Identifiying the Predictors of Self-rated Health with Machine Learning

Varun Satish and Beza Taddess

April 2026

Abstract

Self-rated health captures whether a person self-reports their health as poor, fair, good, very good, or excellent. It is used by population health researchers to understand inequality because it accounts for important outcomes, such as mortality, even after controlling for biological information and expert assessments. Theory suggests many possible explanations for what predicts self-rated health. These explanations represent complex interactions between a vast array of biological, social, and cultural factors. However, empirical studies typically focus their attention on a hand-selected set of these interactions, which while theoretically informed, potentially miss meaningful relationships. In this paper, we leverage supervised machine learning to identify which of the many possible interactions predict self-rated health. Specifically, we use regression trees to estimate predictors from 186 survey questions in the 2024 wave of the Behavioral Risk Factor Surveillance System. These predictors can be interpreted as population groups defined by combinations of survey question responses. Regression trees produced predictors defined by a few variables—general physical health, exercise, difficulty walking, and diabetes—that reveal an important relationship between self-rated health and physical mobility. For example, people with difficulty walking or climbing stairs rate their health 1 standard deviation lower than those without, even at equal levels of general physical health. We then use these predictors to understand why Americans with a college education are twice as likely to report excellent self-rated health and half as likely to report poor self-rated health. We found that college-educated Americans have a dual advantage: college graduates are more likely to be sorted into groups with better self-rated health, on average, but also report systematically better self-rated health than their lesser-educated counterparts within those groups.

1 Introduction

Our understanding of health inequality in the United States is fundamentally shaped by people's subjective perceptions of their own health. Case in point, the answer to the question "Would you say your health is excellent, very good, good, fair, or poor?"—self-rated health—is a measure that has become increasingly important for understanding health inequalities over time (Schnittker and Bacak, 2014). Self-rated health predicts important health outcomes such as mortality, cardiovascular events, and functional decline even after adjusting for clinical assessments and biological markers (DeSalvo et al., 2006; Stewart et al., 2017; Bamia et al., 2017). Moreover, disparities in self-rated health shed light on structural inequalities relating to education, race, and gender that pattern a broad range of health outcomes (Lamidi, 2020; Beck et al., 2014; Boen, 2016; Cummings, 2023). Thus, if we can better understand the factors that shape self-rated health, we can develop a better understanding of which Americans are sick, getting sicker, and among whom inequalities are growing.

People's responses to health surveys reflect complex, embodied, socially situated experiences. These responses are shaped by interactions between diverse biological, social, and cultural factors (Jylhä, 2009). Previous research on self-rated health, for example, has elucidated multifaceted determinants of self-rated health such as race (Assari et al., 2016), aging (Schnittker, 2005; Idler and Cartwright, 2018), gender (Cobb and Assari, 2020), psychological factors (Choi and Jung, 2025), comorbidities (Gumà-Lao and Arpino, 2023), and education Dowd and Zajacova (2010). These determinants are not static across populaton groups. For instance, Dowd and Zajacova (2010) showed that the same self-rated health category corresponded to systematically different biomarker profiles depending on respondents' level of education. In other words the factors that correspond with "very good" health in one group does not correspond with "very good" health in another.

This prior work typically learns about the predictors of self-rated health by interrogating relationships between factors that have been hypothesized *a priori*. Given the vast array of information that is available in most health surveys, could it be possible that there are some relationships, meaningful for predicting self-rated health, that we have not investigated? A growing literature appears to believe that the answer is yes (Clark et al., 2021; Gumà-Lao and Arpino, 2023; Choi and Jung, 2025).

In this study, we take an inductive, data-driven approach to estimating the predictors of self-rated health in the United States. To do so, we leverage regression trees—a supervised machine learning algorithm that partitions the sample into groups within which people have similar levels of self-rated health. Regression trees allow us to let the data identify which combinations of responses (i.e. predictors) distinguish people who rate their health differently, rather than specifying those combinations in advance.

