



Published in final edited form as:

Soc Sci Med. 2015 March ; 128: 356–365. doi:10.1016/j.socscimed.2015.01.011.

Should Age-Period-Cohort Studies Return to the Methodologies of the 1970s?

Eric N. Reither¹, Ryan K. Masters², Y. Claire Yang³, Daniel A. Powers⁴, Hui Zheng⁵, and Kenneth C. Land⁶

¹Department of Sociology and the Yun Kim Population Research Laboratory, Utah State University, 0730 Old Main Hill, Logan UT 84322-0730, eric.reither@usu.edu, 435-797-1217

²Department of Sociology and Institute of Behavioral Science, University of Colorado at Boulder

³Department of Sociology and the Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill

⁴Department of Sociology, Population Research Center, The University of Texas at Austin

⁵Department of Sociology, The Ohio State University

⁶Department of Sociology and Center for Population Health and Aging, Duke University

Abstract

Social scientists have recognized the importance of age-period-cohort (APC) models for half a century, but have spent much of this time mired in debates about the feasibility of APC methods. Recently, a new class of APC methods based on modern statistical knowledge has emerged, offering potential solutions. In 2009, Reither, Hauser and Yang used one of these new methods – hierarchical APC (HAPC) modeling – to study how birth cohorts may have contributed to the U.S. obesity epidemic. They found that recent birth cohorts experience higher odds of obesity than their predecessors, but that ubiquitous period-based changes are primarily responsible for the rising prevalence of obesity. Although these findings have been replicated elsewhere, recent commentaries by Bell and Jones call them into question – along with the new class of APC methods. Specifically, Bell and Jones claim that new APC methods do not adequately address model identification and suggest that “solid theory” is often sufficient to remove one of the three temporal dimensions from empirical consideration. They also present a series of simulation models that purportedly show how the HAPC models estimated by Reither et al. (2009) could have produced misleading results. However, these simulation models rest on assumptions that there were no period effects, and associations between period and cohort variables and the outcome were perfectly linear. Those are conditions under which APC models should never be used. Under more tenable assumptions, our own simulations show that HAPC methods perform well, both in recovering the main findings presented by Reither et al. (2009) and the results

© 2015 Elsevier Ltd. All rights reserved.

Correspondence to: Eric N. Reither.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

reported by Bell and Jones. We also respond to critiques about model identification and theoretically-imposed constraints, finding little pragmatic support for such arguments. We conclude by encouraging social scientists to move beyond the debates of the 1970s and toward a deeper appreciation for modern APC methodologies.

Keywords

age-period-cohort models; cohort effects; research methods; hierarchical modeling; random effects; body mass index; obesity epidemic; social change

Introduction

Shortly after Norman Ryder's classic treatise about the influence of birth cohorts on social change (Ryder, 1965), debates emerged over the simultaneous estimation of age, period and cohort (APC) effects in the social sciences. Some scholars argued that confounding between APC effects (the so-called identification problem) was intractable (Glenn, 1976), requiring the elimination of one of three temporal dimensions in APC analyses (Baltes, 1968). Other scholars challenged such views, contending that identification problems could be reasonably addressed (Pullum, 1980), and that it was not only possible to estimate APC models, but necessary in many instances to avoid erroneous conclusions (K. O. Mason et al., 1973).

These debates persisted with little resolution for over thirty years. But over the past decade, scholars have taken note of the fact that the concepts and tools of statistics have evolved since the 1970s and, by applying these, have developed several innovative methodologies in APC research (Yang & Land, 2013a) that could deliver more reliable tools of discovery to social scientists. Capitalizing on these developments, Reither, Hauser and Yang (2009) examined the separate contributions of secular change (periods of observation), biological age, and birth cohort membership to the U.S. obesity epidemic. Using a methodology called hierarchical APC estimation with cross-classified random effects modeling (HAPC-CCREM), the study found that younger birth cohorts were generally at higher risk for obesity than their predecessors, even after accounting for age and period of observation. However, recent increases in obesity prevalence were primarily attributable to sweeping, period-based secular changes that have affected virtually all Americans. In a commentary on this study, Harding remarked that new APC approaches like HAPC-CCREM "have the potential to reinvigorate research investigating age, period, and cohort effects and increase our confidence in APC analyses" but also cautioned that such techniques "have yet to be widely adopted or evaluated by other researchers or methodologists" (Harding, 2009, 1451).

In a recent issue of *Social Science & Medicine*, Bell and Jones (hereafter, B&J) challenge the widespread adoption of new APC techniques (2014b) through a critical evaluation of the HAPC-CCREM approach used by Reither et al. (2009) in their study of obesity. Specifically, B&J argue that (1) Reither et al. did not adequately address the identification problem, (2) it is plausible that the U.S. obesity epidemic is rooted in cohort-based changes that occurred throughout the 20th century, and (3) the HAPC models estimated by Reither et al. *could* have generated misleading results – although they stop short of asserting that this is actually the case.

Through the following discussion, we address each of these critiques in turn. The arguments and simulation models in our study have important implications, not only for the findings reported by Reither et al. (2009) and the challenges raised by B&J, but also for the future of HAPC modeling and innovative APC methods in general.

The Identification Problem

Discussing the identification problem in APC models, B&J (2013; 2014b, 177) cite the identity: $\text{Age} = \text{Period} - \text{Cohort}$, and state: “As such, if we know the value of two of the terms, we will always know the value of the third.” This reflects the common confusion of the nature and origin of the “identification problem” in APC analysis. It was clearly explicated in the early works of Mason and colleagues (1973) that this problem occurs only when both of two conditions are simultaneously met: 1) age, period, and cohort variables are linearly related to each other ($\text{Age} = \text{Period} - \text{Cohort}$); and 2) each variable is postulated to be linearly related to the outcome (Y). A major deficiency of prior studies is the exclusive focus on condition 1 (as is the case in B&J) and the avoidance of condition 2. That is, in the absence of the latter, the former would not induce the problem by itself. The problem is inevitable only when the three linearly related variables are treated as independent and additive factors in a linear model. Therefore, a critical insight into this old problem is that it is not data specific, but model specific (Fu, 2008).

In the context of the classical single-equation APC accounting/regression model of Mason et al. (1973), conditions 1 and 2 are both satisfied and hence the problem ensues. In this classical APC research design where population level data in terms of rates or proportions are tabulated in an age-by-time period table, widths of the Age and Period intervals are fixed and equal and Cohorts are then arrayed along the diagonals as linear combinations of $\text{Period} - \text{Age}$. It is important to note that there is only one observation per age-by-period cell. Condition 1 is certainly true in this data structure. And the specification of linear models to such data leads to the identification problem by also providing condition 2.

As Yang and Land (2006, 2013a) have emphasized, however, the repeated cross-section sample survey research design is not the same as the classical age-by-time period table of rates and offers opportunities to address both conditions 1 and 2. Specifically, the repeated cross-section survey design allows for the flexibility of defining temporal widths of the time periods and birth cohorts to be not identical to the ages of the individual sample respondents and hence facilitates the loss of the exact algebraic identity stated in condition 1. In addition, the presence of multiple observations instead of one observation in each period-by-cohort cell makes it more transparent that the data follow a multilevel structure that includes individual-level observations on the ages (typically recorded in sample surveys as age in single years at last birthday at the time of the survey) of the sample respondents that are then nested in higher-level units or contexts such as historical periods and birth cohort memberships. Accordingly, the statement from B&J quoted above need not apply to the APC analysis of repeated cross-section surveys if full advantage of this flexible data structure is utilized.

B&J (2014b, 177) next assert that: “Yang and Land’s proposed solution is to use a cross-classified multilevel model, which treats age as a fixed effect and periods and cohort groups as random effects – contexts in which individuals reside.” That is certainly an approach that Yang and Land (2006, 2013a) have taken in developing the HAPC-CCREM class of models. In the context of the above introduction, this new family of models is employed to address condition 2 based on the characteristics of the multilevel data using sample surveys.

