American Sociological Review
76(6) 955–983

DOI: 10.1177/0003122411430940
http://asr.sagepub.com

Variance Function Regression in Hierarchical Age-Period-Cohort Models: Applications to the Study of Self-Reported Health

Hui Zheng,[a] Yang Yang,[b] and Kenneth C. Land[c]

Abstract

Two long-standing research problems of interest to sociologists are sources of variations in social inequalities and differential contributions of the temporal dimensions of age, time period, and cohort to variations in social phenomena. Recently, scholars have introduced a model called Variance Function Regression for the study of the former problem, and a model called Hierarchical Age-Period-Cohort regression has been developed for the study of the latter. This article presents an integration of these two models as a means to study the evolution of social inequalities along distinct temporal dimensions. We apply the integrated model to survey data on subjective health status. We find substantial age, period, and cohort effects, as well as gender differences, not only for the conditional mean of self-rated health (i.e., between-group disparities), but also for the variance in this mean (i.e., within-group disparities)—and it is detection of age, period, and cohort variations in the latter disparities that application of the integrated model permits. Net of effects of age and individual-level covariates, in recent decades, cohort differences in conditional means of self-rated health have been less important than period differences that cut across all cohorts. By contrast, cohort differences of variances in these conditional means have dominated period differences. In particular, post-baby boom birth cohorts show significant and increasing levels of within-group disparities. These findings illustrate how the integrated model provides a powerful framework through which to identify and study the evolution of variations in social inequalities across age, period, and cohort temporal dimensions. Accordingly, this model should be broadly applicable to the study of social inequality in many different substantive contexts.

Keywords

Hierarchical-Age-Period-Cohort-Variance-Function-Regression Model, Hierarchical Age-Period-Cohort Model, Variance Function Regression Model, health disparities

A longstanding core analytic tool of sociology is the study of inequality through regression models (Blau and Duncan 1966; Morris and Western 1999). Another is the study of social change through age-period-cohort (APC) analysis (Mason and Fienberg 1985; Ryder 1965). Use of regression-based models

[a]The Ohio State University
[b]University of North Carolina-Chapel Hill
[c]Duke University

Corresponding Author:
Hui Zheng, Department of Sociology, The Ohio State University, Columbus, OH 43210
E-mail: zheng.64@sociology.osu.edu

in APC analysis closely relates the study of cohort change to the study of inequality—substantively and methodologically—yet there have been no analytic tools for systematic examination of age and temporal (i.e., period and cohort) variations in inequalities beyond those captured by conditional means of regression models.

In this article, we fill this gap by intersecting two recent developments in statistical models for analysis of social and demographic data. Specifically, we embed a Variance Function Regression (VFR) model within a Hierarchical Age-Period-Cohort (HAPC) analysis. This facilitates decomposition not only of *between-group* inequality into age, period, and cohort components (i.e., variations in the conditional mean of an outcome across age, period, and cohort), but also a similar APC decomposition of *within-group* inequality (i.e., variations in the conditional variance or dispersion of an outcome across age, period, and cohort). More generally, the combined model allows integration of theories of social stratification and social change and opens the door to a better, more comprehensive understanding of the dynamics and heterogeneity of social processes by which individual lives unfold over the life course and are shaped by historical time and social context represented by cohort membership.

MODELS TO BE INTEGRATED: A BRIEF OVERVIEW

We begin with a brief overview of the two models we aim to integrate. Standard regression-based approaches to studies of inequality are largely limited to between-group differences. Here, *group* means each category of a covariate. For example, gender has two groups: men and women. In this case, between-group inequality is inequality between men and women; within-group inequality is the remaining inequality within the population of men or within the population of women. Within-group or residual inequality often far exceeds between-group inequality. Scholars often describe residual inequality as due to measurement error or the influence of unobserved or hidden heterogeneity. A recently developed class of statistical models, termed Variance Function Regression (VFR) by Western and Bloome (2009), and, more generally, Heteroscedastic Regression (HR) in statistics (Smyth 1989), can address this limitation by simultaneously modeling the mean and the variance of an outcome variable as functions of covariates; these models take into account both between- and within-group differences. This class of models explicitly targets residual variance in regression models for analysis.

Conventional linear regression models assume that models' residual error terms are independently and identically distributed with constant or homoscedastic variance, and, especially important for small samples for which asymptotic statistical properties of estimators do not apply, that errors have normal probability distributions (see, e.g., Fox 2008). Violations of these assumptions affect estimators of the standard errors of regression coefficients and reduce the statistical efficiency of conventional least-squares estimators. Scholars have developed a variety of statistical methods for diagnosing and correcting nonconstant error variances (Fox 2008). The key feature of VFR/HR is that it treats violations of homoscedasticity as more than a data problem that must be corrected to obtain well-behaved estimators. Instead, it approaches these violations as having potential substantive importance and builds regression models to account for them. In applications to the study of inequality, such as Western, Bloome, and Percheski's (2008) study of trends in family income inequality in the United States from 1975 to 2005, residual variance can be interpreted as measuring within-group risk or insecurity.

Next, consider the three time-related dimensions—age, time period, and birth cohort effects—the distinction of which is crucial for proper inference in studies of temporal change in many social domains. *Age effects* represent variation across different age groups brought about by physiological changes, accumulation of social experience,

or role or status changes. *Period effects* represent variation over time periods that affect all living age groups simultaneously, often resulting from shifts in social, cultural, economic, or physical environments. *Cohort effects* are associated with changes across groups of individuals who experience an initial event, such as birth or marriage, in the same year or years; cohort effects may be due to successive age groups having different formative experiences in successive time periods (Yang 2010).

One common goal of APC analysis is to assess effects of one of the three factors on some outcome of interest net of influences from the other two (Mason and Fienberg 1985). Conventional linear regression models fit to aggregate population rates or proportions suffer from the model identification problem due to the exact linear dependency among age, period, and cohort variables (Period = Age + Cohort) in such data (Mason et al. 1973). A recently developed modeling approach, Hierarchical APC (HAPC) models, avoids this problem by using micro data and multilevel modeling frameworks. The HAPC approach conceptualizes time periods and cohort memberships as social historical contexts within which individuals are embedded and ordered by age and models them as random rather than fixed effects additive to age (Yang 2006; Yang and Land 2006, 2008). This contextual approach broadens the theoretical foundation of APC analysis, helps deal with the identification problem, and also accounts for potentially correlated errors.[1] Empirical applications of the HAPC model to repeated cross-section survey data have estimated distinct contributions of age, period, and cohort to temporal changes across the past few decades in such phenomena as verbal ability (Yang 2006; Yang and Land 2006, 2008), happiness (Yang 2008a), and obesity (Reither, Hauser, and Yang 2009).

These two types of models and the substantive questions they address are related. APC variations represent complex and temporal patterns of inequality, and scholars often need to disentangle sources of social inequality attributable to age, time period, and birth cohort. In his seminal article, Ryder (1965) argued that cohort membership is a structural category that can be as important as other social structural features, such as socioeconomic status (SES), in determining behavior. APC analysis is, in this sense, synonymous with cohort analysis (Smith 2008). To the extent that SES inequalities are defined by both between- and within-group differences (Western et al. 2008), social inequalities by age, period, and cohort should also be assessed in terms of between-group differences and within-group dispersions. Differences in either term can bring about subsequent social demographic change at the population level. However, an integrated model that decomposes these two differences across age, period, and cohort has not been available. Scholars have used VFR/HR models to examine inequalities in both terms but have not distinguished the age, period, and cohort sources of temporal variation. Studies have used HAPC regression models to examine temporal differences in conditional means but not within-group variances. This article illustrates the utility of the intersection of these two modeling frameworks with an application to the analysis of health disparities in the United States from 1984 to 2007.

RESEARCH TOPIC, STRATEGY, AND MODELS

Health Disparities

The empirical analysis in this study concerns health outcomes and disparities or inequalities therein. We use the term *health disparities* to refer to either between- or within-group differences in health and distinguish the two aspects of inequality in specific circumstances. In the context of APC analysis, groups are defined by age, time period, and cohort categories. *Between-group health disparities* refer to variations in the conditional mean (conditional on a set of individual-level sociodemographic variables) of health across age, period, and cohort. *Within-group health*

disparities refer to the conditional variance or dispersion of health within each category of age, period, or cohort. *Changes in within-group health disparities* refer to variations in the conditional variance or dispersion of health across age, period, or cohort.

In addition to a large body of demographic and epidemiologic research on age variation and temporal trends in health and mortality that addresses between-group health disparities, scholars have taken three standard approaches to the study of changes in within-group health disparities: (1) across the life course (e.g., Dannefer 2003; House et al. 1994), (2) across cohorts (e.g., Chen, Yang, and Liu 2010; Warren and Hernandez 2007; Yang and Lee 2009), and (3) across time periods (e.g., Goesling 2007; Pappas et al. 1993). Within each approach, evidence shows significant change in health disparities. For example, gaps in self-rated health, physical functioning, well-being, disease incidence, and mortality by education levels have widened over the life course (see Dupre 2007; Lauderdale 2001; Lynch 2003; Ross and Wu 1996). The gap in self-rated health by education level has also widened across birth cohorts (Chen et al. 2010; Lynch 2003), whereas intracohort gaps by sex and race have been constant across birth cohorts (Yang and Lee 2009). Evidence also shows increasing socioeconomic inequality in health, disability, and life expectancy in the United States over the past several decades (e.g., Crimmins and Saito 2001; Feldman et al. 1989; Goesling 2007; Hummer, Rogers, and Eberstein 1998; Jemal et al. 2008; Liu and Hummer 2008; Meara, Richards, and Cutler 2008; Pappas et al. 1993; Preston and Elo 1995; Schoeni et al. 2005). In addition, recent research documents significant period changes in gender and race inequalities in happiness (Yang 2008a) and period changes in sex differences in cause-specific and total mortality (Yang 2008b). Increasing gender, race, and SES inequalities across the life course, birth cohorts, and time periods conceivably contribute to increasing overall inequality or dispersion across these dimensions. This has not been examined previously, however, and merits a formal test using properly constructed analytic models.

Another limitation of much prior research is that it treats these three approaches separately, yet they are intertwined. For example, an increase in health disparities across time periods may result from either cohort replacement, in which cohorts with larger within-cohort health disparities succeed cohorts with smaller disparities, or an aging society, wherein the elderly, who usually have larger within-age health disparities than do younger people, increase their proportionate share in the population, or from some combination of the two. Similarly, a widening health disparity with age may be confounded with temporal patterns. That is, period patterns in health disparities may affect age variations in health disparities. Cohort patterns may also influence a widening health disparity across age groups. Studies that have tried to disentangle age and cohort patterns in health disparities have found distinct age effects and cohort variations in mean levels of health, as well as changing health disparities by education, income, gender, and race over the life course and across birth cohorts (Chen et al. 2010; Lauderdale 2001; Lynch 2003; Yang and Lee 2009). Lynch (2003) also finds that each pattern is suppressed when the other one is ignored. However, an integrated model that simultaneously assesses effects of age, period, and cohort in between- and within-group health disparities has not heretofore been presented.

The key outcome variable in this article is self-rated health, a widely used measure of general health status. Research shows that self-rated health is highly predictive of mortality and strongly correlated with objective assessments of health, including physician diagnoses (Idler and Benyamini 1997). In fact, self-rated health is a good indicator of objective health and subclinical illness; some studies find it is more predictive of mortality among the elderly than are physician assessments (Hays et al. 1996; Schoenfeld et al. 1994). Close relationships between self-rated

health and objective health indicators also hold across population subgroups (Bosworth et al. 1999; Kennedy, Kasl, and Vaccarino 2001). Recent studies suggest that the gap in self-rated health by education levels has widened across age, time period, or cohort (Goesling 2007; Liu and Hummer 2008; Lynch 2003). These findings are consistent with other studies investigating disease incidence, mortality, or life expectancy gaps by education levels across age, time period, or cohort (e.g., Dupre 2007; Lauderdale 2001; Meara et al. 2008; Pappas et al. 1993). Based on these and related findings regarding its robustness as a single, summary index of an individual's health status, we study self-rated health as a health outcome variable. We caution, though, that health is not a singular condition and our findings may not always generalize to health disparity trends associated with specific health outcomes.

