Descriptive Finding

The role of education in explaining trends in self-rated health in the United States, 1972–2018

Jona Schellekens
Anat Ziv


This open-access work is published under the terms of the Creative Commons Attribution 3.0 Germany (CC BY 3.0 DE), which permits use, reproduction, and distribution in any medium, provided the original author(s) and source are given credit.
See https://creativecommons.org/licenses/by/3.0/de/legalcode.
The role of education in explaining trends in self-rated health in the United States, 1972–2018

Jona Schellekens¹
Anat Ziv²

Abstract

BACKGROUND
The percentage of older adults in the United States reporting being in good health has increased since the 1980s.

OBJECTIVE
This study tries to explain long-term trends in self-rated health in the United States.

METHODS
We used 47 years of repeated cross-sectional data from the National Health Interview Survey to estimate regression models that predict trends in self-rated health.

RESULTS
Our results show that the improvement in self-rated health of men as well as women aged 50–84 is largely explained by gains in educational attainment. Self-rated health has slightly improved among those with post-secondary education, while it did not improve among those without post-secondary education.

CONTRIBUTION
This study is one of the few to try to explain long-term trends in self-rated health. It does so for a much longer period of time than any previous study.

¹ Louis & Pearl Rose Professor in Medical Sociology, Department of Sociology and Anthropology, Hebrew University, Jerusalem, Israel. Email: jona@mail.huji.ac.il.
² Department of Sociology, University of New Brunswick, Canada. Email: aziv@unb.ca.
1. Introduction

Life expectancy has increased over the last century. Whether additional years of life are also accompanied by years in good health has become a subject of intense interest. Little is known about trends in morbidity before the 1960s. From the 1960s until the 1980s morbidity appears to have increased (Riley 1990). Since the 1980s, however, morbidity has declined, at least in the United States (e.g., Crimmins, Zhang, and Saito 2016; Martin et al. 2007; Martin, Schoeni, and Andreski 2010; Salomon et al. 2009).

Very little is known about what factors have influenced trends in morbidity. Medical technology is one factor. Cutler, Landrum, and Steward (2009), for example, estimated that improved medical care for cardiovascular disease explains up to 50% of the reduction in disability after an acute event between 1984 and 1999. On the other hand, if people are treated for a condition that would have caused death but now they stay alive with the condition, morbidity may increase (Gruenberg 1977).

Education is another factor. Adults with higher educational attainment live healthier lives than their less-educated peers (Zajacova and Lawrence 2018). In a cross-sectional analysis of data from 70 countries collected by the 2002/2003 World Health Survey, KC and Lentzner (2010) show that there is a strong correlation between education and health. The correlation is particularly strong in the United States (e.g., Borgonovi and Pokropek 2016; Präg and Subramanian 2017; Prus 2011). Education is associated with many health-related behaviors (e.g., Pampel, Krueger, and Denney 2010), with access to healthcare, and with the ability to navigate the healthcare system (e.g., Lynch 2003). Education is also associated with psychosocial factors such as depression, social isolation, stress, and loss of control, which may impact health (e.g., Cohen, Kaplan, and Salonen 1999; Gallo and Matthews 2003; Mirowsky and Ross 2007; Soskolne and Manor 2010; Hämmig, Gutzwiller, and Kawachi 2014).

Hill and Needham (2006) estimate that in the United States the improvement in self-rated health (SRH) among women, but not among men, between 1972 and 2002 is largely explained by gains in educational attainment. Their conclusions are based on the General Social Survey. Using the much larger National Health Interview Survey (NHIS) over a much more extended period, we show that the rise in SRH among women as well as men between 1972 and 2018 is almost entirely explained by gains in educational attainment.

2. Data

We used the Integrated Health Interview Series, the NHIS harmonized for consistent variable measurement, which we downloaded from the Integrated Public Use Microdata
Series (IPUMS) website at the University of Minnesota (Blewett et al. 2019). The NHIS is a repeated cross-sectional face-to-face interview survey. It is the world’s longest survey time series of health data (King 2011). We used data from 1972, which is when the question on SRH was first asked, through 2018.

