Journal format to use as template

Sandesh Manjunath Raykar   Daniel Lee

This study examines two related questions using data from the 2016–2017 American Health Values Survey. The first investigates whether individuals who report putting greater effort into maintaining their health are less likely to have been diagnosed by a doctor with high blood pressure. The second explores the association between annual household income and the time since an individual's last routine checkup. Using logistic and ordinal regression analyses, the findings suggest that higher self-reported health effort is associated with lower odds of a high blood pressure diagnosis, and that higher income predicts more recent medical checkups. These results highlight socioeconomic and behavioral disparities in health outcomes and access to preventive care.

Introduction and Background

Preventive healthcare plays a critical role in early disease detection and reducing long-term health costs. Regular checkups, screenings, and healthy lifestyle behaviors help identify potential health risks before they become serious, ultimately improving quality of life and reducing the financial burden on individuals and the healthcare system. Despite these benefits, engagement in preventive health practices varies widely among Americans, reflecting differences in awareness, motivation, and access to care.

To better understand patterns in Americans' health, this study takes two complementary perspectives. The first focuses on individual effort to maintain or improve health, capturing personal behaviors such as exercise, healthy eating, and proactive health management. The second examines access to routine medical checkups influenced by income level, representing the structural and socioeconomic conditions that shape preventive healthcare utilization. Together, these perspectives provide a more holistic understanding of how both personal and contextual factors contribute to preventive health outcomes in the United States.

Previous research shows that family income strongly influences access to medical care and preventive services. Adults with lower income levels face greater barriers to care, often forgoing medical visits or prescriptions due to cost, while even brief insurance disruptions reduce preventive services such as routine checkups. Building on this evidence, the present study investigates how socioecono mic factors—particularly annual household income—relate to both

health effort and the time since one's last routine checkup. By analyzing these relationships, this study seeks to highlight socioeconomic disparities in preventive healthcare among U.S. adults and contribute to a deeper understanding of the behavioral and structural determinants of health.

Study Design and Data collection

This study employs a cross-sectional observational design using data from the 2016–2017 American Health Values Survey, a nationally representative dataset of U.S. adults. The analysis focuses on respondents aged 18 and older who provided complete information on key variables, including self-reported effort in maintaining the health, blood pressure diagnosis, annual household income, and time since last routine medical checkup. Data were collected through structured survey questionnaires administered online and by phone, ensuring coverage across diverse demographic and socioeconomic groups. This design allows for examining associations between behavioral and socioeconomic factors and health outcomes without manipulating variables or establishing causality.

This study utilized a cross-sectional observational design drawing upon data from the 2016–2017 American Health Values Survey (AHVS), conducted by NORC at the University of Chicago with support from the Robert Wood Johnson Foundation. The AHVS is a nationally representative survey designed to assess the health beliefs, values, and behaviors of U.S. adults across diverse demographic and regional contexts. The analysis in this study was restricted to respondents aged 18 years and older who provided complete responses to key variables, including self-reported health effort, blood pressure diagnosis, annual household income, and time since last routine medical checkup.

Data were collected using a multimode survey design, incorporating online (CAWI), mail (self-administered paper questionnaires), and telephone interviews (CATI). This mixed-method approach enhanced accessibility for respondents with varying levels of internet access and literacy, ensuring more comprehensive coverage across urban and rural populations. The survey was available in both English and Spanish, and participants received small prepaid and contingent incentives to improve response rates The AHVS and the related Sentinel Community Surveys collected detailed information on participants' health values, personal health priorities, self-efficacy, access to care, and social determinants of health. The structured questionnaire captured both quantitative and categorical data, allowing for the examination of associations between behavioral, social, and economic factors and various health outcomes without any manipulation of variables. Because of its cross-sectional nature, this design allows for identifying correlations and group differences but does not infer causality.

Methods

Use separate sub-sections for data preparation and statistical analysis methods.

Data Preparation

Data for this study were obtained from the 2016–2017 American Health Values Survey. The dataset was first cleaned and filtered to include only respondents aged 18 years and older to focus the analysis on adults. Observations with missing or incomplete responses for key analytical variables—specifically health effort, annual household income, and time since last routine medical checkup—were excluded to ensure data quality and consistency. This filtering process helped minimize bias that could result from incomplete information.

Before computing summary measures, several data-cleaning procedures were implemented. Responses coded as 77 ("Don't know") or 99 ("No answer") were treated as invalid and replaced with missing values (NA) across relevant variables, particularly the health activity items (act1–act7) that contributed to the health effort measure. A new variable, healtheffort, was then created as the mean of act1 through act7 for each respondent, with missing values excluded from the calculation. This variable reflects an individual's overall effort toward maintaining or improving their health.

The income and checkup variables underwent the same cleaning and recoding procedures. Responses coded as 77 ("Don't know") and 99 ("Refused") were treated as missing values (NA). The income variable was categorized into eight ordered levels, ranging from less than $15,000 to $150,000 or more, while the checkup variable was recoded into five ordered groups, from within the past year to never. Data integrity checks were then performed to ensure that all invalid codes were removed and all transformations were accurately implemented. These procedures produced a reliable dataset suitable for statistical analysis and modeling.

Statistical Analysis

The statistical analysis for this study was guided by the analytical framework used in the American Health Values Survey (AHVS) and the Sentinel Community Health Values Surveys (SCHVS). The AHVS employed k-means clustering, a widely used unsupervised classification technique, to identify distinct groups of respondents based on their health values, beliefs, and behaviors. This method partitions the dataset into k mutually exclusive clusters by minimizing the within-cluster variance while maximizing the differences between cluster centroids.

Following the AHVS methodology, several solutions ranging from four to ten clusters were examined to determine the most meaningful group segmentation. The final cluster solution was selected based on statistical indicators such as the cubic clustering criterion and the Pseudo F statistic, ensuring that the identified typology best captured variation in respondents' health attitudes and behaviors. Demographic and descriptive measures were then compared across clusters to confirm the face validity of the typology, ensuring that the observed groupings aligned with known differences in health, political, and socioeconomic characteristics among U.S. adults.

For the present analysis using the cleaned AHVS dataset, descriptive statistics were first computed to summarize respondent demographics and key health indicators (e.g., health effort,