Some prior research has produced evidence of different temporal trends in health disparities by sex. For example, Feldman and colleagues (1989) found that educational differentials in death rates widened for men but were relatively stable for women between 1960 and 1984. The National Center for Health Statistics (1994) reported that the racial gap in life expectancy widened much more for men (from 6.9 to 8.3 years) than for women (from 5.6 to 5.8 years) between 1980 and 1991. Preston and Elo (1995) found that educational disparities in adult mortality widened for men but contracted for working-age women. By contrast, Meara and colleagues (2008) found that educational disparities in life expectancy widened among women in the past two decades. In summary, these studies generally find that within-period health disparities have widened for men but vary for women depending on the time periods and health outcomes examined. That is, substantial evidence shows major gender differences in temporal trends in health disparities. Therefore, we also investigate variations in gender-specific self-rated health disparities across age, time period, and cohort;[2] that is, we investigate how health disparities may change across age, time period, and cohort within each gender.

Our main purpose here is to present a method that facilitates disentanglement of age, period, and cohort variations in health disparities defined by differences in conditional mean levels and conditional dispersions of health (i.e., between- and within-group health disparities). To do so, we integrate the Hierarchical Age-Period-Cohort (HAPC) model with the Variance Function Regression/Heteroscedastic Regression (VFR/HR) model. The HAPC model enables us to disentangle age, period, and cohort effects. The VFR/HR model enables us to separate within-group from between-group health disparities. We integrate these two statistical models and term the result a Hierarchical-Age-Period-Cohort-Variance-Function-Regression Model (HAPC-VFR/HR).

Hierarchical Age-Period-Cohort Analysis

Use of multilevel models aids in estimation of age, period, and cohort components of temporal change because, within these models, cohorts or periods can be conceptualized as higher-level contexts rather than individual attributes similar to age. As such, they do not rest on the assumption of linearity and additivity of the three variables in the conventional linear regression model, which inevitably incurs the identification problem (Yang 2010). Few scholars utilize multilevel models in analysis of temporal change, but some studies exist in demographic research on health (Lynch 2003) and developmental psychology and aging research on cognitive skills (e.g., Alwin 2009). These studies do not explicitly embody a full-blown APC analysis due to the identification problem; instead, they focus on age patterns only in the context of cohorts.

The key contribution in the HAPC approach to cohort analysis developed by Yang and Land (2006, 2008) is the simultaneous modeling of all three factors using micro data and mixed-effects or hierarchical models.

Note that individual-level data available in survey designs allow age intervals to differ from period intervals. Unequal age, period, and cohort intervals then break the exact linear dependency of the three variables in the APC accounting model suited for aggregate population-level data. This solution to the identification problem is unsatisfactory for two reasons (Yang 2010). It is still embedded in the simple linear regression model that assumes linearity and additivity of the three variables and thus does not completely avoid the identification problem. Results may be sensitive to the choice of interval widths, because longer widths may allow a higher degree of overidentification. More importantly, simple linear models do not account for potential correlated errors of individual sample respondents grouped into periods or cohorts. Ignoring multilevel heterogeneity in data may lead to underestimated standard errors.

The HAPC approach utilizes unique features of multilevel survey design and presents a more thorough solution. It begins with the recognition that in this design, respondents are simultaneously nested in and cross-classified by two higher-level social contexts defined by time period and birth cohort. A reasonable alternative to the linear model, then, is a different family of models that do not assume fixed age, period, and cohort effects that are additive. This avoids the identification problem and can statistically characterize contextual effects of historical time and cohort membership. The HAPC model—specifically, the cross-classified random-effects model (CCREM)—satisfies these criteria and can accommodate covariates at individual and contextual levels for better conceptualization of specific social processes generating observed patterns in the data. In addition to the verbal test outcome analysis illustrated in Yang and Land (2006), two examples of applications of such models to social data can be found in Yang's (2008a) study of inequalities in subjective well-being and Reither and colleagues' (2009) study of the obesity epidemic in the United States.

This HAPC-CCREM approach to APC analysis can be illustrated with a linear mixed-effects or hierarchical regression model for data on an outcome variable Y for which we specify variability associated with individuals, cohorts, and periods as follows:

Level-1 or Within-Cell Model:[3]

$$Y_{ijk} = \beta_{0jk} + \beta_1 X_{1ijk} + \beta_2 X_{2ijk} + ... + \beta_P X_{Pijk} + e_{ijk}, e_{ijk} \sim N(0, \sigma^2) \quad (1)$$

Level-2 or Between-Cell Model:

$$\beta_{0jk} = \gamma_0 + u_{0j} + v_{0k}, u_{0j} \sim N(0, \tau_u), \quad v_{0k} \sim N(0, \tau_v) \quad (2)$$

Combined or Mixed-Effects Model:

$$Y_{ijk} = \gamma_0 + \beta_1 X_{1ijk} + \beta_2 X_{2ijk} + ... + \beta_P X_{Pijk} + u_{0j} + v_{0k} + e_{ijk} \quad (3)$$

for $i = 1, 2, \ldots, n_{jk}$ individuals within cohort j and period k;
$j = 1, \ldots, J$ birth cohorts; and
$k = 1, \ldots, K$ time periods (survey years);

where within each birth cohort j and survey year k, respondent i's outcome, Y_{ijk}, is modeled as a function of explanatory variables/covariates X_{1ijk}, X_{2ijk}, . . ., X_{Pijk}, (which are grand mean centered for continuous variables and usually include grand mean centered age and possibly higher-order functions of age such as age-squared), and the intercept varies by birth cohort and time period.[4]

In this CCREM, β_{0jk} is the intercept or cell mean, that is, the mean Y of individuals who belong to birth cohort j and were surveyed in year k; $\beta_1, \ldots \beta_p$ are level-1 fixed effects; e_{ijk} is the random individual effect, that is, the deviation of individual ijk's Y from the cell mean with covariates $X = x$, which are assumed to be normally distributed with mean 0 and a within-cell variance σ^2 ; γ_0 is the expected mean at zero values of all level-1 variables averaged over all periods and cohorts; u_{0j} is the residual random effect of cohort j, that is, the contribution of cohort j averaged over all periods, on β_{0jk} , which is assumed to be normally distributed

with mean 0 and variance τ_u; and v_{0k} is the residual random effect of period *k*, that is, the contribution of period *k* averaged over all cohorts, assumed to be normally distributed with mean 0 and variance τ_v. In addition, $\beta_{0j} = \gamma_0 + u_{0j}$ is the cohort *Y* score averaged over all periods with all individual-level covariates at grand mean level; and $\beta_{0k} = \gamma_0 + v_{0k}$ is the period *Y* score averaged over all cohorts with all individual-level covariates at grand mean level.

The HAPC-CCREM model specified in Equations 1 through 3 is a random intercepts model that specifies that significant random variation across cohorts and periods occurs only in the intercepts and not in the slopes of regressors at the individual level. Specification of such a model for a specific empirical application should be preceded by preliminary testing using standard methods (see, e.g., Raudenbush and Bryk 2002) to determine whether there is evidence of random variation across time periods or cohorts in the level-1 slope coefficients. If evidence suggests such significant variation, the model should be modified to incorporate this variation.

Variance Function Regression

Variance Function Regression/Heteroscedastic Regression[5] has two parts: a regression for an outcome variable, Y_i, and a regression for the logarithm of the residual variances, $\log(\sigma_i^2)$ (Western and Bloome 2009):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_P X_{Pi} + e_i \quad (4)$$

$$\log(\sigma_i^2) = \lambda_0 + \lambda_1 Z_{1i} + \lambda_2 Z_{2i} + ... + \lambda_R Z_{Ri} \quad (5)$$

where observations on individual sample members are indexed by *i*; X_1, X_2, ..., X_P is a set of *P* explanatory variables for Y_i; Z_1, Z_2, ..., Z_R is a set of *R* explanatory variables (possibly equal to X_1, X_2, ..., X_P) for the logarithm of the residual variance $\log(\sigma_i^2)$, with residual random error term e_i for Y_i. The quantity σ_i^2 is the square of corresponding residuals, $\hat{e}_i^2$, from the first regression. From a substantive viewpoint, the first regression describes how covariates affect the Y_i response variable and account for deviations of the within-group sample means from the average or grand mean $\bar{Y}$ (i.e., between-group inequality), while the second regression explains how covariates affect within-group variability of the response variable around the group means (i.e., within-group inequality).

Intersecting the HAPC and VFR/HR Models

We integrate the VFR/HR model with the HAPC model by using the HAPC model to estimate the two equations in the variance function regression model, treating cohort and period as random effects in the context of a repeated cross-section survey research design across a broad range of ages—so that the question of the relative contributions of age, time period, and birth cohort temporal dimensions are relevant. To do so, we engage in a two-step estimation algorithm.

Step 1: Estimate the β regression coefficient vectors for between-group inequality across age, period, and cohort.

We use the restricted maximum likelihood (REML) estimator (Raudenbush and Bryk 2002) of the CCREM to estimate Equation 4 of the VFR regression model. Equations 1 through 3 present the algebra for this algorithm.

This step produces a set of estimates of fixed-effects coefficients (for individual-level explanatory covariates), random-effects coefficients (for cohorts and periods), and a random variance components matrix that evaluates the contributions of these individual-level and period and cohort contextual variables to the explanation of variance in the conditional expected value or conditional mean of the outcome variable. In the context of this study, this step estimates variations in the conditional mean of self-rated health across period, cohort, age, and other individual-level covariates.

Step 2: Estimate the λ regression coefficient vectors for within-group inequality across age, period, and cohort.

We next calculate residuals ($\hat{e}_{ijk} = Y_{ijk} - X'_{ijk}\hat{\beta}$) from the Step 1 regression for each sample respondent *i* and compute the squared residuals, $\hat{e}^2_{ijk}$ or denoted as σ^2_{ijk}. We then apply the residual pseudo-likelihood (RSPL) estimator of the CCREM to estimate Equation 5 of the VFR/HR regression model.[6] For normal distributed errors, squared residuals will have a gamma distribution, and Equation 5 is then estimated in generalized linear mixed model form, that is, as a gamma regression of $\hat{e}^2_{ijk}$ on the X_{ijk} using a log link function (Western and Bloome 2009; see also Nelder and Lee 1991).[7]

The algebra for this algorithm can be stated as follows:

Level-1 or Within-Cell Model:

$$\log(\sigma^2_{ijk}) = \lambda_{0jk} + \lambda_1 X_{1ijk} + \lambda_2 X_{2ijk} + \ldots + \lambda_P X_{Pijk} \quad (6)$$

Level-2 or Between-Cell Model:

$$\lambda_{0j} = \pi_0 + \omega_{0j} + \varphi_{0k},\ \omega_{0j} \sim N(0, \psi_\omega)$$
$$\varphi_{0k} \sim N(0, \psi_\varphi) \quad (7)$$

Combined or Mixed-Effects Model:

$$\log(\sigma^2_{ijk}) = \pi_0 + \lambda_1 X_{1ijk} + \lambda_2 X_{2ijk} + \ldots + \lambda_P X_{Pijk} + \omega_{0j} + \varphi_{0k} \quad (8)$$

where λ_{0jk} is the intercept or cell mean, that is, the mean $\log(\sigma^2)$ of individuals who belong to birth cohort *j* and were surveyed in year *k*; π_0 is the expected mean of $\log(\sigma^2)$ at the zero values of all level-1 variables averaged over all periods and cohorts; ω_{0j} and φ_{0k} are residual random effects of cohort *j* and period *k,* respectively, assumed to be normally distributed with mean 0 and variance ψ_ω and ψ_φ. In addition, $\lambda_{0j} = \pi_0 + \omega_{0j}$ is the cohort $\log(\sigma^2)$ score averaged over all periods with all individual-level covariates at grand mean level; and $\lambda_{0k} = \pi_0 + \varphi_{0k}$ is the period $\log(\sigma^2)$ score averaged over all cohorts with all individual-level covariates at grand mean level.

This step produces a set of estimated fixed-effects coefficients (for individual-level explanatory covariates), random-effects coefficients (for cohorts and periods), and a random variance components matrix that evaluates contributions of these variables to the explanation of variance in the logarithm of the residual variances, $\log(\sigma^2_i)$, for each sample respondent *i*. In the context of this study, this step estimates variations in the variance or dispersion of self-rated health across period, cohort, age, and other individual-level covariates. We should emphasize that the predicted σ^2 for age, period, or cohort represents a general form of dispersion of health across age, period, or cohort, whereas previous empirical research on health disparities focuses on specific inequality by one or two dimensions such as sex, race, or SES. An increase or decrease in any one specific dimension will contribute to the increase or decrease in the general dispersion, but this has not been studied in the current literature. Instead of examining changing effects of each specific dimension on health, this study investigates how the general dispersion of self-rated health may change across age, period, and cohort.

