Trajectories of Subjective Health: Testing Longitudinal Models for Self-rated Health From Adolescence to Midlife

Kenneth A. Bollen and Iliya Gutin

ABSTRACT  Self-rated health (SRH) is ubiquitous in population health research. It is one of the few consistent health measures in longitudinal studies. Yet, extant research offers little guidance on its longitudinal trajectory. The literature on SRH suggests several possibilities, including SRH as (1) a more fixed, longer-term view of past, present, and anticipated health; (2) a spontaneous assessment at the time of the survey; (3) a result of lagged effects from prior responses; (4) a function of life course processes; and (5) a combination of the preceding. Different perspectives suggest different longitudinal models, but evidence is lacking about which model best captures SRH trajectory. Using data from the National Longitudinal Study of Adolescent to Adult Health and the National Longitudinal Survey of Youth, we employ structural equation modeling to correct for measurement error and identify the best-fitting, theoretically guided models describing SRH trajectories. Results support a hybrid model that combines the lagged effect of SRH with the enduring perspectives, fitted with a type of autoregressive latent trajectory (ALT) model. This model structure consistently outperforms other commonly used models and underscores the importance of accounting for lagged effects combined with time-invariant effects in longitudinal studies of SRH. Interestingly, comparisons of this latent, time-invariant autoregressive model across gender and racial/ethnic groups suggest that there are differences in starting points but less variability in SRH trajectories from early life into adulthood.

KEYWORDS  Self-rated health • Trajectories • Longitudinal analyses • Structural equation modeling

Introduction

Self-rated health (SRH) is one of the most common measures in survey research on health. Despite the increasing availability of clinical measures of physiological health, SRH remains valued as a parsimonious survey item that summarizes multiple physical and psychosocial influences on individuals’ overall health (DeSalvo et al. 2006; Jylhä 2009). Given its ease of collection, SRH is often the only consistently available measure of health in longitudinal surveys (Au and Johnston 2014; Boardman 2006). Consequently, numerous studies have leveraged the longitudinal
availability of SRH to model individuals’ trajectories of health over time, yielding foundational studies on life course processes (Shuey and Wilson 2008; Willson et al. 2007). Various sociodemographic factors and health conditions are associated with SRH trajectories, including gender, race, ethnicity, educational attainment, wealth, subjective social status, family structure, religiosity, personality, and various diseases (Bauldry et al. 2012; Berdahl and McQuillan 2018; Brown et al. 2016; Foraker et al. 2011; Hargrove and Brown 2015; Leopold 2019; McCullough and Laurenceau 2004, 2005; McDonough and Berglund 2003; Meadows 2009; Mirowsky and Ross 2008; Shuey and Willson 2008; Takahashi et al. 2018; Willson et al. 2007; Yang and Lee 2009). These same SRH trajectories are used to predict outcomes such as morbidity and premature mortality (Ferraro and Kelley-Moore 2001; Miller and Wolinsky 2000; Stenholm et al. 2016; Wolinsky and Tierney 1998).

Although this large body of research underscores the variety of factors associated with changes in SRH, it has yielded an equally varied set of trajectories. Lacking formal theoretical guidance on the expected trajectories of SRH, researchers make ad hoc decisions about how best to model their longitudinal analyses. Table 1 displays a range of longitudinal models from recent studies, including autoregressive, linear, and other growth trajectories, all providing different conclusions about SRH. We recognize that these studies differ in their study samples; our goal is not to critique extant work but rather to highlight the challenges researchers face when deciding on a modeling strategy. Namely, this choice of model is consequential because it imposes specific assumptions about and constraints on how SRH changes over time.

Moreover, our review suggests that most researchers have not systematically compared these longitudinal models to determine which best fits the data. We find evidence of comparisons within a particular type of model (e.g., testing different polynomial terms in a growth curve) but not across multiple longitudinal model specifications. For instance, including a quadratic term for individuals’ rate of change in SRH in a growth curve improves model fit; however, researchers cannot be confident that the choice to use such a growth model is appropriate to begin with.

The lack of empirical evidence on the appropriateness of SRH trajectory models is especially troublesome when examining individuals across the life course, whose perceptions and experiences of their health may be changing in tandem. Studies of older adults have found a downward trajectory of SRH corresponding with declining health (Liang et al. 2010; Rohlfse and Kronenfeld 2014; Sargent-Cox et al. 2010), but poor health is increasingly observed in early life (Lawrence et al. 2018), suggesting the possibility of significant variation in SRH. To date, few studies have examined trajectories from adolescence to adulthood (Bauldry et al. 2012; Sokol et al. 2017), with mixed conclusions about which model is more appropriate.

