and health behaviours – smoking, diet, exercise and so on – are set. As such, we might expect improvements in diet over time to result from a cohort effect on health – not because any individual's health behaviour actually improves, but because newer cohorts have got into the habit of eating healthier foods than the generations that came before them. We would also expect discrete effects. Being brought up during war is likely to have effects that last long after the war is over, for everyone who experienced war in their formative years. Evidence also suggests that being born or *in utero* during a flu pandemic results in a higher risk of mortality for the rest of your life, because of the negative effects the pandemic has on the development of young children. The idea of cohort effects has been long established in demography (Keiding 2011), sociology (Mannheim 1928; Ryder 1965), and health sciences (Frost 1940), and is now the basis of many ideas in life-course epidemiology (Hanson and Gluckman 2016).

Cohorts are usually measured based on birth year; however often (for some reason, more often than period and age) these are grouped into particular groups. Sometimes these are arbitrary (5-year cohorts are often used in demography), and other times the groupings are based on theoretical generations – for instance, people born between 1980 and 1994 are often described as 'Millennials', a supposedly distinct group separate from Baby boomers (those born between 1946 and 1964), Generation X (born 1965–1979) and Generation Z (born between 1995 and 2012) (e.g. see Dimock 2019). Whilst these often have some theoretical merit, there is often little reason for choosing particular start and end points, and in some cases they are often not much more justified than the 5-year groups that are often assigned, at least in terms of how they might be related to a particular outcome.

2.4. Conceptual difficulties with age, period and cohort effects

Whilst these definitions of APC effects are, at first glance, unassuming, the conceptual basis of them is often difficult

to pin down. First, we have already noted a key difference between continuous effects of APC – caused by a continuous improvement or decline over time – and discrete APC effects – caused by specific events. Continuous effects are not necessarily linear – they might be curved – but form a continuous shape. It is probably true that, for most outcomes, and particularly for biological outcomes, age trends will generally be continuous because we would not expect a higher death rate for a particular discrete age group, although particular age periods – puberty, midlife crises, etc – might produce more discrete effects. Period and cohort effects, in contrast, are more likely to be discrete, whilst also having the potential to have long-run trends. The key point here is that any conceptual framework for APC analysis needs to be clear about what sorts of APC effects it is looking for, and the modelling framework needs to be able to capture that type of effect.

Second, in many respects, APC are not really explanatory variables at all, given that they don't give any explanation of the mechanisms that underlie the variable. For instance, a higher mortality rate related to the second world war was a result of the war, not a result of anything intrinsic about the years the war happened to take place. Improvements over time occur because of specific medical advances, practices, or behaviours not because of the actual act of time passing.

Third, and relatedly, variables associated with one of APC might also be related to another. Medical advances might produce both period and cohort effects, depending on whether it affects all ages simultaneously (period effects) or affects specifically individuals in their formative years that they then take on for the rest of their lives. Indeed, we might often expect a particular phenomenon to have both period and cohort effects – and these might reinforce each other in conceptually quite complex ways. As such, even knowing the variables that drive change is not always enough to help us understand which of APC are the relevant drivers of change.

Fourth, it is quite plausible to imagine interactions between APC effects. For instance, we can think of the effect of a war as often being an interaction between age and period (and gender), since the group most affected by the period effect of war would be young men. This greatly complicates how we might think about, and model, these effects.

2.5. APC with different sorts of data

It is worth briefly considering how APC operate in different data structures, particularly those often used to find, or understand, age effects that are commonly thought about in human biology. In a single cohort design commonly used in human biology, there is no variation in cohorts, and age and period are exactly aligned, meaning there is no way of knowing (without good theory), if change is occurring as a result of periods or a result of cohorts. This is represented by one of the dashed rectangles in Figure 1. Similarly, in cross-sectional studies (the solid rectangles in Figure 1), there is no variation in period, and age and cohort are exactly correlated, with no way of telling whether cohort or age are producing any differences. Alternatively, we could combine multiple cohort studies together (represented by the 3 cohorts with dashed rectangles in Figure 1), allowing us variation across all of APC – for instance the UK 1946, 1958, 1970 and 2000 cohort studies could be used in this way. Some studies also have the format of the dotted rectangle – time-sequential designs like panel data (e.g. the English Longitudinal Study of Ageing, would have this shape, and would also have variation in all of APC).

However, even when we have variation in all of APC, the three are linked together mathematically, making the three effects particularly difficult to tell apart. It is to this so-called identification problem we turn to now.