Iteration and Maximum Likelihood Estimation

Even though both of these steps produce restricted maximum likelihood or residual pseudo-likelihood estimators from the CCREM, they must be iterated to obtain maximum likelihood (ML) estimators for the variance function regression model (Aitkin 1987). As Western and Bloome (2009) indicate, fitted values ($\hat{\sigma}^2_{ijk}$) from an application of the two steps should be saved and used in a weighted regression of Y_{ijk} on X_{1ijk}, X_{2ijk}, . . ., X_{Pijk} with weights $(1/\hat{\sigma}^2_{ijk})$. Estimates of residuals from Step 1 are then updated, Step 2 is computed, and so forth until convergence. Western and Bloome (2009) note that the ML estimator may perform poorly in small samples, in which case an REML or Bayes estimator can be used. In our empirical application, sample sizes are very large, so adjustments in the REML made for the loss of degrees of freedom resulting from estimation of regression parameters will be very small, if not trivial. Therefore, for purposes of empirical application of the HAPC-VFR/HR model in this study, we apply the ML estimator.

DATA AND METHODS

Data

Our analysis is based on annual data from the National Health Interview Survey (NHIS) for the 24-year period from 1984 to 2007.[8] The NHIS is a multistage probability sample survey of the civilian non-institutionalized U.S. population conducted by the National Center for Health Statistics (2009). NHIS collects health information for each member of a family or household sampled, as reported by one primary respondent. To reduce reporting and measurement errors, we limit our analysis to the primary respondent. The sample size for men is about 16,670 each year (in total, 16,670 x 24 = 400,080) and about 12,575 for women (in total, 12,575 x 24 = 301,800).

The NHIS's sampling frame is redesigned every 10 years and was redesigned in 1995. Nonetheless, the fundamental design of the 1995 to 2007 NHIS is similar to that of the 1985 to 1994 NHIS. Three changes in the sampling design and weighting structure are notable. First, since 1995, the number of primary sampling locations increased from 198 to 358. Second, black and Hispanic populations were oversampled in the 1995 to 2007 NHIS, while only blacks were oversampled in the 1985 to 1994 NHIS. Third, weighting structure changed after 1996. These three changes potentially affect variances (i.e., health disparities in our study) among samples after 1995 and 1996. We therefore use sample weights in all analyses reported here to adjust for the multistage sampling design. We also created an indicator/dummy variable named "redesign" (1984 to 1994 = 0; 1995 to 2007 = 1) to adjust regression model estimates for any effects of sampling design changes since 1995.

Variables

The outcome variable, *self-rated health,* remained largely unchanged across periodic revisions of the NHIS questionnaires, which facilitates analysis of trends. It has five response categories: poor, fair, good, very good, and excellent.

Our objective is to examine age-period-cohort variations in self-reported health and disparities therein after controlling for individual-level demographic and social variables that previous research (e.g., Bird et al. 2010) has linked to health, namely, gender, race, marital status, work status, education, and income. The NHIS measured income using several categories. We first calculated the midpoint of each income category, converted the midpoints to 2007 U.S. dollars, and used $10,000 as the metric for the income variable. Education was measured as single years of formal education and ranges from 0 to 18 years. Work status is a binary variable in which 1 = full- or part-time job and 0 = not employed. We also control for several variables that are established correlates of health: sex (1 = male, 0 = female), race (1 = white, 0 = non-white), and marital status (1 = married, 0 = unmarried). Among respondents, 57 percent are men, 82 percent are white, 55 percent are married, and 67 percent are employed. Table 1 presents summary statistics for these data.

Model Specification and Estimation

The nature of the self-rated health outcome variable—in the form of five ordered response categories (poor, fair, good, very good, and excellent)—complicates specification and estimation of the combined HAPC-VFR/HR model. Earlier, we described this model in a linear mixed-effects regression format. We apply this specification to the NHIS data by scaling the self-rated health outcome variable as a five-point scale with responses numbered from 1 to 5. We do this, first, because this choice facilitates comparisons with prior research using similar self-rated health data (see, e.g., Schnittker 2007). Second, the equal-intervals assumption of the five-point scale is, in fact, a good specification for self-rated health responses in the NHIS data. We obtained evidence of this from an ordered logit regression analysis of this outcome variable.[9] Third, after presenting empirical results of this analysis, we will describe various tests

Table 1. Summary Statistics for Self-Reported Health Data from NHIS, 1984 to 2007

Outcome	Description	*N*	Mean	SD	Min	Max
Health	Self-reported health: 1 = poor, 2 = fine, 3 = good, 4 = very good, 5 = excellent	701,888	3.76	1.13	1	5
Level-1 Variables						
Sex	1 = man, 0 = woman	701,888	.57	.49	0	1
Race	1 = white, 0 = other races	701,888	.82	.38	0	1
Age	Respondent's age at survey year	701,888	46.55	17.30	18	85
Education	Respondent's years of schooling	701,888	12.65	3.17	0	18
Marital	1 = married, 0 = others	701,888	.55	.50	0	1
Employed	1 = employed, 0 = others	701,888	.67	.47	0	1
Income/10,000	Household income at survey year	701,888	4.58	2.76	.068	10.41
Redesign	1 = 1995 to 2007, 0 = 1984 to 1994	701,888			0	1
Level-2 Variables						
Cohort	Five-year birth cohorts	18			1899 to 1904	1985 to 1989
Period	Survey year	24			1984	2007

of robustness of the findings to the NHIS dataset and to this model specification. We will also describe extensions of the HAPC-VFR/HR model to more complicated model specifications that more faithfully accommodate the ordered nature of the self-rated health variable.

With respect to the two-step algorithm for estimation of the model stated earlier, analysis of the estimated conditional expectation function or mean outcome variable describes how the age, period, and cohort temporal dimensions affect the reported health outcome. These regressions tell us about differences in mean levels of self-reported health among groups defined by age, time periods, and birth cohorts as well as other measured covariates. Prior studies of health status have examined these differences in group-specific means, and the HAPC model permits decomposition of temporal changes into age, period, and cohort components. By comparison, the integrated HAPC-VFR/HR analysis of the regression model for the logarithm of the residual variances explains how dispersions of self-reported health status change temporally within these groups, that is, how health disparities change across age, period, and cohort. The integrated HAPC-VFR/HR model allows us to detect temporal changes in within-group variations and their decomposition into age, period, and cohort components.

RESULTS

Figure 1 displays sample means of self-rated health in the NHIS for 1984 to 2007 after adjusting for sample weights and smoothing annual estimates with a three-point moving average, but without controlling for individual-level covariates and disentangling age-period-cohort effects. Overall, for the whole sample, self-rated health increased from 1984 to 1990, decreased until the mid-1990s, increased afterward, and decreased again after the late-1990s. Men and women exhibit similar period-to-period trends, but women's self-rated health increased earlier and more

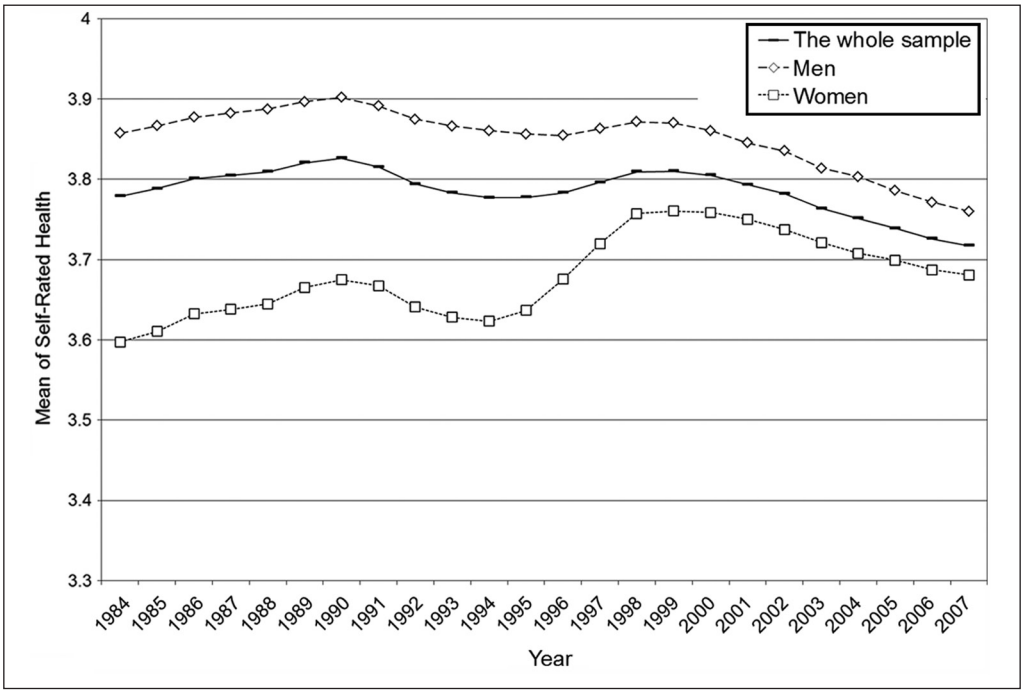

Figure 1. Observed Means of Self-Rated Health, NHIS, 1984 to 2007
Note: Trends are adjusted for sample weights and smoothed by a three-point moving average.

than did men's. These differences in overall trends significantly reduced the self-rated health gap between men and women by the late-1990s, and the reduced gap continued to the end of the series in 2007.

Figure 2 portrays the observed variance in self-reported health in the NHIS from 1984 to 2007 without controlling for individual-level covariates and disentangling age-period-cohort effects, but adjusting for sample weights and applying a three-point moving average to smooth the estimates. Overall, for the whole sample, self-reported health disparity decreased from 1984 to 1990, leveled off until around 1995, decreased afterward, and then rose again after 1998 to 1999. Among each group, men and women show similar period-to-period variations in health disparities.

Variations in Health and Health Disparities by Age, Period, and Cohort, 1984 to 2007

Table 2 reports estimates of parameters, standard errors, and model fit statistics for the HAPC-VFR/HR models of self-rated health in NHIS data from 1984 to 2007. We obtained results using the maximum likelihood estimation method described earlier. The β columns present results for the first-stage regression of the HAPC-VFR/HR model (which estimates variations in mean health across groups), and λ columns present results for the second-stage regression of the HAPC-VFR/HR model (which estimates variations in dispersion of health across groups).

As the β columns show, consistent with findings from previous studies, being male, white, married, more educated, having a job, and having more income are associated with better self-rated health. For example, being a male significantly increases the expected value of self-rated health by .03 points on a five-point scale.[10] Regarding race, being white is associated with a highly significant .165 point increase in expected value of self-rated health. Being married significantly increases the expected value of self-rated health by about .02 points. Regarding education, each year of additional education is associated with a statistically significant .06

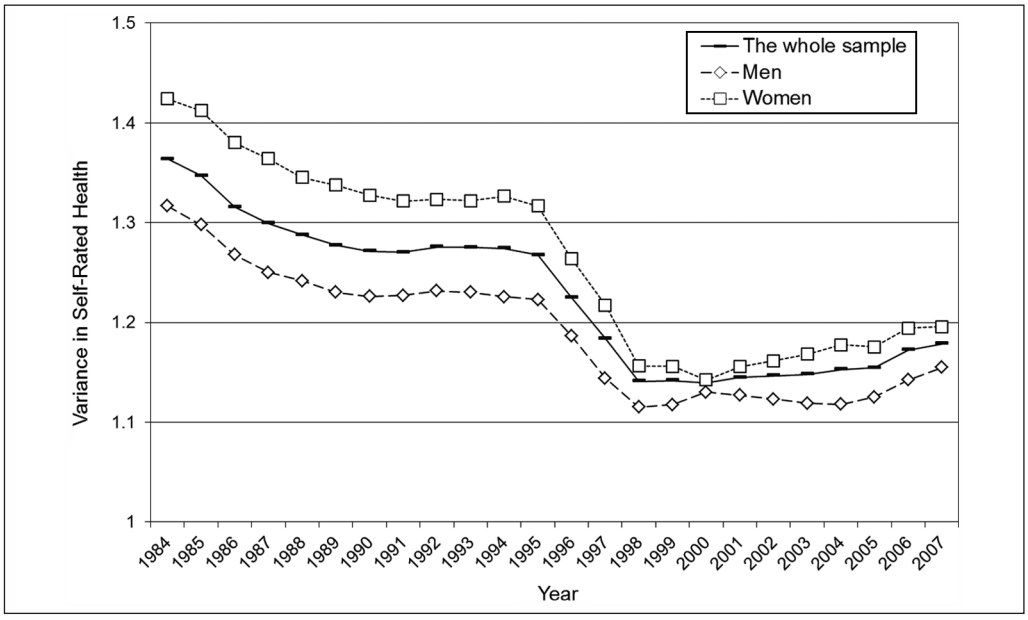

Figure 2. Observed Variances in Self-Rated Health, NHIS, 1984 to 2007
Note: Trends are adjusted for sample weights and smoothed by a three-point moving average.

point increase in the expected value of self-rated health. Being employed has an even larger impact: a statistically significant .39 point increase in the expected value of self-rated health. Finally, every $10,000 increase in household income is associated with a statistically significant .07 point increase in the expected value of self-rated health. Effects of age are curvilinear (quadratic) in that self-rated health declines with age and then begins to increase in late life around age 68. The 1995 NHIS sample redesign significantly decreased the expected value of self-rated health by about .06 points for women and .08 points for men. Absent the sample redesign, we would see an even larger increase in self-rated health in the late-1990s for men and women than that shown in Figure 1. Estimates of residual variance components at level 2 indicate significant period and cohort effects net of effects of individual-level covariates, while the period effect is larger than the cohort effect, as reported in the Variance Components section of Table 2.

