Fitting Age-Period-Cohort Models Using the Intrinsic Estimator: Assumptions and Misapplications

Ryan K. Masters[1], **Daniel A. Powers**[2], **Robert A. Hummer**[3], **Audrey Beck**[4], **Shih-Fan Lin**[5], and **Brian Karl Finch**[6]

Ryan K. Masters: ryan.masters@colorado.edu

[1]Department of Sociology and Institute of Behavioral Science, University of Colorado at Boulder, UCB 327 Ketchum 214, Boulder, CO 80309, USA

[2]Department of Sociology and Population Research Center, University of Texas at Austin, Austin, TX, USA

[3]Department of Sociology and Carolina Population Center, University of North Carolina–Chapel Hill, Chapel Hill, NC, USA

[4]Department of Sociology and Center for Health Equity Research and Policy, San Diego State University, San Diego, CA, USA

[5]Institute for Behavioral and Community Health (IBACH), San Diego State University, San Diego, CA, USA

[6]Department of Sociology and Southern California Population Research Center, University of Southern California, Los Angeles, CA, USA

Abstract

We thank *Demography*'s editorial office for the opportunity to respond to te Grotenhuis et al.'s commentary regarding the methods used and the results presented in our earlier paper (Masters et al. 2014). In this response, we briefly reply to three general themes raised in the commentary: (1) the presentation and discussion of APC results, (2) the fitting of full APC models to data for which a simpler model holds, and (3) the variation in the estimated age, period, and cohort coefficients produced by the intrinsic estimator (IE) (i.e., the "non-uniqueness property" of the IE, as referred to by Pelzer et al. (2015)).

Presenting Results From APC Analyses: Choosing a Reference Category

We remind te Grotenhuis et al. that we were not interested in highlighting the size of black-white differences in mortality rates but rather were interested in examining temporal variation in these rates. As we cautioned readers in that article, "the graphical depictions in all figures are used only to isolate and present the patterns of each temporal dimension; they are not to be interpreted as representative of the actual mortality rates experienced by a specific cohort in a specific period" (Masters et al. 2014:2057). When discussing the results,

Correspondence to: Ryan K. Masters, ryan.masters@colorado.edu.

we focused exclusively on presenting patterns of change, not sizes of differences. That differences exist between U.S. black and white adult mortality rates is well known, and also well known is that the differences have narrowed in recent years (Arias 2014; Harper et al. 2007, 2012). Our analyses were strictly concerned with decomposing the sources of these changes. Thus, when graphing cohort-based variation in black and white men's and women's mortality rates, we chose values of age and period that were close to their respective means. The choice of these values was not to exacerbate or minimize hypothetical difference in the *levels* of the rates, but rather to hold constant the variation in mortality associated with both age and period. The *size* of the mortality gap would, of course, change if one were to center the rates using any other age-period combination, but patterns of the estimated cohort-based changes would not.

Linear and Nonlinear Functional Forms of APC Effects and Data-Generating Processes

Te Grotenhuis et al. refer to Luo's (2013) exercises showing the inability of the IE to recover "true" age, period, and cohort effects behind data-generating processes (DGPs) in her simulated data structures. In our online appendix (Masters et al. 2014: Online Resource 1), we noted that "the circumstances that Luo created to show the IE's inability to retrieve the 'true' age, period, and cohort effects are highly unlikely to transpire in real-world applications" (p. 2). First, we pointed out that many of Luo's DGPs assumed linear functional forms of age, period, and/or cohort effects, which have been shown to be highly problematic for APC analyses (Kupper et al. 1985; Reither et al. 2015; Yang and Land 2013). Second, graphing period-specific age-based variation in outcomes has long been a descriptive tool to check for possible cohort-based variation in outcomes of interest (Frost 1940; Kupper et al. 1985; Yang and Land 2013). Therefore, we also pointed out that the variation in Luo's outcomes exhibited parallel period-based shifts when plotted across age. Applying APC methods to these data structures, in which two-factor models (e.g., age-period or age-cohort) are preferred, will result in biased estimates. In their commentary, te Grotenhuis et al. refer only to the former point, stating, "Masters and colleagues contend that the IE can be applied in their research because the mortality data they use contain nonlinear effects, whereas 'Luo set the functional forms of all APC effects on the outcome Y to be exactly linear' (2014: Online Resource 1). However, Luo (2013) demonstrated that the IE estimates are invalid when the IE's constraint is not satisfied in simulated data with both linear and nonlinear effects (see, e.g., Luo 2013: figure 1, 1963–1964)."