Given these mixed findings, how SRH changes from adolescence into adulthood merits greater understanding. More importantly, identifying the appropriate longitudinal model is critical for ensuring an accurate and unbiased understanding of health during this crucial life course transition. Ideally, longitudinal models would be based on existing theories that dictate whether, say, a growth curve or an autoregressive structure best captures the trajectories of SRH. Although we lack such precise guidance, extant research offers a mix of descriptive data and conceptual ideas that provide plausible assumptions that can be incorporated into an examination of how SRH changes over time. The assortment of trajectories summarized in Table 1 is complemented by
<table>
<thead>
<tr>
<th>Study</th>
<th>Models Considered</th>
<th>Final Model</th>
<th>Comparison</th>
<th>Correct for Measurement Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bauldry et al. (2012)</td>
<td>Random effects (quadratic); fixed effects</td>
<td>Random effects (linear)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Berdahl and McQuillan (2018)</td>
<td>Random effects (linear)</td>
<td>Random effects (linear)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Brown et al. (2016)</td>
<td>Random effects (linear, quadratic, cubic)</td>
<td>Random effects (linear)</td>
<td>Yes (likelihood ratio test)</td>
<td>No</td>
</tr>
<tr>
<td>Ferraro et al. (1997)</td>
<td>Autoregressive</td>
<td>Autoregressive</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Foraker et al. (2011)</td>
<td>Random effects (quadratic)</td>
<td>Random effects (quadratic)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Hargrove and Brown (2015)</td>
<td>Random effects (linear)</td>
<td>Random effects (linear)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Kosloski et al. (2005)</td>
<td>Autoregressive</td>
<td>Autoregressive</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Lazar et al. (2007)</td>
<td>Autoregressive</td>
<td>Autoregressive</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Leopold (2019)</td>
<td>Random effects (linear)</td>
<td>Random effects (linear)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Lersch et al. (2018)</td>
<td>Random effects (linear)</td>
<td>Random effects (linear)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Liang et al. (2010)</td>
<td>Random effects (linear, quadratic, cubic)</td>
<td>Random effects (linear)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>McCullough and Laurenceau (2004)</td>
<td>Random effects (linear, quadratic, cubic)</td>
<td>Random effects (cubic)</td>
<td>Yes (Δ deviance)</td>
<td>No</td>
</tr>
<tr>
<td>McCullough and Laurenceau (2005)</td>
<td>Random effects (linear, quadratic, cubic)</td>
<td>Random effects (cubic)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>McDonough and Berglund (2003)</td>
<td>Random effects (linear, quadratic)</td>
<td>Random effects (linear)</td>
<td>Yes; not shown</td>
<td>No</td>
</tr>
<tr>
<td>McDonough et al. (2005)</td>
<td>Random effects (linear)</td>
<td>Random effects (linear)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Meadows (2009)</td>
<td>Random effects (linear)</td>
<td>Random effects (linear)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Miller and Woinsky (2007)</td>
<td>Linear regression</td>
<td>Linear regression</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Mirowsky and Ross (2008)</td>
<td>Random effects (linear)</td>
<td>Random effects (linear)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Study</td>
<td>Models Considered</td>
<td>Final Model</td>
<td>Comparison</td>
<td>Correct for Measurement Error</td>
</tr>
<tr>
<td>------------------------------</td>
<td>----------------------------------------</td>
<td>--------------------------------------------------</td>
<td>---------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Rohlfisen and Kronenfeld (2014)</td>
<td>Random effects (linear)</td>
<td>Random effects (linear)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Sargent-Cox et al. (2010)</td>
<td>Random effects (linear, quadratic)</td>
<td>Random effects (little variation in slope; linear)</td>
<td>Yes (Δ chi-square)</td>
<td>No</td>
</tr>
<tr>
<td>Shuey and Willson (2008)</td>
<td>Random effects (quadratic)</td>
<td>Random effects (quadratic)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Sokol et al. (2017)</td>
<td>Random effects (quadratic)</td>
<td>Random effects (quadratic)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Takahashi et al. (2018)</td>
<td>Random effects (linear)</td>
<td>Random effects (linear)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Wilkinson et al. (2012)</td>
<td>Random effects (linear, quadratic)</td>
<td>Random effects (little variation in slope; linear)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Willson et al. (2007)</td>
<td>Random effects (quadratic)</td>
<td>Random effects (little variation in slope; linear)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Wolinsky et al. (2008)</td>
<td>Mixed effects (fixed + random effects)</td>
<td>Mixed effects (square root of slope)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Yang and Lee (2009)</td>
<td>Random effects (linear, quadratic)</td>
<td>Random effects (quadratic)</td>
<td>Yes; not shown</td>
<td>No</td>
</tr>
<tr>
<td>Yao and Robert (2008)</td>
<td>Random effects (linear, quadratic)</td>
<td>Random effects (quadratic)</td>
<td>Yes (Δ AIC)</td>
<td>No</td>
</tr>
<tr>
<td>Zajacova and Burgard (2010)</td>
<td>Random effects (linear)</td>
<td>Random effects (linear)</td>
<td>No</td>
<td>Yes; intercept and slope only</td>
</tr>
</tbody>
</table>
Theoretical perspectives on SRH as having enduring and spontaneous components that may influence its stability and change over time (Bailis et al. 2003; Boardman 2006; Gunasekara et al. 2012; Jylhä 2009; Perruccio et al. 2010). From another perspective, growth curve models suggest that SRH tracks with changes in the life course; these changes in status, context, and physical health correspond to different ages, which can alter perceptions of SRH (Jylhä 2009; Shuey and Willson 2008; Willson et al. 2007).

To advance longitudinal research on SRH, we begin by reviewing substantive ideas on its longitudinal patterns. We then match these ideas to different longitudinal models—and comparisons across models—to determine which is most consistent with our empirical data on SRH from adolescence to adulthood across two comparable, nationally representative samples. Our goal is to provide a more solid foundation for studies concerned with SRH trajectories; researchers can draw on these results to determine which modeling assumptions are most appropriate for their analyses. In turn, better models provide more precise descriptive representations of trajectories; more accurate causal interpretation of changes in SRH; and most critically, opportunities to refine theory on SRH that can guide future research.

Background

The popularity of SRH in research has produced a notable body of work commenting on the hypothesized processes by which individuals evaluate their health (Bailis et al. 2003; DeSalvo et al. 2006; Jylhä 2009). Although this research does not explicitly touch on issues of longitudinal modeling, it highlights unique properties of stability and change in SRH that may have implications for such models. Consequently, we draw on this literature to identify five plausible perspectives on SRH, which we refer to as (1) enduring, (2) spontaneous, (3) lagged effects, (4) life course, and (5) hybrid models.

The enduring view hypothesizes individuals’ ratings as a reflection of a more stable, potentially lifelong narrative of health informed by past and present experiences, coupled with future expectations (Jylhä 2009). Individuals develop a self-perception of being generally healthy or not healthy, which persists over time (Bailis et al. 2003). As Bombak (2013), Gunasekara et al. (2012), Huisman and Deeg (2010), and Quesnel-Vallee (2007) noted, SRH is not purely a reflection of one’s “true” health status. Instead, it reflects one’s perceptions and how they create “considerable state dependence” in SRH “even after controlling for observed and unobserved confounding” related to changes in one’s health (Gunasekara et al. 2012:1118). This perception could be shaped by comparisons with others in their cohort or reference group, or it could be driven by personality characteristics. Regardless of its source, there is an enduring and constant perception of SRH that influences responses over time.

The spontaneous perspective offers an alternative explanation. As identified by Bailis et al. (2003) and elaborated by others (Boardman 2006; Jylhä 2009; Perruccio et al. 2010), the spontaneous perspective hypothesizes that individuals’ SRH is a reflection of near-term factors influencing their subjective health, occurring at or around the time when they are asked. Bailis et al. (2003:205) noted that “[m]ost research in medical sociology has implicitly adopted the spontaneous assessment perspective” in assuming that SRH directly captures individuals’ current physical and mental health. From a longitudinal perspective, repeated measures of SRH should
be responsive, in the sense that SRH reflects “a response to one’s current state of well-being or illness” (Perruccio et al. 2010:1637); that is, repeated SRH measures are valid snapshots of health over time. Consequently, we would expect SRH to mirror the fluctuations in context, conditions, and mental and physical health that individuals experience, resulting in considerable variability.

The lagged effects hypothesis departs from both the enduring and spontaneous ones in its emphasis on the role of previous SRH on current SRH. This hypothesis does not emphasize a long-term perspective that influences repeated measures of SRH, nor does it see the immediate context and conditions as the predominant influence. Rather, the lagged influence hypothesis identifies past SRH—or one’s recent history (Jylhä 2009)—as a major influence on current perceptions. It reflects stability in perceptions but not over a period of life as long as that suggested by the enduring perspective. To the extent that individuals implicitly compare current health with past health (Norman 2003; Ross 1989), the lagged effects view suggests that the most recent history of SRH is the predominant effect, even if this past assessment differs from the current one.

The life course perspective is the fourth way to view SRH. This perspective recognizes that individuals’ experiences are shaped by the unique circumstances of their age, period, and cohort. As a person moves through stages of life, major events—for example, schooling, marriage, parenthood, work, and many others—can shape SRH. The complex interactions of changes in both health and life circumstances in early life present some difficulty in identifying a single life course trajectory of SRH.

On the one hand, adolescents and young adults are relatively healthy and free of later-life comorbidities (Mulye et al. 2009; Park et al. 2006). Relatively minor downturns in health may not alter perceptions and, thus, limit the extent to which we observe change (Boardman 2006). However, recent evidence suggests that poor health—such as obesity, hypertension, and hyperglycemia—is increasingly prevalent at younger ages (Harris 2010), challenging the notion that physical health is the same for all in early life.