The top graph in Figure 3 clearly portrays this quadratic age-dependence of the conditional mean of self-rated health. Figure 3 also contains graphs of annual and smoothed estimates of cohort and period effects on mean self-rated health from the HAPC part of the integrated model. These show that late baby boomers born between1955 and 1964 generally have better self-rated health than do earlier or later birth cohorts. An exception is the 1899 to 1904 cohort, whose relatively large positive effect may be due to a selective survival effect as well as the small number of respondents from this early cohort in the NHIS data. In addition, before 1998, period-to-period changes in self-rated health exhibit a very slight increase accompanied by cycles up and down, with a significant decline after 1998. Comparing graphs of the estimated cohort and period effects in Figure 3 to the overall trends in Figure 1 and the number of significant β coefficients and the size of residual variance components by cohort and period in Table 2, it is clear that periods explain

Table 2. Estimated HAPC-VFR/HR Models of Self-Rated Health, NHIS, 1984 to 2007

	β		λ	
	coefficient	*se*	coefficient	*se*
Fixed Effects				
Intercept	3.281***	.009	.403***	.022
Age	−.142***	.002	.071***	.005
Age^2	.034***	.001	−.041***	.001
Male	.030***	.003	−.009*	.005
White	.165***	.003	−.080***	.004
Married	.022***	.003	−.024***	.004
Education	.060***	.000	−.026***	.001
Employed	.388***	.003	−.338***	.004
Income/10,000	.065***	.001	−.042***	.001
Redesign	−.061***	.009	−.068***	.008
Redesign*Male	−.021***	.005	.016*	.007
Random Effects				
Cohort				
1899	.027*	.012	.201***	.033
1905	−.005	.010	.124***	.029
1910	−.008	.009	.030	.027
1915	.002	.008	−.003	.025
1920	−.005	.007	−.021	.024
1925	−.020**	.007	−.029	.023
1930	−.009	.007	−.024	.022
1935	−.001	.006	−.029	.022
1940	.003	.006	−.041	.022
1945	−.009	.006	−.066**	.021
1950	.007	.005	−.088***	.022
1955	.018***	.005	−.103***	.022
1960	.023***	.006	−.097***	.023
1965	.003	.006	−.066**	.024
1970	−.006	.007	−.039	.025
1975	−.026***	.008	.044	.027
1980	−.007	.009	.077**	.029
1985	.012	.011	.129***	.033
Period				
1984	−.011	.007	.018**	.006
1985	−.006	.007	.005	.006
1986	.006	.008	−.005	.006
1987	−.007	.007	−.009	.006
1988	−.013	.007	.006	.006
1989	.008	.007	−.008	.006
1990	.020**	.007	.000	.006
1991	.017*	.007	−.006	.006
1992	.002	.007	−.008	.006
1993	−.005	.007	.006	.006
1994	.004	.007	.001	.006
1995	−.012	.007	.000	.006
1996	−.004	.008	.001	.007
1997	.025***	.007	−.004	.006
1998	.025**	.008	−.005	.006

(continued)

Table 2. *(continued)*

	β		λ	
	coefficient	*se*	coefficient	*se*
1999	.025**	.008	.000	.006
2000	.015	.008	.003	.006
2001	.007	.008	.002	.006
2002	.003	.008	.001	.006
2003	.002	.008	.000	.006
2004	−.023**	.008	−.005	.006
2005	−.021**	.008	−.004	.006
2006	−.011	.008	−.001	.007
2007	−.046***	.008	.011	.007
Variance Components	variance	*se*	variance	*se*
Cohort	.0002*	.000	.008**	.003
Period	.0003**	.000	.000	.000
Model Fit				
BIC	1941250			
−2 Res Log Pseudo-Likelihood			2351732	

*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests).

modestly more of the overall trend in self-rated health from 1984 to 2007 than do cohorts.

As a key output of the VFR/HR part of the integrated model, the λ columns in Table 2 show how individual-level covariates affect within-group health disparities. Estimated within-group health disparities for males, whites, married persons, the more highly educated, employed individuals, and persons with more income are smaller than disparities among their counterparts, that is, females, blacks, unmarried persons, the less educated, unemployed individuals, and those with less income. For example, the predicted log of within-group variance in male's self-rated health is .009 units smaller than that for females. Regarding race, the predicted log of within-white variance in self-rated health is .08 units smaller than the within–non-white variance in self-rated health. Being married significantly decreases the predicted log of within-group variance in self-rated health by about .02 units. Regarding education, each year of additional education is associated with a statistically significant .03 units decrease in the predicted log of within-group variance in self-rated health. Being employed has an even larger impact: a statistically significant .34 units decrease in the predicted log of within-group variance in self-rated health. Finally, every $10,000 increase in household income is associated with a statistically significant .04 units decrease in the predicted log of within-group variance in self-rated health. The 1995 sample redesign significantly decreased the predicted log of variance in self-rated health by about .07 units for women and .05 units for men. Without the sample redesign, we would see a smaller decline in the variance of self-rated health in the years after 1994 than that shown in Figure 2.

In addition, the integrated HAPC-VFR/HR model yields estimates of expected or predicted variations in health disparities across age, period, and cohort (or within-age, within-period, and within-cohort health disparities). Estimates of residual variance components at level 2 indicate significant cohort and non-significant period effects net of effects of

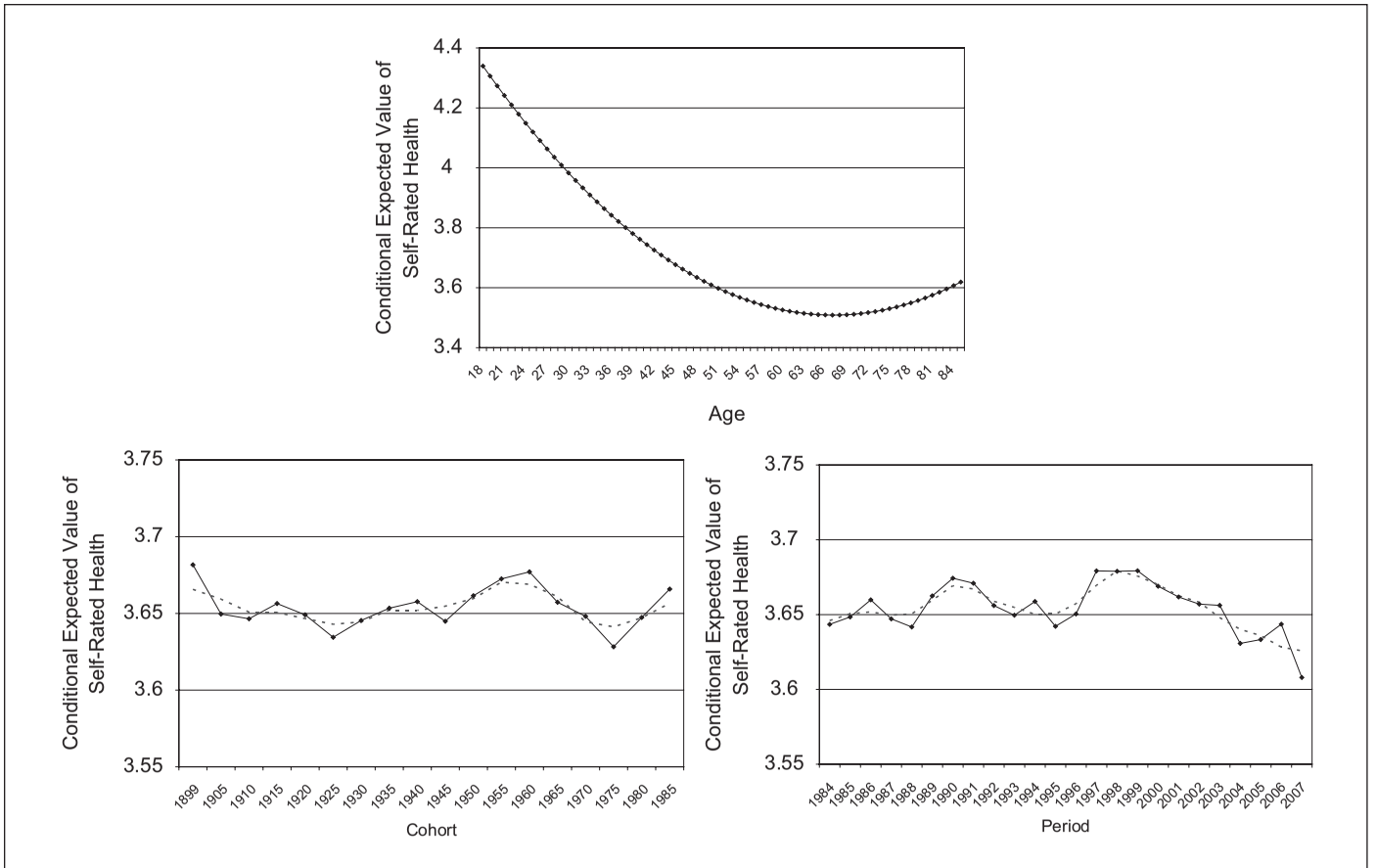

Figure 3. Variations in Conditional Expected Values of Self-Rated Health across Age, Cohort, and Period
Note: Dotted lines in bottom two figures indicate trends smoothed by a three-point moving average.

individual-level covariates as reported in the Variance Components section of Table 2. Figure 4 presents graphs of these estimated effects. After controlling for demographic and socioeconomic statuses, estimated health disparities during young adulthood are relatively small, indicating that almost everyone is relatively healthy. Health disparities increase with age, reaching a peak around age 55 (see the top figure in Figure 4), after which a decline sets in.

Figure 4 also shows that within-cohort health disparities decreased from the 1899–1904 cohort to the 1925–1929 cohort, leveled off in cohorts born during the Great Depression and World War II, and then decreased in baby boomer cohorts followed by substantial increases in post-baby boomer cohorts (this increasing trend is most pronounced in recent cohorts). After controlling for individual-level covariates and age and cohort effects, estimates of within-period health disparities graphed in Figure 4 are very flat between 1984 and 2007. When compared with Figure 2, it appears that cohort effects contribute to fluctuations of crude variance in self-rated health over time. For example, the recent increase in health disparities around 2000 in Figure 2 corresponds to increasing proportions of post-baby boomer cohorts (born after 1964) in the population—that is, cohorts with larger within-cohort health disparities than the preceding cohorts (see Figure 4). The number of significant λ coefficients by cohort and period in Table 2 further confirms this argument. The statistically insignificant variance component for the period effects in Table 2 also implies that variance in self-rated health does not significantly vary across periods.

Variations in Gender-Specific Health and Health Disparities by Age, Period, and Cohort, 1984 to 2007

Table 3 presents estimates of gender-specific models of self-rated health and health disparities. We see that being white, married, more educated, employed, and having more income are associated with better self-rated health for men and women, except married men's self-rated health is not significantly better than that of unmarried men (see β columns in Table 3).

Figure 5 contains graphs of the dependence of self-rated health on age, time period, and birth cohort as estimated in Table 3's models. Age-dependence curves in Figure 5 show that men's and women's self-rated health declines with increasing age, but trends reverse around age 69 for men and age 72 for women. Men report better self-rated health than do women at all ages. The gender gap in health is largest in the early adult years, narrows until around age 61, and widens afterward.