B&J (2014b, 177) go on to say “it has been shown elsewhere that this methodological advance in fact amounts to another constraint” citing as authority an unpublished paper (Luo & Hodges, 2013) that incorrectly assumes that the statistical model for HAPC-CCREM analysis is that of the classical single-equation APC accounting model, and that also tests the HAPC approach by assuming exact linear algebraic (i.e., no stochastic/random component) period-and cohort-based trends in the model.¹ In fact, and as clearly stated by Yang and Land (2006, 2013a), HAPC-CCREMs, as with any class of statistical models, are based on statistical assumptions. In any specific empirical application, these assumptions should be evaluated. For instance, the HAPC-CCREM, as with any mixed (fixed and random) effects model, assumes zero correlation between the individual-level regressors and the random period and cohort effects. As noted by Yang & Land (2008, 2013a), this assumption can be assessed by application of a Hausman-type chi-square test. If the zero correlation assumption is rejected, then the model specification can be modified by treating either the time period or cohort effects (or both) as fixed.

A Perspective on Time Periods and Cohorts as Contexts and as Random Variables

In the classical APC accounting/linear regression model – which evidently is assumed by B&J (2013, 2014a, b) to be the only proper model for APC analysis or at least is unquestionably taken as the basis of their commentaries – the effects of all three temporal dimensions are fixed. In the 1970s, when this model was fully articulated and applied to APC analysis, fixed effects regression models were essentially the only game in town. In the intervening decades, however, statistical methods for estimating and evaluating mixed effects regression models have been extensively developed and applied to demographic, epidemiological, and social data, especially in the form of nested or hierarchical models in which individual-level observations are nested within structural or contextual units. Such models typically specify the effects of individual-level covariates on individual-level outcome variables as fixed with random effects for the contextual units. This is the approach of the HAPC-CCREM analysis of repeated cross-section sample data in which the time periods of the surveys and the cohorts to which individual observations belong are taken as social contexts, which is entirely consistent with Ryder’s (1965) characterization of these structural entities. It is no coincidence that the same conceptualization of period and cohort effects as contexts appeared in recent epidemiologic research as articulated by Suzuki (2012).

¹B&J (2013, 1) also cite an article (Luo, 2013) as “... questioning the capabilities of ... other methodological innovations to disentangle APC effects.” They do not cite the response of Yang and Land (2013b) to Luo (2013), which shows that the Luo application is incorrectly applied and misleading.

In addition, there are methodological reasons that favor the HAPC-CCREM mixed effects specification (Yang & Land, 2013a, 200–201). First, conventional statistical wisdom and methodological guidelines (Goldstein, 2003, 3–4; Snijders & Bosker, 1999, 43–44) indicate that:

1. If the contextual groups are regarded as *unique entities* and the objective of the analysis is primarily to draw conclusions pertaining to each of the groups, then it is appropriate to use the conventional analysis of covariance model.
2. If the groups are regarded as a *sample* from a (real or hypothetical) population and the objective of the analysis is to make inferences about this population, then the random coefficients model is appropriate.

Applied to the age, period, and cohort temporal dimensions of APC analysis of repeated cross-section sample surveys, it follows from this reasoning that, since the range of the age categories for contemporary human populations is essentially fixed at 0 to 125 and most empirical studies utilize only a part of this fixed range, the individual ages or age categories may be regarded as unique entities and it is reasonable to specify the age effects as fixed. On the other hand, the time period and cohort categories available for any specific empirical analysis typically are only a sample of periods and cohorts for any human population. Therefore, it is reasonable to specify the period and cohort effects as random, although corresponding fixed-effects specifications are also available.

Substantively, Ryder (1965) emphasized that a birth cohort moves through life together and encounters the same historical and social events at the same ages. Cohort effects then reflect formative experiences resulting from the intersection of individual biographies and macrosocial influences. However, those macrosocial influences are not mediated by “cohorts” defined by hard and fast birth year boundaries. Some demographers apply broadly construed definitions of cohorts. For instance, a recent analysis of Census 2000 data by Hughes & O’Rand (2004) compared the Baby Boomers (born 1946 – 1964) with their predecessors born earlier in the 20th century that were defined by events and experiences unique to their times: Young Progressives (1906 – 1915), Jazz Age Babies (1916 – 1925), Depression Kids (1926 – 1935), and War Babies (1936 – 1945). Differences in socioeconomic attainment, marriage, fertility, and family structure across these cohorts are substantial and reflect the post-World War II transformation of American society. Other demographers work with cohorts defined by smaller numbers of birth years for two reasons: First, broadly construed cohorts, though relatively easy to convey to the general public and therefore popular in journalism, may miss some interesting internal variations within cohorts defined by longer birth year intervals. Second, the use of cohorts defined by smaller numbers of birth years usually leads to a larger number of birth cohorts in an analysis and thereby increases statistical power for the detection of cohort effects. The application of five-year birth cohorts, as used by Reither et al. (2009), is a conventional practice. In any empirical application of statistical models and methods, however, it is good practice to vary such methodological decisions in order to assess the robustness of empirical findings – i.e., the sensitivity of the empirical estimates to the model specifications. In this way, the concerns of B&J (2014b) about the possible effects of these “equality constraints” (Reither et al., 2009, 1442) on estimates and inferences in a HAPC-CCREM analysis can be

addressed. With respect to the Reither et al. (2009) analyses, the B&J argument (2014b) that the empirical results are highly sensitive to choice of equality constraints is addressed below with simulation models showing similar results for 7-year and 3-year birth cohorts.

Solid Theory — or Questionable Assumptions?

Because B&J assert that it is impossible to separate age, period, and cohort effects using modern statistical techniques (2013), they emphasize the importance of “very solid theory” in APC analyses (2014a, 338). By developing such theories, researchers can make defensible *a priori* assumptions about one of the three temporal dimensions in an APC analysis – thereby addressing the identification problem by effectively removing one dimension from the model. This suggestion is not new and has been previously discussed by K. O. Mason and colleagues (1973) as well as others (Kupper et al., 1985; W. M. Mason & Wolfinger, 2002) who have refuted it due to the potential for model misspecification. In fact, replacing one dimension with a theoretical priori or some substantive variable, while solving an identification problem, makes room for others because it rests on the assumption that the replacement or omission is completely sound (Smith et al., 1982). In the background of modern APC analysis as we introduced above, this assumption does not appear to hold.

Without meaningful supporting evidence, B&J argue that explanations for the U.S. obesity epidemic emphasizing recent secular changes – e.g., those advanced by Cutler and colleagues (2003) – are no more plausible than “the alternative explanation offered by [B&J]” that the historic increase in obesity was initiated during the formative years of successive birth cohorts (Bell & Jones, 2014b, 179). Unfortunately, this “alternative explanation” is not consistent with results from several studies. For example, body mass index ($BMI=kg/m^2$) among white girls in the Fels Longitudinal Study did not change for cohorts born between 1929 and 1964, but increased significantly for cohorts born after 1965 (Demerath et al., 2004). Another study of Fels data found rapid BMI growth among cohorts born after 1970, leading the authors to conclude, “The observed trends reflect the same time course of accelerated BMI gain after 1960 seen in the US pediatric population” (Johnson et al., 2012, 1140). It is worth noting that these studies and the general pediatric trend since 1960 are consistent with the cohort effects reported by Reither et al. (2009) and also by Robinson et al. (2013) in a study of abdominal obesity in the U.S.

Moreover, if the obesity epidemic were rooted in the formative experiences of birth cohorts spaced across the 20th century, then it would be reasonable to anticipate a relatively steady increase in adolescent and adult obesity rates over time. This theoretical expectation is not reflected in national epidemiologic data. For instance, using four large nationally representative studies, Lee et al. (2011) found substantial increases in adolescent and young adult BMI late in the 20th century, but relatively little change between 1959 and 1980. Similarly, between 1960 and 1980 the prevalence of adult obesity climbed at glacial pace, from 13.4 to 15.0 percent (Flegal et al., 2002). Then over the next 20 years, the prevalence of adult obesity doubled, reaching 30.9 percent in the year 2000.

