UC Irvine
UC Irvine Previously Published Works

Title
A Data-Driven Approach to Identify the Predictors of Perceived Health Status Among Chinese and Korean Americans.

Permalink
https://escholarship.org/uc/item/42d3z4t1

Journal
Computers in nursing, 41(9)

Authors
Park, Jung In
Lee, Grace
Lee, Sunmin

Publication Date
2023-09-01

DOI
10.1097/CIN.0000000000000995

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access
Author manuscript
Comput Inform Nurs. Author manuscript; available in PMC 2024 September 01.

Published in final edited form as:
Comput Inform Nurs. ; 41(9): 730–737. doi:10.1097/CIN.0000000000000995.

A Data-Driven Approach to Identify the Predictors of Perceived Health Status Among Chinese and Korean Americans

Jung In Park, PhD, RN[1] **[Assistant Professor]**, **Grace Eunyoung Lee**[2] **[Health Educator]**, **Sunmin Lee, ScD, MPH**[2] **[Professor]**

[1]University of California- Irvine, Sue & Bill Gross School of Nursing, Irvine, CA

[2]University of California-Irvine, Department of Medicine, School of Medicine, Irvine, CA

Abstract

Asian Americans are the country's fastest-growing racial group, and several studies have focused on the health outcomes of Asian Americans, including perceived health status. Perceived health status provides a summarized view of the health of populations for diverse domains, such as the psychological, social, and behavioral aspects. Given its multifaceted nature, perceived health status should be carefully approached when examining any variables' influence because it results from interactions among many variables. A data-driven approach using machine learning provides an effective way to discover new insights when there are complex interactions among multiple variables. To date, there are not many studies available that use machine learning to examine the effects of diverse variables on the perceived health status of Chinese and Korean Americans. This study aims to develop and evaluate three prediction models using logistic regression, random forest, and support vector machines to find the predictors of perceived health status among Chinese and Korean Americans from survey data. The prediction models identified specific predictors of perceived health status. These predictors can be utilized when planning for effective interventions for the better health outcomes of Chinese and Korean Americans.

Keywords

Perceived health status; self-rated health; Asian Americans; Data-driven approach; Machine Learning; prediction modeling

INTRODUCTION

According to the U.S. Census Bureau population estimates, Asian Americans are the country's fastest-growing racial group, now numbering about 19 million nationwide.[1]

Corresponding Author: Jung In Park, 802 W Peltason Dr, 100D Berk Hall, Irvine, CA 92617, 949-824-4118, junginp@hs.uci.edu.

Conflicts of interest/Competing interests: Except the funding source, the authors have no additional conflicts of interest of competing interests to declare. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Ethics approval: This study received ethics approval from the Institutional Review Boards of University of Maryland and University of California, Irvine.

Consent to participate: All respondents completed informed consent prior to participating in the study.

Consent for publication: All of the authors consent to publishing this manuscript.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Because of the growing number of Asian Americans that reside in the United States, several studies have focused on the health outcomes of Asian Americans—and perceived health status is one of the outcomes of interest.[2–4] Perceived health status reflects people's overall perception of their health, and it is predictive of people's future health care use and associated health outcomes, such as various chronic diseases and mortality.[5–7] Moreover, quality of life is directly related to the perception of health because it assesses how people think about their health status and other aspects of their lives.[8] Perceived health status, often used interchangeably with "self-rated health," provides a summarized view of the health of populations for diverse domains, such as the psychological, social, and behavioral aspects.[9–10] Given its multifaceted nature, perceived health status should be carefully approached when examining specific variables' influence because it results from interactions among many variables.