Figure 5 also contains graphs of the estimated period and cohort effects on the conditional expected values of self-rated health status. Estimated cohort effects are relatively flat across cohorts for men, except that baby boomers born between 1950 and 1959 have significantly better self-rated health than do other cohorts. By comparison, conditional expected values of self-rated health changed dramatically across cohorts for women: these values declined from the 1899–1904 cohort to the early baby boomers (i.e., the 1945–1954 cohort) and then rose for middle and late baby boomers and afterward. This results in a widened and then narrowed self-rated health gap between men and women. Estimated period effects on self-rated health for men and women in Figure 5 show similar trends as the observed means of self-rated health seen in Figure 1. In general, estimated β coefficients in Table 3 suggest period effects contribute slightly more than cohort effects to changes in self-rated health for men and women from 1984 to 2007.

Estimated within-group health disparities in Table 3 (in the λ columns) for individuals who are white, married, more highly educated, employed, and have more income are smaller than disparities among their counterparts, for men and women, except that health disparities are not significantly smaller for married women than for unmarried women.

Figure 6 contains graphs of the estimated within-group health disparities by age, time

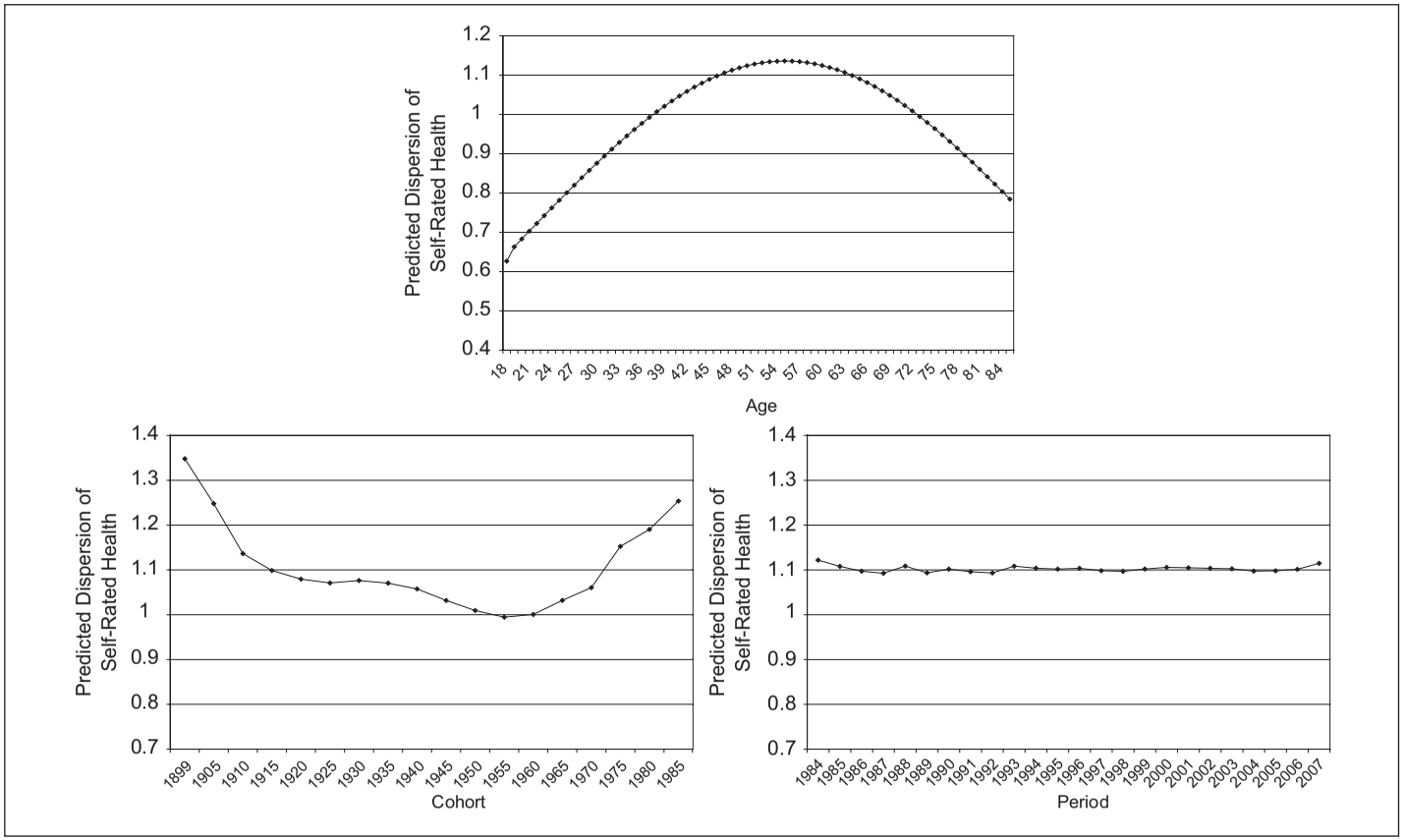

Figure 4. Variations in Predicted Dispersion of Self-Rated Health across Age, Cohort, and Period

Table 3. Estimated HAPC-VFR/HR Models of Self-Rated Health by Gender, NHIS, 1984 to 2007

	β				λ			
	Men	*se*	Women	*se*	Men	*se*	Women	*se*
Fixed Effects								
Intercept	3.318***	.012	3.285***	.016	.394***	.025	.375***	.019
Age	−.162***	.002	−.120***	.004	.082***	.006	.059***	.005
Age^2	.037***	.001	.025***	.001	−.044***	.001	−.032***	.001
White	.115***	.004	.206***	.005	−.078***	.006	−.078***	.006
Married	.004	.004	.028***	.004	−.016**	.005	−.004	.006
Education	.059***	.001	.064***	.001	−.029***	.001	−.023***	.001
Employed	.464***	.005	.316***	.005	−.373***	.006	−.302***	.006
Income/10,000	.061***	.001	.070***	.001	−.045***	.001	−.037***	.001
Redesign	−.076***	.010	−.071***	.011	−.047***	.009	−.076***	.008
Random Effects								
Cohort								
1899	.021	.017	.069**	.025	.203***	.041	.123***	.031
1905	−.015	.015	.032	.021	.131***	.035	.058*	.027
1910	−.035**	.013	.052**	.019	.019	.031	−.010	.024
1915	.006	.011	.031	.017	−.004	.028	−.033	.022
1920	.002	.010	.015	.015	−.002	.026	−.063**	.020
1925	−.020*	.009	−.001	.014	−.001	.025	−.067***	.019
1930	−.008	.009	.000	.014	.014	.024	−.069***	.019
1935	.010	.008	−.027*	.013	−.009	.023	−.028	.018
1940	.021**	.008	−.052***	.013	−.037	.023	−.002	.018
1945	.006	.007	−.072***	.012	−.063**	.023	−.019	.017
1950	.022**	.007	−.060***	.012	−.097***	.023	−.020	.017
1955	.025***	.007	−.037**	.012	−.120***	.023	−.030	.018
1960	.014	.007	−.002	.013	−.111***	.024	−.037*	.018
1965	−.007	.008	−.007	.014	−.083**	.026	−.024	.019
1970	−.010	.009	−.006	.015	−.043	.028	−.019	.021
1975	−.037***	.011	−.004	.017	.045	.030	.046*	.023
1980	−.001	.012	.014	.019	.051	.033	.081**	.026
1985	.007	.015	.055*	.023	.105**	.040	.112***	.032
Period								
1984	−.019*	.008	.002	.011	.019*	.008	.009	.007
1985	−.019*	.008	.021	.011	.008	.008	.001	.007
1986	−.003	.009	.022	.012	−.002	.008	−.005	.008
1987	−.014	.008	.005	.010	−.005	.007	−.009	.007
1988	−.021**	.008	.003	.010	.015*	.007	−.006	.007
1989	.003	.008	.016	.010	−.005	.007	−.009	.007
1990	.014	.008	.029**	.010	−.005	.007	.006	.007
1991	.018*	.008	.014	.010	−.014	.007	.004	.007
1992	.009	.008	−.013	.010	−.007	.007	−.007	.007
1993	.009	.008	−.028**	.010	−.001	.007	.010	.007
1994	.015	.008	−.016	.010	.004	.007	−.001	.007
1995	.001	.008	−.035**	.011	−.008	.008	.005	.007
1996	.005	.010	−.020	.012	.000	.008	.001	.008
1997	.024*	.010	.028**	.010	−.002	.008	−.004	.007
1998	.021*	.010	.028**	.010	−.005	.009	−.003	.007
1999	.030**	.010	.023*	.010	.001	.009	−.002	.007

(continued)

Table 3. *(continued)*

	β				λ			
	Men	*se*	Women	*se*	Men	*se*	Women	*se*
2000	.016	.010	.015	.010	.002	.009	.002	.007
2001	.008	.010	.008	.010	.001	.009	.001	.007
2002	.005	.010	.001	.010	.001	.009	.000	.007
2003	.009	.010	−.006	.010	−.003	.009	.002	.007
2004	−.027**	.010	−.021*	.010	−.002	.009	−.004	.007
2005	−.017	.010	−.026*	.010	−.008	.009	.002	.007
2006	−.014	.011	−.011	.011	.000	.009	−.001	.007
2007	−.054***	.010	−.041*	.011	.014	.009	.007	.007
Variance Components								
Cohort	.0004*	.0002	.002*	.0008	.008*	.0032	.004*	.0015
Period	.0005**	.0002	.001**	.0002	.0001	.0001	.0001	.0001
Model Fit								
BIC	1097607		840963					
−2 Res Log Pseudo-Likelihood					1369748		982421	

*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests).

period, and birth cohort from Table 3. Estimated within-age health disparities have a bell shape that peaks around age 56 for men and women. Compared to men, within-age health distributions for women are slightly more spread out. Women's dispersions are larger than men's at all ages; they increase at a slower rate before age 56 and decrease at a slower rate after age 56.

Similar to results of age variations, within-cohort and within-period heterogeneities are larger for women than for men. Figure 6 shows that for women, within-cohort variances decreased from the 1899–1904 cohort to the 1930–1934 cohort, increased in cohorts born in the late stages of the Great Depression and World War II, then decreased in baby boomer cohorts and increased in recent cohorts. For men, within-cohort variances fluctuate more across cohorts: they decreased from the 1899–1904 cohort to the 1915–1919 cohort and then were relatively flat until they substantially declined again for cohorts born in the Great Depression and World War II and baby boomer cohorts born between 1945 and 1959; this was followed by substantial increases afterward, especially in more recent cohorts. After controlling for individual-level covariates and age and cohort effects, graphs in Figure 6 show that within-period variances are relatively flat from 1984 to 2007 for men and women. The lack of significant random-effects coefficients for period, and the statistically insignificant variance components of period effects for men and women in Table 3, also suggest that variance in self-rated health does not significantly vary across periods. These gender-specific analyses further support the inference that cohort effects contribute more than period effects to changes in health disparities from 1984 to 2007, for both men and women (see Figure 2). The striking gender difference in changes in within-cohort health disparities merits further research.

Robustness Analyses and Extensions of the HAPC-VFR/HR Model

We noted earlier that our application of the linear mixed-effects specification of the HAPC-VFR/HR model is based on its ease of interpretation. But this model is not fully

