Forthcoming, *American Sociological Review*

Aug 14, 2011

VARIANCE FUNCTION REGRESSION IN HIERARCHICAL AGE-PERIOD-COHORT MODELS,

With Applications to the Study of Self-Reported Health

Word count: 12835

Number of Tables: 3

Number of Figures: 6

Hui Zheng, Yang Yang, and Kenneth C. Land

---

*Hui Zheng, Initiative in Population Research and Department of Sociology, The Ohio State University. Yang Yang, Lineberger Comprehensive Cancer Center, Department of Sociology, Carolina Population Center, University of North Carolina at Chapel Hill. Kenneth C. Land, Center for Population Health and Aging, Duke Population Research Institute and Department of Sociology, Duke University. We thank Bruce Western, members of the DuPRI Seminar and the anonymous reviewers for helpful comments on a previous version of the paper. Address all correspondence to Hui Zheng, Initiative in Population Research and Department of Sociology, The Ohio State University, Columbus, OH 43210; email: huizhenghome@gmail.com.

VARIANCE FUNCTION REGRESSION
IN HIERARCHICAL AGE-PERIOD-COHORT MODELS,

With Applications to the Study of Self-Reported Health

Abstract

Two long-standing research problems of interest to sociologists are the sources of variations in social inequalities and the differential contributions of the temporal dimensions of age, time period, and cohort to variations in social phenomena. Recently, a model, called Variance Function Regression, has been introduced to sociologists 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 problem. This paper presents an integration of these two models as a means to study the evolution of social inequalities along distinct temporal dimensions. The integrated model then is applied to survey data on subjective health status. Substantial age, period, and cohort effects as well as gender differences are found 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 the detection of age, period, and cohort variations in the latter disparities that application of the integrated model permits. Net of the effects of age and individual-level covariates, in recent decades cohort differences in the 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 Boomer birth cohorts show significant and increasing levels of within-group disparities. These findings illustrate how the integrated model provides a powerful framework and lens through which to identify and study the evolution of variations in social inequalities across the age, period, and cohort temporal dimensions. Accordingly, this model should be broadly applicable to the study of social inequality in many different substantive contexts.

VARIANCE FUNCTION REGRESSION IN HIERARCHICAL AGE-PERIOD-COHORT MODELS,

With Applications to the Study of Self-Reported Health

One of the long-standing core analytic tools 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 the age-period-cohort (APC) analysis (Mason and Fienberg 1985; Ryder 1965). The use of regression-based models in APC analysis closely relates the study of cohort change to the study of inequality both substantively and methodologically, yet no analytic tools have been available for the systematic examination of age and temporal (i.e., period and cohort) variations in inequalities beyond those captured by the conditional means of regression models.

In this paper, we fill this gap by intersecting two recent developments in statistical models for the 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 the 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 the 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.

THE MODELS TO BE INTEGRATED: A BRIEF OVERVIEW

We begin with a brief overview of the two models which we aim to integrate. Standard regression-based approaches to studies of inequality are largely limited to between-group differences. *Group* here means each category of a covariate. For example, "gender" has two groups: men and women. Between-group inequality in this case is the inequality between men and women. Within-group inequality in this case is the remaining inequality within the population of men or within the population of women. It often is the case that between-group inequality is far exceeded by within-group or residual inequality. Residual inequality often is considered 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 be used to address this limitation by simultaneously modeling both the mean and variance of an outcome variable as functions of covariates and hence takes into account both between-group and within-group differences. This class of models explicitly targets the residual variance in regression models for analysis.

Conventional linear regression models assume that the residual error terms of the models 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 the errors have normal probability distributions (see, e.g., Fox 2008:187-219). Violations of these assumptions affect estimators of the standard errors of regression coefficients and reduce the statistical efficiency of conventional least-squares estimators. A variety of statistical methods have been developed for diagnosing and correcting non-constant error variances (Fox 2008:272-277). The key feature of VFR/HR is that it treats violations of homoscedasticity as more than a data problem that needs to be corrected in order to obtain well-behaved estimators. It rather approaches these violations as being of 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 over the years 1975 to 2005, the residual variance can be interpreted as measuring within-group risk or insecurity.

Consider next the three time-related dimensions, namely, 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 the variation across different age groups brought about by physiological changes, accumulation of social experience, and/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; these may reflect the effects of having different formative experiences for successive age groups in successive time periods (Yang 2010).

One common goal of APC analysis is to assess the effects of one of the three factors on some outcomes of interest net of the influences of 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, Mason, Winsborough and Poole 1973). A recently developed modeling approach, *Hierarchical APC (HAPC)* models, has been used to avoid this problem using micro data and multilevel modeling framework. 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 as opposed to fixed effects additive to that of age (Yang and Land 2006, 2008; Yang 2006). This contextual approach broadens the theoretical foundation of APC analysis, helps to 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 the 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. One often needs to understand sources of social inequality attributable to age, time period, and birth cohort in that the APC variations represent the complex and temporal patterns of inequality. In his seminal article (1965), Ryder argued that cohort membership is a structural category that could 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. The VFR/HR models have been used to examine inequalities in both terms but have not distinguished the age, period, and cohort sources of temporal variation.

The HAPC regression models have been used to examine temporal differences in conditional means but not within-group variances. The rest of the paper illustrates the utility of the intersection of these two modeling frameworks with an application to the analysis of health disparities in the United States over the years 1984 to 2007. This application produces several useful findings.

The next section describes this synthesis of research strategies and models. This is followed by a description of the data on self-reported health to be analyzed and a presentation of the empirical results of the empirical analysis. A discussion and conclusions section ends the paper with general observations on the utility and potential for advances in empirical analyses of inequality of this synthesis of analytical models.

RESEARCH TOPIC, STRATEGY, AND MODELS

Health Disparities

The specific topic of empirical analysis in this paper is health outcomes and disparities or inequalities therein. We use the term *health disparities* to refer to either between-group 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 the age, time period, and cohort categories. *Between-group health disparities* refer to the 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 the 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 which has addressed between-group health disparities, *there are three standard approaches to the study of changes in within-group health disparities*: (1) *across the life course* (e.g., House, Lepkowski, Kinney, Mero, and Kessler 1994; Dannefer 2003), (2) *across cohorts* (e.g., Chen, Yang, and Liu 2010; Yang and Lee 2009; Warren and Hernandez 2007), and (3) *across time periods* (e.g., Pappas, Queen, Hadden, and Fisher 1993; Goesling 2007). Within each approach, there is evidence for 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 (e.g., Ross and Wu 1996; Lauderdale 2001; Lynch 2003; Dupre 2007). The gap in self-rated health by education levels has also widened across birth cohorts (Chen et al. 2010; Lynch 2003), whereas the intracohort gaps by sex and race have been constant across birth cohorts (Yang and Lee 2009). There is also evidence of increasing socioeconomic inequality in health, disability, and life expectancy in the U.S. in the past several decades (e.g., Feldman, Makuc, Kleinman, and Cornoni-Huntley 1989; Pappas et al. 1993; Preston and Elo 1995; Hummer, Rogers, and Eberstein 1998; Meara, Richards, and Cutler 2008; Jemal, Ward, Anderson, Murray, and Thun 2008; Schoeni, Martin, Andreski, and Freedman 2005; Crimmins and Saito 2001; Goesling 2007; Liu and Hummer 2008). In addition, recent research also 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, however, has not been examined previously and merits a formal test using properly constructed analytic models.

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

The key outcome variable in this paper is self-rated health. Self-rated health is a widely used measure of general health status that has been found to be 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, subclinical illness, and has been found in some studies to be more predictive of mortality among the elderly than physician assessments (Schoenfeld, Malmrose, Blazer, Gold, and Seeman 1994; Hays, Schoenfeld, Blazer, and Gold 1996). Close relationships between self-rated health and objective health indicators also hold across population subgroups (Bosworth, Siegler, Brummett, Barefoot, Williams, Clapp-Channing, and Mark 1999; Kennedy, Kasl, and Vaccarino 2001). Recent studies suggest that the gap in self-rated health by educational levels has widened across age, time period or cohort (e.g., Goesling 2007; Liu and Hummer 2008; Lynch 2003). These findings are consistent with other studies investigating disease incidence, mortality, or life expectancy gaps by educational levels across age, time period or cohort (e.g., Dupre 2007; Lauderdale 2001; Pappas et al. 1993; Meara et al. 2008). 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. But we also caution that health is not a singular condition and findings in this paper 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 et al. (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 years to 8.3 years) than for women (from 5.6 years 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. In contrast, Meara et al. (2008) found that educational disparities in life expectancy widened among women in the two most recent decades. In sum, these studies generally have found that within-period health disparities have widened for men, while varying for women depending on the time periods and health outcomes examined. That is, there is substantial evidence of 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] In other words, we investigate how health disparities may change across age, time period and cohort within each gender (i.e., men and women).