A data-driven approach using machine learning provides an effective way to discover new insights when there are complex interactions among multiple variables. This approach is often used for prediction modeling and identifying predictors because this process includes an exhaustive search to find valid and useful patterns from collections of data. To date, not many studies have been conducted to identify the predictors of perceived health status among Asian Americans using a data-driven approach. Many studies focus on the relationships of one or two variables to perceived health status among Asian Americans; however, traditional analytic methods have limitations when employed to investigate multiple interactions among many variables.[11] The present study uses a data-driven approach to unearth new information from the survey data about the predictors of perceived health status among Chinese and Korean Americans. Identifying the predictors of outcome allows timely and proper interventions for those who might benefit.[12]

The overall purpose of this study is to identify predictors of perceived health status among Chinese and Korean Americans. More specifically, in this study, we aim to develop and validate prediction models to find the predictors of perceived health status among Chinese and Korean Americans. We utilized machine learning approaches such as logistic regression (LR), random forest (RF), and support vector machines (SVM). We then evaluated the prediction models and identified the predictors of perceived health status among Chinese and Korean Americans. The findings from this study can help understand the culturally sensitive aspects of Chinese and Korean Americans to inform and improve the nursing care tailored to these populations.

METHODS

Data Source

This study used the data from a randomized controlled trial designed to increase colorectal cancer screening among 400 Chinese and Korean Americans (200 Chinese and 200 Korean Americans) living in the Washington, DC, metropolitan area. Study participants were between the ages of 50 and 75 and were recruited from primary care physicians' clinics. More description of the data is available elsewhere.[7] In the current study, we used the baseline survey data collected from August 2018 to June 2020. Ninety-two percent of the data collection was completed in-person, while the remaining 8% were collected by phone

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

because of the COVID-19 outbreak in March 2020. After providing informed consent, the participants completed a self-administered questionnaire (for in-person data collection) or a researcher-led phone survey (for phone-based data collection) in their preferred language (Mandarin, Korean, or English). Bilingual staff translated the survey questionnaires and independently back translated them into the three available languages. The Institutional Review Board of the University of California, Irvine (#20216482), approved this study.

Predictor Variables and Outcome Variables

The predictor variables were extracted from the survey questionnaires. The survey consists of the following sections: demographic information including gender, age, and ethnicity; marital status; socioeconomic factors including education, income, employment status, and health care coverage; immigration-related factors including years in the United States; English-speaking proficiency and ethnic identity; acculturative stress; perceived stress; distress; social support; sleep disturbance, sleep apnea, and sleep duration; and chronic conditions. Also, the data included weight, height, BMI, blood pressure, and biomarkers from blood work, including levels of cholesterol, glucose, and triglycerides.

The primary outcome variable was the perceived health status of Chinese and Korean Americans. The question was, “Would you say that in general, your health is excellent, very good, good, fair, or poor?” The responses were binarized into two categories: “excellent,” “very good,” and “good” were considered as “good perceived health,” whereas “fair” and “poor” were considered as “poor perceived health.”

Data Preprocessing and Feature Selection

Raw input data were transformed into a format that a machine learning prediction model could effectively recognize and compute. We applied min-max normalization to the ranges of ages and years lived in the United States. The min-max scaler transforms features by scaling each feature to a given range between 0 and 1 to avoid the unintended weighting of certain features. Also, categorical features were pivoted into binary features using one-hot encoding; if a data point belonged to the category, then it was assigned 1 and 0 if not.

Lastly, a variance threshold feature selector was used to select relevant features for the model and eliminate the low variance features below 0.05, because to keep them would only increase the computational cost to the model without adding extra information. We divided the data set into a training set and a test set to build and test the model using an unseen data set.

Prediction Model Development and Evaluation

We trained and compared the performance of three prediction models using LR, RF, and SVM. The LR model uses an odds ratio (OR) to predict the odds of an event’s occurrence. The ORs show the magnitude of each feature’s contribution to the outcome variable. The RF model constructs a large number of individual decision trees at training and operates as an ensemble for more accurate prediction. Each tree outputs a class prediction, and the class with the most votes becomes the RF model’s prediction. The RF model computes each feature’s importance score, which is calculated based on how critical the feature is

Author Manuscript

for model development when removed. The score can be used to interpret which features are important for predicting the outcome. The SVM model is one of the most powerful classification algorithms for predictive accuracy. However, its decision process is difficult to interpret because of the black-box approach. That is, this model shows robust performance in classification but does not report which features contribute to the outcome. Although the SVM model did not identify the predictors in this study, we still included this model to compare the predictive performance with the other interpretable models.