This transition from adolescence into early adulthood is also associated with changes in one’s social context and resources, traversing many critical events, such as completing one’s education, entering the workforce, living independently, and starting a family (Elder et al. 2003; Shanahan 2000). Even if physical health remains stable, this tumultuous and stressful period may affect individuals’ psychosocial health (Burton-Jeangros et al. 2015; Pearlin et al. 2005; Williams and Umberson 2004), which influences reports of subjective health (Jylhä 2009). These momentous events may lead to systematic changes in respondents’ views of their overall health, leading to a more dynamic growth trajectory. Thus, as individuals reassess their subjective health status in light of gained experience and modified expectations, SRH may exhibit a pronounced trajectory over time, indicative of changes in one’s perception of themselves as a healthy individual. Although the life course view anticipates patterns that reflect different experiences throughout the life course, these trajectories are likely to vary from person to person (Elder et al. 2003; Shanahan 2000).

Finally, the enduring, spontaneous, lagged effect, and life course perspectives are not mutually exclusive; individuals’ SRH may be simultaneously informed by all four. Thus, we expect that hybrid models combining multiple perspectives hold promise.
The aforementioned enduring self-concept can coexist with the spontaneous component and lagged effects (Bailis et al. 2003; Boardman 2006; Jylhä 2009), and the mix of these can interact with each other over the life course with various changes in both life events and health. As Jylhä (2010:656) noted, “in their self-ratings people predominantly do focus on their present health, but within the broader context of their health history,” which is “influenced by their earlier health experiences, present health conditions and other health-relevant encounters.” Indeed, although Boardman’s (2006) results demonstrate stability in perceptions of health in adolescence, stratifying the analyses to focus on older respondents suggests a “convergence toward the spontaneous assessment” of SRH as respondents age, whereby SRH is “comprised of both dynamic (spontaneous) and static (enduring) aspects” (p. 407). This result echoes Bailis et al.’s (2003) and Perruccio et al.’s (2010) findings that both processes are likely at play, influencing individuals’ assessments of changes in subjective health over time.

**Hypothesized Trajectories and Longitudinal Models**

As reviewed earlier, several theoretical perspectives are common when it comes to examining subjective health over time. Although these perspectives do not offer explicit recommendations about the most appropriate trajectory, we draw on their insights to develop plausible longitudinal models.

For instance, we invoke Bailis et al.’s (2003) concept of an enduring self-concept of subjective health in developing the enduring hypothesis, which suggests considerable stability in SRH such that individuals’ perceptions of their health do not exhibit much change even if their health status worsens or improves. This perspective holds that perceptions are distinct from an individual’s true health, and we should not expect direct correspondence between the two. Moreover, individuals’ responses may exhibit stability because of a propensity to evaluate one’s health relative to those in their cohort; indeed, this comparative framing provides more consistent responses than alternative interpretations of SRH (Eriksson et al. 2001).

A model with a latent time-invariant variable captures important aspects of the *enduring hypothesis*:

\[
L_{i,t} = \alpha_t + \xi_i + \epsilon_{i,t},
\]

where \(L_{i,t}\) is the latent subjective health variable for individual \(i\) and time \(t\); \(\alpha_t\) is the equation intercept for time \(t\); \(\xi_i\) is a time-invariant variable that endures throughout the period; and \(\epsilon_{i,t}\) is the random error that varies over person and time, with a mean of 0 and uncorrelated with \(\xi_i\). The model allows for differences in the intercept over time that hold for all respondents. The \(\xi_i\) variable corresponds to the idea of stability under the enduring hypothesis.

The *lagged effect hypothesis* places importance on previous SRH influencing current values. This is captured by the following equation:

\[
L_{i,t} = \alpha_t + \rho_{i,t-1} L_{i,t-1} + \epsilon_{i,t},
\]

where the new term \(L_{i,t-1}\) is the lagged value of subjective health; \(\rho_{i,t-1}\) is the autoregressive regression (AR) coefficient that estimates the magnitude of lagged effects;
and \( \epsilon_{ij,t} \) is the random error that varies over person and time; with a mean of 0 and uncorrelated with \( L_{ij,t-1} \). The model suggests that individuals’ current subjective health is primarily driven by their previous assessment, without any underlying *enduring* component.

The *life course hypothesis* suggests that individual SRH follows a trajectory over time that reflects their life experiences, with individuals having different starting points and rates of change. A latent growth curve model captures aspects of this perspective:

\[
L_{ij,t} = \alpha_i + \lambda_t \beta_i + \epsilon_{ij,t},
\]

(3)

where \( \alpha_i \) is the random intercept; \( \beta_i \) is the random slope; \( \lambda_t \) is the time trend variable that counts the number of years since the data series started; and \( \epsilon_{ij,t} \) is the random error, with a mean of 0 and no correlation with \( \alpha_i \) and \( \beta_i \). The random intercepts (\( \alpha_i \)) are the different starting points for each individual; similarly, the random slopes (\( \beta_i \)) differ across individuals so as to permit different rates of change in their life course process. Thus, differences in both intercepts and slopes support a life course perspective. We also test nonlinear parameterizations of this model because the rate of change may exhibit more complex forms.

The *spontaneity hypothesis* is not easily represented in a single model. If spontaneity is viewed as an unpredictable and constantly changing view of SRH (i.e., moment-to-moment changes [Gunasekara et al. 2012; Perruccio et al. 2010]), we would not expect any systematic influence on SRH, as in the earlier models. We contend that the error variable, present in the preceding models, best represents evidence for the spontaneity hypothesis. Namely, we would expect to find very low \( R \)-squared values for the preceding equations because the unpredictability of latent subjective health would not be captured by any of the preceding models.

The enduring, lagged effect, and life course hypotheses all hold promise, and models associated with these perspectives have been used in past research. However, our review of the literature suggests that the integration of multiple perspectives might provide better explanations than any one of them alone. For example, the enduring and the lagged effect hypotheses might operate simultaneously, as might the life course and lagged effect perspectives in offering a more appropriate trajectory. None of the preceding models capture these hybrids, but alternative models allow for such integrative hypotheses.

For instance, one hybrid model is the latent variable-autoregressive latent trajectory (LV-ALT) model (Bianconcini and Bollen 2018):

\[
L_{ij,t} = \alpha_i + \lambda_t \beta_i + \rho_{t-1} L_{ij,t-1} + \epsilon_{ij,t},
\]

(4)

As before, the life course hypothesis is reflected in the first two terms (i.e., the latent intercept and slope), and the third term represents the lagged effect hypotheses and the inclusion of autoregressive terms. Similarly, we can combine the enduring and lagged effect hypotheses (Bollen and Brand 2010):

\[
L_{ij,t} = \alpha_i + \rho_{t-1} L_{ij,t-1} + \xi_j + \epsilon_{ij,t}.
\]

(5)

As before, we can modify these hybrid models to account for nonlinear latent slopes. Thus, we are left with multiple plausible and compelling hypotheses about
which model might best capture the trajectory of SRH. Given that these models are supported in one form or another by extant theory and that substantive knowledge does not provide a clear sense of which is best, we need to compare them systematically to determine which is most appropriate for our longitudinal data.

Before turning to our data and models, we note the importance of controlling for measurement error in SRH. Extant research on the test-retest reliability of SRH suggests that its reliability is modest at best (Boardman 2006; Crossley and Kennedy 2002; Fosse and Haas 2009; Zajacova and Dowd 2011), yet few studies have taken this substantial measurement error into consideration when modeling longitudinal change (Kosloski et al. 2005). The preceding models include the latent variable of SRH ($L_t$), free of measurement error, whereas leaving SRH as “observed” and not correcting measurement error can affect multiple longitudinal properties of interest to our study, such as assessments of SRH’s enduring properties as well as the autoregressive coefficients. Consequently, correcting for measurement error leads to an unbiased assessment of how SRH influences other variables (including SRH) in longitudinal models where SRH is included as a predictor/covariate.

