



Published in final edited form as:

Soc Sci Med. 2015 November ; 145: 125–128. doi:10.1016/j.socscimed.2015.07.013.

Clarifying hierarchical age–period–cohort models: A rejoinder to Bell and Jones

Eric N. Reither^{a,*}, Kenneth C. Land^b, Sun Y. Jeon^a, Daniel A. Powers^c, Ryan K. Masters^d, Hui Zheng^e, Melissa A. Hardy^f, Katherine M. Keyes^g, Qiang Fu^h, Heidi A. Hansonⁱ, Ken R. Smith^j, Rebecca L. Utz^k, and Y. Claire Yang^l

^aDepartment of Sociology and the Yun Kim Population Research Laboratory, Utah State University, 0730 Old Main Hill, Logan UT 84322-0730, USA

^bDepartment of Sociology and Center for Population Health and Aging, Duke University, USA

^cDepartment of Sociology, Population Research Center, The University of Texas at Austin, USA

^dDepartment of Sociology and Institute of Behavioral Science, University of Colorado at Boulder, USA

^eDepartment of Sociology, The Ohio State University, USA

^fDepartment of Sociology and Criminology, Population Research Institute, The Pennsylvania State University, USA

^gDepartment of Epidemiology, Mailman School of Public Health, Columbia University, USA

^hDepartment of Sociology, The University of British Columbia, Canada

ⁱDepartment of Family and Preventive Medicine and the Huntsman Cancer Institute, The University of Utah, USA

^jDepartment of Family and Consumer Studies and the Huntsman Cancer Institute, The University of Utah, USA

^kDepartment of Sociology, The University of Utah, USA

^lDepartment of Sociology and the Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, USA

Abstract

Previously, Reither et al. (2015) demonstrated that hierarchical age–period–cohort (HAPC) models perform well when basic assumptions are satisfied. To contest this finding, Bell and Jones (2015) invent a data generating process (DGP) that borrows age, period and cohort effects from different equations in Reither et al. (2015). When HAPC models applied to data simulated from this DGP fail to recover the patterning of APC effects, B&J reiterate their view that these models provide “misleading evidence dressed up as science.” Despite such strong words, B&J show no curiosity about their own simulated data—and therefore once again misapply HAPC models to data that violate important assumptions. In this response, we illustrate how a careful analyst could

*Corresponding author. eric.reither@usu.edu (E.N. Reither).

have used simple descriptive plots and model selection statistics to verify that (a) period effects are not present in these data, and (b) age and cohort effects are conflated. By accounting for the characteristics of B&J's artificial data structure, we successfully recover the “true” DGP through an appropriately specified model. We conclude that B&J's main contribution to science is to remind analysts that APC models will fail in the presence of exact algebraic effects (i.e., effects with no random/stochastic components), and when collinear temporal dimensions are included without taking special care in the modeling process. The expanded list of coauthors on this commentary represents an emerging consensus among APC scholars that B&J's essential strategy—testing HAPC models with data simulated from contrived DGPs that violate important assumptions—is not a productive way to advance the discussion about innovative APC methods in epidemiology and the social sciences.

Keywords

Age–period–cohort models; Cohort effects; Simulation models; Hierarchical modeling; Random effects; Body mass index; Obesity epidemic; Social change

We would like to thank Bell and Jones (henceforth B&J) for highlighting some areas of agreement in their commentary (2015) on our recent defense of innovative age–period–cohort (APC) methods (Reither et al., 2015). To our understanding, we have achieved consensus on the following points:

- APC data are generated by historical and social processes, not by algebraic functions invented in data laboratories.
- In the presence of exact linear effects, traditional specifications of APC models will fail.
- From a conceptual standpoint, it is appropriate to treat birth cohorts and periods of observations as random effects.
- Although theory and speculation are useful, they cannot replace evidence in any field of scientific inquiry.
- Given patterns of BMI change in the U.S. population, it is unlikely that birth cohort effects are principally responsible for the obesity epidemic.