Performance evaluation metrics that we used included *area under the receiver operating characteristic* (AUROC) *curve* to measure the overall model performance, *sensitivity* to capture the ability of a test to be positive when the condition is present, and *specificity* to measure the proportion of true negatives that are correctly identified as negative. These measures range up to 1, with higher values suggesting better performance. In this study, the parameters of the models were tuned to enhance the AUROC curve and to optimize the overall model performance.

Tenfold cross-validation was used to validate the predictive accuracy of each machine learning model. In tenfold cross-validation, the data are divided randomly into ten equal partitions, and each partition is used for testing, while the remainder is used for training. Python software v.3.8.3 was used for (a) data preprocessing; (b) prediction model development for LR, RF, and SVM (Scikit-learn v0.23.1); and (c) evaluation of prediction models. The Scikit-learn v0.23.1 library in Python was used to optimize the hyperparameters of the prediction methods. The Scikit-learn is commonly used and suitable for many machine learning applications, although it lacks compatibility with deep learning-oriented modeling using high-dimensional data such as medical images.

RESULTS

Data Preprocessing

To preprocess the data for model development, we excluded nonfeature columns such as IDs, comments, and interviewer information. Also, we excluded data that had missing values (8% of the data), such as weight, height, BMI, blood pressure, and biomarkers from blood work including glucose, and cholesterol level. The partial availability of the data is due to a change in data collection method from in-person (92% of data) to phone (8% of data) after the COVID-19 outbreak in 2020. Then, we transformed the features for effective modeling, including categorical features pivoted into one-hot encoding (1 and 0).

Acculturative stress was measured using nine questionnaires, and the responses were recorded as follows: “yes” as 1, “no” as 2, and “n/a” as 9. Distress was measured using 20 questionnaires that fell under one of these categories: practical problems (range 0–5), family problems (range 0–3), emotional problems (range 0–6), physical problems (range 0–5), and spiritual/ religious problems (range 0–1). The responses were converted to 1 and 0 and then aggregated by each category. Perceived stress was measured using 10 questions, and the responses were categorized as “very often,” “fairly often,” “sometimes,” “almost never,” and “never.” The responses about social support were also categorized: “as much as I would like,” “almost as much as I would like,” “some, but would like more,” “less than I

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

would like," and "much less than I would like." Then, we dropped the low-variance features in which the thresholds were less than 0.05.

The data set was randomly divided into two sets: a training set, which contained about 70% of the data and which was used to develop the prediction models; and a test set, which contained about 30% of the data and which was used to test the prediction performance of the trained model.

Feature Selection and Sample Characteristics

After data preprocessing and feature selection, the number of final features available for analysis was 49. The final feature categories included gender, age, ethnicity, highest level of education, marital status, working hours, annual household income, health care coverage, ethnic identity, English proficiency, perceived stress, level of distress in the past week, problems experienced in the past week, social support, snoring during sleep, medical history, and currently taking medication for hypertension, hypercholesterolemia, and diabetes.

Among the 400 participants, 223 participants (55.8%) perceived their health as good, whereas 177 participants (44.2%) perceived their health as poor. Table 1 reports the sample characteristics for the selected features. A higher percentage of females (60.5%) perceived their health as poor compared to males (40%). A slightly higher percentage of Korean Americans (54.2%) perceived their health as poor compared to Chinese Americans (53.4%). People who perceived their health as poor were slightly older (mean: 58.9 years old; SD: 6.3) than those who perceived their health as good (mean: 58.0; SD: 6.4).

People who perceived their health as good had a higher level of education (college graduate or above; 59.2%) than those who perceived their health as poor (37.3%). Annual household income (mean) was higher for those who perceived their health as good ($60,000–69,999) than those who perceived their health as poor ($40,000–49,999). More people who perceived their health as good were married (83.4%) than those who perceived their health as poor (81.4%). A higher percentage of people who perceived their health as good was working full time (64.1%) compared to those who perceived their health as poor (50%). A higher percentage of people who perceived their health as good had health care coverages (83.0%) compared to those who perceived their health as poor (74.6%). The majority of the members of both groups were covered by a private health insurance plan (65.5% and 53.7%, respectively).

A significantly higher percentage of people who perceived their health as poor rated themselves as "very Asian" (71.8%) compared to those who perceived their health as good (51.6%). No participant (0.0%) from the group who perceived their health as poor reported themselves as "westernized." A significantly higher percentage of people who perceived their health as good reported that they speak English well or fluently (32.7%), compared to those who perceived their health as poor (10.1%).