Te Grotenhuis et al.'s sole emphasis on linear versus nonlinear effects in DGPs misses the crucial point that Luo's data structures do not exhibit variation across all three temporal dimensions. For example, Fig. 1 shows the age-specific patterns of Y across periods in the data Luo used to produce her figure 1.

The period-based differences in the age-specific values of Y are almost perfectly equal, suggesting that the patterns do not suffer "a lack of 'parallelism' among these curves" (Kupper et al. 1985:815). Thus, although Luo used quadratic functional forms for all temporal effects in her DGP, the resulting data structure is not appropriate for fitting a full

APC model because no substantive variation is observed in Y beyond age and period. Indeed, when comparing the Bayesian information criteria (BIC) from two-factor models— age-period (AP) and age-cohort (AC)—with the BIC for the full APC model, we see the AP model is preferred for fitting the variation in Y. BIC_{APC} (34 df) = 93,174.7; BIC_{AC} (26 df) = 93,167.8; and BIC_{AP} (18 df) = 93,119.4. In our original article's supplementary material (Online Resource 1), we graphed period-based trends in the age-specific mortality rates of U.S. black and white men and women to highlight nonparallel variation, which motivates a possible application of a full APC model. We did so to contrast the variation we observed in these rates to the parallel period-based trends observed in Luo's (2013) simulations used to critique the IE for APC analyses. The invariance in Luo's simulated data is present in the DGPs created from both linear and nonlinear effects, a point te Grotenhuis et al. sidestep in their comment.

"Alternative Estimates" or Misapplication of IE?

Finally, as to te Grotenhuis et al.'s central claim that the IE can produce wildly different APC effects, we reply by first reemphasizing the fact that the IE does not estimate "the unique" solution, which our original article explicitly acknowledged. We agreed with Luo's position when we wrote that all APC models "provide just *one* possible solution from the infinite number of solutions" (Masters et al. 2014:2066). On the other hand, we did "defend the IE as the preferred solution to estimating APC variation in U.S. adult mortality rates in the National Vital Statistics System data" (Masters et al. 2014:2066). Second, we question te Grotenhuis et al.'s use of dummy coding when applying the IE, given that effect coding is a central component of the IE method. To see how and why, revisit Kupper et al.'s (1985) classic review of APC analysis and Kupper and Janis (1980).

Starting with the APC multiple classification model's specific structure,

$$Y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{a-i+j} + \varepsilon_{ij}, \quad (1)$$

where $i = 1, 2, \ldots, a$ and $j = 1, 2, \ldots, p$; $Y_{ij} = f(\hat{R}_{ij}) = f(O_{ij}/N_{ij})$ represents a function of the observed rate $\hat{R}_{ij}$, μ is the overall mean, α_i is the fixed effect of the ith age category, β_j is the fixed effect of the jth period category, and γ_{a-i+j} is the fixed cohort effect associated with the ith age category and the jth period category.

Kupper et al. (1985:818) noted,

as with all fixed-effect ANOVA-type models, it is natural to reparameterize model (1) to its equivalent form

$$Y_{ij} = \mu^* + \alpha_i^* + \beta_i^* + \gamma_{a-i+j}^* + \varepsilon_{ij}, \quad (2)$$

where

$\mu^* = (\mu + \bar{\alpha} + \bar{\beta} + \bar{\gamma})$, $\alpha_i^* = (\alpha_i - \overline{\alpha})$, $\beta_j^* = \left(\beta_j - \overline{\beta}\right)$, and $\gamma_{a-i+j}^* = (\gamma_{a-i+j} - \overline{\gamma})$; clearly, then, we have

$$\sum_{i=1}^{a} \alpha_i^* = \sum_{j=1}^{p} \beta_j^* = \sum_{k=1}^{a+p-1} \gamma_k^* = 0. \quad (3)$$

It is important to emphasize that the reparameterized model (3) simply reexpresses each effect in model (1) as a deviation from the mean of all effects of that type, and such centering creates no distortion with respect to assessing patterns in estimated effects. In contrast, the use of unnatural constraints like $\alpha_1 = \beta_1 = \gamma_1 = 0$ does not lead to a straightforward equivalent representation of equation (1), and *can produce misleading patterns in estimated coefficients* [emphasis added].