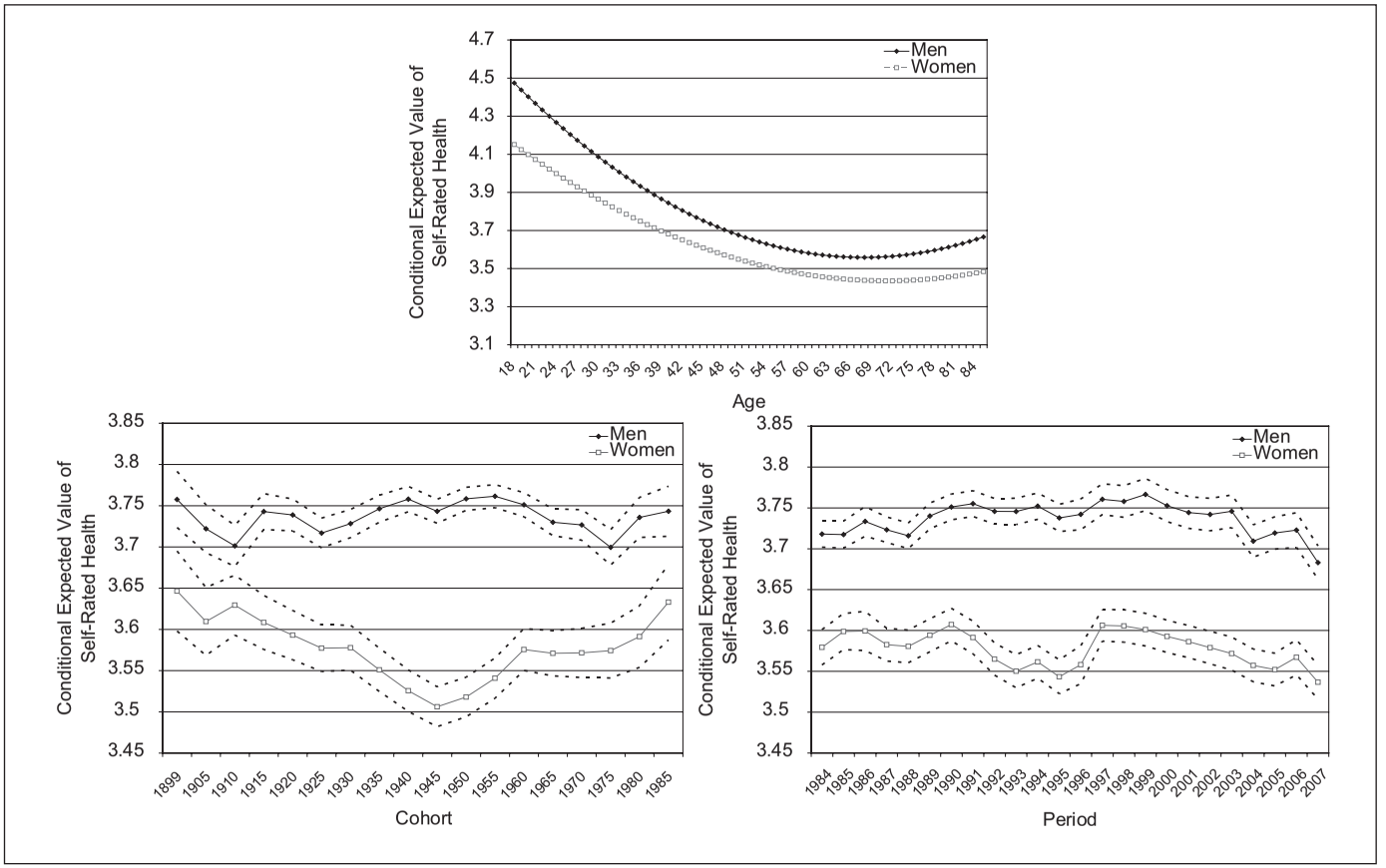

Figure 5. Variations in Conditional Expected Values of Gender-Specific Self-Rated Health across Age, Cohort, and Period, with 95% Confidence Intervals

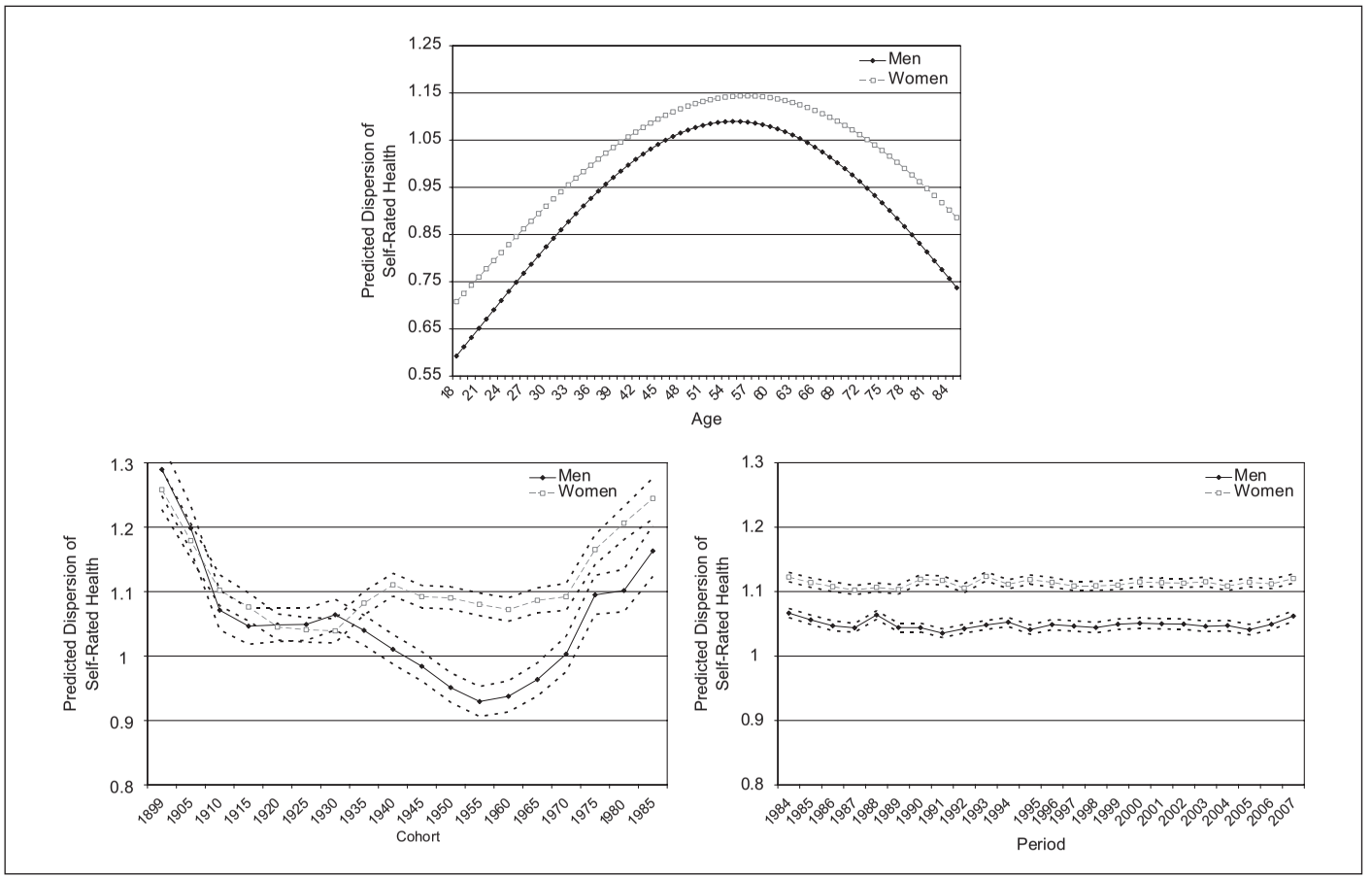

Figure 6. Variations in Predicted Dispersion of Gender-Specific Self-Rated Health across Age, Cohort, and Period, with 95% Confidence Intervals

sensitive to either the ordered categorical nature of survey responses to the self-rated question or to the non-normal frequency distribution of the five-point scaling of responses. Accordingly, we describe several analyses to assess the robustness of our empirical findings and some methodological extensions necessary to adapt the model to the self-reported health outcome variable.

Alternative coding of the response variable. As a first robustness analysis, we considered alternative approaches to the extraction of a measurement signal from the ordered response categories of the self-rated health question. One possibility is to dichotomize responses into, for example, fair or poor versus excellent, very good, or good (see, e.g., Lynch 2003). We conducted HAPC-VFR/HR analyses—in linear probability and logit regression specifications—of a dichotomized self-rated health outcome for NHIS data. Findings regarding effects of individual-level covariates and the period, and cohort random effects on the mean of self-rated health, are consistent with those reported here for the five-point scale and are available from the authors on request.[11] The second step analysis on the variance of self-rated health, however, is not straightforward, because the mean and variance are functionally related for a binary dependent variable, that is, variance = mean times (1-mean). Therefore, the VFR/HR model must be adjusted before it can be applied to a binary variable. Some scholars suggest adding an overdispersion parameter Ψ to capture the extra-Bernoulli variation (i.e., variance = Ψ times mean times [1-mean]) and writing the overdispersion parameter Ψ as a function of covariates in its own regression (Western and Bloome 2009).[12]

Replication of empirical findings using a different dataset. As a second assessment of the robustness of findings from the NHIS data, we analyzed General Social Survey (GSS) data from 1984 to 2008. The GSS is relatively consistent in question format and sampling design. It includes a self-rated health question (with four ordered response categories) in most surveys, except 1986. The downside of the GSS is its relatively small sample size of around 2,000 respondents every year. For the refined HAPC-VFR/HR analyses described here, this produces estimates with relatively large stochastic variability over time. Taking this into account, trends in observed means of self-reported health status and variances therein generally are similar to those reported for the NHIS data. Empirical findings from an application of the linear mixed-effects specification of the HAPC-VFR/HR model are similar as well and are available from the authors on request.

Other model specifications. In brief, the empirical findings reported earlier survive these two robustness analyses. Beyond this, the linear mixed-effects specification of the HAPC-VFR/HR model must be generalized to other forms that take into account the ordered categorical nature of the self-rated outcome variable. This requires specification of a corresponding generalized linear model with a random effects version of the model (Fox 2008). Specification and estimation of such a version of the HAPC-VFR/HR model is not only more difficult to interpret, but it is also methodologically challenging for the second-step regression.[13]

Initial work with ordered logit specifications of these generalized models, however, produces findings consistent with analyses from the linear mixed-effects specification of the HAPC-VFR/HR model. This is also due to the very large pooled NHIS samples analyzed (more than 700,000 sample observations in total; see Table 1). These very large sample sizes bring the asymptotic consistency and normality distributional properties of the maximum likelihood-based estimator for non-normally distributed data into play very strongly.

DISCUSSION AND CONCLUSIONS

The study of social inequality is one of the defining problems of sociology, and social

change in inequality can be represented in temporal variations of inequality across time period and birth cohort. Prior research treats these sources of variations separately. In any given dataset, however, they are potentially confounded with each other as well as with an individual's age. By integrating a variance function regression analysis, which facilitates separation of within- from between-group inequality, with an HAPC model, which facilitates estimation of age, period, and cohort effects in a nonlinear fashion, we achieve a Hierarchical-Age-Period-Cohort-Variance-Function-Regression Model (HAPC-VFR/HR).

We demonstrated the utility of this integration with an analysis of self-reported health disparities in the United States from 1984 to 2007. The core idea of the HAPC-VFR/HR model is to estimate mixed (fixed and random) effects regression specifications of the two equations in the variance function regression model as a function of age, period, and cohort and other individual-level covariates, treating cohort and period as random effects. Thus, the first mixed-effects regression describes how age, period, and cohort affect the mean of self-rated health net of a set of individual-level covariates (i.e., gender, race, marital status, work status, education, and income); the second mixed-effects regression explains how within-group health disparities change across age, period, and cohort. We obtained results using the maximum likelihood estimation method.

We find evidence of all three sets of effects on the conditional mean of self-rated health: (1) self-rated health decreases with age to the late 60s and then increases in late life due to selective mortality (see, e.g., Kulminski et al. 2008); (2) late baby boomers (born 1955 to 1964) report better health than do other cohorts; and (3) self-rated health has significantly declined since the late 1990s. Net of age effects, however, period effects appear to contribute relatively more than cohort effects to changes in the conditional mean of self-rated health in the past two decades.

In terms of variations in within-group health dispersions across age, period, and cohort, we find strong age effects: health dispersion increases with age, peaks at age 55, and diminishes afterward, consistent with prior research that finds convergence in SES health disparities in late life. This convergence may be due to mortality selection, that is, sicker people, who are disproportionately of lower SES, are more likely to die earlier or be disabled, which removes them from the survey (Dupre 2007). But this convergence may also result from universal biological frailty (Yang and Kozloski 2011), diminished socioeconomic differences in exposure to risk factors, postponement of morbidity and functional limitation for higher SES people, or equalization of health care usage and protections through Medicare coverage at age 65 (House, Lantz, and Herd 2005; House et al. 1994).

We also find cohort and period effects. Within-cohort health disparities generally decreased from the 1899–1904 cohort to the baby boomer cohorts and substantially increased for post-baby boomer cohorts. After taking into account age and cohort effects, within-period health disparities are flat since 1984. In contrast to what we find for the conditional mean of self-rated health, cohort effects appear to contribute much more than period effects to changes in the variance of self-rated health over the past two decades. Because post-baby boomer cohorts (especially cohorts born after 1980) have much larger within-cohort health disparities than do preceding cohorts, and within-age health disparities increase with age until around age 55, one can expect that health disparities in the general population will further increase in the next one or even two decades as these cohorts age and replace preceding cohorts.