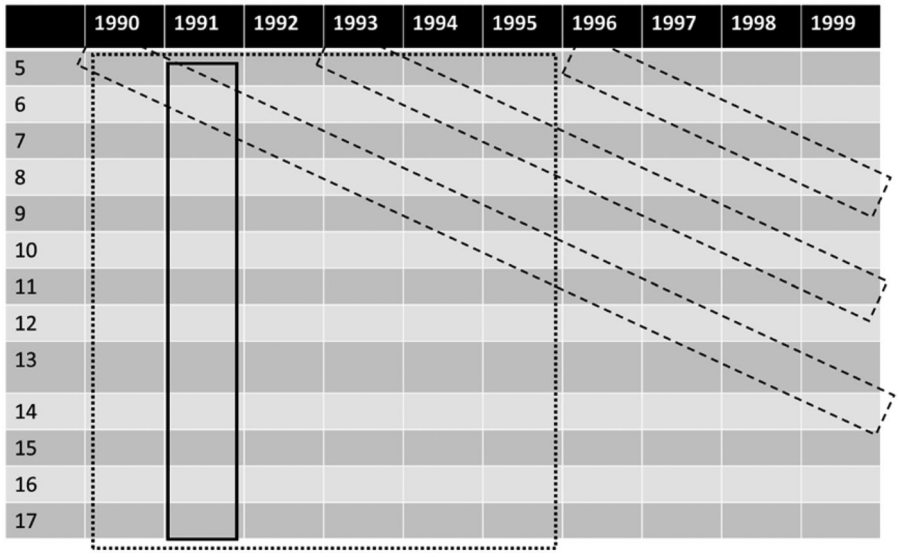

Figure 1. Different data structures commonly used in human biology. Rows represent ages, whilst columns represent years/periods. The dotted line represents a typical panel or repeated cross-sectional design; the dashed lines represent single cohort studies, that could potentially be linked together. The solid line represents a cross-sectional design.

3. The APC identification problem

There is a problem in identifying all three of age, period and cohort effects, because they are exactly collinear – that is,

$$\text{Age} = \text{Period} - \text{Cohort}, \text{ or } \text{Age} = \text{Year} - \text{Birth Year} \quad (1)$$

This is a mathematical problem, but it is also a conceptual one. When talking about causal effects, in particular in experiments, we usually talk of the effect of changing one variable whilst maintaining the other variables at a constant level. However, with age, period and cohort, we cannot change one of them whilst holding the other two constant, because if we hold two constant, we are by definition holding the third constant as well. The only way to break this relationship would be through some kind of ageing technology (allowing age to change without changing year and cohort) or a time machine (allowing year to change without changing age and cohort). None of these seem plausible in the near future and are likely to have other issues with external validity!

However, in my view the inability to hold other variables constant doesn't (contra Goldstein 1979) necessarily negate the possibility of all three having a causal effect. However, it does make it much more difficult to identify those effects. For instance, we might be interested in estimating a regression of all three of age period and cohort, along the lines of:

$$\text{Health} = \beta_0 + \beta_A\text{Age} + \beta_P\text{Period} + \beta_C\text{Cohort} + \text{residual} \quad (2)$$

We would not be able to estimate all of β_A, β_P, & β_C because of the exact collinearity between the three variables. If we were somehow able to produce estimates from these, they could well be distributed across the three variables wrongly. For instance, a world where β_A, β_P, and β_C all equal 1:

$$\text{Health} = \beta_0 + 1*\text{Age} + 1*\text{Period} + 1*\text{Cohort} + \text{residual} \quad (3)$$

Would be exactly equivalent to a different imaginary world where there is only a larger period effect, since:

$$\text{Health} = \beta_0 + 1*(\text{Period} - \text{Cohort}) + 1*\text{Period} + 1*\text{Cohort} + \text{residual}$$

reduces to:

$$\text{Health} = \beta_0 + 2*\text{Period} + \text{residual} \quad (4)$$

These different imaginary worlds (or 'data generating processes') suggest very different social processes, even though they are different ways of describing the exact same data (we are just replacing Age in the equation with Period-Cohort, which are identical as per Equation (1) above). The implications of this are significant: if we are interested in age, we could find a big effect, or no effect (or, in fact, any other effect), depending on which estimates our chosen model finds in the data.

It's worth considering the ways in which this identification problem operates. Firstly, the identification problem only affects linear APC effects. As such, any discrete APC effects such as those described in Section 2 could be identified, as could any non-linear curves around whatever the linear trend might be. Secondly, whilst there are many difficult combinations of APC effects that would produce exactly the same data, it isn't the case that all possible combinations could have produced the data. Indeed, the combination can be taken to be lying somewhere on what is called the 'line of solutions' (Rodgers 1982; O'Brien 2011, 2015a, 2016; Fosse and Winship 2019a, 2019b). Imagine a 3D graph, where the three axes are the size of the age, period, and cohort linear effects. The line of solutions will form a line through this 3D graph, and the data would suggest that the truth lies somewhere on that line. In the above example, we had two solutions: Age effect = 1, Period effect = 1 and Cohort effect = 1; and Age effect = 0, Period effect = 2, and Cohort effect = 0. The line of solutions would include these two solutions, and an infinite combination of other solutions. O'Brien (2011) attempts to display this 3-dimensionally, whilst Fosse and Winship (2019b) displays this 2-dimensionally, as shown in Figure 2.