To fit these trends in BMI and obesity, an explanation based entirely on cohort effects would have to presume that changing childhood conditions during the 20th century resulted in BMI

increases that, for the most part, became manifest near a single point in history. So for example, a cohort born in 1930 would have experienced some unique childhood conditions that were latent for half a century before resulting in substantial BMI gains around 1980. A cohort born in 1970, by contrast, would experience a much shorter latency period. Because such a coincidence in the manifestation of cohort effects is highly implausible, we think it is unreasonable to assume that cohort effects can, by themselves, explain the rapid and ubiquitous changes in body mass observed during the latter part of the 20th century.

An Empirical Assessment of HAPC Models in Reither et al. (2009)

B&J assert that Reither et al.'s (2009) findings pertaining to the APC effects associated with U.S. obesity prevalence could have arisen from “a very different data generating process (DGP)” than the one reported by Reither and colleagues (Bell & Jones, 2014b, 178). To show how, B&J simulate data assuming a linear cohort effect size .04, a linear age effect size .1, a quadratic age effect size $-.001$, and a nonexistent period effect size 0. Random variation about both cohorts and periods are also generated from a normal distribution with a mean 0 and a variance of .01. Results from HAPC models fitted to these simulated data show APC patterns broadly consistent with those reported by Reither et al. (2009), but which are known to be biased estimates of the “true” APC effects in the DGP. That is, the HAPC-CCREM estimates strong period effects where none exist, and nonexistent cohort effects when, in fact, a strong linear effect exists in the DGP. B&J contend that “such a linear trend [in their DGP] is an appropriate means of generating the data for this situation” (2014b, footnote 2).² However, there are two problems with this assumption. First, counter to B&J's contention, such a linear trend is highly improbable in real-world cases. Second, imposing a perfectly linear functional form for the period- and/or cohort-based trend effects in a DGP produces data that assure APC model misspecification (Yang & Land, 2013b). Consequently, B&J's simulation exercise leads to faulty conclusions regarding the HAPC-CCREM analyses of Reither et al. (2009).

To the first problem, the B&J (2014b) simulation study specifies a data generation process (DGP) for the logit of the probability of being obese as a quadratic function of age, a linear function of cohorts, and normally distributed random error terms for cohorts and time periods. B&J (2013, 2014a) thus consider, as an example supporting the argument of the impossibility of separating age, period, and cohort effects, Step 1, a DGP where equal age, period and cohort effects, of magnitude 1, contribute to the generation of an outcome variable (Y):

$$Y = \text{Age} + \text{Period} + \text{Cohort}$$

and argue that, Step 2, because of the dependency $\text{Age} = \text{Period} - \text{Cohort}$ of classical age-by-time period tabular research designs, $\text{Age} + \text{Cohort}$ can be substituted for Period to yield:

²B&J (2014b) states: “... the trend found (by Reither et al.) is very much linear in appearance and interpreted as such (“a monotonic increase over time” – Reither et al. 2009:1443). As such a linear trend is an appropriate means of generating the data for this situation.” Comment: This is a basic error. There are many monotonically increasing mathematical functions that are not linear functions. It is erroneous to equate the two categories or to infer that an empirical finding of a monotonically increasing relationship is appropriately modeled by a linear function.

$$Y=2*Age+2*Cohort$$

in which, Step 3, Period can be substituted for Age + Cohort to give:

$$Y=2*Period$$

and thus all three of these DGPs can generate the same data.

There are problems with this analysis. First, in practice, if there are exact linear algebraic dependencies among the three temporal dimensions, as Yang and Land (2013b) noted in response to Luo (2013), Steps 1 and 2 of the above three-step model specification and estimation process would show that all three dimensions are not operative, and there is no need to specify a three-dimensional model. Table 1 below shows the log likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and degrees of freedom for the age, age-period, age-cohort, and age-period-cohort models fitted to different data. Both the AIC and BIC indicate that the age-cohort models are preferred to all other models, APC included, for the data simulated by B&J's DGP in Eq(4); see the upper left panel of Table 1 (the data structures of the other DGPs included in the other panels of Table 1 are discussed in the subsequent section).

Second, on a more theoretical probability level, this argument portrays the possibility of exact algebraic temporal functions in one or more of the three APC temporal dimensions as realistic. But is it? If we think of the outcome data (the Y variable) and the temporal dimensions of the APC model as having a joint continuous probability distribution, then, in a measure-theoretic analysis (Halmos, 1950; Loève, 1977) of this joint distribution, the measure (probability) of a point (a real number) on a single dimension has probability zero, with non-zero probabilities existing only for intervals of real numbers. Furthermore, in a multidimensional probability space for the parameters of the A, P, and C dimensions of the model, the probability of an exact algebraic linear function of any of the A, P, and C dimensions is zero, and the probability that all three dimensions will have exact (no non-zero error term) functions also is zero. In short, while B&J's algebraic analyses appear to be convincing, they deal with conditions/likelihoods of linear processes that are unlikely, that is, have probability zero, in a continuous probability space.

To demonstrate problems in B&J's approach to testing the HAPC-CCREM, we fitted HAPC-CCREMs on four simulated data sets. First, we began by replicating B&J's (2014b) findings by fitting HAPC-CCREMs to data simulated from the DGP presented as Eq(4) in their letter (and specified below). We used B&J's third "grouping scenario" in which cohort trends in the data are generated using "7-year birth cohorts in the DGP" but the cohort trends are tested using an "HAPC model fitted with 5-year birth cohorts" (2014b, 179). Second, we fitted an HAPC-CCREM to simulated data from a DGP assuming cohort- and period-based variations in Y that are consistent with the DGP specified by B&J in Eq(4), but the period- and cohort-based variation in the DGP are not confined to perfect linear functional forms on Y. Specifically, the new DGP reflects nonsignificant trends in one-year period categorical

measures and a strong, near-linear trend composed of 7-year cohort categorical measures, which combined constitute a cohort-based trend that is consistent with B&J's .04 linear cohort trend. Third, we further tested B&J's concerns about cohort widths in HAPCs needing to match exactly the cohort widths in DGPs by simulating a similar DGP composed of a nonsignificant trend in one-year period categorical measures and a strong, near-linear trend composed of 3-year cohort categorical measures. On data from both of these DGPs, we fitted HAPC-CCREMs that used 5-year cohort measures, thereby evaluating the HAPC models' ability to retrieve true APC effects when groupings of cohorts do not exactly match the DGPs' groupings. Finally, we fitted an HAPC-CCREM using 5-year cohort measures to simulated data from a DGP consistent with Reither et al.'s (2009) findings, but which was generated using 7-year cohorts. This final exercise was to again test B&J's contention that cohort groupings in the HAPC fitted to APC data need to match the groupings of cohorts in the DGP. The measures and functional forms employed by all four simulation exercises are summarized in Table 2.

In conducting these analyses, we show that the results presented by B&J likely stem from their fitting HAPC models to unrealistic DGPs that are generated from perfectly linear functional forms of cohort- and period-based effects. That is, the demonstrated failure of the HAPC models to retrieve the "true" APC effects in B&J's exercises do not appear to reflect a failure of the HAPC method, but rather reflect B&J's inappropriate application of the method to DGPs that violate basic assumptions of HAPC models. Below we demonstrate that the HAPC-CCREMs are able to retrieve the "true" effects in the simulated DGPs presented by B&J by simply breaking the implausible functional forms imposed by B&J. We then show that the HAPC-CCREM estimates of age-, period-, and cohort-based variation from simulated data are consistent with the patterns presented by Reither et al. (2009), even if the cohort measures in the HAPC models do not match the cohort groupings in the DGPs. Below we present the models we fitted and discuss the estimates.

To start, we used Stata (StataCorp, 2013), R (R Development Core Team, 2010) and WinBUGS (Lunn et al., 2000) to simulate data using B&J's (2014b) Eq(4) and fitted an HAPC-CCREM to replicate the results they present in Figure 1 of their paper. Specifically, we modeled the probability of obesity on data generated as follows:

$$\text{Logit}[\text{Pr}(Y = 1)] = -2 + (.1 * \text{Age}) + (-.001 * \text{Age}^2) + (.04 * \text{Cohort}) + U_c + U_p$$

$$U_c \sim N(0, .01), U_p \sim N(0, .01), \quad (1)$$

where the random variation on the cohort and period effects are applied on the individual effect sizes.