The main purpose of this paper is to present a method that facilitates the disentanglement of age, period, and cohort variations in health disparities defined by differences in both conditional mean levels and conditional dispersions of health (i.e., between-group and within-group health disparities). To do so, we intersect 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 intersect these two statistical models and term the result a Hierarchical-Age-Period-Cohort-Variance-Function-Regression Model (HAPC-VFR/HR). We next describe each of these components in more detail.

Hierarchical Age-Period-Cohort Analysis

The application of multilevel models aids the estimation of age, period, and cohort components of temporal change in that, 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). Studies utilizing multilevel models in the analysis of temporal change are few but exist in demographic research on health (Lynch 2003) and developmental psychology and aging research on cognitive skills (e.g., Alwin 2009). These examples did not explicitly embody a full-blown APC analysis due to the identification problem and focused on age patterns in the context of cohorts only.

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. First, note that the 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 alone is unsatisfactory for two reasons (Yang 2010). It is still embedded in the simple linear regression model which assumes linearity and additivity of the three variables and does not completely avoid the identification problem. The results may be sensitive to the choice of interval widths as longer widths may allow a higher degree of overidentification. And 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 the data may lead to underestimated standard errors.

The HAPC approach utilizes unique features of the multi-level survey design and presents a more thorough solution. It begins with the recognition that in this design, respondents are nested in, and cross-classified simultaneously by, the 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 and therefore avoid the identification problem and can statistically characterize contextual effects of historical time and cohort membership. The HAPC model, and specifically, the cross-classified random-effects model (CCREM) form of this model, satisfy these criteria and can accommodate covariates at both individual and contextual levels to aid 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 (2008) study of inequalities in subjective well-being and Reither et al.'s (2009) study of the obesity epidemic in the U.S.

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]

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

Level-2 or "Between-Cell" Model:

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

Combined or Mixed Effects Model:

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

for $i = 1, 2, \ldots, n_{jk}$ individuals within cohort *j* and period *k*;

$j = 1, \ldots,$ J birth cohorts;

$k = 1, \ldots,$ 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}, $X_{2ijk}, \ldots, X_{Pijk}$, (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 the 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 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}, assumed 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 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)-(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 of the individual-level. The 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 or not there is evidence of random variation across time periods or cohorts in the level-1 slope coefficients. If there is evidence of such significant variation, then the model should be modified to incorporate this variation.

Variance Function Regression

Variance Function Regression/Heteroscedastic Regression[5] has two parts, including a regression for an outcome variable, Y_i, and a regression for logarithm of the residual variances, $\log(\sigma_i^2)$ (Western and Bloome 2009:299):

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

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

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 the 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 the deviations of the within-group sample means from the average or grand mean $\bar{Y}$ (which can be termed the *between-group inequality*), while the second regression explains how covariates affect the within-group variability of the response variable around the group means (which can be termed the *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 the age, time period, and birth cohort temporal dimensions are relevant. To do this, 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. The algebra for this algorithm is represented in Equations (1) to (3).

This step produces a set of estimates of fixed-effects coefficients (for the individual-level explanatory covariates), random-effects coefficients (for cohorts and periods), and random variance components matrix which 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 paper, this step estimates the 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 the 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, $\hat{e}_{ijk}^{2}$ or denoted as σ_{ijk}^{2}. We then apply the residual pseudo-likelihood (RSPL) estimator of the CCREM to estimate Equations (5) of the VFR/HR regression model.[6] For normal distributed errors, the squared residuals will have a gamma distribution, and Equation (5) then is estimated in generalized linear mixed model form—as a gamma regression of $\hat{e}_{ijk}^{2}$ on the X_{ijk} using a log link function (Western and Bloome 2009:300; see also Nelder and Lee 1991).[7]

The algebra for this algorithm can be stated as follows:

Level-1 or "Within-Cell" Model:

$$\log(\sigma_{ijk}^2) = \lambda_{0jk} + \lambda_1 X_{1ijk} + \lambda_2 X_{2ijk} + ... + \lambda_P X_{Pijk}, \tag{6}$$

Level-2 or "Between-Cell" Model:

$$\lambda_{0jk} = \pi_0 + \omega_{0j} + \varphi_{0k}, \ \omega_{0j} \sim N(0, \psi_u), \ \varphi_{0k} \sim N(0, \psi_v) \tag{7}$$

Combined or Mixed Effects Model:

$$\log(\sigma_{ijk}^2) = \pi_0 + \lambda_1 X_{1ijk} + \lambda_2 X_{2ijk} + ... + \lambda_P X_{Pijk} + \omega_{0j} + \varphi_{0k} \tag{8}$$

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 the residual random effects of cohort *j* and period *k*, respectively, assumed normally distributed with mean 0 and variance ψ_u and ψ_v. 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 the individual-level explanatory covariates), random-effects coefficients (for cohorts and periods), and random variance components matrix which evaluates the contributions of these variables to the explanation of variance in the logarithm of the residual variances, $\log(\sigma_i^2)$, for each sample respondent *i*. In the context of this paper, this step estimates the variations in the variance or dispersion of self-rated health across period, cohort, age and other individual-level covariates. It merits emphasizing here that the predicted $\widehat{\sigma}^2$ for age, period or cohort represents a general form of dispersion of health across age, period or cohort, whereas previous empirical research in health disparities focused on specific inequality by one or two dimensions such as sex, race, or SES. An increase or decrease in any one of the specific dimension will contribute to the increase or decrease in the general dispersion, which however has not been studied in the current literature. Therefore, instead of examining the 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 each of these two steps produces restricted maximum likelihood or residual pseudo-likelihood estimators from the CCREM, it must be iterated in order to obtain maximum likelihood (ML) estimators for the variance function regression model (Aitkin 1987). As Western and Bloome (2009:301) indicate, the fitted values $(\hat{\sigma}_{ijk}^2)$ from an application of the two-steps should be saved and used in a weighted regression of Y_{ijk} on $X_{1ijk}, X_{2ijk}, \ldots, X_{Pijk}$ with weights $(1/\hat{\sigma}_{ijk}^2)$. Estimates of the residuals from Step 1 then are updated, Step 2 is computed, and so forth until convergence. Western and Bloome (2009:301) note that the ML estimator may perform poorly in small samples, in which case a REML or Bayes estimator can be used. In the empirical application described below, however, the sample sizes are very large, which case the adjustments in the REML made for the loss of degrees of freedom resulting from estimation of the regression parameters will be very small, if not trivial. Therefore, for purposes of the empirical application of the HAPC-VFR/HR model in this paper, the ML estimator is applied.

DATA AND METHODS

Data

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

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

Variables

The outcome variable, *self-rated health*, has remained largely unchanged across periodic revisions of the NHIS questionnaires, which facilitates the analysis of trends. It has *five response categories*: poor, fair, good, very good, and excellent. Findings from previous studies regarding the validity and robustness of this summary health outcome variable were reviewed in the text above.

The objective of this analysis is to examine age-period-cohort variations in self-reported health and disparities therein after controlling for individual-level demographic and social variables that have been linked to health in previous research (e.g., Bird, Conrad, Fremont, and Timmermans 2010), namely, gender, race, marital status, work status, education and income. In the NHIS, income was measured by several income categories. We first calculated the mid-point of each income category, converted the mid-points 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 ranged from 0 to 18. Work status is a binary variable in which 1 = full/part time job and 0 = not employed. We also controlled 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). This data is composed of 57% men, 82% white, 55% married, and 67% employed. Summary statistics for this data are presented in Table 1.