Algebraically, the line of solutions means that the three values of the APC effects are linked, such that if one increases, the others will increase or decrease by the same amount, to offset the increase (Fosse and Winship 2019b; O'Brien 2015b, 2020). So, for given estimates of $\beta_A, \beta_P, \beta_C$, the true values $\beta_A^t, \beta_P^t, \beta_C^t$ are

$$\begin{aligned} \beta_A^t &= \beta_A + v \\ \beta_P^t &= \beta_P - v \\ \beta_C^t &= \beta_C + v \end{aligned} \quad (5)$$

Therefore, from a given set of APC estimates, if we increase the age effect by v=1, we would decrease the period effect by 1, and increase the cohort effect by 1, in order for the new solution to remain on the line of solution. Had we done this to Equation (3), we would have ended up with an age effect of 2, a period effect of 0, and a cohort effect of 2. In Equation (4), we did the opposite: going from Equation (3) to Equation (4), we decreased the age effect by 1 (from 1 to 0), increased the period effect by 1 (from 1 to 2) and decreased the cohort effect by 1 (from 1 to zero). All three of these solutions (1,1,1; 2,0,2; and 0,2,0) lie on the line

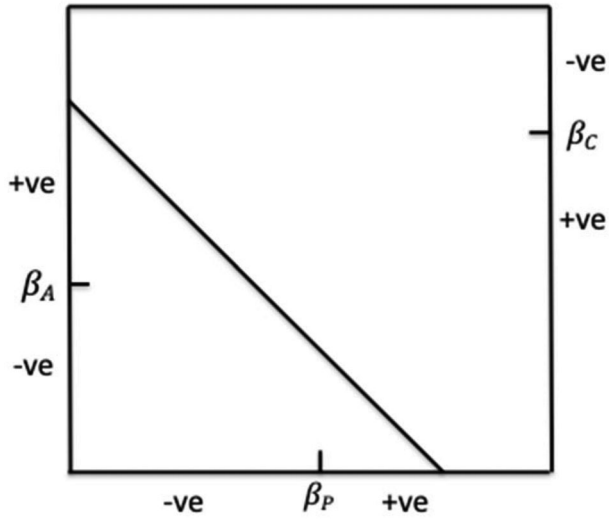

Figure 2. A 2D representation of the line of solutions, adapted from Fosse and Winship (2019a, 2019b). The bottom axis represents the size of the period effect, from negative (left) to positive (right). Similarly the left axis represents the size of the age effect, and the right axis represents the size of the cohort effect.

of solutions. The problem is that, for a given set of estimators we cannot know the value of v – we cannot know where on the line of solutions the true estimates lie. The good news is that, if we happened to know the true value of one of $\beta_A, \beta_P, \beta_C$, we would be able to work out the others. Unfortunately, that is quite a big if.

In sum, with every longitudinal dataset with variation in age, period and cohort, we have clear characteristics. First, there are unknown linear APC effects that lie somewhere on a line of solutions; whilst we cannot know the individual linear effects, we can identify the line of solutions on which those linear effects somewhere lie. Second, there is non-linear variation around that unknown line, which we can identify.

4. What can we do? part 1: without assumptions

Fortunately, there are some things that we as researchers can do, without having to make strong assumptions (beyond the usual assumptions often made in statistical analysis). Unfortunately, those things do not provide a solution to the APC identification problem. Having said that, some of what can be produced are useful, but need to be cautiously interpreted.

4.1. Estimable functions – expressing the line of solutions

The fact that the line of solutions is identifiable, even if the individual APC effects that make up the line are not identifiable, implies that there are attributes of linear effects that can be identified and numerated. Indeed, many have suggested a number of different estimable functions (Clayton and Schifflers 1987; Holford 1983, 1985, 1991; Robertson and Boyle 1998; Robertson et al. 1999; O'Brien 2014, 2015a). In each case, these either describe the line of solutions, or express the deviations from the line of solutions (discussed further below).

For explanations of the line of solution, Firebaugh (1989, 2008) defines two quantities: 'cohort replacement effects', and 'individual change effects'. Cohort replacement effects define differences that result from cohorts replacing each other, but in doing so confound cohort and period effects together. Similarly, individual change effects refer to changes within an individual – and as such confound age and period effects together. These two quantities are estimable: taking the equations in (5), we could add β_C^t and β_P^t together, or β_A^t and β_P^t. In both cases, the v's would cancel out, leaving a quantity equivalent to a cohort replacement or individual change effect, respectively. The problem with this is that we still do not know what is driving the change. For instance, a change that has resulted from time passing is a very different thing from change resulting from ageing, and we cannot know which is driving a particular individual change effect. Because of this, Rodgers (Firebaugh 1990; Rodgers 1990) argues these attempts at decomposition are often meaningless because we cannot know the underlying APC effects that created them.

Other estimable functions relating to the line of solutions exist: net drift, local drift, cross-sectional age trend, longitudinal age trend, and so on. There are even online tools available to automatically calculate such quantities (Rosenberg et al. 2014). All of these are some form of the above: an expression of the line of solutions rather than any specific APC effect. There is nothing explicitly wrong with this, except that they run the risk of being interpreted as APC effects when these cannot be interpreted as such. It would be reasonable, for instance, for a quantity described as a 'Longitudinal Age' effect to be treated as an age effect, even though it is confounded with period.

4.2. Estimable functions – non-linearities

As an alternative to these estimable functions, one could look only at the non-linear components of APC effects, that occur around whatever linear effects might be found. Again, these are another type of estimable function (and are often confusingly reported together with those that refer to the line of solutions) in that they remain the same regardless of the constraint used to identify linear APC trends. However, they are estimating a fundamentally different thing to the line of solutions and can often only be conceptually understood in conjunction with the line of solutions (or in some cases, the true APC effects themselves).