Methods

Data

Data for this study come from the National Longitudinal Study of Adolescent to Adult Health (Add Health), a nationally representative survey of adolescents (in grades 7–12) who were interviewed in school and in person in 1994–1995 (Wave I), 1996 (Wave II; grades 8–12), 2001–2002 (Wave III; aged 18–26), 2008 (Wave IV; aged 24–32), and 2016–2018 (Wave V; aged 32–42) (Harris et al. 2019). At all waves, respondents were asked, “In general, how is your health?,” with response options of excellent (1), very good (2), good (3), fair (4), and poor (5). Responses were reverse-coded and treated as ordinal indicators of latent subjective health. A distinct advantage of Add Health is the availability of two near-contemporaneous measures of SRH at Wave I. These measures are separated by an average of six months, allowing us to estimate individuals’ baseline reliability of this measure and to compare estimates of reliability at subsequent waves. The final analytic sample consists of 12,300 respondents across all five waves.

Aware of mixed findings in past research, we want to be confident that our results are not specific to Add Health and that we would observe a similar pattern in a comparable sample. The National Longitudinal Survey of Youth 1997 (NLSY97) provides an ideal data set for such a comparison while permitting additional flexibility in our models. NLSY97 data are nationally representative, the cohort is similar (ages of 12–17 in 1997), and the sample size of 8,984 is large. NLSY97 respondents were surveyed annually between 1997 and 2011 and biennially until 2015. These 17 waves of data help account for any effects introduced by the unequal spacing between waves in Add Health. Thus, in identifying the most appropriate models, we draw on results from both data sets to avoid biases arising from differences in survey design.

Our analyses are adjusted for individuals’ ages at the first wave because these differences might affect initial SRH and subsequent trajectories. We also incorpo-
rate information on gender and race/ethnicity when assessing group differences in
the appropriateness of these models. We categorize race/ethnicity as non-Hispanic
White, non-Hispanic Black, and Hispanic to be consistent across the two samples
and to focus on racial/ethnic groups that are large enough to yield stable estimates.

Given our specification of self-rated health as an ordinal variable (detailed later), we
use Mplus’s diagonally weighted least squares (DWLS) estimator with its missing data
procedures, as described in Asparouhov and Muthén (2010). In brief, this approach uses
a multistage procedure in which a maximum likelihood method estimates univariate
and bivariate statistics for all available cases; these means, variances, and covariances
are then assembled as input into the DWLS procedure (for details, see Asparouhov
and Muthén 2010). We use the R structural equation modeling (SEM) package lavaan
(Roseel 2012) and Mplus to estimate our models (Muthén and Muthén 1998–2017).
An annotated R script with selected model syntax is available in the online appendix.

Analytic Strategy

The goal of our study is to find the best longitudinal models for SRH consistent with
extant theory. Assessing the fit statistics of different models will allow us to iden-
tify assumptions that best describe longitudinal changes in SRH. SEM has a number
of fit statistics. We use chi-square tests and other measures to assess how closely
the hypothesized models fit the Add Health and NLSY97 data. Both data sets have
very large sample sizes, so the statistical power of testing is generally high; even
minor specification errors could lead to statistically significant chi-square tests. To
supplement the chi-square test, we also use a Bayesian information criterion (BIC)
comparison statistic that compares the fit of the saturated and hypothesized models
and approximates the Bayes factor. Negative values of this BIC comparison statistic
provide evidence favoring the hypothesized over the saturated model (Raftery
1995).1 The comparative fit index (CFI), Tucker-Lewis index (TLI), and 1 minus the
root mean square error of approximation (RMSEA) are other common fit statistics
(Bentler 1990; Steiger and Lind 1980; Tucker and Lewis 1973). Across all three, val-
ues closer to 1 represent better fit, and values less than .9 are considered inadequate.

In treating SRH as an ordinal variable, we fix the first two response thresholds at
1 and 2, respectively;2 intriguingly, the freely estimated third and fourth thresholds in
all models are close to 3 and 4. Although this is evidence that could support analyzing

1 We calculate the BIC = (chi square) – (degrees of freedom) × (natural log of sample size) to enable this
interpretation. Through the remainder of the study, we refer to these BIC comparisons of the saturated and
hypothesized model as “BIC” to avoid confusion with BIC differences across nested models, discussed
later.

2 We set the first two thresholds of the ordinal SRH measures to 1 and 2 for several reasons. First, this
permits us to estimate the means and variances of the continuous variables underlying the ordinal SRH
measures, which is helpful for interpreting the models. Second, assuming that the first two thresholds are
the same over time is more substantively plausible than assuming that the means or variances of the under-
lying variables are constant over time. Third, by setting the first two thresholds to 1 and 2, we can estimate
the remaining two thresholds to see how close they are to the ordinal codings of 3 and 4; indeed, they are
close for all waves of the SRH measures. This helps to explain why treating these ordinal variables as if
they were continuous leads to results quite similar to the analysis in which we take account of their ordinal
SRH as continuous (Fisher and Bollen 2020), we follow convention in treating SRH as ordinal with underlying continuous variables. Furthermore, we allow for random measurement error in these underlying indicators of SRH (Boardman 2006; Crossley and Kennedy 2002; Fosse and Haas 2009; Zajacova and Dowd 2011). For most years, we have only a single indicator of SRH. Thus, we assume that the variance of the random measurement error is the same over all waves of data to help to identify the model (see Heise 1969; Werts et al. 1971; Wiley and Wiley 1970).

Figures 1–3 display path diagrams of models that incorporate the different properties of SRH trajectories in Add Health. One can imagine corresponding diagrams for NLSY97 containing additional time points but only one measure at the first observation. Panel a of Figure 1 represents a hypothesized model in which the enduring

coding. See Jöreskog (2005), Wu and Estabrook (2016), and Fisher and Bollen (2020) for more discussion of these alternative methods to code ordinal variables.
Fig. 2 Growth curve models. To simplify the diagrams, we do not show the continuous underlying variables that correspond to the ordinal $Z_j$ to $Z_5$ variables. The $\varepsilon$s refer to the errors of the continuous underlying variables, as do the correlations of the errors. $L_j$ = Latent subjective health at wave $j$. $Z_j$ = measures of self-reported health at wave $j$: $j = 1, 2, 3, 4, 5$. In-School = in-school self-reported health, measured at Wave I. $\alpha$ = latent intercept. $\beta$ = latent slope. $(x)$ = freely estimated loading factor with $x$ starting value. Latent variables for subjective health, intercept, and slopes are influenced by age at Wave I.
Fig. 3 Autoregressive latent trajectory (linear) growth model. To simplify the diagrams, we do not show the continuous underlying variables that correspond to the ordinal $Z_1$ to $Z_5$ variables. The $\varepsilon$s refer to the errors of the continuous underlying variables, as well as the errors for the latent variables. $L_j =$ latent subjective health at wave $j$. $Z_j =$ measures of self-reported health at wave $j; j = 1, 2, 3, 4, 5$. In-School = in-school self-reported health, measured at Wave I. $\alpha =$ latent intercept. $\beta =$ latent slope. Latent variables for subjective health, intercept, and slopes are influenced by age at Wave I.

perception of SRH corresponds with Eq. (1). The latent time-invariant variable intercept plays a role in determining SRH in each wave. Variability around this enduring perception enters via possible changes in the constant intercept of each wave and in the error term for latent subjective health. By contrast, panel b of Figure 1 is an autoregressive longitudinal model; trajectories of subjective health ($L_i$) are a function of the lagged effect of SRH ($L_{i-1}$) from one wave to the next.