Our approach builds on classic work in sociology which recognizes that the categories through which people organize their experience should emerge from data rather than be imposed in advance (Glaser and Strauss, 2017). Prior work that has used this inductive logic with respect to health demonstrates its power: research on chronically ill individuals revealed an identity hierarchy, from supernormal social identity to restored self, contingent personal identity, and salvaged self, that pre-specified categories would have missed entirely (Charmaz, 1990). It is also distinct from other work which has used machine learning to understanding self-rated health because we use an algorithm that allows us to estimate a specific conceptual estimand.

Namely, we conceive of a predictor as being a population group. This framing has an empirical payoff because, as we discuss below, it allows us to make sense of differences in self-rated health within and between sociologically meaningful groups.

We train a regression tree to predict self-rated health from 186 health-related survey items in the 2024 wave of the Behavioral Risk Factor Surveillance System (BRFSS). The tree identified seven groups of respondents whose self-rated health differed substantially, on average (Table 1). These seven groups are defined by just four variables: the number of bad physical health days in the past month, serious difficulty walking or climbing stairs, a prior diabetes diagnosis, and any exercise in the past month. This suggests that despite hundreds of available items, the algorithm uncovers a set of interactions centered on what the body can and cannot do. Thus our approach reveals that self-rated health looks less like a comprehensive summary of health information and more like an evaluation of physical health and mobility.

Because the predictors estimated by regression trees are population groups, they allow us to ask: (1) how social position shapes which Americans are most exposed to the predictors of self-rated health; and (2) how social position affects how Americans rate their health conditional on those exposures. We exploit these properties to explore a consequential empirical phenomena: the educational gradient in self-rated health. This gradient refers to the fact that college graduates report substantially higher self-rated health than Americans with a high school education or less; roughly one standard deviation higher on the ordinal scale.

To this end, we use the seven tree-defined groups to study differences in two quantities across levels of education. The first is compositional—it refers to the share of Americans at a given level of education who fall into each of the predictor groups. The second is systematic—it refers to the average level of self-rated health within a predictor group at a given level of education. This analysis reveals that college-educated Americans have a dual-advantage with respect to self-rated health: (a) they are disproportionately concentrated in high-self-rated-health groups characterized by good physical health, no diabetes, and regular exercise; and (b) they report higher self-rated health than their lower-educated counterparts within those groups.

These findings contribute to the study of health inequality by specifying two mechanisms behind the educational gradient in self-rated health. The compositional mechanism reflects something already familiar: college-educated Americans are, on average, in better physical health and report higher self-rated health partly because of it. The systematic mechanism, on the other hand, is particularly noteworthy. Within the tree-defined groups, college graduates still rate their health more favorably than less-educated Americans in the same physical state. This extends the biomarker-based analysis from Dowd and Zajacova (2010) to a broader BRFSS feature space and quantifies how much of the gradient is attributable to differences in evaluation rather than body. Self-rated health is therefore not a transparent index of physical health that happens to correlate with education. It is an evaluation in which the same body is mapped onto different response categories depending on social position. Disparities in self-rated health are not only disparities in health; they are also disparities in the evaluative standard that respondents apply. This reframing has consequences for how researchers interpret educational gradients in subjective health and, more broadly, for what subjective health measures can be taken to reveal.

2 Data and Methods

2.1 Behavioral Risk Factor Surveillance System (BRFSS)

Data come from the 2024 wave of the Behavioral Risk Factor Surveillance System (BRFSS), a national survey administered by state health departments in collaboration with the Centers for Disease Control and Prevention (CDC). Survey data and detailed methodology are publicly available at www.cdc.gov/brfss.

Sample Selection. Our analytic sample consists of 457,665 respondents with non-missing values on the main outcome variable, self-rated health, and includes self-identified Black women across all 50 states and the District of Columbia. We split this sample into training and test sets comprising 90% and 10% of observations, respectively.

Independent Variables. The BRFSS includes 300 survey questions beyond self-rated health, organized into topical modules covering health status, chronic conditions, health care access, health behaviors (e.g., physical activity, tobacco and alcohol use, diet), demographics, disability, cancer screening, immunization, sexual and reproductive health, social determinants, adverse childhood experiences, and cognitive decline, among others. We exclude questions from administrative and logistical modules—record identification, landline and cell phone introductions, and weighting variables—as well as calculated variables and calculated race variables, since the latter are derived from existing survey questions. This leaves 186 survey questions. For each of these 186 questions, we generate a binary missing data indicator equal to 1 if the respondent did not answer that question. Missing values are then recoded to -999. Our final set of independent variables comprises 372 variables: 186 survey questions and a corresponding set of missing data indicators.