3. Variables

SRH is the outcome variable in the analysis. Beginning with the 1982 survey, both the question wording and the response categories changed. Before 1982 the question wording was: “Compared to other persons his age, would you say that (subject’s name)’s health is excellent, good, fair or poor?” Thereafter, the wording is: “Would you say (subject’s name)’s health (in general) is excellent, very good, good, fair, or poor?” To facilitate comparability over time, we created a dichotomous measure of whether respondents rated their health as good, very good, or excellent, as compared to poor or fair. To control for the changed wording of the question we added a dummy variable to the preliminary analysis indicating whether the survey was conducted before 1982. However, the effect of the indicator variable was small (0.002) and very insignificant (0.521). Hence, it has been omitted from the final analysis. Only a small number of respondents failed to answer this question. We have omitted these from the analysis.

Figures 1 and 2 present trends in the percentage of women and men who reported being in good health for selected age groups. Morbidity declined in all age groups. However, the decline did not occur simultaneously in each age group. The older the age group, the later the decline, suggesting the presence of a cohort effect. The rise in education has contributed to differences between cohorts (Ryder 1965: 854), suggesting that gains in educational attainment may explain part of the decline in morbidity.

Predictors include gender, age, race, and education. The analysis includes a variable indicating whether the respondent is female. In several years there is an open-ended age group starting at age 85. Therefore, the analysis has been limited to respondents aged 50–84. Following Choi, Schoeni, and Martin (2016), the statistical model includes a quadratic term of age. There are racial differences in self-assessed health (e.g., Beck et al. 2014; Cummings and Jackson 2008). Thus, the analysis includes a variable indicating whether the respondent is white.
Figure 1: Percentage of women who reported being in good health for selected age groups, United States 1972–2018

Figure 2: Percentage of men who reported being in good health for selected age groups, United States 1972–2018
The analysis includes a variable indicating whether the respondent finished at least grade 8 but did not go to college, and one indicating whether the respondent had at least one year of college. Unlike some other measures of socioeconomic status such as occupation and income, education is easily measured, and generally is fixed for each respondent relatively early in life. As a result, health problems that emerge in late life are unlikely to influence educational attainment (Freedman and Martin 1999). In general, there are few missing cases in the NHIS. Most of the missing data is due to nonresponse on education. When only a small fraction of the data is missing for a categorical variable, a simple approach to imputation is to add an extra category for the variable indicating missingness (Gelman and Hill 2007: 533). Only 2.4% of the respondents did not answer the question on education. Hence, a variable indicating nonresponse on education has been included in the analysis.

The age declines in health may not be the same across education and ethnic groups (e.g., Lynch 2003; Ross and Wu 1996; Willson, Shuey, and Elder 2007). Moreover, the relationship of health with race and education may vary by gender (e.g., Pinillos-Franco and García-Prieto 2017). Therefore, we have added interactions of age with gender, race, and education, and of gender with race and education.

4. Statistical methods

In preliminary analyses, logistic regression models were used to assess the effects of the independent variables on the probability that a respondent reported being in good health. However, in the final analysis we used linear regression models in order to facilitate the comparison of coefficients across models (e.g., Mood 2010). Hellevik (2009) has made a case for using linear regression when the dependent variable is a dichotomy. A logistic model should be used if it fits the data much better than a linear model. However, often the linear model fits just as well as the logistic model, especially when probabilities are not close to zero or to one.

To estimate the contribution of better education to the rise in the percentage of respondents who reported being in good health, we first regressed SRH on dummy variables indicating five-year survey periods. (The first period (1972–1974) and the last one (2015–2018) only cover three to four years.) Next, we controlled for education to test whether the relationship between survey period and SRH is mediated by education. A noticeable reduction in the size of the coefficients of the period dummy variables would suggest mediation (Hill and Needham 2006).

The NHIS used a complex sampling design. Where sampling weights are solely a function of independent variables included in the model, unweighted estimates of the coefficients are preferred because they are unbiased, consistent, and have smaller
standard errors than weighted estimates (Winship and Radbill 1994). Thus, we used weights when estimating trends in SRH (see Figures 1 and 2), but did not use them in the regression analyses.

5. Results

Table 1 presents three linear regression models of the probability that a respondent reported being in good health in the period 1972–2018. All three models control for gender, age, race, and the interactions between them. Controlling for age, race, and survey period, women are less likely to report being in good health, whereas whites are more likely to report being in good health.