Results from the HAPC-CCREMs fitted to these simulated data are presented in Figure 1. For this and the next three exercises, period and cohort estimates are predicted probabilities of obesity for an adult 25 years of age (i.e., results are "centered" on age 25). Also, all of our results are median estimates from 100 simulations using Markov chain Monte Carlo

(MCMC) on HAPC-CCREMs with 20,000 iterations, following a 2,000 iteration burn-in. Our results in this first exercise closely match the findings presented by B&J (2014b) in column 2 of (B&J's) Figure 1. In their interpretation of this result, B&J (2014b, 179) state, "As can be seen, the typical median result does not match the DGP at all. No cohort trend is found, an erroneous period trend is found, and the age effect is underestimated." Our findings shown in Figure 1 are consistent with this conclusion.

Next, we simulated data from a DGP equation that assumes period- and cohort-based variation in obesity that are very similar to B&J's Eq(4), but does not specify that these variations are *exactly* linear in the DGP. Specifically, we simulated 27 1-year categorical period effects that generate negligible period-based trends in obesity, and 14 7-year categorical cohort effects that generated a strong positive cohort-based trend in obesity starting in the early twentieth century. That is, we simulate data from a DGP counter to results found by Reither et al. (2009) and consistent with B&J's Eq(4), but which are not generated from a DGP stipulating perfectly linear functional forms of period- and cohort-based trends. We simulated these data using the following DGP:

$$\begin{aligned} \text{Logit}[\text{Pr}(Y=1)] = & -2 + (.1*\text{Age}) + (-.001*\text{Age}^2) + (-.28*\text{C1}) + (-.30*\text{C2}) + (-. \\ & 28*\text{C3}) + (-.22*\text{C4}) + \\ & (-.09*\text{C5}) + (-.01*\text{C6}) + (.04*\text{C7}) + (.06*\text{C8}) + (.17*\text{C9}) + (.28*\text{C10}) + (. \\ & 49*\text{C11}) + (.52*\text{C12}) + \\ & (.56*\text{C13}) + (.59*\text{C14}) + (-.02*\text{P1}) + (.03*\text{P2}) + (.04*\text{P3}) + (.04*\text{P4}) + (. \\ & 02*\text{P5}) + (-.03*\text{P6}) + (- \\ & .03*\text{P7}) + (-.05*\text{P8}) + (-.05*\text{P9}) + (-.05*\text{P10}) + (-.05*\text{P11}) + (-.05*\text{P12}) + (-. \\ & 06*\text{P13}) + (-.05*\text{P14}) \\ & + (-.05*\text{P15}) + (-.02*\text{P16}) + (0*\text{P17}) + (.02*\text{P18}) + (-.02*\text{P19}) + (-.02*\text{P20}) + \\ & (0*\text{P21}) + (-.02*\text{P22}) \\ & + (.02*\text{P23}) + (.05*\text{P24}) + (.08*\text{P25}) + (.1*\text{P26}) + (.1*\text{P27}) + U_c + U_p \end{aligned}$$

$$U_c \sim N(0, .01), U_p \sim N(0, .01), \quad (2)$$

where the random variation on the cohort and period effects are applied on the individual effect sizes.

Median estimates from 100 simulations using MCMC on HAPC-CCREMs fitted to these data using linear and quadratic age terms, 27 1-year categorical periods, and 19 5-year categorical cohorts are presented in Figure 2. Overall, Figure 2 illustrates the HAPC-CCREM's ability to estimate patterns in each APC effect that are not significantly different from the true APC patterns in the DGP. That is, unlike B&J's findings, which show a strong period-based trend similar to Reither et al.'s (2009) findings, the results here show the HAPC-CCREM correctly estimates strong cohort-based trends consistent with the DGP in Eq(2) above, as well as the true age patterns and true nonsignificant period-based trend in the DGP. This encouraging result is obtained despite the fact that a complete APC model in

which all three temporal dimensions are operative is not the best-fitting option for the second DGP (see Table 1).

These results also counter B&J's contention that only on the rare occasion that one fits an HAPC model using cohort measures that precisely match the cohort groupings in the DGP (e.g., 5-year cohorts in the HAPC model and 5-year cohorts in the DGP) can HAPC models retrieve true APC effects. To further test this contention, we generate two additional datasets with different DGPs. Above, we showed that an HAPC model fitted with 19 5-year cohort measures estimates APC effects that are consistent with the DGP, even though the DGP was generated from 14 7-year cohorts. Here, we further evaluate B&J's claim by simulating data from a DGP with 34 3-year birth cohorts and fitting an HAPC model with 19 5-year birth cohorts. We simulated these data using the following DGP:

$$\begin{aligned} \text{Logit}[\text{Pr}(Y=1)] = & -2 + (.1*\text{Age}) + (-.001*\text{Age}^2) + (-.28*C1) + (-.30*C2) + (-. \\ & 27*C3) + (-.28*C4) + \\ & (-.29*C5) + (-.22*C6) + (-.25*C7) + (-.1*C8) + (-.09*C9) + (-.03*C10) + (-. \\ & 01*C11) + (.03*C12) + \\ & (.04*C13) + (.05*C14) + (.04*C15) + (.06*C16) + (.08*C17) + (.17*C18) + (. \\ & 24*C19) + (.28*C20) + \\ & (.37*C21) + (.43*C22) + (.49*C23) + (.51*C24) + (.52*C25) + (.56*C26) + (. \\ & 56*C27) + (.57*C28) + \\ & (.57*C29) + (.59*C30) + (.62*C31) + (.61*C32) + (.64*C33) + (.66*C34) + (-. \\ & 02*P1) + (.03*P2) + \\ & (.04*P3) + (.04*P4) + (.02*P5) + (-.03*P6) + (-.03*P7) + (-.05*P8) + (-. \\ & 05*P9) + (-.05*P10) + (- \\ & .05*P11) + (-.05*P12) + (-.06*P13) + (-.05*P14) + (-.05*P15) + (-.02*P16) + \\ & (0*P17) + (.02*P18) \\ & + (-.02*P19) + (-.02*P20) + (0*P21) + (-.02*P22) + (.02*P23) + (.05*P24) + (. \\ & 1*P25) + (.05*P26) \\ & + (.04*P27) + U_C + U_P \end{aligned}$$

$$U_C \sim N(0, .01), U_P \sim N(0, .01), \quad (3)$$

where the random variation on the cohort and period effects are applied on the individual effect sizes.

Median estimates from 100 simulations using MCMC on HAPC-CCREMs fitted to these data are presented in Figure 3 and, similar to results in Figure 2, show that the age, period and cohort estimates from HAPC models are consistent with the true APC patterns in the DGP. Both tests, therefore, provide strong and consistent evidence suggesting that the inability of HAPC models to recover the "true" APC effects in B&J's DGPs stem not from a failure to meet a special condition of equally-sized cohort measures in the HAPCs and DGPs, but rather from B&J's application of HAPC models to highly restrictive DGPs that

violate important assumptions in APC analysis. As before, these results are obtained despite the fact that the full APC specification is not the best-fitting model (see Table 1).

Finally, we conclude by simulating data from a DGP consistent with the findings reported by Reither et al. (2009), but use 13 7-year birth cohort effects instead of the 18 5-year birth cohort effects used by Reither et al. (2009):

$$\begin{aligned} \text{Logit}[\text{Pr}(Y=1)] = & -1.988 + (.059*\text{Age}) + (-.001*\text{Age}^2) + (.100*C1) + (.086*C2) \\ & + (.040*C3) + \\ & (-.028*C4) + (-.047*C5) + (-.049*C6) + (-.067*C7) + (-.083*C8) + (-.101*C9) \\ & + (-.122*C10) + \\ & (.039*C11) + (.138*C12) + (.1*C13) + (-.524*P1) + (-.474*P2) + (-.443*P3) + \\ & (-.423*P4) + \\ & (-.377*P5) + (-.362*P6) + (-.349*P7) + (-.314*P8) + (-.275*P9) + (-.214*P10) \\ & + (-.169*P11) + \\ & (-.135*P12) + (-.117*P13) + (-.054*P14) + (-.039*P15) + (.029*P16) + (. \\ & 099*P17) + (.172*P18) + \\ & (.212*P19) + (.256*P20) + (.299*P21) + (.410*P22) + (.463*P23) + (.529*P24) \\ & + (.555*P25) + \\ & (.605*P26) + (.640*P27) + U_C + U_P \end{aligned}$$

$$U_c \sim N(0, .01), U_p \sim N(0, .01), \quad (4)$$

where the random variation on the cohort and period effects are applied on the individual effect sizes.