Outcome Variable. Self-rated health is measured with the question: "In general, how would you rate your overall health?" Response options are "poor", "fair", "good", "very good", and "excellent". Because it is ordinal, we operationalize it numerically, with "poor" self-rated health being represented with 1 and "excellent" being represented with 5.

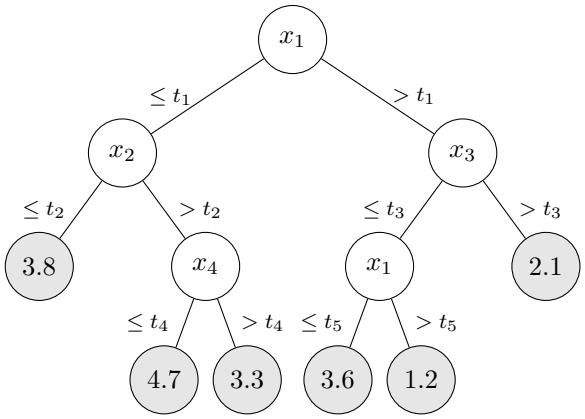
2.2 Regression Tree

A regression tree predicts self-rated health by partitioning a sample into mutually exclusive and exhaustive groups based on values of the observed characteristics—that is, the 186 BRFSS survey questions. Partitioning is achieved through a grouping function $\text{tree}(\cdot)$ that is learned algorithmically through supervised learning. Supervised learning works by learning groups that minimize within-group variance in self-rated health.

Regression trees learn predictor groups that are interpretable. This is because the grouping function $\text{tree}(\cdot)$ represents a tree structure that can be conceived of as a series of decision rules. These rules operate as logical "if-then" statements that can be applied to the observed characteristics of each observation.

Once $\text{tree}(\cdot)$ is learned on the training data, it can be extrapolated to the test data. To be concrete, the predictor group that individual i is assigned to can be determined by applying the grouping function to their individual characteristics: $\text{predictor}_i = \text{tree}(\mathbf{x}_i)$. The application of this function involves evaluating one of the 186 BRFSS survey questions against a threshold,

directing the individual left or right until it reaches a terminal node[1] which determines the it's final group allocation (Figure 1).



Predictor Group	Test Mean
`x1` $\leq$ `t1 & x2` $\leq$ `t2`	3.8
`x1` $\leq$ `t1 & x2` $>$ `t2 & x4` $\leq$ `t4`	4.7
`x1` $\leq$ `t1 & x2` $>$ `t2 & x4` $>$ `t4`	3.3
`x1` $>$ `t1 & x3` $\leq$ `t3 & x1` $\leq$ `t5`	3.6
`x1` $>$ `t1 & x3` $\leq$ `t3 & x1` $>$ `t5`	1.2
`x1` $>$ `t1 & x3` $>$ `t3`	2.1

Figure 1: Example of a regression tree and its equivalent table representation using fake data. In the tree, circles show the variable used for splitting, edges show the threshold conditions (t_i denotes the threshold value for variable x_i). Shaded circles represent the 6 unique predictor groups defined by tree($\cdot$) and the numbers inside them represent the mean self-rated health among people in the test set who were assigned to the group. Each row of the table corresponds to one predictor, expressed as the conjunction of split conditions along a path on the tree.

3 Empirical Results: Estimating Predictors of Self-rated Health

The regression tree identified 7 predictor groups. These were: (1) people who reported no bad physical health days in the past 30 days, do not have diabetes, and who exercise; (2) people who reported no bad physical health days in the past 30 days, do not have diabetes, and who do not exercise; (3) people who report between 1-13 bad physical health days in the past 30 days and who do not have difficulty walking; (4) people who reported no bad physical health days in the past 30 days and who have diabetes; (5) people who report between 1-13 bad physical health days in the past 30 days and have difficulty walking; (6) people who report between 14-30 bad physical health days in the past 30 days and do not have difficulty walking; and (7) people who report between 14-30 bad physical health days in the past 30 days and have difficulty walking (Table 1).