(Table 1 about here)

Model Specification and Estimation

The nature of the self-rated health outcome variable—in the form of a five ordered response categories (poor, fair, good, very good, and excellent)—complicates the specification and estimation of the combined HAPC-VFR/HR model. This model was described above in a linear mixed effects regression format. For several reasons, 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. First, 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 the self-rated health responses in the NHIS data. Evidence of this was obtained from an ordered logit regression analysis of this outcome variable. [9] Third, after presenting the empirical results of this analysis, we will describe various tests of robustness of the findings to the NHIS dataset and to this model specification. We also describe extensions of the HAPC-VFR/HR model to more complicated model specifications that more faithfully accommodate the ordered nature of the self-rate health variable.

With respect to the two-step algorithm for estimation of the model stated above, 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. These differences in group-specific means are the topic of study in prior studies of health status and the HAPC model permits the decomposition of temporal changes therein 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, health disparities changes across age, period and cohort. It is the detection of these temporal changes in within-group variations, and their decomposition into age, period, and cohort components that are made possible by the integrated HAPC-VFR/HR model.

RESULTS

Figure 1 displays the sample means of self-rated health in the NHIS for the years from 1984 to 2007 after adjusting for sample weights and smoothing the 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 afterwards, 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 than men's. These differences in the overall trends significantly reduced the self-rated health gap between men and women by the late-1990s, with the reduced gap continuing to the end of the series in 2007.

(Figure 1 about here)

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 afterwards, and then rose again after 1998-

1999. Men and women show similar period-to-period variations in health disparities among each group of them.

(Figure 2 about here)

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

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. The results were obtained using the maximum likelihood estimation method described above. The "β" column presents the results for the first-stage regression of the HAPC-VFR/HR model (which estimates variations in mean health across groups), and "λ" column presents the results for the second-stage regression of the HAPC-VFR/HR model (which estimates variations in dispersion of health across groups).

(Table 2 about here)

As shown in "β" column, consistent with findings from previous studies, being male, white, married, more educated, having a job and more income are associated with better self-rated health. For example, being a male significantly increases the expected value of self-rated health by 0.03 points on a five-point scale.[10] Regarding race, being white is associated with a 0.165 points increase in expected value of self-rated health and this is highly statistically significant. Being married significantly increases the expected value of self-rated health by about 0.02 points. Regarding education, each year of additional education is associated with a statistically significant 0.06 points increase in the expected value of self-rated health. Being employed has an even larger impact: a statistically significant 0.39 points increase in the expected value of self-rated health. Finally, every $10,000 increase in household income is associated with a statistically significant 0.07 points increase in the expected value of self-rated health. The effects of age are curvilinear (quadratic) in that the 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 for about 0.06 points for women and 0.08 points for men. In other words, in the absence of the sample redesign, we would have seen an even larger increase in self-rated health in the late-1990s for both men and women than shown in Figure 1. The estimates of residual variance components at Level-2 indicate significant period and cohort effects net of the effects of individual-level covariates, while the period effect is larger than cohort effect as reported in the "Variance Components" section.

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 earlier or later birth cohorts. An exception is the 1899-1904 cohort, whose relatively large positive effect may be due to the selective survival effect as well as the small number of respondents from this early cohort in the NHIS data. In addition, before 1998, the 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 the 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 modestly more than cohorts of the overall trend in self-rated health from 1984 to 2007.

(Figure 3 about here)

As a key output of the VFR/HR part of the integrated model, the "λ" column in Table 2 shows how individual-level covariates affect within-group health disparities. The estimated within-group health disparities for males, whites, married persons, the more highly educated, employed individuals, and those more income are smaller than those of their counterparts, i.e., females, blacks, unmarried persons, the less educated, unemployed individuals, and those with less income. For example, the predicted log of within-group variance in self-rated health of male is 0.009 units smaller than that of female. Regarding race, the predicted log of within-white variance in self-rated health is 0.08 units smaller than the within-nonwhite variance in self-rated health. Being married significantly decreases the predicted log of within-group variance in self-rated health by about 0.02 units. Regarding education, each year of additional education is associated with a statistically significant 0.03 units decrease in the predicted log of within-group variance in self-rated health. Being employed has an even larger impact: a statistically significant 0.34 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 0.04 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 0.07 units for women and 0.05 units for men. In the other words, without the sample redesign, there would have been a smaller decline in the variance of self-rated health in the years after 1994 than 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). The estimates of residual variance components at Level-2 indicate significant cohort and non-significant period effects net of the effects of individual-level covariates as reported in the "Variance Components" section. Graphs of these estimated effects are show in Figure 4. After controlling for demographic and socioeconomic statuses, estimated health disparities in the young adult ages are relatively small, indicating that most everyone is relatively healthy. But health disparities increase with age, reaching a peak around age 55 as shown in 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 in 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 trends is most pronounced in recent cohorts). After controlling for individual-level covariates and age and cohort effects, the 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 the fluctuations of crude variance in self-rated health over time. For example, the recent increase in health disparities after circa the year 2000 in Figure 2 corresponds to the increasing proportions of post-baby boomer cohorts (born after 1964) in the population—cohorts that have larger within-cohort health disparities than the preceding cohorts as seen in 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 the variance in self-rated health does not significantly vary across periods.

(Figure 4 about here)

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

Estimates of the gender-specific models of self-rated health and health disparities are given in Table 3. These show that being white, married, more educated, employed and having more income are associated with better self-rated health for both men and women except that married men do not significantly have better self-rated health than unmarried men (see the “β” column in Table 3).

(Table 3 about here)

Figure 5 contains graphs of the dependence of self-rated health on age, time period, and birth cohort as estimated in the models of Table 3. The age-dependence curves in Figure 5 shows that both men and women’s self-rated health declines with increasing age, but the trends reverse after around age 69 for men and age 72 for women. It can be seen that men report better self-rated health than women at all ages. The gender gap in health is largest in the early adult years, narrows until around age 61, and widens afterwards.

Figure 5 also contains graphs of the estimated period and cohort effects on the conditional expected values of self-rated health status. The estimated cohort effects are relatively flat across cohorts for men, except that baby boomers born between 1950 and 1959 significantly have better self-rated health than other cohorts. By comparison, the conditional expected values of self-rated health changed dramatically across cohorts for women. They continued declining from the 1899-1904 cohort to the early baby boomers born 1945-1954 and then rose for the middle and late baby boomers and afterwards, which result in a widened and then narrowed self-rated health gap between men and women. The estimated period effects on self-rated health for men and women in Figure 5 show similar trends to the observed means of self-rated health shown in Figure 1. In general, the estimated β coefficients in Table 3 suggest period effects contribute slightly more than cohort effects to the changes in self-rated health from 1984 to 2007 for both men and women.

(Figure 5 about here)

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

Figure 6 contains graphs of the estimated within-group health disparities by age, time period, and birth cohort from Table 3. These show that estimated within-age health disparities have a bell shape that peaks around age 56 for both men and women. Compared to men, the within-age health distributions for women are slightly more spread out. The dispersions are larger than those of men at all ages. They increase at a slower rate before age 56 and decrease at a slower rate after age 56.

(Figure 6 about here)

Similar to the results of age variations, within-cohort and within-period heterogeneities are larger for women than men. For women, Figure 6 shows that within-cohort variances decreased from the 1899-1904 cohort to the 1930-1934 cohort, increased in cohorts born in the late stage of Great Depression and World War II followed by decreases in baby boomer cohorts and increases 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 in cohorts born in the Great Depression, World War II and baby boomer cohorts born between 1945 and 1959 followed by substantial increases afterwards, especially in more recent cohorts. After controlling for individual-level covariates and age and cohort effects, the graphs in Figure 6 show that within-period variances are relatively flat from 1984 to 2007 for both men and women. The lack of significant random-effects coefficients for period and the statistically insignificant variance components of the period effects for both men and women in Table 3 also suggest the 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 the changes in health disparities from 1984 to 2007, for both men and women shown in Figure 2. The striking gender difference in the 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 was based on its ease of interpretation. But this model is not fully sensitive to either the ordered categorical nature of the survey responses to the self-rated question or to the non-normal frequency distribution of the five-point scaling of these responses. Accordingly, we describe several analyses to assess the robustness of the empirical findings reported above from application of this model and some methodological extensions necessary to adapt it to the nature of the self-reported health outcome variable.