Median estimates from 100 simulations using MCMC on HAPC-CCREMs fitted to these data are presented in Figure 4, highlighting the ability of HAPC models to estimate APC effects that are not significantly different from the true APC effects in the DGP. That is, HAPC models fitted with 5-year cohort measures are able to retrieve the true APC effects associated with obesity prevalence in the DGP generated from 7-year cohorts. Complete HAPC models retrieve the true effects, even though three-dimensional models are not the best fit to this fourth set of simulated data (Table 1). This is contrary to actual NHIS data used by Reither et al. (2009) and subsequently by Masters et al. (2013), which are clearly three-dimensional in nature. That is, the best specification for HAPC models of obesity based on NHIS data is to include age, period and cohort effects. The summary indicators of model fit show that it is inappropriate to “assume away” any one of these effects, as B&J (2014b) do in their simulation exercises.

Together these exercises show that HAPC-CCREMs using different cohort groupings from the actual cohort groupings in the DGPs estimate APC patterns that are consistent with the true APC effects in the DGPs. These findings directly counter the results presented by B&J (2014b), and suggest that B&J’s results likely reflect their decision to subject HAPC models to very unusual data structures that reflect unrealistic DGPs. Namely, the fundamental

limitation of B&J's exercises was their decision to impose exact linear functional forms of period- and cohort-based trends in their DGPs – data characteristics that are highly improbable in real populations. Under these unrealistic circumstances HAPC models do not perform well, and proper diagnostics using model fit criteria would have indicated from the outset that full APC models should not have been fitted to such data structures. But, as we show above, under conditions that directly test the claims made by B&J, the HAPC models perform quite well.

Discussion

Ever since the publication of Norman Ryder's seminal paper on the influence of birth cohorts on societal change (1965), social scientists have sought to develop and implement methods that could separate age, period, and cohort effects. In the ensuing years, debates emerged over these APC methods, pitting advocates of APC estimation (e.g., K. O. Mason et al., 1973; Pullum, 1980) against scholars who emphasized identification problems (Glenn, 1976) and the sensitivity of research findings to equality constraints. Because early APC methods were rooted in simple linear modeling techniques, these debates persisted without much progress for several decades.

In recent years, critical advancements in statistical methodologies and new data structures have enabled scholars to create a new suite of techniques for APC estimation, including use of the intrinsic estimator for tabular age-by-period arrays of aggregate data, and hierarchical methods for use with repeated cross-sections of individual-level data (Yang & Land, 2013a). Reither et al. (2009) employed hierarchical models (HAPC-CCREMs) in their repeated cross-sectional study of the U.S. obesity epidemic, which was hailed at the time as a potential breakthrough in APC methodology (Harding, 2009). The hierarchical approach is also applicable to longitudinal data that tracks the same individuals over time, as clearly illustrated for a hypothetical data structure in Figure 2 of Suzuki (2012) and in HAPC analyses of NHIS-linked mortality files (Masters et al., 2012). In both cases, annual observations are embedded in time-varying periods and individuals of differing ages, who are further embedded in time-invariant cohorts. Despite early praise for these innovations, skepticism regarding hierarchical approaches to APC modeling persists, as most clearly illustrated through B&J's recent critiques that (1) Reither et al. (2009) did not adequately address the identification problem, (2) birth cohort effects are an equally plausible alternative explanation for the U.S. obesity epidemic, and (3) the HAPC models estimated by Reither et al. may have generated misleading results.

Through the preceding discussion, we have addressed each of these critiques. First, B&J make the common error of equating "the identification problem" only with linear dependencies between age, period and cohort variables without acknowledging that these variables must also be linearly related to the outcome, as shown by Mason and colleagues several decades ago (1973). This is not the case in Reither et al. (2009), where apparent nonlinearities are present, especially for age and cohort effects.

As we have argued throughout the paper, data like those generated by B&J (2014b) are theoretically improbable and empirically rare. Indeed, we are unaware of any three

dimensional data structures produced by real historical processes that have resulted in an exact algebraic model (e.g., linear, quadratic, logarithmic) for time period and cohort trends. However, we acknowledge that such a data structure would pose a challenge to “traditional” HAPC-CCREMs, like those estimated by Reither et al. (2009), which are usually specified only via the normally distributed random error components for periods and birth cohorts. As Yang and Land (2013a) have emphasized, it is important for researchers to engage in exploratory data analysis (i.e., graphical study and model selection statistics) prior to model specification and estimation. Such exploratory analyses may reveal that a complete three-dimensional APC model is not the most appropriate for modeling the data, as was the case reported above for the B&J DGP, and the researcher can proceed accordingly with a reduced model. In the unlikely event that preliminary analyses suggest that a complete APC model is appropriate to the data and that one of the time period or cohort trends is unfolding according to an exact algebraic model in addition to the random error component, like that in B&J (2014b), an algebraic specification can be included to capture the process at work in the level-2 period or cohort groups.

Second, we have shown that “the alternative explanation offered by [B&J]” that cohort effects are an equally plausible explanation for the U.S. obesity epidemic is rooted in neither theory nor data, but rather some fairly questionable assumptions (Bell & Jones, 2014b, 179). Based on the patterns of BMI and obesity change in the U.S. over the past few decades, there is good reason to suspect that ubiquitous secular changes have played a central role in the U.S. obesity epidemic. But even if the alternative explanation advocated by B&J were more compelling, it would be necessary to have some empirical method to adjudicate between competing theoretical perspectives. Although intriguing and necessary, compelling speculation can never replace evidence in any field of scientific inquiry. This highlights the need for methods that are capable of separating age, period and cohort effects in the social sciences.

Third, we have shown that B&J’s simulation models, ostensibly showing weaknesses in the HAPC-CCREM method used by Reither et al. (2009), were applied to a situation where APC modeling should *never* be used. Specifically, B&J apply three-dimensional HAPC methods to data generated under the assumptions that (a) net period effects were zero (i.e., obesity change in the U.S. is really a two-dimensional process), and (b) the functional form of associations between obesity and period and cohort variables was perfectly linear. This combination of assumptions all but ensures failure of any and all APC approaches, not just the HAPC-CCREM approach outlined by Yang and Land (2013a). However, our simulation models clearly demonstrate 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. Furthermore, contrary to the claims of B&J, findings from these models are stable across different ways of measuring the width of birth cohorts. On balance, these simulation exercises provide additional evidence supporting Reither et al.’s (2009) findings that increases in U.S. obesity prevalence primarily reflect strong period-based factors, with cohort-based increases in obesity seen only among more recent birth cohorts.

Cohort Effects in the U.S. Obesity Epidemic

The title and text of the B&J commentary (2014b) insinuate that Reither et al. (2009) failed to detect meaningful birth cohort effects in the U.S. obesity epidemic. This revisionist interpretation blithely sweeps aside the conclusions reached by Reither and his colleagues. Based on their findings, Reither et al. contend that “newer birth cohorts have suffered a form of double jeopardy in which ubiquitous secular changes and birth cohort membership have independently contributed to increased odds of obesity. In the context of secular changes that have caused obesity rates to increase sharply across social and demographic groups in the U.S., the trends we observed among recent birth cohorts are deeply troubling” (1444). From these conclusions, Reither et al. go on to suggest that “it could prove useful to develop cohort-specific strategies that limit weight gain early in the life course” (1447).