People who perceived their health as poor reported higher distress levels (4.14) in the past week than those who perceived their health as good (3.25). More people who perceived their health as poor reported that they had some problems in their life in the past week than those

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

who perceived their health as good. The problems included (a) family problems (0–3 range; 0.62 poor perceived health vs. 0.35 good perceived health), (b) emotional problems (0–6 range; 1.32 poor perceived health vs. 0.70 good perceived health), (c) physical problems (0–5 range; 1.94 poor perceived health vs. 0.97 good perceived health), and (d) spiritual/religious problems (0–1 range; 0.08 poor perceived health vs. 0.05 good perceived health).

Significantly higher percentages of people who perceived their health as good reported that they had social support. The responses included (a) "I have people who care what happens to me (71.8% good perceived health vs. 61.0% poor perceived health)," (b) "I get love and affection (79.8% good perceived health vs. 67.2% poor perceived health)," (c) "I get chances to talk to someone about problems at work or with my housework (80.2% good perceived health vs. 61.0% poor perceived health)," (d) "I get chances to talk to someone I trust about my personal or family problems (76.3% good perceived health vs. 63.8% poor perceived health)," (e) "I get chances to talk about money matters (72.7% good perceived health vs. 55.9% poor perceived health)," (f) "I get invitations to go out and do things with other people (75.4% good perceived health vs. 58.2% poor perceived health)," (g) "I get useful advice about important things in life (74.0% good perceived health vs. 59.3% poor perceived health)," and (h) "I get help when I am sick in bed (79.5% good perceived health vs. 63.2% poor perceived health)."

More people who perceived their health as poor reported that they snored during sleep (53.1%) compared with people who perceived their health as good (46.2%). Higher percentages of people who perceived their health as poor had a medical history, including high blood pressure (41.8% poor perceived health vs. 27.8% good perceived health), high cholesterol (53.7% poor perceived health vs. 31.8% good perceived health), heart attack or any other heart disease (9.0% poor perceived health vs. 3.6% good perceived health), diabetes (29.4% poor perceived health vs. 12.6% good perceived health), anxiety or depression (7.9% poor perceived health vs. 4.0% good perceived health), obesity (20.9% poor perceived health vs. 9.0% good perceived health), and any other health problem (15.8% poor perceived health vs. 8.5% good perceived health). More people who perceived their health as poor reported that they were taking medication, including hypertension medication (36.2% poor perceived health vs. 25.1% good perceived health), cholesterol medication (34.5% poor perceived health vs. 15.7% good perceived health), and diabetes medication (20.3% poor perceived health vs. 9.0% good perceived health).

Prediction Model Development and Evaluation

Three prediction models were developed using LR, RF, and SVM. For the LR model, we used standard implementation from the Scikit-learn library in Python. We used L2-norm for the penalty and the L-BFGS optimizer with enough iterations to full converge under the given tolerance (0.0001). For the RF model, the hyperparameters were tuned with tenfold cross validation for given parameter ranges to maximize AUROC curve. For the RF model, the number of trees and the maximum depth of a tree were tuned. We found that a maximum depth of 4 and 40 trees were optimal for our data. For the SVM model, we used 1.0 for the regularization parameter and the linear kernel for the model fitting.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Predictive performance was evaluated using the AUROC curve, sensitivity, and specificity (Table 2). The RF model slightly outperformed the LR and SVM models in AUROC curve (0.76) and recall (0.88), whereas the SVM model had the highest precision (0.78).