Indeed, in their prior work that preceded and influenced the development of the IE, Kupper and Janis (1980:10) reparameterized Eq. (1) using effect coding "so as to deal *directly* with the constraint" they imposed in their solution to the APC identification problem. Their proposed solution, fundamental equations, and coding of the model design matrix became the standard in later approaches (e.g., Fu 2000). The original approach adopts a centered-effects coding, with the last category used as the reference. This may have seemed innocuous at the time, but it was a convention that made sense. Rather than impose a reference effect of 0 for each of the APC factors, one opts for grand-mean centering of effects. In a model with collinear predictors, a given dummy-variable design will strongly privilege a particular solution in the solution space. Centering the design tends to counter the tendency for alternative solutions to traverse a large area within that space. One cannot work around the parameter invariance, but one can constrain this invariance to a smaller region of the solution space by centering predictors. In this respect, the centered-effects coding is analogous to a form of standardization or rescaling of the input variables, such as what is required when performing principal components regression, ridge regression, and other related methods. In fact, it has long been known that the IE can be motivated from these different perspectives (see, e.g., Fu 2000; Kupper and Janis 1980). It is also well known that principal components and ridge regression estimators are sensitive to scaling of the inputs; it has been shown that the IE is as well (Luo et al. forthcoming). Thus, it makes more sense to center/scale the inputs than to adopt a cornered-effects parameterization of the APC design matrix. As Kupper and Janis (1980:10) noted,

> the philosophy behind such methods is that it is worthwhile to introduce a slight bias into an estimation procedure if there is an accompanying large decrease in variance…such a philosophy becomes all the more appealing when one accepts the fact that the form of the true underlying population model is never actually known, and that the validity of the claim of unbiasedness rests on such unavailable knowledge. It is in this spirit that we propose an estimation technique.

Thus, reparameterizing the model using effect coding was a deliberate strategy and a central feature of the method itself. Other data-dependent properties are likely to be at work here that cannot be revealed by simulations involving tabular data with only a few age and period

categories, such as the type of simulations done by Luo (2013) and others, which consist of minimal data configurations. Kupper and Janis (1980:12) acknowledged this point, admitting that their estimator "is not an unbiased estimator…unless the constraint actually… holds among the population parameter values. However, limited numerical evaluations have suggested that the bias is small if a and p both exceed 5 in value." Indeed, as the number of age and period categories grows to realistic counts present in data sets spanning long periods, results tend to be less sensitive to alternative design coding. This very much appears to be the case with our results, as we demonstrated in our original article's supplementary material.

In short, taking into consideration the known problems with dummy coding and the inherent strategy of using effect coding in the IE, it is unclear *why* a researcher would choose to set a specific APC reference category effect to 0, as te Grotenhuis et al. did in their comment. This choice appears to reify a particular reference group and would thus require strong theoretical justification to explain why the particular chosen contrasts are more informative than others. The critique by te Grotenhuis et al. presents results from coding strategies that probably should not be used for comparative purposes.

The IE was designed with effect coding as essential to its estimation strategy. Therefore, te Grotenhuis et al.'s attempt to show a high degree of variation in the IE model's estimates by moving away from effect coding and applying the IE using dummy coding is puzzling. The inconsistency in model estimates that is central to their critique stems *not from the IE model itself*, but instead from their use of inappropriate coding schemes. To show the consistency of model estimates from the application of the IE as it was intended, contrast the variation in estimates in Pelzer et al.'s (2015) figure 2—estimated using effect coding—with the variation in estimates in their figure 3—estimated using dummy coding. The variation in the estimates in their figure 3 stems not from the IE, but rather from the use of dummy coding. Thus, although te Grotenhuis et al. present their results as "alternative estimates" produced by the IE itself, they are anything but. The results they present are derived from a misapplication of the IE resulting from the use of dummy coding in their estimation strategy, which has been known to produce "misleading patterns in estimated coefficients" (Kupper et al. 1985:818). Estimates of age, period, and cohort coefficients from IE models using effect coding are quite consistent with one another, irrespective of the choice of reference category; this is shown clearly in both our supplementary analyses (Masters et al. 2014: Online Resource 1) and in Pelzer et al.'s (2015) own figure 2.