These areas of agreement could help narrow the divide between innovators and antagonists of APC research. However, we continue to have significant concerns with certain arguments and data structures advanced by B&J, and we thank the editors at *Social Science & Medicine* for granting us this opportunity to address them.

1. Another inappropriate data structure for APC models

In our recent assessment of B&J's critiques (Reither et al., 2015), we found that the purported failures of hierarchical APC (HAPC) models occur not because of inherent problems with the method, but rather because B&J subject them to unrealistic data structures that violate key assumptions. We concluded that “when more reasonable assumptions are employed – that is, when associational nonlinearities are permitted in a data structure that is

truly three-dimensional – the HAPC-CCREM approach consistently performs well” (Reither et al., 2015, 9). Ostensibly to test our claim, B&J (2015) invent a data structure that borrows the period trend from our second equation and the age trend from our fourth equation. They also create a cohort trend that, curiously, is based on the pattern of period effects shown in Figure 4 of (Reither et al., 2015). The impression B&J may create for some readers is that, by borrowing age, period and cohort trends from our models, they have created a data structure that satisfies the assumptions of APC modeling, thereby providing an opportunity to “scrutinize” the HAPC method via the specification of a complete HAPC model.

But what kind of a data structure have they actually created? To address this question, we use the data generating process (DGP) provided by B&J to simulate 100 datasets, each with 20,000 observations. We then use simple tools (descriptive and model selection statistics) that are available to any analyst to help inform our modeling choices. To gather a preliminary sense of the data structure, we plot descriptive estimates of obesity prevalence by age, period and birth cohort for each dataset. As shown in Fig. 1, the age and cohort trends are mirror images of each other, and the period trends are flat, consistent with the second equation in Reither et al. (2015). The obvious similarity in age and cohort patterns is obfuscated in the way that B&J present the age DGP, which is centered at the grand mean (generally about age 49 in the simulated data) in their DGP equation but – for reasons that are obscure – at age 40 in their Fig. 1 (2015). When the quadratic age DGP is graphically represented in the same way that it is modeled (i.e., centered at the grand mean), it also resembles the cohort DGP that B&J invented (Fig. 2). Indeed, a quadratic equation explains 99.8% of the variation in the cohort DGP, further suggesting that the cohort effects may not be distinct from the age effects in this data structure. Therefore, given these features of the DGP, it is inappropriate to simply move forward and specify a HAPC model (or any APC model) with three operative temporal dimensions – and B&J’s commentary (2015) rests upon a house of cards.

2. The appropriate use of model selection statistics

If B&J had incorporated model selection statistics and descriptive plots of the APC effects, they should have noticed that their DGP is not three-dimensional. For the 100 datasets we simulated from the B&J DGP, we estimated AIC and BIC statistics for various APC models (Table 1). *In no instance do these model selection statistics point to a fully three-dimensional data structure.* Instead, for a large majority of datasets that we simulated, both AIC and BIC indicate that a model with polynomial age and cohort terms best fits the data. This is consistent with our preliminary observations of the descriptive age–period–cohort plots and the DGPs that B&J invented. Although the data structure appears to be essentially one-dimensional and dominated by the age process – as suggested by BIC for 11 out of 100 simulated datasets in Table 1, and for all 100 datasets when the quadratic cohort term is omitted – random variability around the cohort DGP produces sufficient differences from the age DGP in most datasets to warrant inclusion of this second quadratic dimension.¹

¹This point is affirmed by deviance statistics ($-2 \log$ likelihood) for 10 datasets with age and cohort estimates that lie near the overall median estimates (see Fig. 2). In comparison to age-only models, models that include the quadratic specification of cohort effects reduce the deviance by an average of 28.6 (range = 24.7–32.8), which is significant at 2 degrees of freedom ($p < 0.001$).