Despite the fact that these groups could be defined based on 186 variables, regression trees ultimately reflected interactions between just four: the number of bad physical health days reported in the last 30 days, having been diagnosed with diabetes, exercising, and having difficulty walking (Table 2). Of these four, the number of bad physical health days reported in the last 30 days appears particularly important given that it appears in the decision rules that generate of all of the predictor groups. Generally speaking, this fact demonstrates a strong connection between physical and self-rated health.

More substantively, the predictor groups shed light on the importance of a specific domain of physical health: physical mobility. First, two of the other variables used to define predictor

[1] All individuals within a group have the same predicted self-rated health. For the purposes of our analysis, these predicted values are used to calculate metrics such as MSE, but they do not play a role in any other analyses.

groups, exercising, and having difficulty walking are specifically related to physical mobility. Second, the average self-rated health of individuals with better physical mobility is higher across the board. This is demonstrated by two comparisons. Conditional on reporting 1-13 bad physical health days, those who have difficulty walking (Row 5, Table 1) report 0.6 (1.1 std. deviations) lower self-rated health than those who do not (Row 3, Table 1). This gap remains when comparing those who report 14-30 bad physical health days: those with difficulty walking (Row 6, Table 1) report 0.65 (1.14 std. deviations) lower self-rated health than those who did not (Row 7, Table 1).

The predictors also highlighted an important relationship between self-rated health and chronic disease, specifically diabetes. Conditional on reporting no bad physical health days, people who had diabetes (Row 4, Table 1) reported between 0.65 and 0.27 lower self-rated health (1.14 and 0.47 std. deviations) compared to people who do and do not exercise (Rows 1-2, Table 1).

	Predictor	Group Size	Mean SRH
1	No bad physical health days, no diabetes, exercises	0.436	3.83
2	No bad physical health days, no diabetes, no exercises	0.090	3.45
3	Some bad physical health days (1-13), no walking difficulty	0.210	3.34
4	No bad physical health days, diabetes	0.065	3.18
5	Some bad physical health days (1-13), walking difficulty	0.051	2.74
6	Many bad physical health days (14-30), no walking difficulty	0.069	2.57
7	Many bad physical health days (14-30), walking difficulty	0.080	1.92

Table 1: This table displays the predictors estimated by the regression tree. Here, the predictors have been represented by natural language labels we assigned based on the BRFSS codebook. Table 2 links these labels to the corresponding BRFSS variable names, survey questions, and numeric responses. Table A1 in the Appendix presents predictors in variable name form (i.e without natural language labels). The "Group Size" column contains the proportion of the test set that falls into the predictor group. The "Mean SRH" column contains the mean self-rated health among test set observations who fall into that predictor group.

Variable Name	Interpretation	Survey Question	Responses
PHYSHLTH	Number of bad physical health days	Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good?	1–30 (days); 88 (none)
DIABETE4	Diagnosed with diabetes	Ever told you had diabetes?	1 (Yes); 2 (Yes, pregnancy only); 3 (No); 4 (Pre-diabetes)
EXERANY2	Exercises	During the past month, other than your regular job, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?	1 (Yes); 2 (No)
DIFFWALK	Difficulty walking	Do you have serious difficulty walking or climbing stairs?	1 (Yes); 2 (No)

Table 2: Variables used to define predictors in Table 1. "Variable Name" gives the name of the variable in the BRFSS codebook; "Interpretation" gives the natural-language label we assigned to each variable, as used in Table 1; "Survey Question" gives the exact question wording; and "Responses" lists how responses were coded by the BRFSS enumerators.

4 Empirical Results: Explaining Educational Gradients in Self-rated Health

College-educated Americans report better self-rated health than their lesser-educated counterparts. College graduates are 1.6 times as likely to report excellent health and about half as likely to report poor health compared to those with a high school education or less. When operationalized as an ordinal variable, self-rated health is roughly one standard deviation higher among college-educated Americans than their less-educated peers.