Alternative Coding of the Response Variable. As a first robustness analysis, alternative approaches to the extraction of a measurement signal from the ordered response categories of the self-rated health question should be considered. One possibility is to dichotomize the responses into, say, fair or poor versus excellent, very good, or good (see, e.g., Lynch 2003). We conducted HAPC-VFR/HR analyses—in both linear probability and logit regression specifications—of a dichotomized self-rated health outcome for the NHIS data. Findings regarding the effects of the individual-level covariates and the period, and cohort random effects on the mean of self-rated health from these analyses 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, i.e., 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 form of assessment of the robustness of the findings from the NHIS data reported above, we also analyzed General Social Survey (GSS) data for the period 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 2000 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 the observed means of self-reported health status and variances therein generally are similar to those reported above for the NHIS data. And 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 above have survived the foregoing two robustness analyses. Beyond this, the linear mixed effects specification of the HAPC-VFR/HR model needs to be generalized to other forms that take into account the ordered categorical nature of the self-rated outcome variable. This requires the specification of a corresponding generalized linear model with random effects version of the model (Fox 2008:335-424). The specification and estimation of such a version of the HAPC-VFR/HR model not only is more difficult to interpret, but also is methodologically challenging for the second-step regression. [13]

Initial work with ordered logit specifications of these generalized models has, however, produced findings that are consistent with the 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 (over 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 has treated these sources of variations separately. In any given set of data, they are, however, potentially confounded with each other as well as individual age. By intersecting a variance function regression analysis, which facilitates the separation of within-group inequality from between-group inequality, with a HAPC model, which facilitates the estimation of age, period, and cohort effects in a nonlinear fashion, the result is a Hierarchical-Age-Period-Cohort-Variance-Function-Regression Model (HAPC-VFR/HR).

We have shown the utility of this intersection with an application to the analysis of self-reported health disparities in the United States over the years 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 affects mean of self-rated health net of a set of individual-level covariates (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. The results were obtained by application of 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, Ukraintseva, Kulminskaya, Arbeev, Land, and Yashin

2008); 2) late baby boomers (born 1955-64) report better health than 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 the 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 afterwards, which is consistent with the findings of convergence in SES health disparities in late life in prior research. This convergence may be a result of mortality selection, that is, sicker people, who are disproportionately of lower SES, are more likely to die earlier and be disabled which removes them from the survey (Dupre 2007). But this 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, and equalization of health care usage and protections through Medicare coverage at age 65 (House et al. 1994; House, Lantz, and Herd 2005).

We also find cohort and period effects. Within-cohort health disparities generally decreased from the 1899-1904 cohort to the baby boomer cohorts and have substantially increased for post-baby boomer cohorts. After taking into account age and cohort effects, within-period health disparities are flat since 1984. By contrast to what we find for the conditional mean of self-rated health, cohort effects appear to contribute much more than period effects to the changes in the variance of self-rated health over the past two decades. As post-baby boomer cohorts (especially cohorts born after 1980) have much larger within-cohort health disparities than preceding cohorts, and within-age health disparities increase with age until around age 55, it can be expected 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 the post-baby boomer cohorts. First, income inequality has increased dramatically in the last three decades in the U.S., which strengthens 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 the recent cohorts as well. Second, since 1980, young adult cohorts have had an increased prevalence of immigrants (documented and undocumented). This may be associated with both the 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 the increasing heterogeneity of the cohorts). Third, compared to the “healthy migrant” effects, which contribute to the upper tail (higher residual self-reported health) of the conditional dispersions, dramatically increased obesity risk in cohorts born after 1970 (Reither et al. 2009) may contribute to the lower tail (lower residual self-reported health) of the conditional dispersions. The presence of either or both of these effects of these population composition changes 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 obesity risk further increase the conditional health dispersions of recent cohorts. Fourth, as part of this increased cohort heterogeneity, these relatively young cohorts have grown up in the information age of the Internet and the World Wide Web. While information on the Web is generally available to everyone, the past two decades have seen a “digital divide” in access to, and use of, the Web and its information sources (Norris 2001) that may also be related to the increasing recent cohort variances in health knowledge and consequently health outcome (Brodie, Flournoy, Altman, Blendon, Benson, and Rosenbaum 2000). These postulated explanations are readily testable using the HAPC-VFR/HR model in future research.

This paper also examines the 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 (Case and Deaton 2003; Arber and Cooper 1999). Our results suggest a narrowing of the self-reported health gap until around age 61 and then a widening afterwards. This is consistent with a process by which men are more likely 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 than women (Case and Paxson 2005), which reduces the self-reported health gap in the first place and then widens it afterwards due to the selective survival, leaving relatively stronger men but a relatively larger share of women with poorer health alive in older ages.

Compared to the relative 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 early baby boomers cohort and then rises afterwards. 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? Is it because female infants were more negatively affected by WWI, the Great Depression and WWII, or because female infants were more likely to survive but were in bad health during hardship and male infants were more likely to die from these negative social events which left relatively healthier male infants? Future research is needed to solve this puzzle. By comparison, the gender-specific period trend in health is consistent with the one for the whole population, that is, both men and women have increasingly reported 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, especially for later ages. Men tend to report better health than women in early adulthood (MacIntyre, Hunt, and Sweeting 1996), which makes their within-age health disparities smaller in early adulthood. However, they are also more likely to get severe chronic diseases than women in middle age (Case and Paxson 2005), which possibly results in smaller difference 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 in later ages, which leads to a faster drop in within-age health disparities than women in late life.

Men also have smaller within-cohort heterogeneity than women for all cohorts, especially for cohorts born in late stage of Great Depression, WWII, and baby boomers. This raises a similar puzzle as the distinct cohort patterns in gender-specific health: What causes distinct cohort patterns in gender-specific health disparities? It appears that men and women have significantly different experiences and endure the different impacts from these several historical events in the mid 20th century. These puzzles lead to a potentially important project in the future. In terms of within-period health disparities, they are relatively flat for both men and women as for the whole population. These gender-specific analyses further support that cohort effects contribute more than period effects to the changes in health disparities from 1984 to 2007.

By using the HAPC-VFR/HR model, we have been able to give a more complete picture of the evolution of changes in self-rated health and health disparities in the U.S. from 1984 to 2007. More 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 with each other but have not been systematically disentangled in the existent studies due to the lack of an integrated model. By using the HAPC-VFR/HR model, this study demonstrates that changes in self-rated health disparities in the last two decades have been much more of a cohort than a period story. Therefore, further research in this area should pay more attention to the "cohort" perspective. Second, prior literature has focused on the changes in health inequality defined by specific aspects of social stratification system such as gender, race, and SES across age, period or cohort, without an overall picture of the changes in 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, which provides the basis for further research to examine the contribution to the general dispersion by each specific aspect of social stratification system.

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

Notes

1. To estimate HAPC models, Yang and Land (2006) applied the statistical methodology of mixed (fixed and random) effects or hierarchical regression models, treating the 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 and/or periods for reliable statistical estimation and the assumption that the period and cohort effects are independent of individual-level regressors. Yang and Land (2008) examined these assumptions and compared 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 showed 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. Respondents in the repeated cross-section sample surveys are cross-classified by both 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 the 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 biological process of aging *internal to individuals*. By contrast, period and cohort effects reflect the influences of *forces that are external to individuals* and operate in different ways. It is for these reasons that we 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 the attention of sociologists the heteroscedastic regression model, we use the term variance function regression throughout this paper although this model has been termed *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's 2005). There are a number of differences in the model specifications and estimation algorithms developed in the DHGLM and GAMLSS approaches to heteroscedastic regression. 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, Milliken, Stroup, Wolfinger, and Schabenberger 2006), but RSPL is consistent and asymptotically normally distributed for non-normal data as well.

7. The outcome variable in the analyses described below, self-rated health is not normally distributed; it is skewed to the left. The 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 the coefficients of 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 the fixed and random effects of a mixed model that are consistent and asymptotically normally distributed; even the identically distributed assumption can be relaxed (Demidenko 2004:647).