In Figure 2, the linear growth (panel a), quadratic growth (panel b), and freed loading growth (panel c) correspond to possible linear and nonlinear trends in SRH. Each of these growth curve models allows individuals different starting points (i.e., intercepts) and rates of change (i.e., slopes) in SRH. The preassigned loadings on the slope for the linear and quadratic growth models correspond with the number of years between waves (Bollen and Curran 2006).4

Finally, Figure 3 corresponds with the LV-ALT model, which includes an additional autoregressive component in the commonly used growth curves shown earlier (Bollen and Curran 2004). The LV-ALT model permits SRH at the previous point in time to influence current SRH at the same time that its trajectory is partly governed by the intercepts and slopes (Bollen and Curran 2004:346). One can interpret these

---

3 This model is similar to the commonly used fixed-effects model, although it does not include any time-varying determinates of the repeated variable. Our emphasis is on the time-invariant nature of the variable in that it closely corresponds to the idea of an enduring influence. See Bollen and Brand (2010) for modeling fixed effects in a structural equation model.

4 We use one-fifth of a year in the quadratic models to reduce the magnitude of the slope parameters and help with model convergence.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>12,300</td>
<td>3.910 (0.918)</td>
<td>3.911 (0.880)</td>
<td>3.923 (0.885)</td>
<td>3.995 (0.857)</td>
<td>3.708 (0.915)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>6,973</td>
<td>3.785 (0.907)</td>
<td>3.827 (0.889)</td>
<td>3.818 (0.892)</td>
<td>3.897 (0.863)</td>
<td>3.642 (0.919)</td>
</tr>
<tr>
<td>Male</td>
<td>5,327</td>
<td>4.049 (0.910)</td>
<td>4.004 (0.861)</td>
<td>4.042 (0.862)</td>
<td>4.106 (0.835)</td>
<td>3.783 (0.905)</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>6,910</td>
<td>3.907 (0.902)</td>
<td>3.925 (0.869)</td>
<td>3.924 (0.874)</td>
<td>4.025 (0.834)</td>
<td>3.768 (0.892)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>2,415</td>
<td>3.945 (0.971)</td>
<td>3.894 (0.932)</td>
<td>3.969 (0.914)</td>
<td>3.969 (0.904)</td>
<td>3.580 (0.940)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1,773</td>
<td>3.875 (0.961)</td>
<td>3.848 (0.872)</td>
<td>3.900 (0.896)</td>
<td>3.931 (0.929)</td>
<td>3.521 (0.946)</td>
</tr>
</tbody>
</table>

Notes: N = 12,300; N = 11,098 in race/ethnicity–specific models. For these descriptive statistics, we treat self-rated health as continuous. Means use Wave V cross-sectional survey weights and casewise maximum likelihood (also called FIML) estimation to account for missing data on self-rated health (Arbuckle et al. 1996).

models as indicative of both processes net of each other; we can assess the presence and strength of a growth curve after accounting for the lagged effect of SRH, and vice versa (Bianconcini and Bollen 2018:805–806). Figure 3 serves as a general LV-ALT framework that we modify based on the additional growth curves included in our analyses.

After assessing the best-fitting longitudinal models, we conduct tests of invariance for this model across gender and race/ethnicity. Research has documented differences between men and women on SRH (Benyamini et al. 2003; McCullough and Laurenceau 2004; Rohlfisn and Kronenfeld 2014; Zajacova et al. 2017) and across racial/ethnic groups (Ferraro et al. 1997; Franks et al. 2003; Hargrove and Brown 2015; Liang et al. 2010; Shuey and Willson 2008; Willson et al. 2007; Yao and Roberts 2008). Although we do not have explicit hypotheses about the nature of group differences given the mixed results of past research, this literature suggests that variation in model structure and parameter estimates should be assessed. The extent to which we observe invariance across groups has implications for group-specific longitudinal models; one can compare estimates only if there is sufficient equivalence in the model across groups (Cheung and Rensvold 2002).

Finally, given that our goal is to maximize the sample size for our analyses, we do not include longitudinal weights, which reduce the sample size by ∼40%. We checked whether the exclusion of these weights may have biased our analyses. Model fit is comparable when we included the survey weights, which suggests that our unweighted estimates do not bias conclusions regarding the most appropriate SRH trajectory.
Results

Our first step is to obtain a descriptive understanding of SRH over time. Table 2 displays SRH means and variances across the five waves of Add Health, exhibiting relative stability from Waves I through III (μ≈3.9)—that is, adolescence and young adulthood—and then a small average decline at Wave IV (μ≈3.71) and into Wave V (μ≈3.54), corresponding with ages 32–42. Figure 4 shows this mean over time, overlaid with six randomly chosen cases to demonstrate the variability in the sample.

This pattern is similar for female and male respondents, although male respondents’ SRH is approximately 0.1 to 0.2 higher for Waves I–III (μ≈4.0) and then converges with female respondents’ SRH again by Waves IV (μ≈3.7) and V (μ≈3.5). Similar patterns hold across racial/ethnic groups, with SRH stable at Waves I and II, peaking at ~4.0 at Wave III, and then showing small declines across Waves IV and V. White and Black respondents have slightly higher SRH at Waves I and II (μ≈3.93) than their Hispanic counterparts (μ≈3.87); by Waves IV and V, White adults have the highest SRH, at 3.7, whereas Black and Hispanic adults have an SRH closer to 3.5.

Table 3 summarizes fit statistics for longitudinal models corresponding with the enduring, lagged effect, and life course hypotheses. We consider five models: a latent time-invariant model, an autoregressive model, and three variants of a growth model with different slopes. Although we observe good measures of fit for multiple models, the freed loading growth curve is the best. The CFI, TLI, and 1-RMSEA values are close to 1 and are generally higher than other models. The BIC of −75.969 is negative.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>BIC</th>
<th>CFI</th>
<th>TLI</th>
<th>1-RMSEA</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latent Time-Invariant</td>
<td>2,014.645</td>
<td>15</td>
<td>1,873.385</td>
<td>0.928</td>
<td>0.899</td>
<td>0.896</td>
<td>Negative error variance for subjective health</td>
</tr>
<tr>
<td>With Freely Estimated Intercepts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LV at Wave II</td>
</tr>
<tr>
<td>Autoregressive (AR)</td>
<td>57.817</td>
<td>12</td>
<td>−55.191</td>
<td>0.998</td>
<td>0.997</td>
<td>0.982</td>
<td>—</td>
</tr>
<tr>
<td>Linear Growth Curve</td>
<td>274.902</td>
<td>16</td>
<td>124.224</td>
<td>0.991</td>
<td>0.988</td>
<td>0.964</td>
<td>—</td>
</tr>
<tr>
<td>Quadratic Growth Curve</td>
<td>48.552</td>
<td>12</td>
<td>−64.456</td>
<td>0.999</td>
<td>0.998</td>
<td>0.984</td>
<td>Negative error variance for subjective health</td>
</tr>
<tr>
<td>Freed Loading Growth Curve</td>
<td>46.457</td>
<td>13</td>
<td>−75.969</td>
<td>0.999</td>
<td>0.998</td>
<td>0.986</td>
<td>Negative error variance for subjective health</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LV at Wave II</td>
</tr>
</tbody>
</table>

Notes: $N=12,300$. Results are based on the DWLS estimator with its missing data procedures, as described in Asparouhov and Muthén (2010). All models control for age at the first wave, and observed SRH variances are constrained to be equal.

and high in absolute magnitude, exceeding the guidelines for “very strong” evidence of good model fit compared with the saturated model (Raftery 1995). Although other models also have a negative BIC (e.g., autoregressive and quadratic growth), the BIC magnitude is lower than the freed loading model. However, the good fit for the autoregressive model—without any negative error variances—suggests that hybrid models accounting for the lagged effect perspective are worth investigating.