Chauvel et al. (2016, 2020) present a method for uncovering these – they remove the APC linear trends, and then use the remaining data to establish what non-linear patterns remain. They then extend this to consider the extent of the non-linearity, measuring how the extent of fluctuations vary at different values of APC. They use the method to find that cohort fluctuations in suicide around the line of solutions have declined in a number of countries, meaning that, whilst the rate of suicide cannot be known, the extent of variation from cohort to cohort can be.

Jones et al. (2018) take a similar approach in studying mortality in the 20th century – they use a random effects model that sets the cohort linear trend to zero but ignores the linear APC trends and looks only at non-linear trends. They find a number of patterns common to a number of countries – such as period effects associated with the first and second world wars, and cohort effects associated with the 1919 flu pandemic.

A key point to be aware of, here, is that non-linearities can take two forms – discrete effects related to a particular year, year of birth, or continuous non-linearities that might be better summarised with, for instance, a quadratic relationship. For Jones et al. the focus was on the latter, and other non-linearities, although present, were deliberately ignored. The problem with continuous non-linearities is that their meaning can be different depending on the linear APC trends, which cannot be known (Holford 2006). For instance, a U-shaped pattern in the deviations from the unknown age linear trend could represent a U-shaped age effect (if the true linear age trend is zero) or a slight curve on a strong linear effect. Without knowing what that linear effect is, those

non-linear continuous curves are at risk of being over- or mis-interpreted.

4.3. Visualisation with line graphs

It is often commented, both in APC analysis and in social statistics in general, that data visualisation is an important and valuable way to understand data and the patterns that it produces. One can do this in a number of ways.

First, one could plot two of APC against a particular outcome (one as lines, one as groups), for example as shown in Figure 3 (alternatively, we could produce a 3D graph with two of APC plotted with the outcome). Here, we are able to reveal patterns in APC; the problem is that the way we plot the data leads to particular assumptions being made in the interpretation of those plots. In Figure 3(a), we connect the data by year of measurement (or period) and appear to see a positive age effect – but this is only the case because of the assumptions we are making – in that case assuming there are no cohort effects. Figure 3(b) joins up the same data using lines by cohorts rather than year – and this innocuous-looking decision produces a very different impression – of a negative age effect. Effectively, this interpretation of Figure 3(b) assumes that there are no period effects. To reiterate, the data is the same in both cases. For more examples of this see Bell and Jones (2014b). The equivalent in a 3D plot would be that looking at a 3D plot from different angles would imply different interpretations.

The issue here is not that the data have been plotted wrongly, or that any assumption has been made when plotting. However, the nature of those plots invites the reader to make particular assumptions, potentially without realising. And that is a problem when those assumptions can have a strong effect on the results and can turn out to be incorrect. Indeed Holford argues that one advantage of a modelling approach (rather than simply descriptive plotting) is that it "forces us to recognise a fundamental problem … that you should remember even when trying to understand a graphical display" (Holford 1991, p. 452–453).

4.4. Visualisation with lexis plots

Another commonly used visualisation is the Lexis plot – this plots a graph of year against age (or occasionally, cohort against age), with the values of an outcome variable indicated with variable colouring. Lexis diagrams have a long history in demography (back to Wilhelm Lexis in 1875) but are often relatively underutilised in other disciplines. Yet they can be effective in revealing patterns in APC data. For instance, Minton (2016) use contour plots (a smoothed version of a lexis plot) to reveal age-cohort hotspots in suicide rates, with particular combinations of cohort and age having different suicide rates. Similarly, Jones et al. (2018) use Lexis plots to reveal particular patterns in mortality across the 20th century, in particular relating to particular events. Such plots reveal discrete cohort effects particularly well, with 'scars' appearing emanating from particular years of measurement or years of birth. These can even be plotted as 3D-printed graphs, where physical height from an age-by-period surface indicates the value of an outcome (Minton 2016).

However, whilst this is useful for identifying particular scars across the lexis plot indicative of discrete APC effects, lexis plots do not help to uncover linear APC effects because, as with the line graphs above, multiple data generating processes would produce the same data. For instance, a period effect would appear in a lexis plot as a change from low to high values across years, with no changes with age apparent. However, a combination of age and cohort effects of the same size would produce exactly the same pattern.

Clearly, none of the options presented in this section solve the problem of where on the line of solutions the true estimates of APC effects lie. In order to do that, researchers need to make some assumptions in order to get estimates of APC that are contingent on those assumptions.

5. What can we do? part 2: with assumptions

Assumptions form an important part of any statistical modelling strategy. However, for APC analysis, if we want to estimate APC linear effects, we need to make particularly strong assumptions. We will see that these assumptions tend to be