8. We did not use waves of NHIS data prior to 1984 for two reasons: 1) the self-rated health question contained four responses 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 2008:363-368). 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 the thresholds which then can be used to assess whether the equal intervals assumption for the categorical variable is violated. In an application to the NHIS self-reported health variable, we found that the variability of the distances from threshold to threshold is about five 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 the 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 found that a dichotomous coding of self-rated health responses produced 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. A 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. And 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 in order 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 there are other possible specifications as well. A separate paper 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* 36:332-9.

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: 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(6): 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 Difference 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* 75:126-150.

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”. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 58: S327-S337.

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

Dupre, Mattew 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): 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(5): 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(4): 1621-44.

Goldstein, Harvey. 2003. *Multilevel Statistical Models*. 3d ed. London: Oxford University Press.

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 2002/2002 (American's 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(3):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(5): 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(4): 551-61.

Lee, Youngjo and John A. Nelder. 2006. "Double Hierarchical Generalized Linear Models." *Journal of the Royal Statistical Society Series C* 55:139-185.

Littell, Ramon C., George A. Milliken, Walter W. Stroup, Russell D. Wolfinger, and Oliver Schabenberger. 2006. *SAS for Mixed Models,* 2^{nd} 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: 1898-906.

Lynch, Scott M. 2003. "Cohort and Life-Course Patterns in the Relationship between Education and Health: A Hierarchical Approach." *Demography* 40(2): 309-31.

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

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.

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(2): 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: Pub. Health Serv.

Nelder, John A. and Youngjo. Lee. 1991. "Generalized Linear Models for the Analysis of Taguchi-Type Experiments." *Applied Stochastic Models and Data Analysis* 7:101-20.

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: 103-15.

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(4): 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 U.S.” *Social Science & Medicine* 69:1439-1448, with Commentary.

Rigby, R. A. and D. M. Stasinopoulos. 2005. “Generalized Additive Models for Size, Location, 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: 104-20.

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:M109-15.

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(11): 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(3): 392-9.

Smith, Herbert L. 2008. “Advances in age-period-cohort analysis.” *Sociological Methods & Research* 36: 287 – 296.

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* 11:836-847.

Smyth, Grodon K., A. Frederik Huele, and Arunas P. Verbyla. 2001. “Exact and Approsimate REML for Heteroscedastic Regression.” *Journal of Graphical and Computational Statistics* 11:836-847.

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* 73:903-20.

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 Kyriakos S. Markides. Sage Publications.

Yang, Yang. 2008a. “Social Inequalities in Happiness in the United Stated, 1972-2004: An Age-Period-Cohort Analysis.” *American Sociological Review* 73:204-26.

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.): Chapter 2*, edited by B. Binstock and L.K. George. Academic Press.

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(February): 297-326.

Yang, Yang and Linda C. Lee. 2009. “Sex and Race Disparities in Health: Cohort Variations in Life Course Patterns.” *Social Forces* 87: 2093-124.

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.

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.

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

Outcome	***Description***	***N***	***Mean***	***SD***	***Min***	***Max***
Health	Self-reported health. 1=poor, 2=fine, 3=good, 4=very good, 5=excellent	701888	3.76	1.13	1	5
Level-1 Variables						
Sex	1=man, 0=woman	701888	0.57	0.49	0	1
Race	1=white, 0=other races	701888	0.82	0.38	0	1
Age	Respondent's age at survey year	701888	46.55	17.30	18	85
Education	Respondent's years of schooling	701888	12.65	3.17	0	18
Marital	1=married, 0=others	701888	0.55	0.50	0	1
Employed	1=employed, 0=others	701888	0.67	0.47	0	1
Income/10000	Household income at survey year	701888	4.58	2.76	0.068	10.41
Redesign	1=1995-2007, 0=1984-1994	701888			0	1
Level-2 Variables						
Cohort	Five-year birth cohorts	18			1899-1904	1985-1989
Period	Survey year	24			1984	2007

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

	β		*λ*	
Fixed Effects	***coefficient***	***se***	***coefficient***	***se***
Intercept	3.281***	*0.009*	0.403***	*0.022*
Age	-0.142***	*0.002*	0.071***	*0.005*
Age2	0.034***	*0.001*	-0.041***	*0.001*
Male	0.030***	*0.003*	-0.009*	*0.005*
White	0.165***	*0.003*	-0.080***	*0.004*
Married	0.022***	*0.003*	-0.024***	*0.004*
Education	0.060***	*0.000*	-0.026***	*0.001*
Employed	0.388***	*0.003*	-0.338***	*0.004*
Income/10000	0.065***	*0.001*	-0.042***	*0.001*
Redesign	-0.061***	*0.009*	-0.068***	*0.008*
Redesign*Male	-0.021***	*0.005*	0.016*	*0.007*
Random Effects				
Cohort	***coefficient***	***se***	***coefficient***	***se***
1899	0.027*	*0.012*	0.201***	*0.033*
1905	-0.005	*0.010*	0.124***	*0.029*
1910	-0.008	*0.009*	0.030	*0.027*
1915	0.002	*0.008*	-0.003	*0.025*
1920	-0.005	*0.007*	-0.021	*0.024*
1925	-0.020**	*0.007*	-0.029	*0.023*
1930	-0.009	*0.007*	-0.024	*0.022*
1935	-0.001	*0.006*	-0.029	*0.022*
1940	0.003	*0.006*	-0.041	*0.022*
1945	-0.009	*0.006*	-0.066**	*0.021*
1950	0.007	*0.005*	-0.088***	*0.022*
1955	0.018***	*0.005*	-0.103***	*0.022*
1960	0.023***	*0.006*	-0.097***	*0.023*
1965	0.003	*0.006*	-0.066**	*0.024*
1970	-0.006	*0.007*	-0.039	*0.025*
1975	-0.026***	*0.008*	0.044	*0.027*
1980	-0.007	*0.009*	0.077**	*0.029*
1985	0.012	*0.011*	0.129***	*0.033*
Period				
1984	-0.011	*0.007*	0.018**	*0.006*
1985	-0.006	*0.007*	0.005	*0.006*
1986	0.006	*0.008*	-0.005	*0.006*
1987	-0.007	*0.007*	-0.009	*0.006*
1988	-0.013	*0.007*	0.006	*0.006*
1989	0.008	*0.007*	-0.008	*0.006*
1990	0.020**	*0.007*	0.000	*0.006*
1991	0.017*	*0.007*	-0.006	*0.006*
1992	0.002	*0.007*	-0.008	*0.006*
1993	-0.005	*0.007*	0.006	*0.006*
1994	0.004	*0.007*	0.001	*0.006*
1995	-0.012	*0.007*	0.000	*0.006*
1996	-0.004	*0.008*	0.001	*0.007*
1997	0.025***	*0.007*	-0.004	*0.006*
1998	0.025**	*0.008*	-0.005	*0.006*
1999	0.025**	*0.008*	0.000	*0.006*
2000	0.015	*0.008*	0.003	*0.006*
2001	0.007	*0.008*	0.002	*0.006*
2002	0.003	*0.008*	0.001	*0.006*
2003	0.002	*0.008*	0.000	*0.006*
2004	-0.023**	*0.008*	-0.005	*0.006*
2005	-0.021**	*0.008*	-0.004	*0.006*
2006	-0.011	*0.008*	-0.001	*0.007*
2007	-0.046***	*0.008*	0.011	*0.007*
Variance Components	***variance***	***se***	***variance***	***se***
Cohort	0.0002*	*0.000*	0.008**	*0.003*
Period	0.0003**	*0.000*	0.000	*0.000*
Model Fit				
BIC	1941250			
-2 Res Log Pseudo-Likelihood			2351732	