Subsequent research by contributors to the Reither et al. (2009) study has further emphasized the importance of birth cohort effects in the U.S. obesity epidemic. For example, in a study appearing in *Health Affairs*, Reither, Olshansky and Yang (2011) argue that obesity trends among newer birth cohorts could have devastating long-term health implications. More recently, Reither and Yang helped unearth compelling evidence in favor of this argument, detecting much stronger effects of obesity on U.S. adult mortality when birth cohort trends in BMI are taken into consideration (Masters et al., 2013).

Both Reither et al. (2009) and subsequent studies by these scholars clearly assert that obesity trends among newer birth cohorts are cause for serious concern in the public health community. We are unsure why B&J felt compelled to paint a different picture, but it is important to recognize that their portrayal of this research is erroneous and misleading.

Conclusion

The development of new APC methods presents an opportunity for social scientists to move beyond the methods and debates of the 1970s. But contrarians to these advancements in APC research have enjoyed a resurgence, as illustrated by the string of B&J commentaries (2013, 2014a, b). And so the social science research community has come to this important crossroad. Down one road lie the misconceptions of the past, where comfortably familiar but fundamentally flawed refrains about “identification problems” and other purportedly insurmountable hurdles to APC research echo over and over again, conspiring to dissuade the application of innovative techniques. Down the other road lies the embrace of new statistical approaches, advances in data structures, and the recognition that APC modeling has become available to anyone with an interest in Ryder’s crucial observations about the importance of birth cohort effects. To be sure, researchers must exercise caution in ensuring that their data meet the basic assumptions of APC methods – but this is true of *any* statistical application. B&J have succeeded in showing the limitations of APC methods when fabricated data violate these assumptions. But we have succeeded in showing that, when these basic assumptions are satisfied, hierarchical approaches to APC modeling perform remarkably well. We believe that this should provide researchers with the evidence necessary to leave the 1970s behind and anticipate a more hopeful and robust future for APC modeling in the social sciences.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Baltes PB. Longitudinal and cross-sectional sequences in the study of age and generation effects. *Hum Dev.* 1968; 11:145–171. [PubMed: 5663535]
- Bell A, Jones K. The impossibility of separating age, period and cohort effects. *Soc Sci Med.* 2013; 93:163–165. [PubMed: 23701919]
- Bell A, Jones K. Another ‘futile quest’? A simulation study of Yang and Land’s Hierarchical Age-Period-Cohort model. *Demographic Research.* 2014a; 30:333–360.
- Bell A, Jones K. 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.* 2014b; 101:176–180. [PubMed: 24094683]
- Cutler DM, Glaeser EL, Shapiro JM. Why Have Americans Become More Obese? *Journal of Economic Perspectives.* 2003; 17:93–118.
- Demerath EW, Li J, Sun SS, Chumlea WC, Remsburg KE, Czerwinski SA, et al. Fifty-year trends in serial body mass index during adolescence in girls: the Fels Longitudinal Study. *Am J Clin Nutr.* 2004; 80:441–446. [PubMed: 15277168]
- Flegal KM, Carroll MD, Ogden CL, Johnson CL. Prevalence and trends in obesity among US adults, 1999–2000. *JAMA.* 2002; 288:1723–1727. [PubMed: 12365955]
- Fu WJ. A Smoothing Cohort Model in Age–Period–Cohort Analysis With Applications to Homicide Arrest Rates and Lung Cancer Mortality Rates. *Sociological Methods & Research.* 2008; 36:327–361.
- Glenn ND. Cohort Analysts Futile Quest - Statistical Attempts to Separate Age, Period and Cohort Effects. *Am Sociol Rev.* 1976; 41:900–904.
- Goldstein, H. *Multilevel Statistical Models.* London: Edward Arnold; 2003.
- Halmos, PR. *Measure Theory.* Princeton, NJ: Van Nostrand Co., Inc; 1950.
- Harding DJ. Recent advances in age-period-cohort analysis. A commentary on Dregan and Armstrong, and on Reither, Hauser and Yang. *Social Science & Medicine.* 2009; 69:1449–1451. [PubMed: 19766371]
- Hughes, ME.; O’Rand, AM. *The Lives and Times of the Baby Boom.* Census 2000 Monograph. New York: Russell Sage, Population Reference Bureau; 2004.
- Johnson W, Soloway LE, Erickson D, Choh AC, Lee M, Chumlea WC, et al. A changing pattern of childhood BMI growth during the 20th century: 70 y of data from the Fels Longitudinal Study. *Am J Clin Nutr.* 2012; 95:1136–1143. [PubMed: 22418089]
- Kupper LL, Janis JM, Karmous A, Greenberg BG. Statistical age-period-cohort analysis: A review and critique. *Journal of Chronic Diseases.* 1985; 38:811–830. [PubMed: 4044767]
- Lee H, Lee D, Guo G, Harris KM. Trends in body mass index in adolescence and young adulthood in the United States: 1959–2002. *J Adolesc Health.* 2011; 49:601–608. [PubMed: 22098770]
- Loève, M. *Probability Theory I.* New York: Springer-Verlag; 1977.
- Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing.* 2000; 10:325–337.
- Luo L. Assessing Validity and Application Scope of the Intrinsic Estimator Approach to the Age-Period-Cohort Problem. *Demography.* 2013; 50:1945–1967. [PubMed: 24072610]
- Luo, L.; Hodges, J. *Population Association of America.* New Orleans: Louisiana; 2013. *The Cross-Classified Age-Period-Cohort Model as a Constrained Estimator.*
- Mason KO, Mason WM, Winsborough HH, Poole WK. Some Methodological Issues in Cohort Analysis of Archival Data. *Am Sociol Rev.* 1973; 38:242–258.
- Mason, WM.; Wolfinger, NH. Cohort analysis. In: Smelser, NJ.; Baltes, PB., editors. *International encyclopedia of the social and behavioral sciences.* New York: Elsevier; 2002. p. 2189–2194.

- Masters RK, Hummer RA, Powers DA. Educational Differences in U.S. Adult Mortality: A Cohort Perspective. *Am Sociol Rev.* 2012; 77:548–572. [PubMed: 25346542]
- Masters RK, Reither EN, Powers DA, Yang YC, Burger AE, Link BG. The impact of obesity on US mortality levels: the importance of age and cohort factors in population estimates. *Am J Public Health.* 2013; 103:1895–1901. [PubMed: 23948004]
- Pullum TW. Separating Age, Period, and Cohort Effects in White United-States Fertility, 1920–1970. *Social Science Research.* 1980; 9:225–244.
- R Development Core Team. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2010. Retrieved from <http://www.R-project.org>
- Reither EN, Hauser RM, Yang Y. Do birth cohorts matter? Age-period-cohort analyses of the obesity epidemic in the United States. *Soc Sci Med.* 2009; 69:1439–1448. [PubMed: 19773107]
- Reither EN, Olshansky SJ, Yang Y. New forecasting methodology indicates more disease and earlier mortality ahead for today's younger Americans. *Health Aff (Millwood).* 2011; 30:1562–1568. [PubMed: 21700600]
- Robinson WR, Utz RL, Keyes KM, Martin CL, Yang Y. Birth cohort effects on abdominal obesity in the United States: the Silent Generation, Baby Boomers and Generation X. *Int J Obes.* 2013; 37:1129–1134.
- Ryder NB. The cohort as a concept in the study of social change. *Am Sociol Rev.* 1965; 30:843–861. [PubMed: 5846306]
- Smith HL, Mason WM, Fienberg SE. More chimeras of the age-period-cohort accounting framework: Comment on Rodgers. *Am Sociol Rev.* 1982; 47:787–793.
- Snijders, T.; Bosker, R. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling.* London: SAGE; 1999.
- StataCorp. *Stata Statistical Software: Release 13.* College Station, TX: StataCorp LP; 2013.
- Suzuki E. Time changes, so do people. *Social Science & Medicine.* 2012; 75:452–456. [PubMed: 22591827]
- Yang Y, Land KC. 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. *Sociological Methodology.* 2006; 36:75–97.
- Yang Y, Land KC. Age-Period-Cohort Analysis of Repeated Cross-Section Surveys: Fixed or Random Effects? *Sociological Methods & Research.* 2008; 36:297–326.
- Yang, Y.; Land, KC. *Age-Period-Cohort Analysis: New Models, Methods, and Empirical Applications.* Boca Raton, FL: CRC Press; 2013a.
- Yang Y, Land KC. Misunderstandings, Mischaracterizations, and the Problematic Choice of a Specific Instance in Which the IE Should Never Be Applied. *Demography.* 2013b; 50:1969–1971. [PubMed: 24132743]

Research highlights

- We clarify the nature of the identification problem in conventional linear APC analysis.
- Cohorts and periods are social contexts; eliminating either via “solid theory” is often problematic.
- Like any statistical method, APC models fail when basic assumptions are violated.
- Hierarchical APC models perform well on data structures characteristic of real-world questions.
- Simulations are useful tools when they rest on tenable assumptions and plausible data structures.