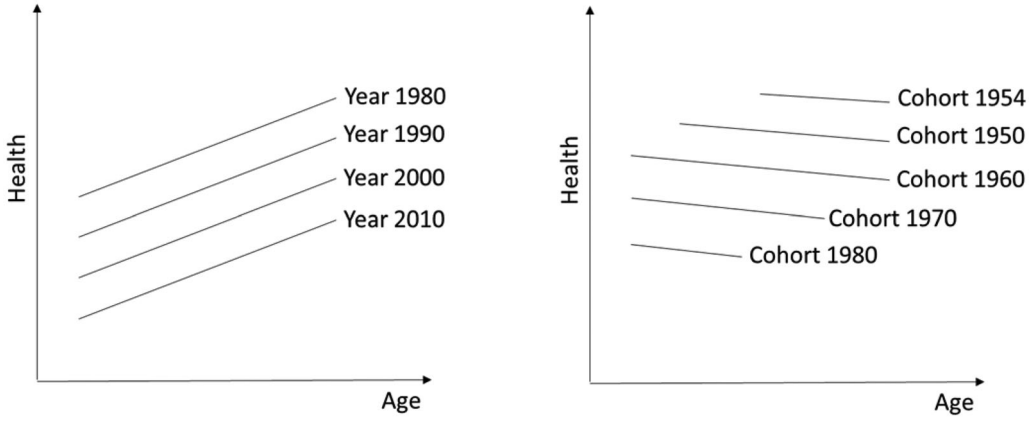

Figure 3. Plots of a hypothetical dataset, plotting health by (a) age and year, and (b) age and cohort. Although the data that produced these plots are the same, the interpretation of the two graphs is very different.

quite big, even if they don't look that way at first glance. In order to establish the true point on the line of solutions, we need to know a constraint that is true in the data generating process, which in practice often means we have to know the true linear effect of one of APC in order to establish the value of the other two.

5.1. Apparently innocuous assumptions

One apparent way around this problem is to treat APC as categorical, rather than continuous, variables. Because the identification problem only applies to linear effects, categorical variables can be easily manipulated so that they are no longer estimating exactly linearly dependent effects. There might be a case of grouping one of APC in courser groups than the others. In the sociological literature, it is common to see cohorts grouped in 5-year intervals, whilst period and age are grouped in single years (for instance, see Yang and Land 2006). Alternatively, one might group based on theory: for instance, group cohorts based on groupings like Baby Boomers, Generation X, Millennials and so on, that it is postulated share characteristics. One doesn't even need to go that far. Simply grouping two adjacent years in one of APC together would have the desired effect of breaking the linear dependency between APC and so breaking the identification problem. For instance, one could simply group the ages 18 and 19 together and label them both 19 – introducing that small amount of measurement error breaks the identification problem and allows estimates to be calculated, either using linear effects as in Equation (2), or by treating the variables as categorical with a dummy variable for each category of APC less a reference category. This approach was suggested (although he warned of the dangers of it) by Mason et al. (1973), and subsequently by Robertson and Boyle (1986).

The problem is that apparently innocuous assumptions such as these are actually not innocuous at all, and changing which of the APC categories are constrained to equal each other can have a vast effect on the results found (Glenn 1976; Osmond and Gardner 1989). Fundamentally, different constraints will result in different solutions being found – different points on the line of solutions being chosen as the estimates for APC – and will produce identical fit statistics (Fosse and Winship 2019a).

However, in some cases this approach can be useful (Thijs et al. 2020). For instance, Mason and Smith (1985; Smith 2020), in their study of tuberculosis, assume that age groups age 5–9 and age 10–14 are equal to each other, on the basis of theory. However, in other studies these constraints are arbitrary and atheoretical (Fosse and Winship 2019a), meaning that the results that are found are arbitrary and atheoretical also. The danger of these assumptions is that by appearing innocuous, they imply to readers that the assumption is harmless if broken slightly, which is incorrect. It is therefore often better to make a stronger assumption, to be explicit about exactly what the assumption is that is being made.

5.2. Strong assumptions about one of APC

An alternative assumption is one which states the value of the linear effect of one of APC. Thus we might assume, for instance, that the period trend is flat (that β_P in Equation (2) is equal to zero). Whilst often in these situations we assume a zero trend for one of APC, one could just as easily assume a particular value (for instance that β_A in Equation (2) is equal to 1, for instance), perhaps because of a medical understanding of what we would expect the age trend to be.

Such an assumption would not assume that there are no period effects – we could still estimate discrete period effects around a zero trend, for instance. However, it remains a strong assumption that we might often be unwilling to make. Nevertheless, it is often better to make an assumption like this and state it explicitly, rather than pretend you have not made an assumption when in reality you have. For instance, Bell (2014) assumes a zero trend in his study of mental health over the life course, and shows how that assumption changes the estimated life course effect of mental health from a U-shape (where mental health worsens until midlife, then improves) to a more continuous decline in mental health across the life course, which is argued to be more plausible. Similarly, Delaruelle et al. (2015) make the same assumption of no period effect in their study of self-rated health, finding a strong age effect, but no cohort effects.

This approach is good, but again only if one is willing to make a big assumption about one of APC. In the cases identified above, big assumptions were made and stated explicitly, and those assumptions were not uncontroversial. Ideally, we would want to make less strong assumptions. We discuss an approach to this below.

5.3. Partial assumptions about one or more of APC

In some situations, we may not be able to say exactly what we expect the value of the linear age, period or cohort effect estimates to be. However, we might be able to say, on the basis of theory, something about those effects as an inequality. For instance, we might be able to say that we expect the age effect to be greater than or equal to zero, where it would make little theoretical sense for the effect of age to be negative. It may be that we can say something like this about more than one of APC. In the process, we can potentially limit the range of the line of solution. This approach is suggested by Fosse and Winship (2019b). For example, taking the line of solutions depicted in Figure 1, if we knew that both the age and period effects were less than zero (i.e. negative), that would mean we could restrict the possible values of the line of solutions to a relatively small range of values for the APC effects. In a health context, we might be willing to assume that as we get older, our mortality rate will not fall (i.e. the age effect is negative). We might also assume that period and cohort effects are non-negative, given progress in health over time. The problem, of course, is getting those assumptions right, and having enough theoretical

knowledge to be able to make them. Taking the mortality example, a worsening of diet in the western world accompanying the obesity epidemic might have an effect of increasing mortality, which would result in us breaking the assumptions made above and getting incorrect estimates for the APC linear effects.