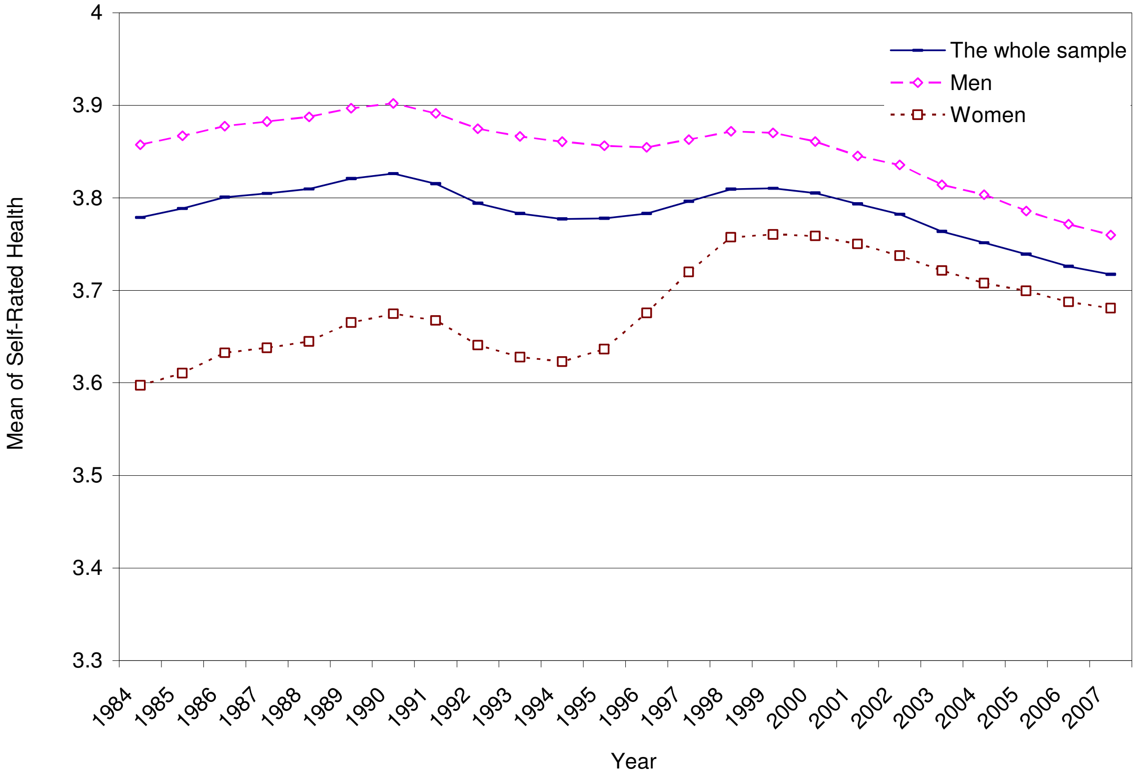
* indicates p<.05; ** indicates p<.01; *** indicates p<.001.

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

	β				***λ***			
Fixed Effects	***Men***	***se***	***Women***	***se***	***Men***	***se***	***women***	***se***
Intercept	3.318***	*0.012*	3.285***	*0.016*	0.394***	*0.025*	0.375***	*0.019*
Age	-0.162***	*0.002*	-0.120***	*0.004*	0.082***	*0.006*	0.059***	*0.005*
Age2	0.037***	*0.001*	0.025***	*0.001*	-0.044***	*0.001*	-0.032***	*0.001*
White	0.115***	*0.004*	0.206***	*0.005*	-0.078***	*0.006*	-0.078***	*0.006*
Married	0.004	*0.004*	0.028***	*0.004*	-0.016**	*0.005*	-0.004	*0.006*
Education	0.059***	*0.001*	0.064***	*0.001*	-0.029***	*0.001*	-0.023***	*0.001*
Employed	0.464***	*0.005*	0.316***	*0.005*	-0.373***	*0.006*	-0.302***	*0.006*
Income/10000	0.061***	*0.001*	0.070***	*0.001*	-0.045***	*0.001*	-0.037***	*0.001*
Redesign	-0.076***	*0.010*	-0.071***	*0.011*	-0.047***	*0.009*	-0.076***	*0.008*
Random Effects								
Cohort								
1899	0.021	*0.017*	0.069**	*0.025*	0.203***	*0.041*	0.123***	*0.031*
1905	-0.015	*0.015*	0.032	*0.021*	0.131***	*0.035*	0.058*	*0.027*
1910	-0.035**	*0.013*	0.052**	*0.019*	0.019	*0.031*	-0.010	*0.024*
1915	0.006	*0.011*	0.031	*0.017*	-0.004	*0.028*	-0.033	*0.022*
1920	0.002	*0.010*	0.015	*0.015*	-0.002	*0.026*	-0.063**	*0.020*
1925	-0.020*	*0.009*	-0.001	*0.014*	-0.001	*0.025*	-0.067***	*0.019*
1930	-0.008	*0.009*	0.000	*0.014*	0.014	*0.024*	-0.069***	*0.019*
1935	0.010	*0.008*	-0.027*	*0.013*	-0.009	*0.023*	-0.028	*0.018*
1940	0.021**	*0.008*	-0.052***	*0.013*	-0.037	*0.023*	-0.002	*0.018*
1945	0.006	*0.007*	-0.072***	*0.012*	-0.063**	*0.023*	-0.019	*0.017*
1950	0.022**	*0.007*	-0.060***	*0.012*	-0.097***	*0.023*	-0.020	*0.017*
1955	0.025***	*0.007*	-0.037**	*0.012*	-0.120***	*0.023*	-0.030	*0.018*
1960	0.014	*0.007*	-0.002	*0.013*	-0.111***	*0.024*	-0.037*	*0.018*
1965	-0.007	*0.008*	-0.007	*0.014*	-0.083**	*0.026*	-0.024	*0.019*
1970	-0.010	*0.009*	-0.006	*0.015*	-0.043	*0.028*	-0.019	*0.021*
1975	-0.037***	*0.011*	-0.004	*0.017*	0.045	*0.030*	0.046*	*0.023*
1980	-0.001	*0.012*	0.014	*0.019*	0.051	*0.033*	0.081**	*0.026*
1985	0.007	*0.015*	0.055*	*0.023*	0.105**	*0.040*	0.112***	*0.032*
Period								
1984	-0.019*	*0.008*	0.002	*0.011*	0.019*	*0.008*	0.009	*0.007*
1985	-0.019*	*0.008*	0.021	*0.011*	0.008	*0.008*	0.001	*0.007*
1986	-0.003	*0.009*	0.022	*0.012*	-0.002	*0.008*	-0.005	*0.008*
1987	-0.014	*0.008*	0.005	*0.010*	-0.005	*0.007*	-0.009	*0.007*
1988	-0.021**	*0.008*	0.003	*0.010*	0.015*	*0.007*	-0.006	*0.007*
1989	0.003	*0.008*	0.016	*0.010*	-0.005	*0.007*	-0.009	*0.007*
1990	0.014	*0.008*	0.029**	*0.010*	-0.005	*0.007*	0.006	*0.007*
1991	0.018*	*0.008*	0.014	*0.010*	-0.014	*0.007*	0.004	*0.007*
1992	0.009	*0.008*	-0.013	*0.010*	-0.007	*0.007*	-0.007	*0.007*
1993	0.009	*0.008*	-0.028**	*0.010*	-0.001	*0.007*	0.010	*0.007*
1994	0.015	*0.008*	-0.016	*0.010*	0.004	*0.007*	-0.001	*0.007*
1995	0.001	*0.008*	-0.035**	*0.011*	-0.008	*0.008*	0.005	*0.007*
1996	0.005	*0.010*	-0.020	*0.012*	0.000	*0.008*	0.001	*0.008*
1997	0.024*	*0.010*	0.028**	*0.010*	-0.002	*0.008*	-0.004	*0.007*
1998	0.021*	*0.010*	0.028**	*0.010*	-0.005	*0.009*	-0.003	*0.007*
1999	0.030**	*0.010*	0.023*	*0.010*	0.001	*0.009*	-0.002	*0.007*
2000	0.016	*0.010*	0.015	*0.010*	0.002	*0.009*	0.002	*0.007*
2001	0.008	*0.010*	0.008	*0.010*	0.001	*0.009*	0.001	*0.007*
2002	0.005	*0.010*	0.001	*0.010*	0.001	*0.009*	0.000	*0.007*
2003	0.009	*0.010*	-0.006	*0.010*	-0.003	*0.009*	0.002	*0.007*
2004	-0.027**	*0.010*	-0.021*	*0.010*	-0.002	*0.009*	-0.004	*0.007*
2005	-0.017	*0.010*	-0.026*	*0.010*	-0.008	*0.009*	0.002	*0.007*
2006	-0.014	*0.011*	-0.011	*0.011*	0.000	*0.009*	-0.001	*0.007*
2007	-0.054***	*0.010*	-0.041*	*0.011*	0.014	*0.009*	0.007	*0.007*
Variance Components								
Cohort	0.0004*	*.0002*	0.002*	*.0008*	0.008*	*.0032*	0.004*	*.0015*
Period	0.0005**	*.0002*	0.001**	*.0002*	0.0001	*.0001*	0.0001	*.0001*
Model Fit								
BIC	1097607		840963					
-2 Res Log Pseudo-Likelihood					1369748		982421	