Education	Poor	Fair	Good	Very Good	Excellent	Ordinal Mean
HS or less	0.08	0.21	0.37	0.23	0.11	3.08
Some college	0.06	0.16	0.36	0.31	0.12	3.27
College graduate	0.03	0.09	0.31	0.39	0.18	3.61

Table 3: The educational gradient in self-rated health. Each row reports the distribution of self-rated health responses for a given education group. The "Ordinal Mean" is the group mean when response categories are coded 1 (Poor) through 5 (Excellent); remaining columns show the proportion selecting each category. The full sample mean and standard deviation are 3.38 and 0.57, respectively.

In this section, we explain why this educational gradient in self-rated health exists. To do so, we use the estimated predictors to shed light on compositional and systematic differences in self-rated health across levels of education. We will first define what we mean by compositional and systematic differences. First note that average self-rated health within an educational subsample can be expressed as a sum of two components across estimated predictor groups (via the law of iterated expectations):

$$\mathbb{E}[y \mid \text{education}] = \sum_{\text{all } g} \mathbb{P}(\text{tree}(\mathbf{x}) = g \mid \text{education}) \times \mathbb{E}[y \mid \text{tree}(\mathbf{x}) = g, \text{education}].$$

The first component of the sum, $\mathbb{P}(\text{tree}(\mathbf{x}) = g \mid \text{education})$, refers to the compositional component. It is compositional in the sense that, comparing across levels of education, it captures differences in the composition of the sample across each of the predictor groups. The second component of the of the sum, $\mathbb{E}[y \mid \text{tree}(\mathbf{x}) = g, \text{education}]$, refers to the systematic component. It is systematic in the sense that, comparing across levels of education, it captures differences in the average self-rated health of observations within the same predictor group.

To explain what studying these components might tell us about educational gradients in self-rated health, it is useful to consider situations occurring at two ends of a spectrum. On one end of this spectrum, differences in self-rated health could be explained entirely by compositional differences. This situation would imply that lower-educated Americans within the same predictor groups would report the same self-rated health, on average. In this case, the educational gradient would be explained by the fact that lower-educated Americans are sorted into groups that report low self-rated health on average (e.g. those with low physical mobility).

On the other end of this spectrum, differences in self-rated health could be explained entirely by systematic differences. This situation would imply that that the same share of lower- and

higher-educated Americans were sorted into each of the predictor groups, respectively. In this case, the educational gradient would be explained by the fact that higher-educated Americans report higher self-rated health even within the same predictor groups. For example, the self-rated health of Americans with no bad physical health days, no diabetes and who exercise being higher among those who had a college education versus those who did not.

These compositional and systematic components across each of the predictor groups within each level of education (high school or less, some college, and college graduate) are displayed in Figure 2.

Figure 2, Panel A illustrates that college-educated Americans are more likely to be in the following groups, relative to those with lower levels of education: those with no bad physical health days, who do not have diabetes and who exercise and those with some bad physical health days but who face no walking difficulties. Conversely, they are less likely to be in these the following groups: those with no bad physical health days, who do not have diabetes and who do not exercise, and those with many bad physical health days as well as walking difficulty. The predictor group that contains the largest share of Americans across all levels of education is the group that reports no bad physical health days, who do not have diabetes and who exercises. However, this share is much larger for the college-educated—0.5 compared to 0.4 for some college and 0.35 for high school or less. Figure 2, Panel B demonstrates that this predictor group reports the highest self-rated health on average.

Importantly, Figure 2, Panel B also illustrates that college-educated Americans report systematically higher self-rated health across all predictor groups. College graduates report self-rated health that is between 0.22 and 0.47 (0.38 and 0.83 std. deviations) higher than those who have a high school education or less.

Taken together, these results reveal that college-educated Americans experience a dual-advantage with respect to their self-rated health. On the one hand, they are more likely to sort into groups that have high self-rated health on average—case in point, the group reporting no bad physical health days, no diabetes, and who exercises. On the other hand, they also report systematically higher self-rated health compared to people in with the same levels of physical health. Taking again the example of the group reporting no bad physical health days, no diabetes, and who exercises, college-educated Americans report 0.17 (0.29 std. deviations) higher self-rated health than those with some college and 0.33 (0.58 std. deviations) higher self-rated health than those with a high school education or less.