5.4. Mechanism-based approaches

An alternative, taking the view that APC is an accounting model, providing a framework for the variables that actually cause APC effects, is to actually model based on the variables that the researcher believes are behind the effects. That is, if you believe that differences between different cohorts are driven by medical advances over time, one could include a measure of medical technology instead of the measure of cohorts. Because these variables are (usually) not exactly correlated with any of APC, this avoids the identification problem. Not only that, but if these variables measuring the APC mechanisms explain the entirety of one of APC, then one can identify the other two of APC as well. This approach is suggested by Winship and Harding (2008).

Whilst this works when we know exactly what variables are likely to produce effects of at least one of APC, it isn't so useful in situations where we don't already understand at least some of the APC processes fairly well. For instance, it is often very difficult to know whether a variable that has caused an improvement in an outcome does so through change across periods or across cohorts. If we fail to explain correctly how all of one of A, P or C effects are produced, we will not be able to estimate the APC effects accurately.

6. What we shouldn't we do: arbitrary and hidden assumptions

We need to be careful to avoid making assumptions that are not clear to the reader. That is, we want the assumptions that we make to be open to scrutiny and need to be clear that the assumptions have a big effect on the results we find. At this point, the paper diverges into somewhat controversial academic debates – whilst I give my view of these debates, I encourage the reader to read alternative viewpoints and make up their own mind (see table 1 of Bell and Jones 2018 for examples).

Unfortunately, in my view, there have been a number of APC models proposed that, often shrouded in statistical legerdemain, claim to provide a mechanical solution to the identification problem that is not reliant on theoretically driven assumptions. These include the Hierarchical APC model (Yang and Land 2006, 2013), the Intrinsic Estimator (Fu 2018; Yang et al. 2004, 2008), Partial Least Squares (Tu et al. 2011), and a number of others. A number of these have (in my view) been shown not to work in plausible data driven scenarios, and (given it is difficult to prove the plausibility of simulated data) with real data (e.g. Bell and Jones 2018). Crucially, they have been shown to make assumptions that are not driven by the processes that produced the data, or any theory. The Hierarchical APC model usually assumes that there are no cohort effects (Bell and Jones 2014a, 2014c, 2018, 2020; Luo and Hodges 2016, 2019). The Intrinsic Estimator and the Partial Least Squares approaches have been shown to produce results that take the point closest to zero for each of the three effects – that is the point on the line of solutions that is perpendicular to the (0,0,0) point on a 3D graph (Luo 2013; O'Brien 2015b; Pelzer et al. 2015; Te Grotenhuis et al. 2016). In each case, the results produced will likely be correct if the assumption is correct. However, the assumption is not only left untested, but is actually hidden by the statistical modelling that the authors use. It is for this reason that these methods are somewhat dangerous to applied researchers who might be seduced by the statistical complexity of the models suggested. The search for a solution to the identification problem that requires no assumptions is a "futile" (Glenn 1976) or "unholy" (Fienberg 2013) quest, and "one of the most bizarre instances in the history of science of repeated attempts to do something that is logically impossible" (Glenn 2005, p. 6).

7. Conclusions

This article may seem odd – taking a long time discussing relatively well-known APC concepts without the usual mathematics, and then confining some of the most used APC methods to a single, rather dismissive paragraph. The reason for this is that poorly conducted APC methods are not the result of individuals having an incomplete grasp of complex statistical methods – rather it is because of a lack of understanding of the conceptual basis of those methods – what APC effects are, how they are linked, and what they mean. When those are understood, the method used becomes less important, and some claims about the suitability of some methods become harder to justify.

It is worth considering, before going on, why APC analysis operates in a particularly weird space. In most social science and scientific disciplines, APC have been fundamental to understanding society. And yet, methods for empirically understanding and disaggregating APC effects are rarely taught, often seen as an 'advanced' method beyond the abilities of the average applied social science and health researcher. The result is analyses that consider one or more of APC, but either ignore the identification problem entirely, or look for 'false prophets' – magical solutions to the identification problem. Often these magical solutions are veiled in the language of statistics – difficult to decipher even for a relatively well-trained quantitative researcher – making it difficult to assess whether those methods are appropriate or not (as argued by Carstensen 2007). The result is misleading and incorrect, for which the researcher can hardly be blamed. As such, I call for APC analysis to become a central part of research methods in any discipline involving the study of humans – not in the form of complex equations (note the lack of such complex equations in this paper!) but an understanding of the concepts that underlie these fundamental building blocks of society.

The overall conclusion of this paper is somewhat sobering for both researchers and statisticians. First, there is no

solution to the identification problem that doesn't require a strong degree of theoretical knowledge about at least one of the APC effects that are being estimated. Second, many of the methods proposed, that 'work' when judged on the basis of conventional model fit statistics, are actually extremely problematic. And third, APC analysis is not something just for the statisticians – it's something all applied researchers with an interest in how things change over time need to take seriously, to ensure they make correct inference.