Table 4 contains the LV-ALT models that integrate the lagged effect and life course perspectives. All three models exhibit good fit across the CFI, TLI, and 1-RMSEA. Only the LV-ALT linear and quadratic growth models have negative BICs (−45.035 and −9.237). However, they have negative autoregressive coefficients, and neither is an improvement over the freed loading growth curve model. Interestingly, evidence of nonsignificant slopes in the LV-ALT quadratic growth model suggests that we can improve the fit by estimating a model with only a latent intercept and autoregressive relations.

We estimate two such models in Table 5, corresponding with the view of subjective health as having both enduring and lagged effect components, whereby individuals have a persistent view of their subjective health complemented by a lingering effect that carries over all waves. Latent subjective health at Wave I is predetermined and correlated with the latent intercept. The LV-ALT intercept-only model assumes individual differences in starting points for the trajectory from Wave II to V, with latent subjective health predicted by the lagged autoregressive effects and time-invariant intercept and no variability in the constant of the intercept at each wave. Conversely, the latent time-invariant with AR model allows for freely estimated intercepts for latent subjective health at Waves II to V, net of the time-invariant intercept and autoregressive regressions. Both models have good CFI, TLI, and 1-RMSEA values; they also have large and negative BICs. However, based on Raftery’s (1995) guidelines for adjudicating between nested models, the difference of −19 strongly favors the LV-ALT intercept-only model.
The *best* model in the Add Health data is somewhat ambiguous: the freed loading growth curve and LV-ALT intercept-only models are comparable in fit. The slightly higher negative BIC favors the freed loading growth curve model, but the difference of −7 compared with the LV-ALT intercept-only model is not definitive. Given this uncertainty, we turn to the NLSY97 data to see how well these same models fit this sample.

Table 6 presents a summary of the model fit statistics across the same 10 models. Many of the models have good fit based on CFI, TLI, and 1-RMSEA; thus, we rely on the BIC to differentiate among models. Similar to results based on Add Health data, the autoregressive, LV-ALT linear and quadratic growth, and LV-ALT intercept-only and latent time-invariant with AR models fit well, with large and negative BICs ranging from $-729.703$ (autoregressive) to $-967.837$ (LV-ALT intercept only). By contrast, the quadratic and freed loading growth curve models using NLSY97 data exhibit worse fit and are improved with the inclusion of lagged effects. Critically, the NLSY97 data are less ambiguous about the choice of the best model: the BIC for the LV-ALT intercept-only model is 27 points lower than that for the next best-fitting model (LV-ALT linear growth).

Given this outcome, the superior fit of the LV-ALT intercept-only model encourages us to interpret the resulting coefficients. We are primarily interested in the autoregressive terms, providing a sense of how much SRH at one time influences subsequent SRH. As shown in Tables A1 and A2 (online appendix), we find a strong association across most waves. Autoregressive values above 0.8 indicate a high degree of lagged variable influence (i.e., “stability”; Kosloski et al. 2005). Although none of the autoregressive coefficients in Add Health reach this threshold, these are estimates net of the latent intercept. Nevertheless, we observe autoregressive coefficients close to 0.7; the only deviation from this pattern is between Waves II and III, where the value of 0.584 is still relatively high. This lower value is likely attributable to the differing number of years between waves. Indeed, when we examine the NLSY97 data, which are separated by one to two years, the autoregressive coefficients are higher and closer in value, at an average of 0.844.
We are also interested in $R$-squared; these values indicate the reliability of the underlying SRH measures as well as how much of the variation in the latent subjective health (L-SRH) is explained across waves. The reliability is modest in both Add Health and NLSY97 ($\sim .6$), indicating considerable measurement error in SRH. The LV-ALT intercept-only model also explains approximately one-half the variation in latent subjective health in Add Health. Although we do not have an $R$-squared estimate at Wave II because of a small and negative but not statistically significant variance, L-SRH at Wave II and age at Wave I explain 48% of the variation in L-SRH at Wave III. The age regressions and autoregressive relationships between L-SRH at Wave III to Wave IV and Wave IV to Wave V explain, respectively, 55% and 62% of the variation in those measures. Again, this pattern is likely explained by the different number of years between waves, as we can see when looking at NLSY97. In these data, $\sim 87\%$ of the variation in the L-SRH measures is explained by the model structure, which is not surprising given the larger autoregressive coefficients and fewer years separating observations. One can interpret the lower autoregressive coefficients and lower $R$-squared for L-SRH in Add Health as indicative of the autoregressive effect gradually diminishing over time. For example, in the well-fitting autoregressive model (not shown), 92% of the variance in L-SRH at Wave II is explained by the autoregressive relationship with Wave I: they are separated by only 1.5 years.

Finally, we assess whether the LV-ALT intercept-only model is equally applicable for both female and male respondents as well as across racial and ethnic groups. We fit a series of increasingly constrained nested models, modifying the model structure to hold parameters equal across groups. This series of constraints loosely follows the weak to strict invariance structure traditionally used in SEM, but we make modifications in keeping with our model structure. Tracking changes in model fit demonstrates whether these equality constraints are appropriate and thus helps identify similarity in the trajectory across groups.\footnote{Remember that the large sample sizes enhance statistical power and the probability of detecting substantively minor differences.}

Table A3 (online appendix) summarizes the invariance tests between female and male respondents in Add Health, finding good model fit even when constraining multiple parameters to equality. The best-fitting model, model E, constrains all the regres-