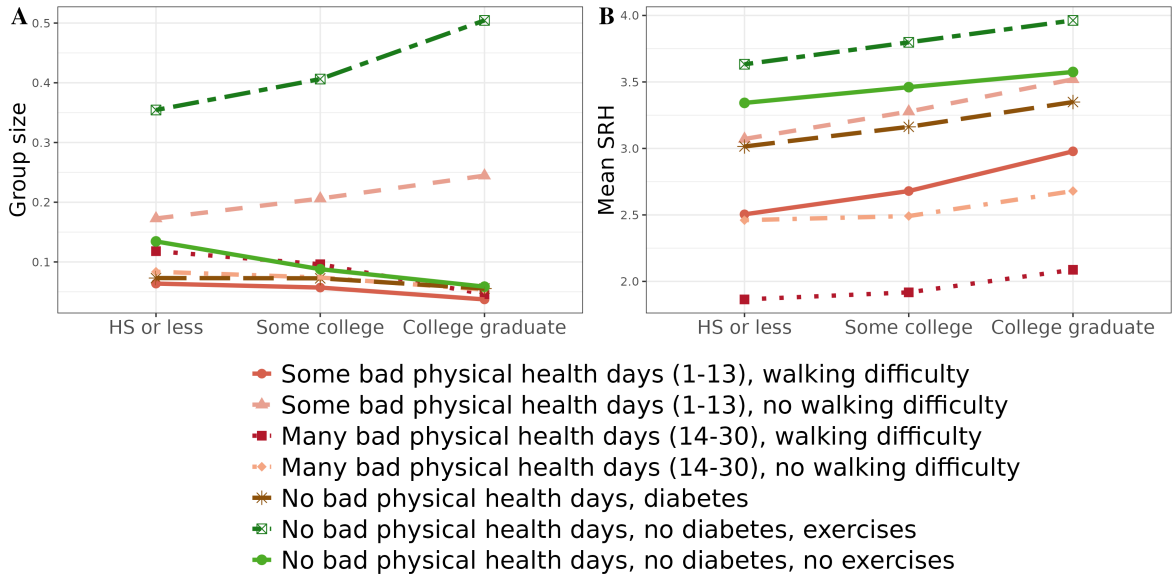


Figure 2: Panel A plots the proportion of individuals in the test set that were assigned to each of the predictor groups, separately by level of education. Panel B plots the mean self-rated health within those groups.

5 Next Steps

In order to scaffold our next steps, it is worth highlighting two key findings from our work. The first is the strong relationship between physical mobility and self-rated health. The second is that college educated Americans experience a dual advantage with respect to the relationship between education and self-rated health (in terms of the compositional and systematic components of it).

A first next step is to ask whether the dual advantage we document for college-educated Americans operates with the same magnitude across racial and ethnic groups. We will quantify the contributions of the compositional and systematic components to the educational gradient for Black and White Americans, drawing on well-established stratification techniques such as the Kitagawa-Oaxaca-Blinder decomposition. These techniques allow us to estimate, in an accounting sense, the share of the educational gap in self-rated health attributable to differences in composition versus within-group rating differences. This question speaks directly to a literature documenting that the health returns to education are smaller for racial minorities than for White Americans on outcomes like mortality, but that has largely not analyzed subjective measures (Farmer and Ferraro, 2005; Hayward et al., 2000; Hummer and Hernandez, 2013; Hendi, 2017).

Our second next step will focus on exploring the mechanisms that underpin education-based sorting into population groups with different levels of self-rated health, on average. To this end, we plan on assessing the role of work and occupations in shaping lower-educated Americans' exposure to poor physical health. Lower- and higher-educated Americans differ in the everyday tasks they carry out at their jobs. Some of these tasks, such as high levels of physical activity, lifting, and stopping postures, have been shown to shape physical mobility (Andrasfay et al., 2023). Following on from this, a natural question is whether the cumulative

effects of exposure to adverse work conditions can explain the strong relationship between physical mobility and self-rated health that we found. This next step leverages a practical advantage of our methods: they estimate predictors that can be leveraged in other datasets. The BRFSS does not contain occupational data, and so we plan to construct comparable predictor groups in the Health and Retirement Study (HRS).