Fortunately, there are a number of theoretically grounded, sensible ways forward for the applied researcher. None solve the identification problem; but they do provide frameworks through which researchers can apply their theoretical understanding of the research questions at hand to produce valid and useful inference.

Acknowledgements

Thanks to Will Johnson, Phil Jones, and the audience at the SSHB conference in Oxford, September 2019, for their comments on earlier iterations of this paper.

Disclosure statement

The author reports no conflict of interest.

ORCID

Andrew Bell http://orcid.org/0000-0002-8268-5853

References

Bell A. 2014. Life course and cohort trajectories of mental health in the UK, 1991–2008: a multilevel age-period-cohort analysis. Soc Sci Med 120:21–30.

Bell A, Jones K. 2014a. Another 'futile quest'? A simulation study of Yang and Land's Hierarchical Age-Period-Cohort model. Dem Res 30: 333–360.

Bell A, Jones K. 2014b. Current practice in the modelling of age, period and cohort effects with panel data: a commentary on Tawfik et al (2012), Clarke et al (2009), and McCulloch (2012). Qual Quant 48(4): 2089–2095.

Bell A, Jones K. 2014c. Don't birth cohorts matter? A commentary and simulation exercise on Reither, Hauser and Yang's (2009) age-period-cohort study of obesity. Soc Sci Med 101:176–180.

Bell A, Jones K. 2018. The hierarchical age–period–cohort model: why does it find the results that it finds? Qual Quant 52(2):783–799.

Bell A, Jones K. 2020. Multilevel models for age-period-cohort analysis. In: Bell A, editor. Age, period and cohort effects: the identification problem, and what to do about it. London Routledge.

Carstensen B. 2007. Age–period–cohort models for the Lexis diagram. Statist Med 26(15):3018–3045.

Chauvel L, Leist AK, Ponomarenko V. 2016. Testing persistence of cohort effects in the epidemiology of suicide: an age-period-cohort hysteresis model. Plos One 11(7):e0158538–20.

Chauvel L, Leist AK, Smith HL. 2020. Detecting the 'Black Hole' of age-period excess mortality in 25 countries: age-period-cohort residual analysis. In: Bell A, editor. Age, period and cohort effects: the identification problem, and what to do about it. London: Routledge.

Clayton D, Schifflers E. 1987. Models for temporal variation in cancer rates. II: age–period–cohort models. Statist Med 6(4):469–481.

Delaruelle K, Buffel V, Bracke P. 2015. Educational expansion and the education gradient in health: a hierarchical age-period-cohort analysis. Soc Sci Med 145:79–88.

Dimock M. 2019. Where millennials end and generation Z begins | Pew Research Center. [accessed 2019 Aug 19]. https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/.

Fienberg SE. 2013. Cohort analysis' unholy quest: a discussion. Demography 50(6):1981–1984.

Firebaugh G. 1989. Methods for estimating cohort replacement effects. Sociol Methodol 19:243–262.

Firebaugh G. 1990. Replacement effects, cohort and otherwise: response to Rodgers. Sociol Methodol 20:439–446.

Firebaugh G. 2008. Seven rules for social research. Princeton, NJ: Princeton University Press.

Fosse E, Winship C. 2019a. Analyzing age-period-cohort data: a review and critique. Annu Rev Sociol 45(1):467–492.

Fosse E, Winship C. 2019b. Bounding analyses of age-period-cohort models. *Demography* in press. [accessed 2018 Oct 31)]. https://q-aps.princeton.edu/sites/default/files/q-aps/files/apcanalysis.pdf.

Frost WH. 1940. The age selection of mortality from tuberculosis in successive decades. Milbank Mem Fund Q 18(1):61–66.

Fu WJJ. 2018. A practical guide to age period cohort analysis. Boca Raton, FL: CRC Press.

Glenn ND. 1976. Cohort analysts' futile quest: statistical attempts to separate age, period and cohort effects. Am Soc Rev 41(5):900–904.

Glenn ND. 2005. Cohort analysis. 2nd ed. London: Sage.

Goldstein H. 1979. Age, period and cohort effects – a confounded confusion. Appl Stat 6(1):19–24.

Hanson M, Gluckman P. 2016. Commentary: developing the future: life course epidemiology, DOHaD and evolutionary medicine. Int J Epidemiol 45(4):993–996.

Holford TR. 1983. The estimation of age, period and cohort effects for vital rates. Biometrics 39(2):311–324.

Holford TR. 1985. An alternative approach to statistical age-period-cohort analysis. J Chronic Dis 38(10):831–840.

Holford TR. 1991. Understanding the effects of age, period, and cohort on incidence and mortality rates. Annu Rev Public Health 12(1): 425–457.

Holford TR. 2006. Approaches to fitting age-period-cohort models with unequal intervals. Statist Med 25(6):977–993.

Jones PM, Minton J, Bell A. 2018. Period and cohort changes in mortality risk over the twentieth century in the UK: an exploratory analysis. OSF Preprints. [accessed 2020 Jan 2]. https://osf.io/4f7jr/.

Keiding N. 2011. Age-period-cohort analysis in the 1870s: diagrams, stereograms, and the basic differential equation. Can J Statistics 39(3): 405–420.