Predictors of good perceived health status for Chinese and Korean Americans were identified from the two interpretable models, the LR and RF models. The 10 selected features with the highest ORs from the LR models are presented in Table 3. It includes (1) higher annual household income, (2) not having medical history: obesity, (3) not having medical history: high cholesterol, (4) highest level of education: attended graduate or professional school, (5) westernized ethnic identity, (6) not having medical history: diabetes, (7) having social support: “I get useful advice about important things in life,” (8) not having medical history: any other health problems, (9) not taking cholesterol medication, and (10) not having medical history: high blood pressure. Table 4 reports the 10 features that showed the highest importance in the RF model for better perceived health status. These include (1) better English proficiency, (2) problems experienced in the past week: physical problems, (3) higher annual household income, (4) highest level of education: attended graduate or professional school, (5) not having medical history: obesity, (6) having social support: “I get chances to talk to someone I trust about my personal or family problems,” (7) having social support: “I get useful advice about important things in life,” (8) not having medical history: high cholesterol, (9) having social support: “I get chances to talk about money matters,” and (10) younger age. The higher annual household income, not having medical history: high cholesterol and obesity, the highest level of education: attended graduate or professional school, and having social support: “I get useful advice about important things in life,” features were found in both LR and RF models.

DISCUSSION

Perceived health status is a subjective measure, but it is a strong predictor of patient outcomes, especially mortality.[13] Multiple factors affect an individual’s perception of their health status, and this should be considered when analyzing the data. A data-driven approach using machine learning provides an effective prediction model that can identify factors derived from complicated data. To date, there are not many studies available that use machine learning to examine the effects of diverse variables on the perceived health status of Chinese and Korean Americans. This study developed three accurate prediction models using machine learning, two of which identified predictors of better perceived health status among Chinese and Korean Americans that can be implemented to guide quality health care planning and tailored nursing interventions.

The LR and RF models identified multiple predictors for the perceived health status of Chinese and Korean Americans. According to these models, higher annual household income and the highest level of education: attended graduate or professional school were two predictors of good perceived health status. This result supports the findings from previous studies that identified household income and education are related, and people who have higher social standing will have better health[14–16]; in particular, education was associated with better self-rated health in both Chinese and Korean Americans.[17–18] Several studies have claimed that education and income are linked through occupation.[19–20] Thus, it

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

is likely that the participants with higher education had highly paid occupations with better health benefits, and this could lead to better perceived health by enabling them to check their health more frequently. It is also supported by the sample characteristics that a higher percentage of people who perceived their health as good had health care coverage than those who perceived their health as poor.

Furthermore, Phelan et al. noted[21] that income and education are considered 'fundamental causes of health.' Socioeconomic status embodies an array of resources such as money, knowledge, prestige, power, and beneficial social connections that protect health no matter what mechanisms are relevant at any given time. Therefore, the underlying fact is that those from low socioeconomic status lack resources to protect and/or improve their health.

Westernized ethnic identity and better English proficiency were also important predictors of perceived health status. A significant portion of people who perceived their health as poor rated themselves as "very Asian," and no participants of those with poor perceived health thought they were "westernized." Relevantly, this group of people was less confident about their English proficiency. It is possible that because of the ethnic identity and English proficiency, some people might have experienced difficulty accessing primary physicians as well as important health information that contribute to health promotion and disease prevention, and eventually this could have led to a poor perception of their health status. This finding is consistent with previous literature indicating that individuals with lower levels of acculturation or living in ethnic enclaves with high spatial clustering of Asian Americans were likely to have poorer self-rated health.[2–4, 22]

Perceived stress related to physical problems, such as constipation, diarrhea, fatigue, memory/concentration, or sleep, was also identified as the predictor for perceived health status. A significantly higher percentage of people with poor health perception reported they had some problems in their life, including family, emotional, physical, and spiritual/religious problems. These findings support the conclusions of prior literature that good relationships with families and friends influence self-rated health.[14] Consequentially, people with poor health perception showed higher distress levels than those with good health perception. This finding suggests that effective intervention strategies for perceived health outcomes have to consider how to manage people's stress and lower their distress levels.

The study also found that having social support was a predictor for better-perceived health, which is consistent with previous findings.[22, 23] It is possible that people who had opportunities to talk to someone about problems or important things in their lives were less stressed, and this support positively affected their perceived health status.

Younger age was another predictor for the better perceived health status. Not having medical conditions, such as obesity, high cholesterol, high blood pressure, diabetes, or other health problems, and not taking cholesterol medication were also predictors for the better health perception. It is likely that the young and people without health conditions tend to think they are healthy. This finding is congruent with previous studies that identified how people's health conditions influenced their self-rated health.[3, 24] It will add value to see in future

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

work if this machine learning modeling is replicated for other races/ethnicity and findings are compared.