<table>
<thead>
<tr>
<th>Model</th>
<th>χ²</th>
<th>df</th>
<th>BIC</th>
<th>CFI</th>
<th>TLI</th>
<th>1-RMSEA</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latent Time-Invariant With</td>
<td>10,905.672</td>
<td>150</td>
<td>9,540.192</td>
<td>0.922</td>
<td>0.920</td>
<td>0.911</td>
<td>Age regressions constrained for convergence; negative variances for LVs</td>
</tr>
<tr>
<td>Freely Estimated Intercepts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autoregressive (AR)</td>
<td>353.578</td>
<td>119</td>
<td>−729.703</td>
<td>0.998</td>
<td>0.998</td>
<td>0.985</td>
<td></td>
</tr>
<tr>
<td>Linear Growth Curve</td>
<td>2,780.945</td>
<td>147</td>
<td>1,442.775</td>
<td>0.981</td>
<td>0.980</td>
<td>0.955</td>
<td>Negative variances for LVs; slope not significant</td>
</tr>
<tr>
<td>Quadratic Growth Curve</td>
<td>1,698.567</td>
<td>143</td>
<td>396.809</td>
<td>0.989</td>
<td>0.988</td>
<td>0.965</td>
<td>Negative variances for LVs</td>
</tr>
<tr>
<td>Freed Loading Growth Curve</td>
<td>2,797.066</td>
<td>132</td>
<td>1,595.444</td>
<td>0.981</td>
<td>0.978</td>
<td>0.953</td>
<td>Negative variances for LVs</td>
</tr>
<tr>
<td>LV-ALT Freed Loading Growth</td>
<td>212.863</td>
<td>114</td>
<td>−824.902</td>
<td>0.999</td>
<td>0.999</td>
<td>0.990</td>
<td>Mean, variances, and covariances of quadratic slope nonsignificant</td>
</tr>
<tr>
<td>LV-ALT Linear Growth</td>
<td>224.678</td>
<td>128</td>
<td>−940.532</td>
<td>0.999</td>
<td>0.999</td>
<td>0.991</td>
<td>Mean, variances, and covariances of quadratic slope nonsignificant</td>
</tr>
<tr>
<td>LV-ALT Quadratic Growth Curve</td>
<td>215.460</td>
<td>123</td>
<td>−904.234</td>
<td>0.999</td>
<td>0.999</td>
<td>0.991</td>
<td></td>
</tr>
<tr>
<td>LV-ALT Intercept Only</td>
<td>233.785</td>
<td>132</td>
<td>−967.837</td>
<td>0.999</td>
<td>0.999</td>
<td>0.991</td>
<td></td>
</tr>
<tr>
<td>Latent Time-Invariant With AR</td>
<td>213.755</td>
<td>117</td>
<td>−851.319</td>
<td>0.999</td>
<td>0.999</td>
<td>0.990</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** N=8,984. Results are based on the DWLS estimator with its missing data procedures as described in Asparouhov and Muthén (2010). All models control for age at the first wave, and observed SRH variances are constrained to be equal.
sion coefficients to be equal across female and male respondents as well as the error variances of observed and latent variables. CFI, TLI, and 1-RMSEA values near 1 and a BIC of $-188$ indicate very good absolute fit and provide strong evidence of better fit compared with the other models. This model structure suggests that the main differences in SRH trajectories are in the starting means at Wave I (3.721 for males vs. 3.447 for females per the model estimate), given that the latent time-invariant intercepts are similar ($\sim 1.14$); after Wave I, the two groups follow similar longitudinal patterns.

Table A4 (online appendix) summarizes a similar set of invariance tests across respondents from different racial/ethnic backgrounds, again finding a high level of invariance across groups. Model F constrains the regression coefficients as well as the error variance of observed and latent SRH to be equal across groups. A closer examination of observed SRH error variances in the unconstrained estimates from Model A suggests that the SRH error variances of White adults are distinct from those of their Black and Hispanic counterparts. Thus, constraining the SRH error variances to be equal for these two groups yields a very well-fitting model, with a negative BIC of $-544$. As shown in Tables A5 and A6, gender and racial/ethnic invariance are also upheld in the NLSY97 sample when the same set of constraints are used.

Beyond these measures of fit, we also confirm that constrained values of the autoregressive coefficients from the best-fitting models are not drastically different from their unconstrained values (i.e., Model A; shown in Tables A7 and A8). The percentage change in autoregressive coefficients from Model A to Model E for female and male respondents is never greater than 20%, and most are <10%, suggesting that these constraints are plausible when both Add Health and NLSY97 data are used. The percentage change in AR coefficients from Model A to Model F across racial/ethnic groups exhibits slightly greater variability but still suggests that Model F provides trustworthy estimates.

**Discussion and Conclusions**

As evidenced by the numerous studies using SRH, the measure is and will remain a key variable in population health research. This large body of literature speaks to its value as a parsimonious summary of individuals’ overall health (DeSalvo et al. 2006; Jylhä 2009), potentially capturing information above and beyond what researchers can explain with physiological biomarkers (Dowd and Zajacova 2007; Franks et al. 2003; Idler and Benyamini 1997; Singh-Manoux et al. 2007). Precisely because of the research importance of SRH, we argue that the convenience, ease of collection, and simplicity of interpretation do not compensate for the fact that SRH continues to be among the most poorly understood measures (Huisman and Deeg 2010; Jylhä 2009), especially with respect to changes in SRH over time.

Extant theories on SRH do not dictate clear functional forms, but they do provide useful perspectives on what researchers may expect. The spontaneous perspective emphasizes the specific circumstances at the time of the survey as the primary influence (Bailis et al. 2003; Perruccio et al. 2010). Conversely, the enduring perspective stresses continuity, with individuals holding a stable view of their health (Bailis et al. 2003; Boardman 2006). The lagged effect view recognizes that recent SRH has a strong effect on current SRH (Jylhä 2009; Kosloski et al. 2005). Finally, the life course orientation recognizes that more systematic shifts in SRH are possible as indi-
individuals age and experience changes in both health and life circumstances that may influence their overall sense of healthiness and well-being (Elder et al. 2003; Shanas 2000). Most importantly, these perspectives are not mutually exclusive, leaving the question of what trajectory is most suitable open to empirical analysis.

We recognize that identifying the most appropriate trajectory is often not the objective of a given study. However, the assumptions that researchers make in their models can affect their conclusions. With this concern in mind, we conducted the first systematic comparison of different SRH trajectories from adolescence to midlife while also accounting for its limited reliability. Despite our examination of linear and nonlinear models that vary in complexity, we find that a hybrid model that combines the enduring and lagged effect hypotheses best matches our data. More specifically, an LV-ALT intercept-only model, which allows for a latent time-invariant variable as well as autoregressive relations, best fits longitudinal data from Add Health and NLSY97. We find that this is true for both females and males as well as across racial/ethnic groups; other than differences in starting points for SRH, the trajectories across groups are very similar.

What, then, are the implications of our results for SRH? Consider first the SRH measure’s moderate degree of reliability (∼.6). In most studies, SRH is treated as if it were free of error. Researchers interested in the causal effects of SRH on health or social outcomes who ignore measurement error will have biased assessments of the impact of SRH and other explanatory variables (see Bollen 1989:151–178). In one longitudinal study, the researchers corrected for measurement error (Kosloski et al. 2005) and estimated reliability at .64, similar to our results; however, they ultimately fixed the reliability of the measure at 1.0 when estimating their models. Thus, we advise researchers to allow for a reliability of .5 to .7 when including SRH as a predictive variable.

From a descriptive perspective, it is important to recognize that an inappropriate model can lead to incorrect conclusions about the trajectory of SRH over time, even if we ignore the issue of measurement error. Researchers can fit a variety of models to their data, all of which may give the appearance of capturing the observed trajectory. However, the assumptions underlying these models—that is, the equations that explain relationships among variables—lead to different interpretations. The reliance on linear and quadratic growth curve models in past literature assumes the existence of latent intercepts and slopes that characterize change or stability in SRH over the entire observation period, which is fundamentally different from models incorporating autoregressive relations. A linear growth curve, for example, suggests that SRH follows a linear growth pattern over time, whereas an autoregressive model directs attention to the prior value of SRH to predict the current value. Our analyses reveal the importance of a hybrid model that combines time-invariant and autoregressive effects, which would be difficult to identify based purely on a descriptive interpretation of SRH over time.