Luo L. 2013. Assessing validity and application scope of the intrinsic estimator approach to the age-period-cohort problem. Demography 50(6):1945–1967.

Luo L, Hodges JS. 2016. Block constraints in age-period-cohort models with unequal-width intervals. Soc Method Res 45(4):700–726.

Luo L, Hodges JS. 2019. Constraints in random effects age-period-cohort models. *arXiv*. [accessed 2019 Aug 19]. http://arxiv.org/abs/1904.07672.

Mannheim K. 1928. The problem of generations. In: Kecskemeti P (ed.) Karl mannheim: essays on the sociology of knowledge. London: Routledge; p. 276–322. [accessed 2019 Aug 21] http://marcuse.faculty.history.ucsb.edu/classes/201/articles/27MannheimGenerations.pdf.

Mason K O, Mason W M, Winsborough H. H, Poole W. K. 1973. Some methodological issues in cohort analysis of archival data. Am Sociol Rev 38(2):242–258.

Mason WM, Smith HL. 1985. Age-period-cohort analysis and the study of deaths from pulmonary tuberculosis. In: Cohort analysis in social research. New York, NY: Springer New York; p. 151–227.

Minton J. 2016. Visualising data in 3D: Handling complexity through visceral and tactile experiences of data. [accessed 2019 Aug 19]. https://blogs.lse.ac.uk/impactofsocialsciences/2016/07/01/visualising-data-through-3d-sculptures/.

O'Brien RM. 2011. Constrained estimators and age-period-cohort models. Sociol Method Res 40(3):419–452.

O'Brien RM. 2014. Estimable functions in age-period-cohort models: a unified approach. Qual Quant 48(1):457–474.

O'Brien RM. 2015a. Age-period-cohort models: approaches and analyses with aggregate data. Boca Raton, FL: CRC Press.

O'Brien RM. 2015b. Age–period–cohort models and the perpendicular solution. Epidemiol Method 4(1):87–99.

O'Brien RM. 2016. Model misspecification when eliminating a factor in age-period-cohort multiple classification models. Sociol Methodol 46(1):358–372.

O'Brien RM. 2020. The line of solutions and understanding age-period-cohort models. In: Bell A, editor. Age, period and cohort effects: the identification problem, and what to do about it2. London: Routledge.

Osmond C, Gardner MJ. 1989. Age, period, and cohort models: non-overlapping cohorts dont resolve the identification problem. Am J Epidemiol 129(1):31–35.

Pelzer B, Te Grotenhuis M, Eisinga R, Schmidt-Catran AW. 2015. The non-uniqueness property of the intrinsic estimator in APC models. Demography 52(1):315–327.

Robertson C, Boyle P. 1986. Age, period and cohort models: the use of individual records. Statist Med 5(5):527–538.

Robertson C, Boyle P. 1998. Age-period-cohort analysis of chronic disease rates. I: modelling approach. Statist Med 17(12):1305–1323.

Robertson C, Gandini S, Boyle P. 1999. Age-period-cohort models: a comparative study of available methodologies. J Clin Epidemiol 52(6): 569–583.

Rodgers WL. 1982. Estimable functions of age, period, and cohort effects. Am Sociol Rev 47(6):774–787.

Rodgers WL. 1990. Interpreting the components of time trends. Sociol Methodol 20(1990):421–438.

Rosenberg PS, Check DP, Anderson WF. 2014. A web tool for age-period-cohort analysis of cancer incidence and mortality rates. Cancer Epidemiol Biomarkers Prev 23(11):2296–2302.

Ryder NB. 1965. The cohort as a concept in the study of social-change. Am Sociol Rev 30(6):843–861.

Smith HL. 2020. Age-period-cohort analysis: what is it good for? In: Bell A, editor. Age, period and cohort effects: the identification problem, and what to do about it. London: Routledge.

Te Grotenhuis M, Pelzer B, Luo L, Schmidt-Catran AW. 2016. The intrinsic estimator, alternative estimates, and predictions of mortality trends: a comment on masters, hummer, powers, Beck, Lin, and Finch. Demography 53(4):1245–1252.

Thijs P, Te Grotenhuis M, Scheepers P. 2020. The Pros and Cons of Constraining Variables. In: Bell A, editor. Age, period and cohort effects: the identification problem, and what to do about it. London: Routledge.

Tu YK, Smith GD, Gilthorpe MS. 2011. A new approach to age-period-cohort analysis using partial least squares regression: the trend in blood pressure in the Glasgow alumni cohort. Plos One 6(4):e19401.

Winship C, Harding DJ. 2008. A mechanism-based approach to the identification of age-period-cohort models. Sociol Method Res 36(3): 362–401.

Yang Y, Land KC. 2006. A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores. Sociol Methodol 36(1):75–97.

Yang Y, Land KC. 2013. Age-period-cohort analysis: new models, methods, and empirical applications. Boca Raton, FL: CRC Press.

Yang Y, Fu WJJ, Land KC. 2004. A methodological comparison of age-period-cohort models: the intrinsic estimator and conventional generalized linear models. Sociol Methodol 34(1):75–110.

Yang Y, Schulhofer-Wohl S, Fu WJ, Land KC. 2008. The intrinsic estimator for age-period-cohort analysis: what it is and how to use it. Am J Sociol 113(6):1697–1736.