In their commentary, B&J (2015) assert that discrepancies across model selection statistics means that they have a tendency to “find the incorrect answer” (332). Although it is true that model selection statistics such as AIC and BIC will occasionally disagree and/or point to the incorrect model, this is not particularly surprising or unique to APC research.² Moreover, B&J fail to appreciate that relatively small random period and/or cohort effects can be absorbed into the error term of the level-2 specification. Indeed, this is what is done in any hierarchical model application when it is concluded that the level-2 effects are not significant. The essential point is that model selection statistics are imperfect but nevertheless very useful tools in APC research as elsewhere; when used in conjunction with descriptive plots, they can help researchers determine which temporal dimensions are *most likely present* in the data. If model selection statistics consistently indicate that fewer than three temporal dimensions are present, then specification of a complete APC model will sometimes yield invalid results if other important assumptions of APC models are violated (e.g., the presence of algebraic effects and/or strong collinearity between temporal dimensions). To make judicious modeling decisions, researchers should carefully cross-assess/cross-validate model selection statistics with descriptive age–period–cohort plots to gather preliminary information about the data structure, as we have done here.

3. Modeling data generated from B&J's DGP

When we apply a complete hierarchical APC model to these data, we obtain results similar to those presented by B&J and fail to detect the true DGP. However, when we make wiser modeling choices based on the information we gather about the data structure through the descriptive plots and model selection statistics, we can successfully replicate the true DGP. Using the `glm` function in R (R Core Team, 2015), we estimate fixed-effects models with quadratic terms for both age and cohort, as detected by AIC and BIC for 9 out of every 10 datasets that we simulated. As shown in Fig. 2, these age–cohort models successfully capture the “true” DGP invented by B&J.

4. Conclusions

There are other areas where we take exception to B&J's commentary, such as their obfuscation of algebraic effects and their continued misrepresentation of identification issues. Unfortunately, we do not have space to address all such issues here. The most important point for APC researchers is that, contrary to the title of their commentary, B&J (2015) provide no meaningful “scrutiny” of innovative APC methods. Instead, B&J once again misapply HAPC models to a homebrewed data structure that not only violates the assumptions of APC research but also finds no resemblance in empirical data that have actually been used in cohort studies. As in their previous work, B&J succeed primarily in showing numerically that APC models—like any statistical method—will fail when basic assumptions are violated. Moreover, it is worth reminding practitioners of statistical analysis of a basic rule: simulation studies involving numeric examples can be arbitrary and are never sufficient to validate (or invalidate) an estimation method in the absence of formal

²Model selection statistics are statistics – that is, random variables subject to probability distributions, which thus can take on numerical values that have low probabilities of occurring.

algebraic proof. Unless an estimator violates statistical assumptions and is thus biased, inefficient or asymptotically inconsistent, results from data simulation should not be accepted as the yardstick against which the validity of a method per se is evaluated.

The most useful contribution that B&J have provided for the APC research community is to remind us that it is inadvisable to model more temporal dimensions than are present in the data (though sometimes the true DGP can still be recovered, as shown by our previous simulations) and that APC models will fail (a) in the presence of exact algebraic effects, and (b) when highly collinear temporal dimensions are included without taking special care in the modeling process (as we have done here through the quadratic cohort specification). Researchers can avoid these pitfalls that B&J willingly fall into by carefully examining descriptive age–period–cohort plots and model selection statistics to help ensure that the specification of complete APC models is appropriate in their applications.

References

- Bell A, Jones K. Should age-period-cohort analysts accept innovation without scrutiny? A response to Reither, Masters, Yang, Powers, Zheng and Land. *Soc Sci Med.* 2015; 128:331–333. [PubMed: 25641207]
- R Core Team. R Foundation for Statistical Computing; Vienna, Austria: 2015. R: A Language and Environment for Statistical Computing. <http://www.R-project.org>
- Reither EN, Masters RK, Yang YC, Powers DA, Zheng H, Land KC. Should age-period-cohort studies return to the methodologies of the 1970s? *Soc Sci Med.* 2015; 128:356–365. [PubMed: 25617033]