Several mechanisms may be contributing to the enlarged health disparities for post-baby boomer cohorts. First, income inequality increased dramatically in the past three decades in the United States, strengthening the protective effects of advantageous social status (e.g., family income, employment, college education, and marriage) on health (Zheng and George 2009). This may contribute

to the enlarged health disparities in recent cohorts. Second, since 1980, young adult cohorts have had an increased prevalence of immigrants (documented and undocumented). This may be associated with increased conditional means of self-rated health in recent cohorts (due to "healthy migrant" effects; Singh and Siahpush 2001) and with our finding of increased cohort variances in these conditional means (due to cohorts' increasing heterogeneity). Third, compared to "healthy migrant" effects, which contribute to the upper tail of conditional dispersions (i.e., higher residual self-reported health), the dramatically increased obesity risk in cohorts born after 1970 (Reither et al. 2009) may contribute to the lower tail (i.e., lower residual self-reported health). The presence of either or both of these effects would produce the increases in the conditional variance measures (the lambdas) for the post-1975 birth cohorts that we reported in Tables 2 and 3. Substantial SES, gender, and race differences in the risk for obesity further increase recent cohorts' conditional health dispersions. Fourth, relatively young cohorts have grown up in the information age of the Internet. Information on the Web is generally available to everyone, but the past two decades have seen a digital divide in access to and use of the Web (Norris 2001); this may be related to the increasing recent cohort variances in health knowledge and, consequently, health outcomes (Brodie et al. 2000). Future research could readily test these postulated explanations using the HAPC-VFR/HR model.

This article also examines age, period, and cohort variations in self-rated health and within-group health disparities by gender. Previous research suggests a significant self-reported health gap between men and women in early adulthood (MacIntyre, Hunt, and Sweeting 1996), a gap that narrows and even disappears in old age (Arber and Cooper 1999; Case and Deaton 2003). Our results suggest a narrowing of the self-reported health gap until around age 61 and then a widening afterward. This is consistent with a process by which men are more likely than women to experience more severe forms of some chronic conditions (e.g., cardiovascular disease and certain lung disorders) and are more likely to die from these chronic conditions at earlier ages (Case and Paxson 2005). This first reduces the self-reported health gap and then widens it later due to selective survival, leaving relatively stronger men but a relatively larger share of women with poorer health alive at older ages.

Compared to the relatively flat curve of the conditional mean of self-rated health across cohorts for men, the cohort trend for women follows a V-shape, that is, the conditional mean of self-rated health declines from earlier cohorts to the early baby boomers cohort and then rises afterward. This results in an enlarged and then shrunken health gap between men and women, which may contribute to the narrowed gender health gap in recent periods found in other studies (e.g., Schnittker 2007). This also raises a puzzle in our study: What causes distinct cohort patterns in gender-specific health? Were female infants more negatively affected than male infants by WWI, the Great Depression, and WWII? Or, were female infants more likely to survive hardship but were in bad health, while male infants were more likely to die from these negative social events, leaving relatively healthier male infants? Further research is needed to solve this puzzle. By comparison, the gender-specific period trend in health is consistent with the trend for the whole population, that is, both men and women increasingly report worse health after the late 1990s.

In terms of variations in within-age health disparities, men are more homogeneous than women in self-rated health for all ages, and especially for later ages. Men tend to report better health than women in early adulthood (MacIntyre et al. 1996), which makes their within-age health disparities smaller at that life stage. However, men are also more likely than women to get severe chronic diseases in middle age (Case and Paxson 2005), which possibly results in smaller differences in within-age health disparities between men

and women in middle adulthood. But men are also more likely to die from these chronic diseases (Case and Paxson 2005), thus leaving relatively healthier men at later ages, leading to a faster drop in within-age health disparities than seen in women in late life.

Men also have smaller within-cohort heterogeneity than do women for all cohorts, especially for cohorts born in the late stage of the Great Depression, WWII, and baby boomers. This raises a puzzle similar to the distinct cohort patterns in gender-specific health: What causes distinct cohort patterns in gender-specific health disparities? Men and women appear to have had significantly different experiences and endured different effects from several historical events in the mid-twentieth century. These puzzles point to a potentially important future project. In terms of within-period health disparities, they are relatively flat for men and women and for the whole population. These gender-specific analyses further support the idea that cohort effects contribute more than period effects to changes in health disparities from 1984 to 2007.

By using the HAPC-VFR/HR model, we have provided a more complete picture of the evolution of changes in self-rated health and health disparities in the United States from 1984 to 2007. Specifically, this study has made two significant contributions to the health disparities literature. First, health disparities across age, time period, and birth cohort are intertwined but have not been systematically disentangled in existent studies due to lack of an integrated model. Using the HAPC-VFR/HR model, this study demonstrates that changes in self-rated health disparities in the past two decades are much more of a cohort than a period story. Future research in this area should thus pay more attention to the cohort perspective. Second, prior literature has focused on changes in health inequality defined by specific aspects of social stratification such as gender, race, and SES across age, period, or cohort, without an overall picture of changes in the general dispersion of health across these three dimensions. The HAPC-VFR/HR model offers an analytic tool to capture the general dispersion of health across these three dimensions. This provides the basis for future research to examine the contribution to the general dispersion made by each aspect of the social stratification system.

Inequality or disparity in statuses occurs in many domains of social life, such as income, wealth, education, and heath care access, to name but a few. The HAPC-VFR/HR model provides a powerful framework through which to identify and study the evolution of variations and social inequalities in these outcomes across age, period, and cohort temporal dimensions. Accordingly, this model should be broadly applicable to the study of social inequality in many different substantive contexts.

Acknowledgments

We thank Bruce Western, members of the DuPRI Seminar, and the anonymous reviewers for helpful comments on a previous version of the article.

Notes

1. To estimate HAPC models, Yang and Land (2006) apply the statistical methodology of mixed (fixed and random) effects or hierarchical regression models, treating effects of individual-level covariates as fixed and cohort and period effects as random. An alternative approach to model specification could be based on a purely fixed-cohort-and-period-effects regression model, which does not require large numbers of cohorts or periods for reliable statistical estimation and assumes that period and cohort effects are independent of individual-level regressors. Yang and Land (2008) examine these assumptions and compare the fixed- versus random-effects model specifications for APC analysis.
2. Case and Paxson (2005) found that men and women are equally likely to report poor health when they have the same condition. They also found that men are more likely to report better health but die earlier, because men experience more fatal diseases but women suffer more from chronic conditions. This implies that self-rated health reflects underlying objective health conditions without regard to gender and is a valid measurement to investigate gender-specific health disparities.
3. We cross-classify respondents in the repeated cross-section sample surveys by the time periods of the surveys in which they responded and the birth cohorts

to which they belong. Each cell is an intersection of a cohort and a period.

4. It is technically possible to group data from an age by period matrix with cohort year as the within-cell predictor. But as Yang (2007) notes, the age variable in APC analyses is associated with the biological process of aging *internal* to individuals. By contrast, period and cohort effects reflect influences of forces that are *external* to individuals and operate in different ways. We thus believe that the most substantively sensible specification is one that treats age as an individual or within-cell explanatory variable, with period and cohort treated as contextual or level-2 variables.
5. Following the recent work of Western and Bloome (2009) in bringing to sociologists' attention the heteroscedastic regression model, we use the term *variance function regression* throughout this article, although this model has been called *double generalized linear models* (DGLM) (Smyth 2002; Smyth, Huele, and Verbyla 2001), *double hierarchical generalized linear models* (DHGLM) (Lee and Nelder 2006), and *generalized additive models for location, scale, and shape* (GAMLSS) (Rigby and Stasinopoulos 2005). Model specifications and estimation algorithms developed in the DHGLM and GAMLSS approaches to heteroscedastic regression have a number of differences. Neither modeling approach has been applied to the estimation of heteroscedastic regression models for ordered logit models (see note 13).
6. We used the SAS PROC MIXED and PROC GLIMMIX procedures to estimate the first- and second-step regressions, respectively. The default estimator of GLIMMIX is RSPL, which maximizes the residual log pseudo-likelihood and provides unbiased predictors of the random effects. The pseudo-maximum likelihood estimator uses a consistent and asymptotically normal estimator rather than a maximum likelihood estimator for the variance parameters. In models for a normally distributed outcome variable with an identity link, RSPL is equivalent to REML (Littell et al. 2006), but RSPL is consistent and asymptotically normally distributed for non-normal data as well.
7. The outcome variable in our analyses, self-rated health, is not normally distributed; it is skewed to the left. Residuals calculated from the Step 1 regression have a symmetric distribution that has short tails compared to a normal distribution. Because of the very large sample size, estimates of coefficients from the Step 2 regression still have good statistical properties. This is due to the fact that for independently and identically distributed data, the RSPL method produces estimates of fixed and random effects of a mixed model that are consistent and asymptotically normally distributed; even the identically distributed assumption can be relaxed (Demidenko 2004).
8. We do not use waves of NHIS data prior to 1984 for two reasons: (1) the self-rated health question contains four response categories (poor, fair, good, excellent) instead of the five categories in the analyses reported here, and (2) a sampling design change after 1984 further complicates the data.
9. The ordered logit regression model posits an underlying latent continuous variable corresponding to an ordered categorical response variable along which sample responses can be arrayed (Fox 2008). The continuous variable is conceived of as dissected into m regions by $m - 1$ thresholds or boundaries of varying width. The ordered logit model permits estimation of thresholds that can be used to assess whether the equal-intervals assumption for the categorical variable is violated. Examining the NHIS self-reported health variable, we found that the variability of distances from threshold to threshold is about 5 percent of the width of the estimated distances. This is not sufficient to produce empirical findings that differ substantively from the analyses described here that are based on the equal-intervals specification.
10. It merits emphasizing here that expected effects are conditional on all other covariates in the model. For the sake of space limits, this is not reiterated in the text.
11. Manor, Matthews, and Power (2000) similarly find that a dichotomous coding of self-rated health responses produces similar results to those obtained with alternative statistical methods that accommodate the ordered nature of self-rated health (e.g., polytomous regression, cumulative odds, continuation ratio, and adjacent categories models).
12. An anonymous reviewer also suggested this adjustment.
13. Neither of the two approaches to statistical estimation of VFR/HR models for generalized linear mixed models with non-interval outcome variables (see note 5) has been applied to the specification and estimation of ordered logit models for ordinal response variables. Thus, this application requires a careful methodological development. The second step of estimation of VFR/HR ordered logit models is complicated by the fact that the conventional estimation of such models fixes the error variance to 1 to set the scale of the latent variable. To estimate the second step of a VFR/HR, there must be an error term with an unconstrained variance at this step. One possible approach is to add overdispersion parameters to capture extra-Bernoulli variance in responses; see the discussion in the text around note 12. But other possible specifications exist as well. A separate article will develop and present alternative specifications of the HAPC-VFR/HR model for ordinal response variables complete with software code for model estimation.

References

Aitkin, Murray. 1987. "Modelling Variance Heterogeneity in Normal Regression Using *GLIM*." *Applied Statistics* 36:332–39.

Alwin, Duane F. 2009. "History, Cohorts, and Patterns of Cognitive Aging." Pp. 9–38 in *Aging and Cognition: Research Methodologies and Empirical Advances,* edited by H. B. Bosworth and C. Hertzog. Washington, DC: American Psychological Association.

Arber, Sara and Helen Cooper. 1999. "Gender Differences in Health in Later Life: The New Paradox?" *Social Science and Medicine* 48:61–76.

Bird, Chloe, Peter Conrad, Allen Fremont, and Stefan Timmermans. 2010. *Handbook of Medical Sociology,* 6th ed. Nashville, TN: Vanderbilt University Press.

Blau, Peter M. and Otis Dudley Duncan. 1966. *The American Occupational Structure*. New York: Wiley.

Bosworth, Hayden B., Ilene C. Siegler, Beverly H. Brummett, John C. Barefoot, Redford B. Williams, Nancy E. Clapp-Channing, and Daniel B. Mark. 1999. "The Association between Self-Rated Health and Mortality in a Well-Characterized Sample of Coronary Artery Disease Patients." *Medical Care* 37:1226–36.

Brodie, Mollyann, Rebecca E. Flournoy, Drew E. Altman, Robert J. Blendon, John M. Benson, and Marcus D. Rosenbaum. 2000. "Health Information, the Internet, and the Digital Divide." *Health Affairs* 19:255–65.

Case, Anne and Angus Deaton. 2003. "Broken Down by Work and Sex: How Our Health Declines." Working Paper 9821, National Bureau of Economic Research, Cambridge, MA.

Case, Anne and Christina Paxson. 2005. "Sex Differences in Morbidity and Mortality." *Demography* 42:189–214.

Chen, Feinian, Yang Yang, and Guangya Liu. 2010. "Social Change and Socioeconomic Disparity in Health over the Life Course in China." *American Sociological Review* 75:126–50.

Crimmins, Eileen M. and Yasuhiko Saito. 2001. "Trends in Healthy Life Expectancy in the United States, 1970–1990: Gender, Racial, and Educational Differences." *Social Science and Medicine* 52:1629–41.