The size of the coefficients of the period dummy variables in the first model increases over time, indicating an improvement in SRH. To test whether the relationship between survey period and SRH is mediated by education, the second model adds education and the interactions of education with age and gender. Education has a positive effect on SRH. The effect of nonresponse resembles that of the omitted category (9–12 years). The interactions between education and gender indicate that the effect of education is slightly different for men and women. The negative effect of low education (<9 years) is stronger among women, whereas the positive effect of higher education (13+ years) is slightly stronger among men.

In the second model the size of the coefficients of all the period dummy variables is much smaller than in the first model. To illustrate this, Figure 3 compares the coefficients of the period dummy variables in both models. The first model shows that the probability of reporting being in good health rose by 0.107 between 1972–1974 and 2015–2018 (solid line). After controlling for education, the probability of reporting being in good health only rises by 0.009 (dashed line). Thus, better education explains most of the rise in the percentage of respondents who reported being in good health. In particular, in the second model the difference in SRH between 2010–2014 and 1972–1974 is very close to zero. Moreover, the difference is very insignificant (0.847). Thus, better education explains the rise in the probability of reporting being in good health between 1972–1974 and 2010–2014 entirely.
Table 1: Linear regression models of the probability of being in good health, United States 1972–2018

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (female = 1)</td>
<td>−0.088***</td>
<td>−0.090***</td>
<td>−0.093***</td>
</tr>
<tr>
<td>Race (white = 1)</td>
<td>0.036***</td>
<td>0.043***</td>
<td>0.046***</td>
</tr>
<tr>
<td>Interaction of race with gender</td>
<td>0.027***</td>
<td>0.030***</td>
<td>0.030***</td>
</tr>
<tr>
<td>Age</td>
<td>−0.015***</td>
<td>−0.014***</td>
<td>−0.014***</td>
</tr>
<tr>
<td>Age squared / 1000</td>
<td>0.059***</td>
<td>0.059***</td>
<td>0.061***</td>
</tr>
<tr>
<td>Interaction of age with gender</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
</tr>
<tr>
<td>Interaction of age with race</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.000***</td>
</tr>
<tr>
<td>Period (1972–74 = 0):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975–79</td>
<td>0.013***</td>
<td>0.000</td>
<td>0.003</td>
</tr>
<tr>
<td>1980–84</td>
<td>0.015***</td>
<td>−0.010***</td>
<td>−0.003</td>
</tr>
<tr>
<td>1985–89</td>
<td>0.046***</td>
<td>0.007***</td>
<td>0.015***</td>
</tr>
<tr>
<td>1990–94</td>
<td>0.060***</td>
<td>0.010***</td>
<td>0.017***</td>
</tr>
<tr>
<td>1995–99</td>
<td>0.070***</td>
<td>0.010***</td>
<td>0.012***</td>
</tr>
<tr>
<td>2000–04</td>
<td>0.076***</td>
<td>0.005**</td>
<td>0.007*</td>
</tr>
<tr>
<td>2005–09</td>
<td>0.082***</td>
<td>0.002</td>
<td>−0.001</td>
</tr>
<tr>
<td>2010–14</td>
<td>0.088***</td>
<td>0.000</td>
<td>−0.009**</td>
</tr>
<tr>
<td>2015–18</td>
<td>0.107***</td>
<td>0.009***</td>
<td>0.003</td>
</tr>
<tr>
<td>Education (9–12 years = 0):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;9 years</td>
<td>0.321***</td>
<td>−0.308***</td>
<td></td>
</tr>
<tr>
<td>13+ years</td>
<td>0.110***</td>
<td>0.094***</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>0.310***</td>
<td>0.219***</td>
<td></td>
</tr>
<tr>
<td>Interaction of education with gender:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;9 years</td>
<td>−0.025***</td>
<td>−0.025***</td>
<td></td>
</tr>
<tr>
<td>13+ years</td>
<td>−0.006***</td>
<td>−0.006***</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>−0.010*</td>
<td>−0.012*</td>
<td></td>
</tr>
<tr>
<td>Interaction of education with age:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;9 years</td>
<td>0.003***</td>
<td>0.003***</td>
<td></td>
</tr>
<tr>
<td>13+ years</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>−0.005***</td>
<td>−0.005***</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.296***</td>
<td>1.246***</td>
<td>1.254***</td>
</tr>
<tr>
<td>R²</td>
<td>0.028</td>
<td>0.069</td>
<td>0.070</td>
</tr>
<tr>
<td>Number of cases</td>
<td>1,248,763</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < 0.05; ** p < 0.01; *** p < 0.001

Note: interactions between education and period in third model not shown.
Non-zero and unidentified period effects remain in the second model, but these are small compared to those in the first model. Our results indicate that without better education, SRH would initially have declined in 1980–1984, perhaps due to changes in the question wording in 1982. If this is correct, then we may have slightly underestimated the rise in SRH, especially among older respondents, and slightly overestimated the contribution of better education. Better education only explains about half of the small rise in SRH that occurred after 2010–2014.