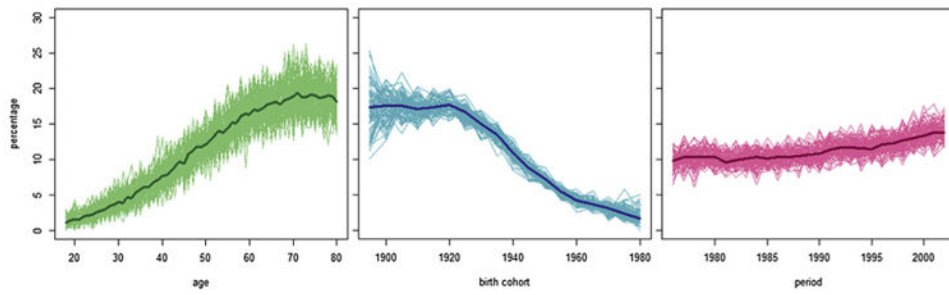


Fig. 1. Descriptive estimates of obesity prevalence by age, period and birth cohort in 100 datasets simulated from the DGP invented by B&J (median estimates in bold).

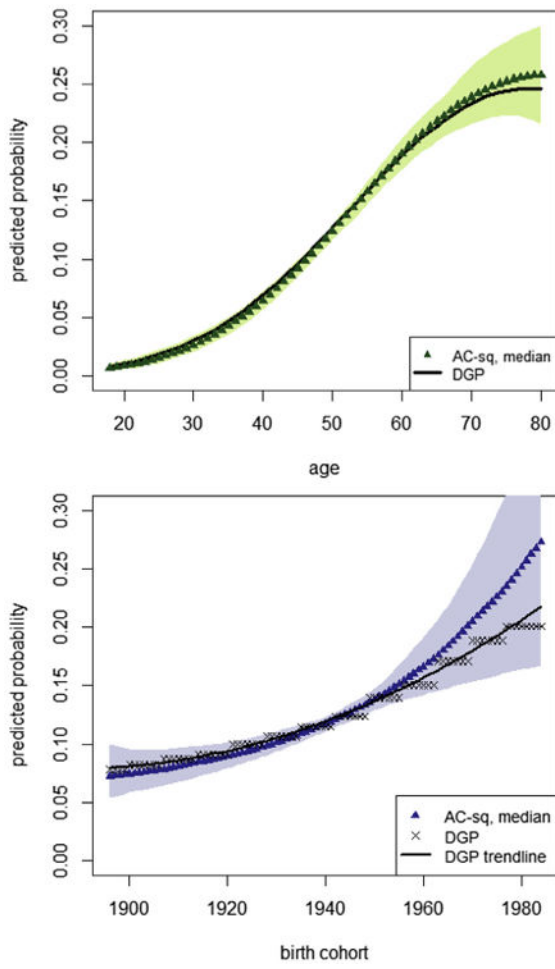


Fig. 2. Estimates from quadratic age-cohort models applied to data simulated from the DGP invented by B&J, with 95% point intervals in shaded areas.

Summary findings from model selection statistics for 100 datasets simulated from B&J's DGP.

Table

Traditional models of obesity (only age is quadratic)							Age + cohort (both quadratic)	
Age	Period	Cohort	Age + cohort	Age + period	Period + cohort	Age + period + cohort		
AIC	0	0	2	4	0	0	0	94
BIC	11	0	0	0	0	0	0	89
<i>df</i>	3	27	18	20	29	44	46	5
N	20,000	20,000	20,000	20,000	20,000	20,000	20,000	20,000

Notes: AIC is the Akaike Information Criterion and is estimated to be $-2 \cdot \log(L) + df \cdot 2$. BIC is the Bayesian Information Criterion and is estimated to be $-2 \cdot \log(L) + df \cdot \log(N)$. For AIC/BIC, each cell in the table represents the total number of simulated datasets that AIC/BIC identify as the "best fitting" model specification.