Dannefer, Dale. 2003. "Cumulative Advantage/Disadvantage and the Life Course: Cross-Fertilizing Age and Social Science Theory." *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 58:S327–S337.

Demidenko, Eugene. 2004. *Mixed Models: Theory and Applications*. Hoboken, NJ: Wiley.

Dupre, Matthew E. 2007. "Educational Differences in Age-Related Patterns of Disease: Reconsidering the Cumulative Disadvantage and Age-As-Leveler Hypotheses." *Journal of Health and Social Behavior* 48:1–15.

Feldman, Jacob J., Diane M. Makuc, Joel C. Kleinman, and Joan Cornoni-Huntley. 1989. "National Trends in Educational Differences in Mortality." *American Journal of Epidemiology* 129:919–33.

Fox, John. 2008. *Applied Regression Analysis and Generalized Linear Models*. Los Angeles: Sage.

Goesling, Brian. 2007. "The Rising Significance of Education for Health?" *Social Forces* 85:1621–44.

Hays, Judith C., David Schoenfeld, Dan G. Blazer, and Deborah T. Gold. 1996. "Global Self-Ratings of Health and Mortality: Hazard in the North Carolina Piedmont." *Journal of Clinical Epidemiology* 49:969–79.

House, James S., Paula M. Lantz, and Pamela Herd. 2005. "Continuity and Change in the Social Stratification of Aging and Health over the Life Course: Evidence from a Nationally Representative Longitudinal Study from 1986 to 2001/2002 (Americans' Changing Lives Study)." *Journals of Gerontology: Series B 605B* (special issue II):15–26.

House, James S., James M. Lepkowski, Ann M. Kinney, Richard P. Mero, Ronald C. Kessler, and A. Regula Herzog. 1994. "The Social Stratification of Aging and Health." *Journal of Health and Social Behavior* 35:213–34.

Hummer, Robert A., Richard G. Rogers, and Isaac W. Eberstein. 1998. "Sociodemographic Approaches to Differentials in Adult Mortality: A Review of Analytic Approaches." *Population and Development Review* 24:553–78.

Idler, Ellen L. and Yael Benyamini. 1997. "Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies." *Journal of Health and Social Behavior* 38:21–37.

Jemal, Ahmedin, Elizabeth Ward, Robert N. Anderson, Taylor Murray, and Michael J. Thun. 2008. "Widening of Socioeconomic Inequalities in U.S. Death Rates, 1993–2001." *PLOS ONE* 3:e2181.

Kennedy, Byron S., Stanislav V. Kasl, and Viola Vaccarino. 2001. "Repeated Hospitalizations and Self-Rated Health among the Elderly: A Multivariate Failure Time Analysis." *American Journal of Epidemiology* 153:232–41.

Kulminski, Alexander M., Svetlana V. Ukraintseva, Irina V. Kulminskaya, Konstantin G. Arbeev, Kenneth C. Land, and Anatoli Y. Yashin. 2008. "Cumulative Deficits Better Characterize Susceptibility to Death in the Elderly than Phenotypic Frailty: Lessons from the Cardiovascular Health Study." *Journal of the American Geriatrics Society* 56:898–903.

Lauderdale, Diane S. 2001. "Education and Survival: Birth Cohort, Period, and Age Effects." *Demography* 38:551–61.

Lee, Youngjo and John A. Nelder. 2006. "Double Hierarchical Generalized Linear Models." *Journal of the Royal Statistical Society Series C* 55:139–85.

Littell, Ramon C., George A. Milliken, Walter W. Stroup, Russell D. Wolfinger, and Oliver Schabenberger. 2006. *SAS for Mixed Models,* 2nd ed. Cary, NC: SAS Institute.

Liu, Hui and Robert A. Hummer. 2008. "Are Educational Differences in U.S. Self-Rated Health Increasing? An Examination by Gender and Race." *Social Science & Medicine* 67:1898–1906.

Lynch, Scott M. 2003. "Cohort and Life-Course Patterns in the Relationship between Education and Health: A Hierarchical Approach." *Demography* 40:309–331.

MacIntyre, Sally, Kate Hunt, and Helen Sweeting. 1996. "Gender Differences in Health: Are Things Really as Simple as They Seem?" *Social Science and Medicine* 42:617–24.

Manor, Orly, Sharon Matthews, and Chris Power. 2000. "Dichotomous or Categorical Response? Analysing Self-Rated Health and Lifetime Social Class." *International Journal of Epidemiology* 29:149–57.

Mason, Karen O., William H. Mason, H. H. Winsborough, and W. Kenneth Poole. 1973. "Some Methodological Issues in Cohort Analysis of Archival Data." *American Sociological Review* 38:242–58.

Mason, William M. and Stephen E. Fienberg. 1985. *Cohort Analysis in Social Research: Beyond the Identification Problem.* New York: Springer-Verlag.

Meara, Ellen R., Seth Richards, and David M. Cutler. 2008. "The Gap Gets Bigger: Changes in Mortality and Life Expectancy, by Education, 1981–2000." *Health Affairs* 27:350–60.

Morris, Martina and Bruce Western. 1999. "Inequality in Earnings at the Close of the Twentieth Century." *Annual Review of Sociology* 25:623–57.

National Center for Health Statistics. 1994. *Excess Deaths and Other Mortality Measures for the Black Population: 1979–81 and 1991.* Hyattsville, MD: National Center for Health Statistics.

National Center for Health Statistics. 2009. *National Health Interview Survey: 1984–2007 Survey Years.* Hyattsville, MD: National Center for Health Statistics.

Nelder, John A. and Youngjo Lee. 1991. "Generalized Linear Models for the Analysis of Taguchi-Type Experiments." *Applied Stochastic Models and Data Analysis* 7:101–120.

Norris, Pippa. 2001. *Digital Divide? Civic Engagement, Information Poverty and the Internet Worldwide.* Cambridge: Cambridge University Press.

Pappas, Gregory, Susan Queen, Wilbur Hadden, and Gail Fisher. 1993. "The Increasing Disparity in Mortality between Socioeconomic Groups in the United States, 1960 and 1986." *New England Journal of Medicine* 329:103–115.

Preston, Samuel H. and Irma T. Elo. 1995. "Are Educational Differentials in Adult Mortality Increasing in the United States?" *Journal of Aging and Health* 7:476–96.

Raudenbush, Stephen W. and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Thousand Oaks, CA: Sage.

Reither, Eric, Robert Hauser, and Yang Yang. 2009. "Do Birth Cohorts Matter? Age-Period-Cohort Analyses of the Obesity Epidemic in the United States." *Social Science & Medicine* 69:1439–48.

Rigby, R. A. and D. M. Stasinopoulos. 2005. "Generalized Additive Models for Location, Scale, and Shape." *Journal of the Royal Statistical Society, Series C* 54:507–554.

Ross, Catherine E. and Chia-ling Wu. 1996. "Education, Age, and the Cumulative Advantage in Health." *Journal of Health and Social Behavior* 37:104–120.

Ryder, Norman B. 1965. "The Cohort as a Concept in the Study of Social Change." *American Sociological Review* 30:843–61.

Schnittker, Jason. 2007. "Working More and Feeling Better: Women's Health, Employment, and Family Life, 1974–2004." *American Sociological Review* 72:221–38.

Schoenfeld, David E., Lynda C. Malmrose, Dan G. Blazer, Deborah T. Gold, and Teresa E. Seeman. 1994. "Self-Rated Health and Mortality in the High-Functioning Elderly—A Closer Look at Healthy Individuals: MacArthur Field Study of Successful Aging." *Journal of Gerontology: Medical Sciences* 49:M109–115.

Schoeni, Robert F., Linda G. Martin, Patricia Andreski, and Vicki A. Freedman. 2005. "Persistent and Growing Socioeconomic Disparities in Disability among the Elderly: 1982–2002." *American Journal of Public Health* 95:2065–70.

Singh, Gopal K. and Mohammad Siahpush. 2001. "All-Cause and Cause-Specific Mortality of Immigrants and Native Born in the United States." *American Journal of Public Health* 91:392–99.

Smith, Herbert L. 2008. "Advances in Age-Period-Cohort Analysis." *Sociological Methods & Research* 36:287–96.

Smyth, Gordon K. 1989. "Generalized Linear Models with Varying Dispersion." *Journal of the Royal Statistical Society, Series B* 51:47–60.

Smyth, Gordon K. 2002. "An Efficient Algorithm for REML in Heteroscedastic Regression." *Journal of Graphical and Computational Statistics* 11:836–47.

Smyth, Gordon K., A. Frederik Huele, and Arunas P. Verbyla. 2001. "Exact and Approximate REML for Heteroscedastic Regression." *Journal of Graphical and Computational Statistics* 11:836–47.

Warren, John R. and Elaine M. Hernandez. 2007. "Did Socioeconomic Inequalities in Morbidity and Mortality Change in the United States over the Course of the Twentieth Century?" *Journal of Health and Social Behavior* 48:335–51.

Western, Bruce and Deidre Bloome. 2009. "Variance Function Regression for Studying Inequality." *Sociological Methodology* 39:293–325.

Western, Bruce, Deidre Bloome, and Christine Percheski. 2008. "Inequality among American Families with Children, 1975 to 2005." *American Sociological Review* 73:903–920.

Yang, Yang. 2006. "Bayesian Inference for Hierarchical Age-Period-Cohort Models of Repeated Cross-Section Survey Data." *Sociological Methodology* 36:39–74.

Yang, Yang. 2007. "Age/Period/Cohort Distinctions." Pp. 20–22 in *Encyclopedia of Health and Aging,* edited by K. S. Markides. Los Angeles, CA: Sage Publications.

Yang, Yang. 2008a. "Social Inequalities in Happiness in the United Stated, 1972 to 2004: An Age-Period-Cohort

Analysis." *American Sociological Review* 73:204–226.

Yang, Yang. 2008b. "Trends in U.S. Adult Chronic Disease Mortality: Age, Period, and Cohort Variations." *Demography* 45:387–416.

Yang, Yang. 2010. "Aging, Cohorts, and Methods." Pp. 17–30 in *The Handbook of Aging and the Social Sciences,* 7th ed., edited by B. Binstock and L. K. George. San Diego, CA: Academic Press.

Yang, Yang and Michael Kozloski. 2011. "Sex Differences in Age Trajectories of Physiological Dysregulation: Inflammation, Metabolic Syndrome, and Allostatic Load." *Journals of Gerontology: Biological Sciences* 66A:493–500.

Yang, Yang and Kenneth C. Land. 2006. "A Mixed Models Approach to the Age-Period-Cohort Analysis of Repeated Cross-Section Surveys, With an Application to Data on Trends in Verbal Test Scores." *Sociological Methodology* 36:75–98.

Yang, Yang and Kenneth C. Land. 2008. "Age-Period-Cohort Analysis of Repeated Cross-Section Surveys: Fixed or Random Effects?" *Sociological Methods and Research* 36:297–326.

Yang, Yang and Linda C. Lee. 2009. "Sex and Race Disparities in Health: Cohort Variations in Life Course Patterns." *Social Forces* 87:2093–2124.

Zheng, Hui and Linda K. George. 2009. "Rising U.S. Income Inequality and the Changing Gradient of Socioeconomic Status on Physical Functioning and Activity Limitations, 1984–2007." Paper presented at the American Sociological Association Annual Meeting, San Francisco, CA.

Hui Zheng is Assistant Professor of Sociology at The Ohio State University. His scholarship applies advanced statistical and demographic methods and interdisciplinary perspectives to study the causes and consequences of health, aging, and mortality. His current research focuses on four areas: the interplay of historical changes, life course, and population dynamics on trends in health disparities; individual trajectories, aggregate trends, and population heterogeneity in obesity, aging, vitality, and mortality; the interplay of social institution/structure and biological factors on health; and the implication of medicalization on population health.

Yang Yang is Associate Professor at the Department of Sociology and Lineberger Cancer Center, Faculty Fellow of the Carolina Population Center at the University of North Carolina at Chapel Hill. Her research interests include integrative sociodemographic and biological approaches to explaining social differentials in health and longevity, aging, and life course trajectories of health and co-morbidity, social behavioral links to cancer biology, and new models and methods for cohort analysis.

Kenneth C. Land is the John Franklin Crowell Professor of Sociology and Demography at Duke University. His research interests are in the development of mathematical and statistical models and methods for substantive applications in demography, criminology, and social indicators/quality-of-life studies. Current research projects include studies of heterogeneity in vitality decline and mortality, unemployment and crime rate fluctuations, structural covariates of crime rates, trends in child well-being, and new models for age, period, and cohort analysis.