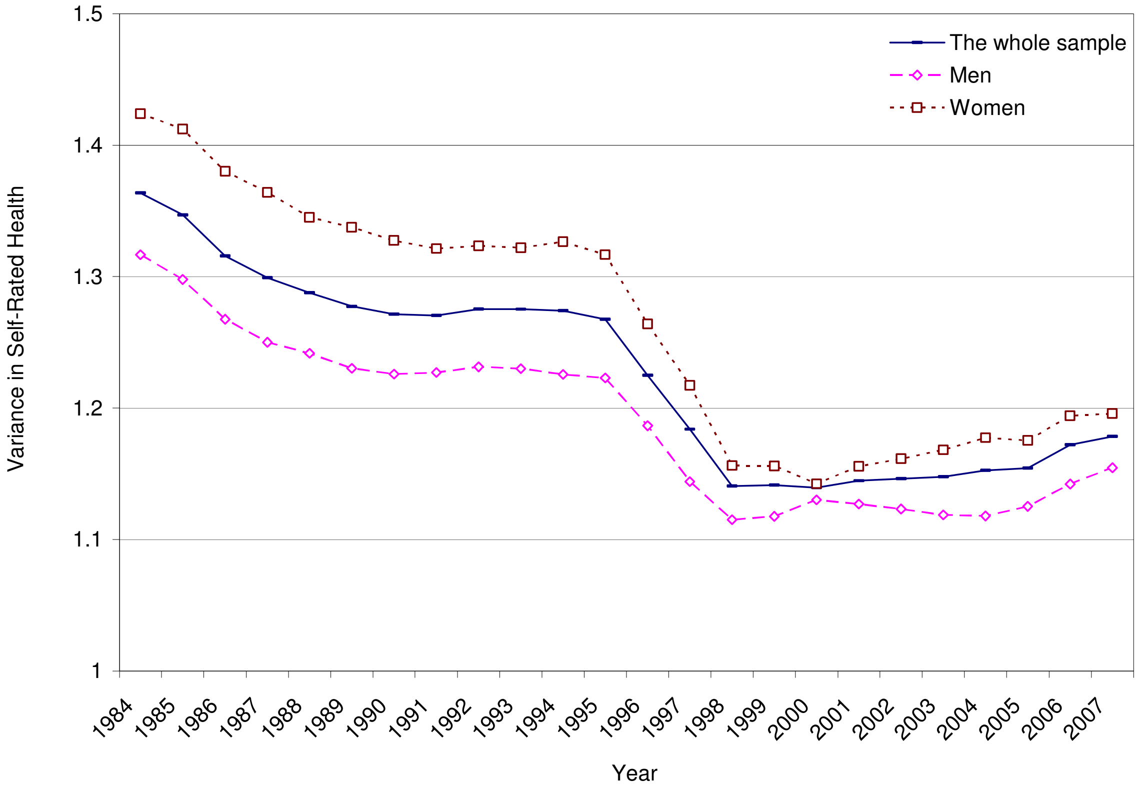
* indicates p<.05; ** indicates p<.01; *** indicates p<.001.

Figure 1. Observed Means of Self-Rated Health, NHIS, 1984 to 2007.



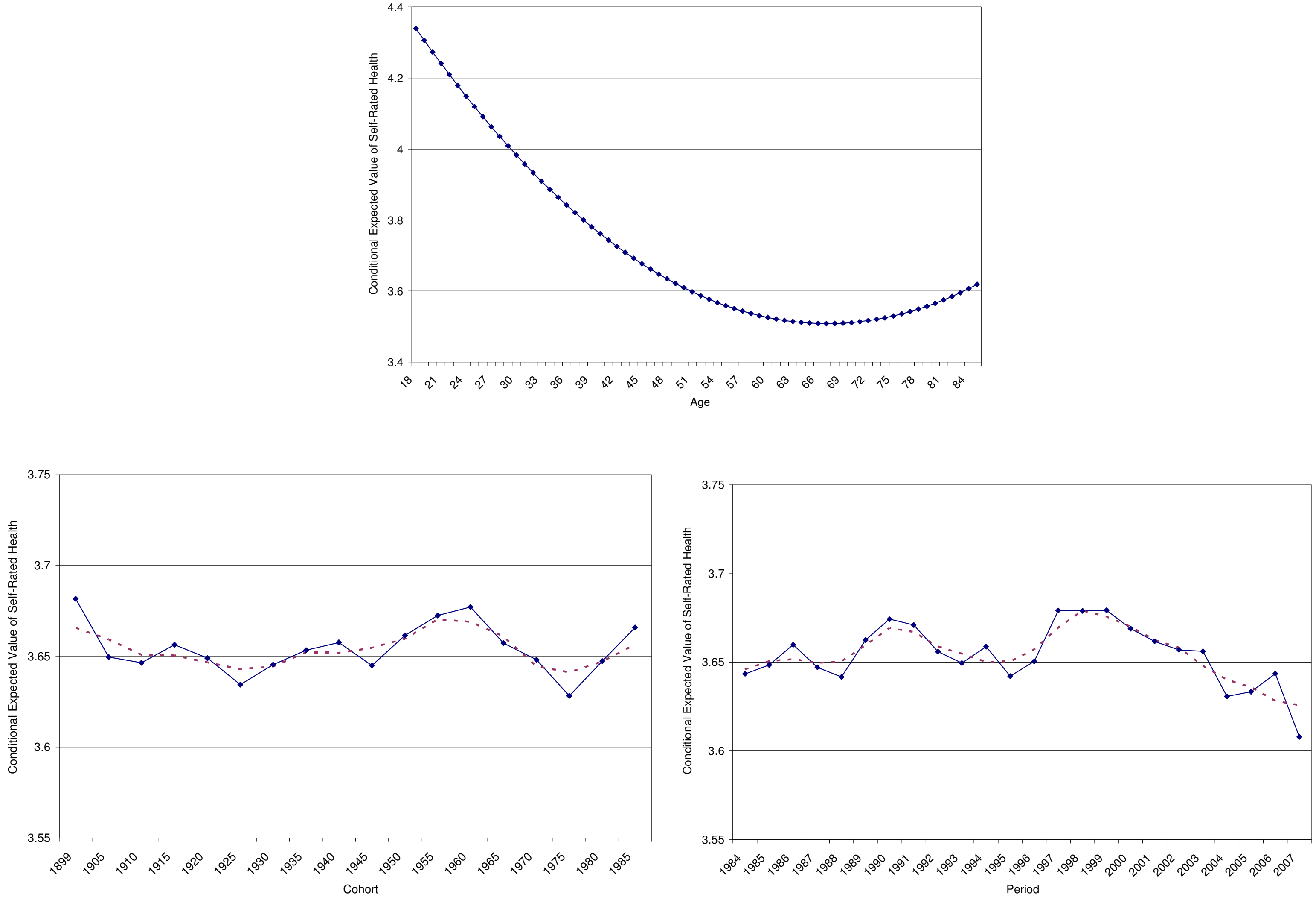
* The trends are adjusted for sample weights and smoothed by a three-point moving average.

Figure 2. Observed Variances in Self-Rated Health, NHIS, 1984 to 2007.



* The trends are adjusted for sample weights and smoothed by a three-point moving average.

Figure 3. Variations in Conditional Expected Values of Self-Rated Health across Age, Cohort and Period.