Appendix

	Predictor	Group Size	Mean SRH
1	`PHYSHLTH >= 82.5 & DIABETE4 >= 1.5 & EXERANY2 < 1.5`	0.436	3.83
2	`PHYSHLTH >= 82.5 & DIABETE4 >= 1.5 & EXERANY2 >= 1.5`	0.090	3.45
3	`PHYSHLTH < 82.5 & PHYSHLTH < 14.5 & DIFFWALK >= 1.5`	0.210	3.34
4	`PHYSHLTH >= 82.5 & DIABETE4 < 1.5`	0.065	3.18
5	`PHYSHLTH < 82.5 & PHYSHLTH < 14.5 & DIFFWALK < 1.5`	0.051	2.74
6	`PHYSHLTH < 82.5 & PHYSHLTH >= 14.5 & DIFFWALK >= 1.5`	0.069	2.57
7	`PHYSHLTH < 82.5 & PHYSHLTH >= 14.5 & DIFFWALK < 1.5`	0.080	1.92

Table A1: This table displays the predictors estimated by the regression tree using the original variable names. The "Group Size" column contains the proportion of the test set that falls into the predictor group. The "Mean SRH" column contains the mean self-rated health among test set observations who fall into that predictor group. `PHYSHLTH` refers to the number of bad physical health days and corresponds with the answer to: *Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good?*. `DIABETE4` refers to whether an individual has been diagnosed with diabetes and corresponds with the answer to: *Ever told you had diabetes?*. `EXERANY2` refers to whether an individual exercises and corresponds with the answer to: *During the past month, other than your regular job, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?*. `DIFFWALK` refers to whether an individual has difficulty walking and corresponds with the answer to: *Do you have serious difficulty walking or climbing stairs?*

References

Andrasfay T, Pebley AR, Goldman N (2023) Physical Work Exposures of Older Workers: Does Measurement Make a Difference? Work, Aging and Retirement 9(2):179–189. https://doi.org/10.1093/workar/waab014, URL https://doi.org/10.1093/workar/waab014

Assari S, Lankarani MM, Burgard S (2016) Black White Difference in Long Term Predictive Power of Self-Rated Health on All-Cause Mortality in United States. Annals of epidemiology 26(2):106–114. https://doi.org/10.1016/j.annepidem.2015.11.006, URL https://pmc.ncbi.nlm.nih.gov/articles/PMC4792698/

Bamia C, Orfanos P, Juerges H, et al (2017) Self-rated health and all-cause and cause-specific mortality of older adults: Individual data meta-analysis of prospective cohort studies in the CHANCES Consortium. Maturitas 103:37–44. https://doi.org/10.1016/j.maturitas.2017.06.023

Beck AN, Finch BK, Lin SF, et al (2014) Racial disparities in self-rated health: Trends, explanatory factors, and the changing role of socio-demographics. Social Science & Medicine 104:163–177. https://doi.org/10.1016/j.socscimed.2013.11.021, URL https://linkinghub.elsevier.com/retrieve/pii/S0277953613006242

Boen C (2016) The role of socioeconomic factors in Black-White health inequities across the life course: Point-in-time measures, long-term exposures, and differential health returns. Social Science & Medicine 170:63–76. https://doi.org/10.1016/j.socscimed.2016.10.008, URL https://linkinghub.elsevier.com/retrieve/pii/S0277953616305676

Charmaz K (1990) 'Discovering' chronic illness: Using grounded theory. Social Science & Medicine 30(11):1161–1172. https://doi.org/10.1016/0277-9536(90)90256-R, URL https://linkinghub.elsevier.com/retrieve/pii/027795369090256R

Choi JHS, Jung DH (2025) The role of psychological factors in predicting self-rated health: implications from machine learning models. Psychology, Health & Medicine 30(6):1158–1170. https://doi.org/10.1080/13548506.2025.2450546, URL https://doi.org/10.1080/13548506.2025.2450546, _eprint: https://doi.org/10.1080/13548506.2025.2450546