CONCLUSION

This study developed three accurate prediction models for the perceived health status of Chinese and Korean Americans using machine learning, and two of these models identified specific predictors for self-rated health status. Given the numerous factors that affect the perception of one's health, it is necessary to analyze the influence of and the interactions among multiple variables. The predictors of perceived health status identified in this study can be utilized when planning for effective nursing interventions for the better health outcomes of Chinese and Korean Americans.

Funding:

Research reported in this publication was supported by the National Institute On Minority Health And Health Disparities of the National Institutes of Health under Award Number R01MD012778.

REFERENCES

1. United States Census Bureau. 2020 Decennial Census Data. https://data.census.gov/cedsci/profile?q=United%20States&g=0100000US. Accessed January 17, 2022.
2. Lee S, Chen L, He X, Miller MJ, Juon HS. A cluster analytic examination of acculturation and health status among Asian Americans in the Washington DC metropolitan area, United States. Social science & medicine. 2013 Nov 1;96:17–23. [PubMed: 24034947]
3. Jang Y, Huang YC, Yoon H, Lin S. Correlates of self-rated health and self-rated mental health in older Chinese Americans. Social work in public health. 2016 Jun 6;31(4):309–15. [PubMed: 27104949]
4. Zhang Q, Metcalf SS, Palmer HD, Northridge ME. Spatial Analysis of Chinese American Ethnic Enclaves and Community Health Indicators in New York City. Frontiers in Public Health. 2022;10.
5. OECD, Perceived health status. OECD-library, 2018. https://www.oecd-ilibrary.org/docserver/health_glance-2017-13-en.pdf?expires=1609480607&id=id&accname=guest&checksum=214CADD790DFC553D90E4A0B340B49A6. Accessed January 17, 2022.
6. Goldberg P, Guéguen A, Schmaus A, Nakache JP, Goldberg M. Longitudinal study of associations between perceived health status and self reported diseases in the French Gazel cohort. J Epidemiology & Community Health. 2001 Apr 1;55(4):233–8.
7. Lee S, Ryu S, Lee GE, Kawachi I, Morey BN, Slopen N. The association of acculturative stress with self-reported sleep disturbance and sleep duration among Asian Americans. Sleep. 2022 Apr;45(4):zsab298. [PubMed: 34922392]
8. Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. Jama. 1994 Aug 24;272(8):619–26. [PubMed: 7726894]
9. Mavaddat N, Valderas JM, Van Der Linde R, Khaw KT, Kinmonth AL. Association of self-rated health with multimorbidity, chronic disease and psychosocial factors in a large middle-aged and older cohort from general practice: a cross-sectional study. BMC family practice. 2014 Dec;15(1):1–1. [PubMed: 24387712]
10. Singh-Manoux A, Martikainen P, Ferrie J, Zins M, Marmot M, Goldberg M. What does self rated health measure? Results from the British Whitehall II and French Gazel cohort studies. J Epidemiology & Community Health. 2006 Apr 1;60(4):364–72.
11. Zhao CM, Luan J. Data Mining: Going beyond Traditional Statistics. New Directions for Institutional Research. 2006;131:7–16.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

12. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, Kripalani S. Risk prediction models for hospital readmission: a systematic review. Jama. 2011 Oct 19;306(15):1688–98. [PubMed: 22009101]
13. Idler E, Leventhal H, McLaughlin J, Leventhal E. In sickness but not in health: self-ratings, identity, and mortality. J health and social behavior. 2004 Sep;45(3):336–56.
14. Knäuper B, Turner PA. Measuring health: improving the validity of health assessments. Quality of Life Research. 2003 Jan;12(1):81–9. [PubMed: 12803314]
15. Cai J, Coyte PC, Zhao H. Determinants of and socio-economic disparities in self-rated health in China. International Journal for Equity in Health. 2017 Dec;16(1):1–27. [PubMed: 28057005]
16. Wagner DC, Short JL. Peer reviewed: longitudinal predictors of self-rated health and mortality in older adults. Preventing chronic disease. 2014;11.
17. John DA, De Castro AB, Martin DP, Duran B, Takeuchi DT. Does an immigrant health paradox exist among Asian Americans? Associations of nativity and occupational class with self-rated health and mental disorders. Social science & medicine. 2012 Dec 1;75(12):2085–98. [PubMed: 22503561]
18. Assari S, Kumar A. Social determinants of physical self-rated health among Asian Americans; comparison of six ethnic groups. Societies. 2018 Jun;8(2):24.
19. Fiscella K, Franks P, Gold MR, Clancy CM. Inequality in quality: addressing socioeconomic, racial, and ethnic disparities in health care. Jama. 2000 May 17;283(19):2579–84. [PubMed: 10815125]
20. Lynch J, Kaplan G. Socioeconomic position. In Berkman LF, & Kawachi I (Eds.), Social epidemiology. New York: Oxford University Press; 2000 Mar 9.
21. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. J health and social behavior. 2010 Mar;51(1_suppl):S28–40.
22. Krokavcova M, van Dijk JP, Nagyova I, Rosenberger J, Gavelova M, Middel B, Gdovinova Z, Groothoff JW. Social support as a predictor of perceived health status in patients with multiple sclerosis. Patient education and counseling. 2008 Oct 1;73(1):159–65. [PubMed: 18467067]
23. Hampton NZ. Subjective well-being among people with spinal cord injuries: The role of self-efficacy, perceived social support, and perceived health. Rehabilitation Counseling Bulletin. 2004 Oct;48(1):31–7.
24. Garbarski D Research in and prospects for the measurement of health using self-rated health. Public Opinion Quarterly. 2016 Dec 1;80(4):977–97. [PubMed: 27833212]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Sample Characteristics for the Selected Features