To conclude, we stand by the results of our article as originally published. We hope that researchers who are interested in better understanding temporal changes in demographic, social, and economic phenomena will continue to carefully apply the IE and other updated APC methods to do so.

References

Arias, E. National Vital Statistics Reports. Vol. 63. Hyattsville, MD: National Center for Health Statistics; 2014. United States life tables, 2010.

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

Fu WJ. Ridge estimator in singular design with applications to age-period-cohort analysis of disease rates. Communications in Statistics Theory and Method. 2000; 29:263–278.

Harper S, Lynch J, Burris S, Davey Smith G. Trends in the black-white life expectancy gap in the United States, 1983–2003. JAMA. 2007; 297:1224–1232. [PubMed: 17369405]

Harper S, Rushani D, Kaufman JS. Trends in the black-white life expectancy gap, 2003–2008. JAMA. 2012; 307:2257–2259. [PubMed: 22706828]

Kupper, LL., Janis, JM. The multiple classification model in age, period, and cohort analysis: Theoretical considerations (Institute of Statistics Mimeo No. 1311). Chapel Hill: Department of Biostatistics, University of North Carolina; 1980.

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]

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, Powers DA, Winship C. The sensitivity of the intrinsic estimator to coding schemes: A comment on Yang, Schulhofer-Wohl, Fu, and Land. American Journal of Sociology. Forthcoming.

Masters RK, Hummer RA, Powers DA, Beck A, Lin SF, Finch BK. Long-term trends in adult mortality for U.S. blacks and whites: An examination of period- and cohort-based changes. Demography. 2014; 51:2047–2073. [PubMed: 25403151]

Pelzer B, te Grotenhuis M, Eisinga R, Schmidt-Catran AW. The non-uniqueness property of the intrinsic estimator in APC models. Demography. 2015; 52:315–327. [PubMed: 25550143]

Reither EN, Masters RK, Yang YC, Powers DA, Zheng H, Land KC. Should age-period-cohort studies return to the methodologies of the 1970s? Social Science & Medicine. 2015; 128:356–365. [PubMed: 25617033]

Yang, YC., Land, KC. Age-period-cohort analysis: New models, methods, and empirical applications. Boca Raton, FL: CRC Press; 2013.
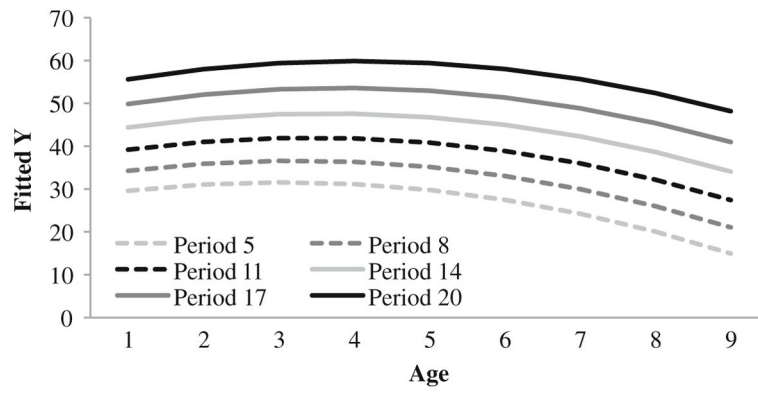
Fig. 1.
Period-specific age-based variation in predicted values of Y from Luo's (2013) figure 1 data-generating process: $Y = 10 + 2(\text{age}) - .5(\text{age}^2) + 1(\text{period}) - .015(\text{period}^2) + .15(\text{cohort}) + .03(\text{cohort}^2)$