We also performed separate analyses for men and women. However, the results of the regression analyses are nearly identical. Therefore, we do not present separate results for men and women.

The period dummies in the second model with education show the average trend in SRH among educational groups. However, SRH may have improved among the more educated, while it deteriorated among the less educated. To check this possibility, the third model in Table 1 adds interactions between period dummies and educational group. The omitted category is 9–12 years of education. Therefore, the main effect of the period dummies in Table 1 shows the trend in SRH among those with 9–12 years of education. By changing the omitted category to less than 9 years and to 13 years or more, we also obtained the trend in SRH for these two educational groups. Figure 4 presents the result for all three educational groups. It shows that SRH among those with
13 years of education or more slightly improved, while SRH among the other two educational groups did not improve from 1972–1974 to 2015–2018.

Figure 4: Regression coefficients of period dummy variables in the first model without education and in the third model with education and interactions between education and period, United States 1972–2018

Figures 1 and 2 show that trends in SRH depend on age. However, the period dummies in the regression models show the trend in SRH for the average age group. To check whether our conclusion that gains in educational attainment explain the improvement in SRH applies to all age groups, we estimated an additional set of models with interactions between age and the period dummies. We replicated Figure 3 for ages 50, 60, 70, and 80 (results not shown). The results were similar to those presented in Figure 3. Thus, our conclusion applies to all age groups in the analysis.

6. Conclusion and discussion

Our results confirm that SRH has steadily improved over the 46-year period under study. These improvements are largely explained by gains in educational attainment. Hill and Needham (2006) reached a similar conclusion for women during the first 30 years covered by our study, but not for men. Our results show that this is not only true for women, but also for men. It is not clear why our results regarding men differ from
those of Hill and Needham (2006). Our results also show that SRH among those with post-secondary education has slightly improved, while it did not do so among those without post-secondary education.

Education is associated with many health-related behaviors, with access to health care, and with psychosocial factors. Unfortunately, our analysis is not very informative about the relative importance of these mechanisms in explaining trends in SRH.

There are also some caveats. There may be a problem with the measurement validity of SRH. However, even though SRH is a subjective measure of morbidity, it is considered to be a valid and reliable measure among those without cognitive impairment (Bombak 2013). SRH has been found to be very predictive of mortality and strongly correlated with objective assessments of health (Idler and Benyamini 1997). Moreover, the predictive validity of SRH with respect to mortality is improving over time (Schnittker and Bacak 2014). Schellekens (2019) reports a similar finding for the decline in disability, which is considered to be a more objective measure of health. In any case, whatever its validity, it is important to understand why SRH is changing over time (e.g., Waller 2015).

There may also be a problem with internal validity. We cannot give the education effects a causal interpretation. Previous research has shown evidence of a causal relationship running from more schooling to better health (e.g., Siles 2009). However, there may also be reverse causation, health problems that emerge in early life influencing educational attainment. In addition, education and health may be jointly determined by other factors (e.g., Ding et al. 2009). For example, socioeconomic status in childhood may affect educational attainment as well as health in later life (e.g., Haas 2008; Pakpahan, Hoffmann, and Kröger 2017). Thus, the rise in living standards in childhood may explain the correlation between trends in SRH and those in educational attainment. If this is correct, then educational attainment serves as a proxy for socioeconomic status in childhood.

Finally, there may be a problem with external validity, because the analysis was limited to respondents aged 50–84. Figures 1 and 2, however, show that SRH also improved among respondents aged 85+, even though with a delay. Thus, our results may also apply to older ages. In addition, there may be sample selection bias, because the analysis excluded the institutional population (e.g., Berk 1983). Crimmins, Zhang, and Saito (2016), however, show that trends in the percentage of years after age 65 with any disability resemble those in the percentage of years in an institution after age 65.
7. Acknowledgements

The authors wish to thank the three anonymous reviewers for their helpful comments. An earlier version was presented at the Dutch Demography Day, Utrecht, November 2019.
References