	Perceived health status: Good (n=223, 55.8%)	Perceived health status: Poor (n=177, 44.2%)
Q1. Gender, n (%)		
Male	119 (53.4%)	70 (40%)
Female	104 (46.6%)	107 (60%)
Q2. Age at the time of interview (continuous variable), mean (SD)	58.0 (6.4)	58.9 (6.3)
Q3. Ethnicity, n (%)		
Korean	104 (46.6%)	96 (54.2%)
Chinese	119 (53.4%)	80 (45.2%)
Q4. Highest level of education, n (%)		
1) Less than high school	20 (9.0%)	23 (13.0%)
2) High school graduate or GED	37 (16.6%)	54 (30.5%)
3) Business or vocational school	14 (6.3%)	17 (9.6%)
4) Some college	20 (9.0%)	17 (9.6%)
5) College graduate	55 (24.7%)	46 (26.0%)
6) Attended graduate or professional school	77 (34.5%)	20 (11.3%)
Q6. Annual household income (ordinal variable), mean (SD)	$60,000-$69,999	$40,000-$49,999
Q7. Marital Status, n (%)		
Married	186 (83.4%)	144 (81.4%)
Divorced	13 (5.8%)	11 (6.2%)
Q8. Employment Status, n (%)		
1) Working full time	143 (64.1%)	88 (50%)
2) Working part time	39 (17.5%)	45 (25.4%)
3) Keeping house	23 (10.3%)	24 (13.6%)
4) Retired	17 (7.6%)	18 (10.2%)
Q9. Health care coverage, n (%)		
Yes	185 (83.0%)	132 (74.6%)
No	38 (17.0%)	45 (25.4%)
Q10. Health care coverage, n (%)		
1) Private health insurance plan	146 (65.5%)	95 (53.7%)
2) Medicare	32 (14.3%)	29 (16.4%)
3) Medicaid	16 (7.2%)	14 (7.9%)
Q12.Ethnic Identity, n (%)		
1) Very Asian	115 (51.6%)	127 (71.8%)
2) Mostly Asian	36 (16.1%)	26 (14.7%)
3) Bicultural	66 (29.6%)	24 (13.6%)
4) Mostly Westernized	5 (2.2%)	0 (0.0%)
5) Very Westernized	1 (0.4%)	0 (0.0%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	Perceived health status: Good (n=223, 55.8%)	**Perceived health status: Poor (n=177, 44.2%)**
Q13. English Proficiency, n (%)		
1) Fluent like a native speaker	11 (4.9%)	2 (1.1%)
2) Well	62 (27.8%)	16 (9.0%)
3) So-So	89 (39.9%)	59 (33.3%)
4) Poorly	51 (22.9%)	76 (42.9%)
5) Not at all	10 (4.5%)	24 (13.6%)
Q20. Perceived Stress, Often, n (%)		
How often have you felt that you were effectively coping with important changes that were occurring in your life?	26 (13.6%)	37 (20.8%)
Q21. Level of distress in the past week,		
(0–10 range; higher the score, higher the stress), mean (SD)	3.25 (2.29)	4.14 (2.49)
Q22. Problems experienced in the past week, mean (SD)		
1) Family Problems (0–3 range)	0.35 (0.67)	0.62 (0.88)
2) Emotional Problems (0–6 range)	0.70 (1.13)	1.32 (1.66)
3) Physical Problems (0–5 range)	0.97 (1.07)	1.94 (1.36)
4) Spiritual/religious problems (0–1 range)	0.05 (0.23)	0.08 (0.27)
Q23. Social support, yes, n (%)		
1) I have people who care what happens to me.	160 (71.8%)	108 (61.0%)
2) I get love and affection.	178 (79.8%)	89 (67.2%)
3) I get chances to talk to someone about problems at work or with my housework.	179 (80.2%)	108 (61%)
4) I get chances to talk to someone I trust about my personal or family problems.	170 (76.3%)	113 (63.8%)
5) I get chances to talk about money matters.	162 (72.7%)	99 (55.9%)
6) I get invitations to go out and do things with other people.	168 (75.4%)	103 (58.2%)
7) I get useful advice about important things in life.	165 (74.0%)	105 (59.3%)
8) I get help when I am sick in bed.	175 (79.5%)	112 (63.2%)
Q24. Snoring (Yes/No), n (%)		
Snoring (Yes)	103 (46.2%)	94 (53.1%)
Q28. Medical history, n (%)		
1) High blood pressure	62 (27.8%)	74 (41.8%)
2) High cholesterol	71 (31.8%)	95 (53.7%)
3) Heart attack, or any other heart disease	8 (3.6%)	16 (9.0%)
4) Diabetes	28 (12.6%)	52 (29.4%)
5) Anxiety or depression	9 (4.0%)	14 (7.9%)
6) Obesity	20 (9.0%)	37 (20.9%)
7) Any other health problem	19 (8.5%)	28 (15.8%)
Currently taking medication, n (%)		
1) Hypertension medication	56 (25.1%)	64 (36.2%)
2) Cholesterol medication	35 (15.7%)	61 (34.5%)

	Perceived health status: Good (n=223, 55.8%)	Perceived health status: Poor (n=177, 44.2%)
3) Diabetes medication	20 (9.0%)	36 (20.3%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Prediction Model Evaluation

	Precision	Recall	AU ROC curve
LR	0.76	0.84	0.75
RF	0.76	0.88	0.76
SVM	0.78	0.79	0.75

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Table 3.

Odds Ratio from the Logistic Regression Model for Better Perceived Health Status

	Variable	Odds Ratio
1	Higher annual household income	1.55
2	Not having medical history: obesity	1.53
3	Not having medical history: high cholesterol	1.50
4	Highest level of education: attended graduate or professional school	1.48
5	Westernized ethnic identity	1.39
6	Not having medical history: diabetes	1.30
7	Having social support: “I get useful advice about important things in life.”	1.27
8	Not having medical history: any other health problems	1.26
9	Not taking cholesterol medication	1.20
10	Not having medical history: high blood pressure	1.16

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Table 4.

Feature Importance from the Random Forest Model for Better Perceived Health Status

	Variable	**Importance**
1	Better English proficiency	0.12
2	Problems experienced in the past week: Physical problems	0.09
3	Higher annual household income	0.08
4	Highest level of education: attended graduate or professional school	0.05
5	Not having medical history: obesity	0.04
6	Having social support: “I get chances to talk to someone I trust about my personal or family problems.”	0.04
7	Having social support: “I get useful advice about important things in life.”	0.04
8	Not having medical history: high cholesterol	0.04
9	Having social support: “I get chances to talk about money matters.”	0.04
10	Younger age	0.03