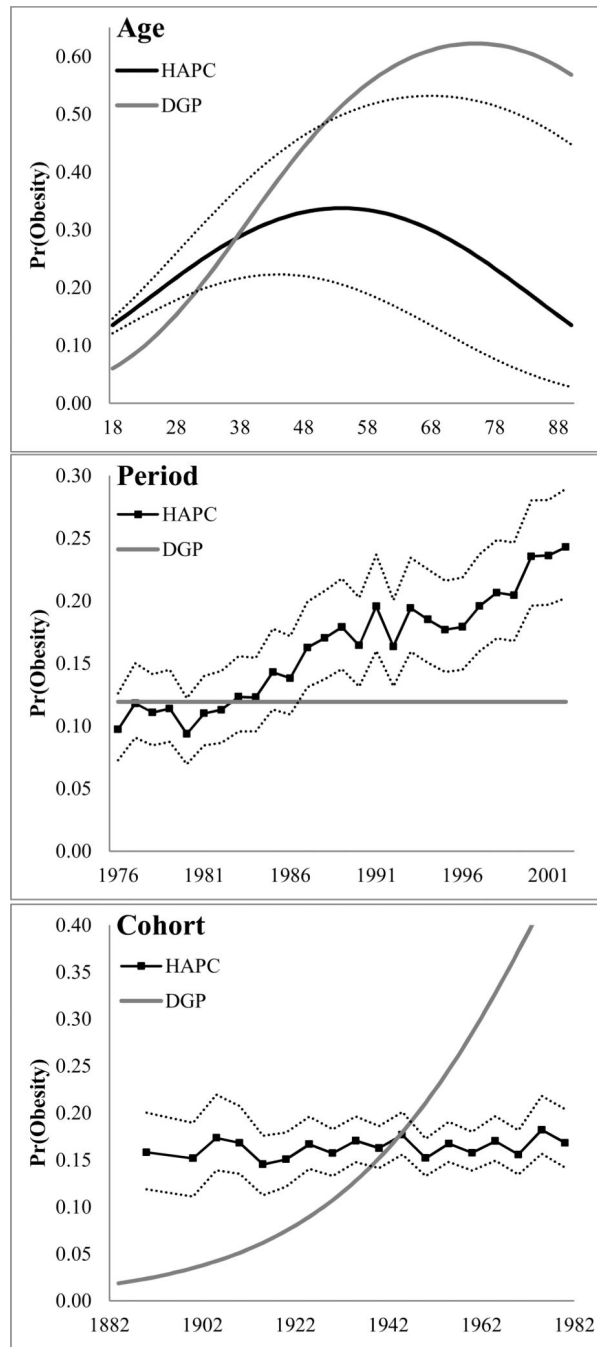


Figure 1. Replicated Results from Bell & Jones (2014b). HAPC-CCREM Fitted to Simulated Data from DGP with Exact Linear 7-year Cohort-based Trends in Obesity, with 95% credible intervals (dotted lines)

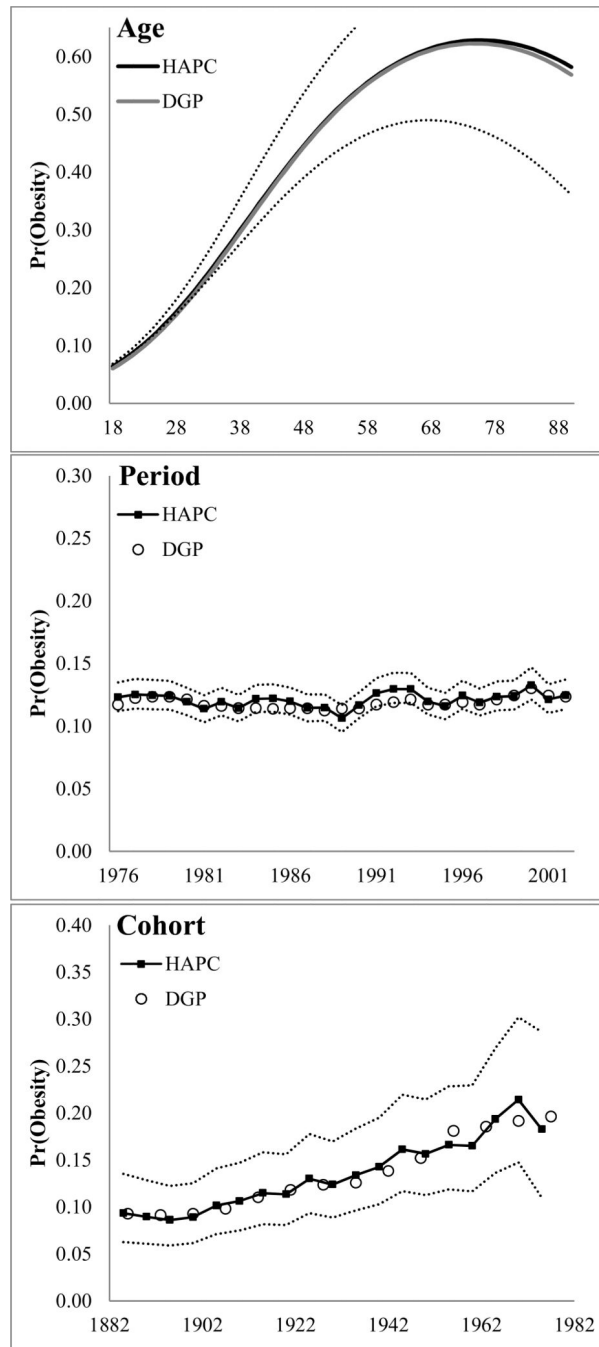


Figure 2. Results from HAPC-CCREM Fitted to Simulated Data from DGP with Near Linear 7-year Cohort-based Trends in Obesity, with 95% credible intervals (dotted lines).