Furthermore, although this analysis primarily focuses on modeling trajectories of SRH, it has implications for other types of research questions centered on longitudinal models and causal analysis. For instance, researchers often use a fixed-effects model that assumes a time-invariant variable that influences SRH for each wave. Such analyses may include other time-varying variables (e.g., income or mental health) to evaluate their effects on SRH over time. The latent time-invariant variable captures
and controls for the unobserved enduring effect on SRH, but failing to account for an autoregressive effect (as our model suggests) will result in distorted estimates of the impact of the other time-varying variables on SRH as long as those other time-varying variables correlate with the lagged SRH. More broadly, our results call into question any fixed-effects analysis of SRH that does not include the lagged value of SRH as an explanatory variable. An analogous argument holds if a researcher chooses a latent linear or quadratic growth curve and attempts to assess the causal impact of other variables on SRH: the latent intercept and slope are insufficient in capturing all influences on SRH at a given point in time.

Perhaps most critically, our comparison of longitudinal models provides new knowledge on the spontaneous, enduring, lagged effect, and life course perspectives on SRH. First and foremost, the spontaneous perspective is perhaps the most weakly supported by the data. Because this perspective suggests that individuals’ responses reflect present health circumstances (Bailis et al. 2003; Perruccio et al. 2010), we would expect greater fluctuation owing to increased variability in health over time. Yet, if this were true, it would be hard to explain why prior SRH is such a strong predictor of the current SRH, as seen in the large autoregressive coefficients and high $R^2$ values for latent subjective health, especially in the more closely spaced NLSY97 data (AR coefficients = −0.85; $R^2$-squared = −85%). Critically, the finding that subjective health at a prior time combined with the latent time-invariant effect explains a large proportion of subjective health to the stable qualities of this measure.

The enduring hypothesis complemented by the lagged effect viewpoint appears more consistent with the stability observed in our results. Many researchers contend that SRH reflects an enduring perspective on health (i.e., a “self-concept” of health [Bailis et al. 2003; Boardman 2006; Huisman and Deeg 2010; Jylhä 2009]); the inclusion of a latent time-invariant variable in our LV-ALT intercept-only model suggests that this is an appropriate assumption. The mechanism of this enduring concept is not something we can ascertain from this model, but past research suggests that it reflects individuals’ propensity to retain a more or less fixed view of their health over time, which explains some proportion of SRH independent of actual changes in health. Individual differences in the interpretation can bias their responses; when asked to evaluate overall health compared with their peers (e.g., on the basis of age), respondents assess their health as consistently better or worse despite improvement or decline in their own health (Baron-Epel and Kaplan 2001; Eriksson et al. 2001; Manderbacka et al. 2003; Vuorisalmi et al. 2006). Kaplan and Baron-Epel (2003) noted differences in interpretation based on individuals’ perceived health, such that younger adults in good health are more likely to employ this comparative perspective than those in worse health. Moreover, our results underscore the importance of lagged effects in modeling SRH. Although the lagged effect hypothesis was not explicitly theorized in past literature, we examine this novel hypothesis in emphasizing how individuals’ current perception of SRH are strongly determined by lingering factors and beliefs from recent history (Jylhä 2009, 2010; Norman 2003; Ross 1989). These lingering attributes are consistent with the significant lagged coefficients, net of any influence associated with the time-invariant latent intercept.

Finally, the life course perspective would lead us to expect growth curves in SRH over time. Our examination of different functional forms does not find any that are
consistently superior to our LV-ALT intercept-only model. It is possible that the portion of the life course we examine has relative stability, leading to large lagged effects. However, adolescence to midlife is a demographically important stage of life (Elder et al. 2003; Shanahan 2000), such that disruptions at these ages should break the close relations between lagged and contemporaneous SRH. Greater consistency between the life course perspective and our empirical findings would require an explanation for why SRH maintains such close relationships to earlier values despite the changes in individuals’ life stages, especially in NLSY97.

Furthermore, our examination of group differences provides evidence of why accounting for measurement error and choosing the appropriate trajectories is important. Using the LV-ALT intercept-only model leads to fewer group differences in SRH trajectories than noted in past literature (Bauldry et al. 2012; Sokol et al. 2017). This does not imply that gender and racial/ethnic health disparities are not real; rather, SRH may not be ideal for measuring these disparities over time. Objective measures of health may be more advantageous for identifying group differences, although we recognize the challenge in observing poor physiological health in early life. Nevertheless, given the widespread use of SRH to track health over time, researchers should be attuned to its limitations as a measure more influenced by enduring and lagged perceptions than current health.

We close by acknowledging the limitations of our analyses as well as recommendations for future research. Ideally, we would have more waves of data with which to model trajectories. Add Health data are not equally spaced, which explains some of the variation in the autoregressive coefficients and R-squared values across waves. More data would allow us to better examine additional and more complex trajectories or to use alternate specifications of time such as individuals’ ages. However, the consistency in results based on the NLSY97—which is nearly continuous in terms of both time and age—suggests that the wave-based analysis in Add Health is not necessarily an issue.

Relatedly, we recognize that longitudinal models should be compared in samples across other contexts in time and geography, different age ranges, additional racial/ethnic groupings, and other relevant populations because the meaning of SRH varies across groups (Grol-Prokopczyk et al. 2011). Age differences are particularly important, especially because much of the research on SRH focuses on older adults (Ferraro and Kelley-Moore 2001; Miller and Wolinsky 2007). SRH trajectories might change as the relative healthiness of youth fades in middle to late adulthood, and age-related declines more strongly influence SRH. Interestingly, however, Kosloski et al. (2005) found similar autoregressive coefficients for an older sample (aged 50+), suggesting that our results may be consistent throughout the life course. Their analysis did not assess alternate models, so we cannot conclude an autoregressive model is best, but their model had good absolute fit. Finally, our focus is to determine the best longitudinal model for SRH; most researchers studying trajectories build more complete models that incorporate additional covariates, which can influence the estimates. This research can and should be done, but our findings suggest that researchers’ baseline assumptions for these models should account for the importance of time-invariant and autoregressive components.

More broadly, this research highlights the need for comprehensive assessments of trajectories for other health measures (Bauldry and Bollen 2018). Despite the grow-
ing availability of health data in surveys, researchers continue to default to a particular model without considering alternate specifications, implicitly assuming that one indicator of health has similar measurement properties to another. Not unlike with SRH, much of this is attributable to the lack of theoretical reasoning to support a given model. Therefore, we encourage not only more rigorous empirical analyses of health trajectories but also a more robust body of theory to guide future studies and help researchers better understand the unique attributes of the many health measures at their disposal.

Acknowledgments This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by Grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. We also acknowledge support from a Population Research Training grant (T32 HD007168) and the Population Research Infrastructure Program (P2C HD050924) awarded to the Carolina Population Center at The University of North Carolina at Chapel Hill by the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

Note Authors are listed in alphabetical order, reflecting their equal contribution.

References


Iliya Gutin (corresponding author)
igutin@live.unc.edu

Bollen • Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; https://orcid.org/0000-0002-6710-3800

Gutin • Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; https://orcid.org/0000-0003-4216-8963