Clark CR, Ommerborn MJ, Moran K, et al (2021) Predicting Self-Rated Health Across the Life Course: Health Equity Insights from Machine Learning Models. Journal of General Internal Medicine 36(5):1181–1188. https://doi.org/10.1007/s11606-020-06438-1, URL https://doi.org/10.1007/s11606-020-06438-1

Cobb S, Assari S (2020) Investigation of the Predictors of Self-rated Health of Economically Disadvantaged African American Men and Women: Evidence for Sponge Hypothesis. International journal of epidemiologic research 7(1):25–34. https://doi.org/10.34172/ijer.2020.05, URL https://pmc.ncbi.nlm.nih.gov/articles/PMC7213594/

Cummings JL (2023) Entangled Inequalities: U.S. Trends in Self-Rated Health at the Intersection of Gender and Race, 1972–2018. Social Indicators Research 169(3):759–774. https://doi.org/10.1007/s11205-023-03180-z, URL https://link.springer.com/10.1007/s11205-023-03180-z

DeSalvo KB, Bloser N, Reynolds K, et al (2006) Mortality prediction with a single general self-rated health question. A meta-analysis. Journal of General Internal Medicine 21(3):267–275. https://doi.org/10.1111/j.1525-1497.2005.00291.x

Dowd JB, Zajacova A (2010) Does self-rated health mean the same thing across socioeconomic groups? Evidence from biomarker data. Annals of Epidemiology 20(10):743–749. https://doi.org/10.1016/j.annepidem.2010.06.007

Farmer MM, Ferraro KF (2005) Are racial disparities in health conditional on socioeconomic status? Social Science & Medicine 60(1):191–204. https://doi.org/10.1016/j.socscimed.2004.04.026

Glaser BG, Strauss AL (2017) The discovery of grounded theory: strategies for qualitative research, first issued in hardback edn. Routledge, Taylor & Francis Group, London New York

Gumà-Lao J, Arpino B (2023) A machine learning approach to determine the influence of specific health conditions on self-rated health across education groups. BMC Public Health 23(1):131. https://doi.org/10.1186/s12889-023-15053-8, URL https://doi.org/10.1186/s12889-023-15053-8

Hayward C, Varady S, Albano AM, et al (2000) Cognitive-behavioral group therapy for social phobia in female adolescents: results of a pilot study. Journal of the American Academy of Child and Adolescent Psychiatry 39(6):721–726. https://doi.org/10.1097/00004583-200006000-00010

Hendi AS (2017) Trends in Education-Specific Life Expectancy, Data Quality, and Shifting Education Distributions: A Note on Recent Research. Demography 54(3):1203–1213. https://doi.org/10.1007/s13524-017-0574-2

Hummer RA, Hernandez EM (2013) The Effect of Educational Attainment on Adult Mortality in the United States. Population Bulletin 68(1):1–16

Idler E, Cartwright K (2018) What Do We Rate When We Rate Our Health? Decomposing Age-related Contributions to Self-rated Health. Journal of Health and Social Behavior 59(1):74–93. https://doi.org/10.1177/0022146517750137

Jylhä M (2009) What is self-rated health and why does it predict mortality? Towards a unified conceptual model. Social Science & Medicine 69(3):307–316. https://doi.org/10.1016/j.socscimed.2009.05.013, URL https://linkinghub.elsevier.com/retrieve/pii/S0277953609002925

Lamidi EO (2020) Trends in self-rated health by union status and education, 2000-2018. SSM - population health 11:100616. https://doi.org/10.1016/j.ssmph.2020.100616

Schnittker J (2005) When mental health becomes health: age and the shifting meaning of self-evaluations of general health. The Milbank Quarterly 83(3):397–423. https://doi.org/10.1111/j.1468-0009.2005.00407.x

Schnittker J, Bacak V (2014) The Increasing Predictive Validity of Self-Rated Health. PLoS ONE 9(1):e84933. https://doi.org/10.1371/journal.pone.0084933, URL https://dx.plos.org/10.1371/journal.pone.0084933

Stewart RAH, Hagström E, Held C, et al (2017) Self-Reported Health and Outcomes in Patients With Stable Coronary Heart Disease. Journal of the American Heart Association 6(8):e006096. https://doi.org/10.1161/JAHA.117.006096