* The dotted lines in the bottom two figures indicate trends smoothed by a three-point moving average.

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

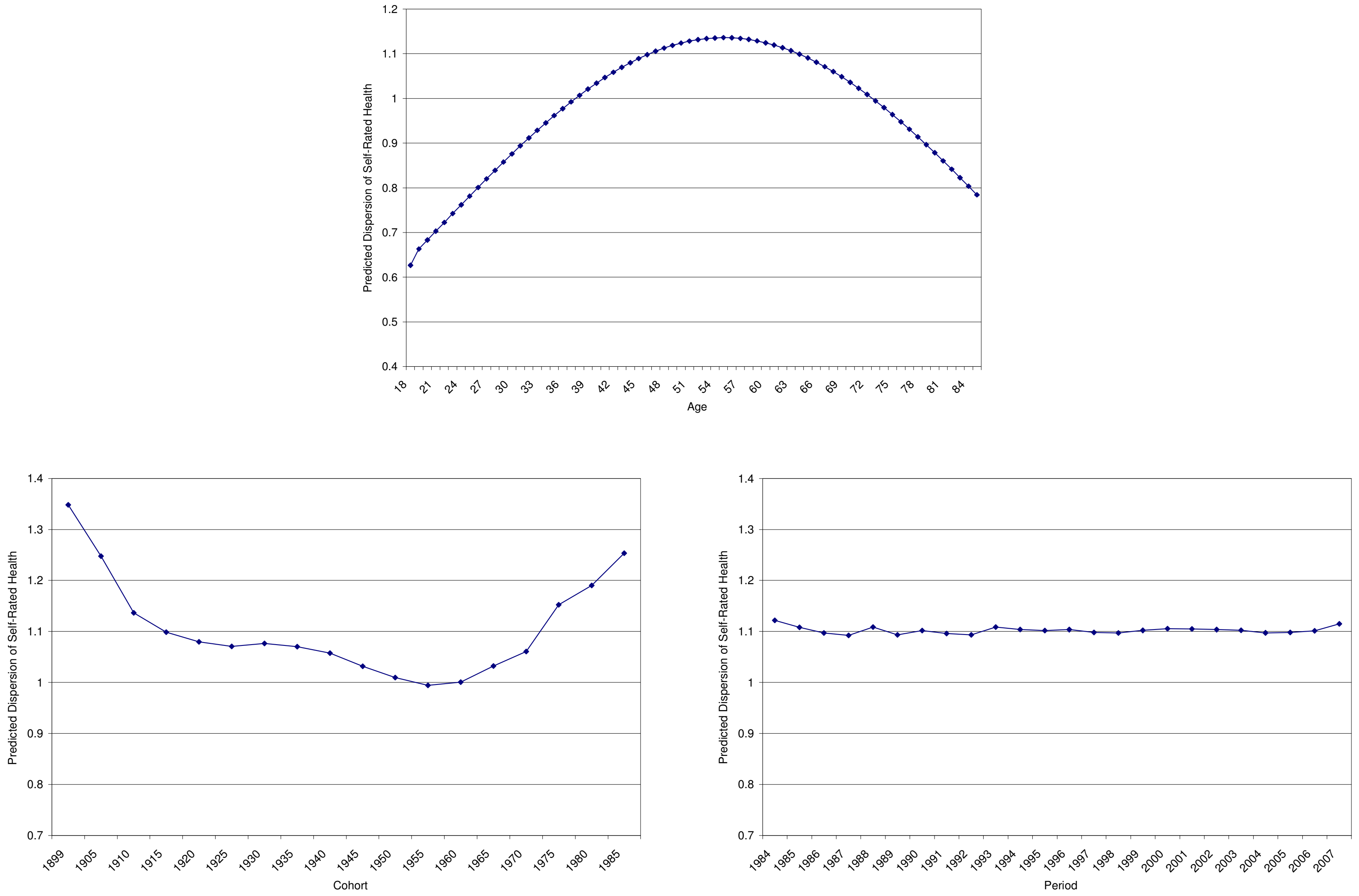


Figure 5. Variations in Conditional Expected Values of Gender-Specific Self-Rated Health across Age, Cohort and Period, with 95% Confidence Intervals.

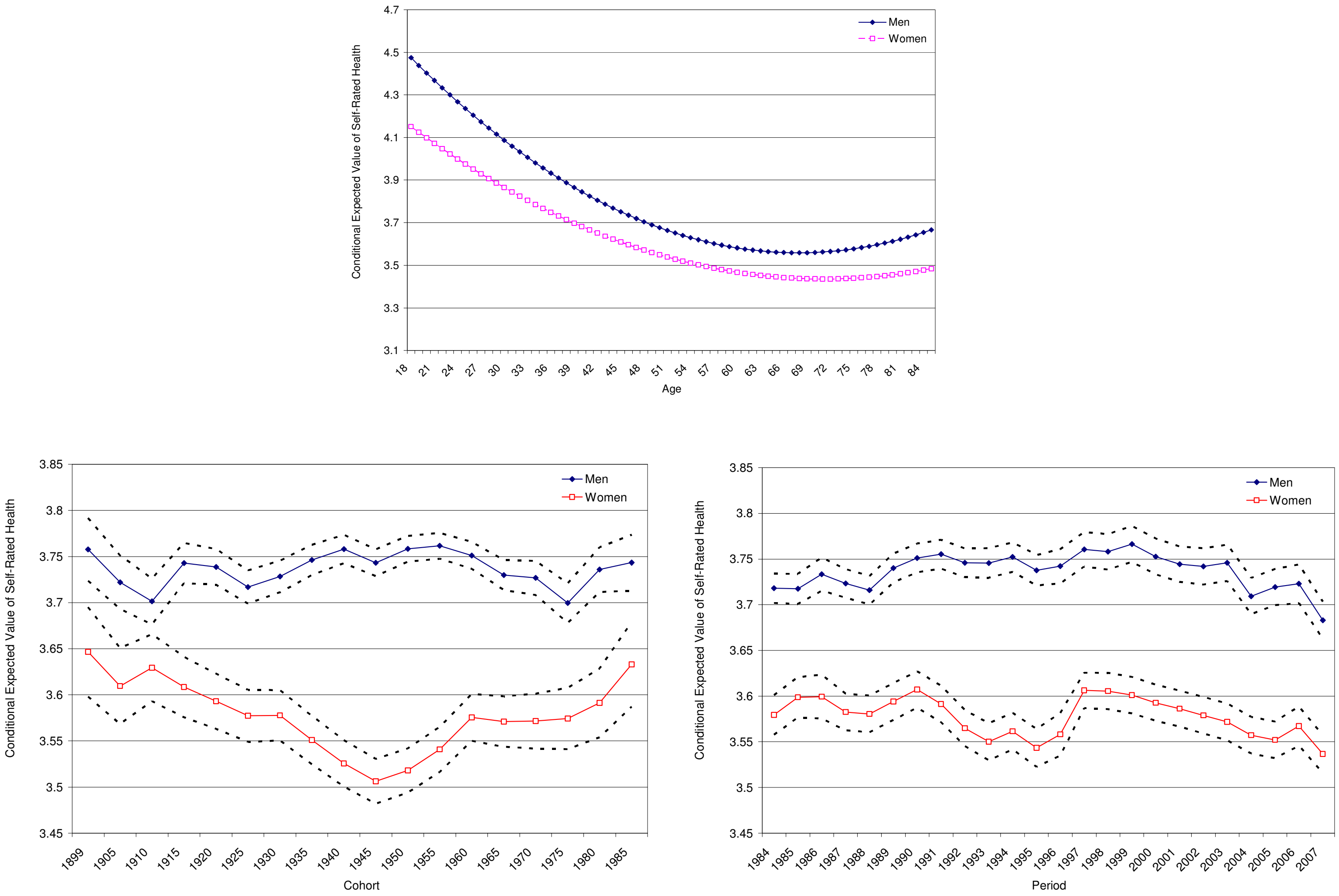


Figure 6. Variations in Predicted Dispersion of Gender-Specific Self-Rated Health across Age, Cohort and Period, with 95% Confidence Intervals.