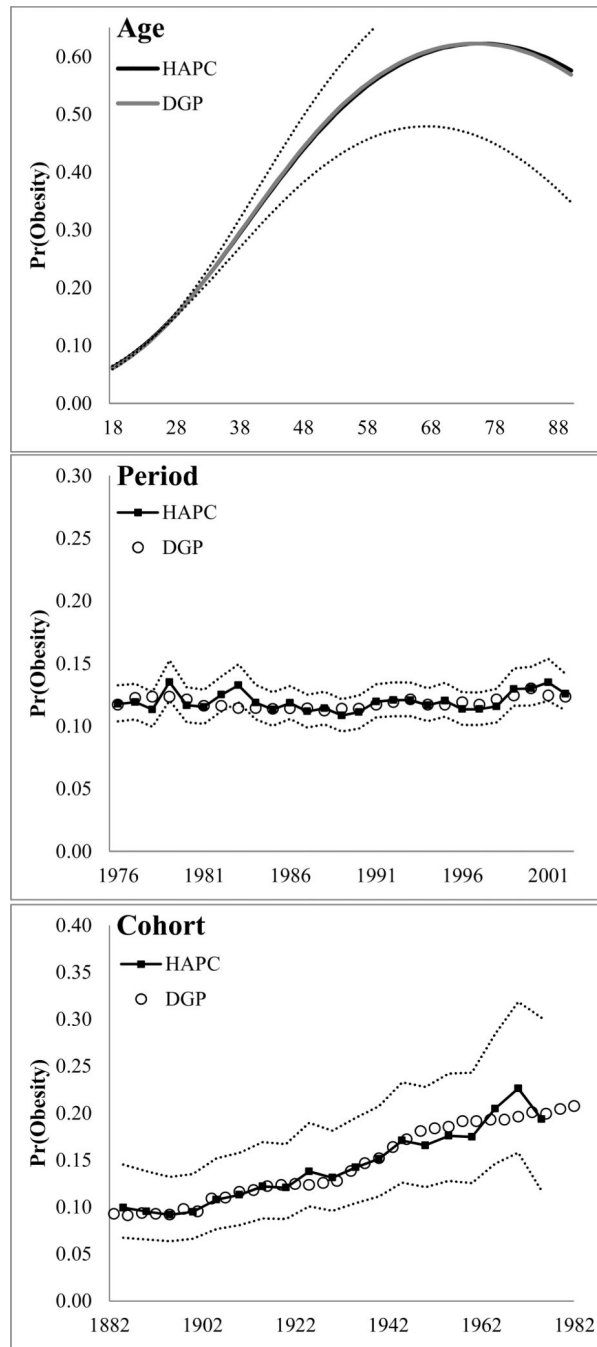


Figure 3. Results from HAPC-CCREM Fitted to Simulated Data from DGP with Near Linear 3-year Cohort-based Trends in Obesity, with 95% credible intervals (dotted lines).

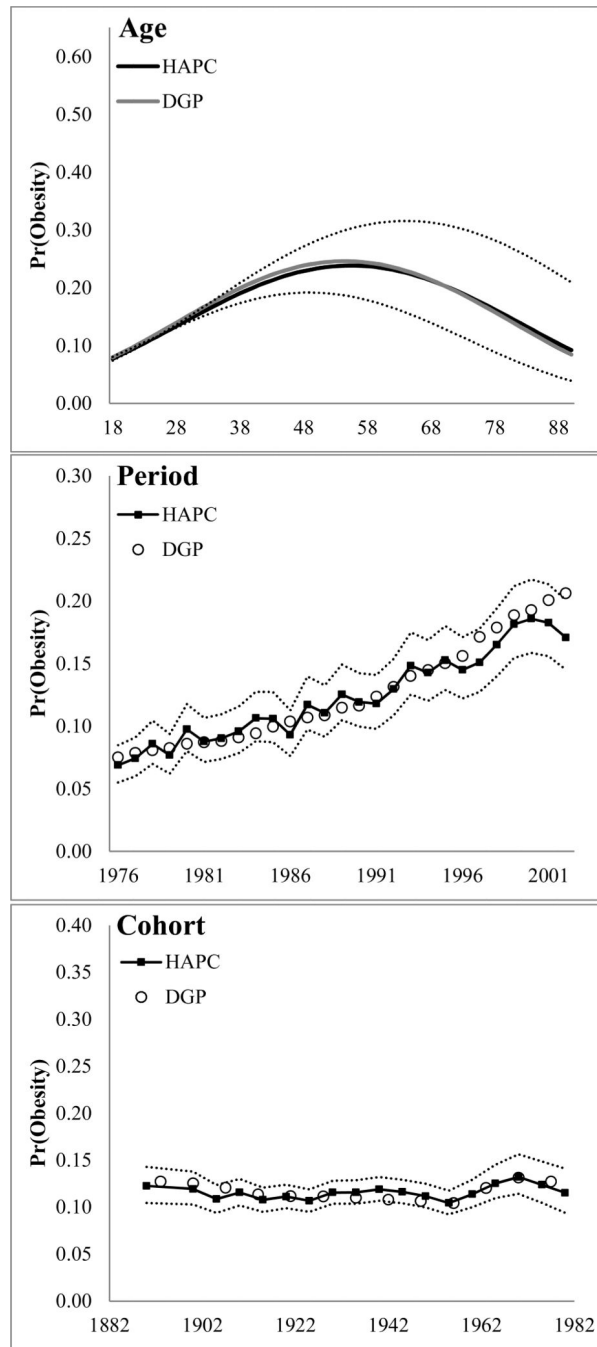


Figure 4. Replicated Results from Reither et al. (2009). HAPC-CCREM Fitted to Simulated Data with 7-year Cohorts in DGP, with 95% credible intervals (dotted lines).

Table 1

Goodness-of-Fit Statistics for Models Fitted to Simulated Data and NHIS Data

	<u>DGP for Simulation 1</u>				<u>DGP for Simulation 4</u>			
	<u>B&J Eq(4)</u>				<u>Reither Replicate, with 7-Year Cohorts</u>			
	A	AP	AC	APC	A	AP	AC	APC
Log Likelihood	-7546.2	-7450.0	-8230.5	-8230.5	-13797.1	-13544.2	-13554.6	-13531.6
AIC	15098.3	14958.1	14950.9	14979.9	27600.2	27146.3	27149.2	27155.2
BIC	15123.2	15199.1	15117.1	15362.1	27625.1	27387.3	27315.4	27537.4
df	3	29	20	46	3	29	20	46
N	30,000	30,000	30,000	30,000	30,000	30,000	30,000	30,000
	<u>DGP for Simulation 2</u>							
	7-Year Categorical Cohorts							
	A	AP	AC	APC	A	AP	AC	APC
Log Likelihood	-13013.7	-12973.5	-12982.9	-12963.8	-229337.3	-228055.6	-227994.5	-227873.3
AIC	26033.3	26005.0	26007.7	26019.7	458680.7	456153.1	456023.0	455816.7
BIC	26058.2	26246.0	26182.2	26410.2	458713.4	456382.3	456208.5	456198.5
df	3	29	20	46	3	21	17	35
N	30,000	30,000	30,000	30,000	404,782	404,782	404,782	404,782
	<u>DGP for Simulation 3</u>							
	3-Year Categorical Cohorts							
	A	AP	AC	APC	A	AP	AC	APC
Log Likelihood	-19054.0	-18997.4	-18995.1	-1876.0	-764779.4	-751395.2	-751787.8	-751065.7
AIC	38113.9	38052.7	38032.2	38046.1	1529565.0	1502848.0	1503616.0	1502223.0
BIC	38138.8	38293.7	38206.7	38436.6	1529602.0	1503206.0	1503862.0	1502790.0
df	3	29	20	46	3	29	20	46
N	30,000	30,000	30,000	30,000	1,645,884	1,645,884	1,645,884	1,645,884

Note: BIC is the Bayesian Information Criterion and is estimated to be $-2*\log(L) + df*\log(N)$. The model fit statistics were estimated for unweighted HAPC-CCFEMs. Results from the NHIS using weighted HAPC-CCFEMs and results from the NHIS using HAPC-CCREMs are consistent with those above.

Table 2
Differences in Measurement and Functional Form across Four Simulation Exercises

	Variable Grouping				Functional Form		
	DGP	HAPC-CCREM	DGP	HAPC-CCREM	DGP	HAPC-CCREM	HAPC-CCREM
Simulation 1	Age Effects	continuous	continuous	continuous	quadratic		quadratic
	Period Effects	27 1-year	27 1-year	27 1-year	none		functional free
	Cohort Effects	13 7-year	13 7-year	18 5-year	exact linear		functional free
	Age Effects	continuous	continuous	continuous	quadratic		quadratic
Simulation 2	Period Effects	27 1-year	27 1-year	27 1-year	none		functional free
	Cohort Effects	14 7-year	14 7-year	19 5-year	near linear		functional free
	Age Effects	continuous	continuous	continuous	quadratic		quadratic
	Period Effects	27 1-year	27 1-year	27 1-year	none		functional free
Simulation 3	Cohort Effects	34 3-year	34 3-year	19 5-year	near linear		functional free
	Age Effects	continuous	continuous	continuous	quadratic		quadratic
	Period Effects	27 1-year	27 1-year	27 1-year	monotonic increase		functional free
	Cohort Effects	13 7-year	13 7-year	18 5-year	functional free		functional free

Note: Simulations 1 and 4 are strict interpretations of Reither et al. (2009), who collapsed 1877–1899 into a single cohort to ensure a sufficient number of participants in this group. Simulations 2 and 3 relax this slightly by introducing an additional cohort prior to 1900 to improve the consistency of cohort widths.