Published in final edited form as:

Med Care. 2007 January ; 45(1): 80–87. doi:10.1097/01.mlr.0000241114.90614.9c.

Self-Rated Health among Foreign- and US-Born Asian Americans: A Test of Comparability

Elena Erosheva, PhD^{*,†}, Emily C. Walton, MA[§], and David T. Takeuchi, PhD^{†,§}

^{*} Department of Statistics, University of Washington, Seattle, Washington

[†] School of Social Work, University of Washington, Seattle, Washington

[§] Department of Sociology, University of Washington, Seattle, Washington

Abstract

Objectives—We investigated differences between foreign- and US-born Asian Americans in self-rating their physical and mental health. In particular, we tested whether the foreign-born respondents underreport the extreme categories of the scale as compared to US-born respondents.

Methods—We analyzed data from the National Latino and Asian American Study to examine whether immigrants are less likely to use the extreme ends of the 5-category self-rated health scales than their US-born counterparts. We used propensity score matching to derive groups of US- and foreign-born Asian Americans who share similar demographic and health characteristics. We defined propensity scores as predicted probabilities of being US-born given individual background characteristics. The propensity score framework allowed us to make descriptive comparisons of self-rated health responses controlling for background characteristics. We used log-linear symmetry models to examine cross-tabulations of self-rated health reports in matched pairs by the two (extreme and non-extreme) and five (“excellent”, “very good”, “good”, “fair”, and “poor”) categories.

Results—Controlling for background characteristics, we found no evidence that foreign-born Asian Americans are less likely to endorse extreme categories in self-rated physical or mental health than US-born, as well as no evidence of imbalances in endorsement of any particular self-rated health category between the two groups.

Conclusions—Controlling for demographic and health characteristics, we find no systematic differences between foreign- and US-born Asian Americans in reporting self-rated physical and mental health on the 5-category scales from “excellent” to “poor”.

Keywords

extreme response style; immigrants; propensity score matching; regression; self-rated physical and mental health; log-linear symmetry models

Please direct all correspondence to: Elena Erosheva, Department of Statistics, University of Washington, Box 354320, Seattle, WA 98195; Fax: 206-685-7419; Telephone: 206-685-0166; erosheva@u.washington.edu.

Elena Erosheva, PhD, Department of Statistics, University of Washington, Box 354320, Seattle, WA 98195, Fax: 206-685-7419, Telephone: 206-685-0166, erosheva@u.washington.edu

Emily C. Walton, MA, Department of Sociology, University of Washington, Box 353340, Seattle, WA 98195, Fax: 206-543-2516, Telephone: 206-543-5882, waltone@u.washington.edu

David T. Takeuchi, PhD, School of Social Work, University of Washington, Box 354900, Seattle, WA 98195, Fax: 206-221-1228 Telephone: 206-543-5133, dt5@u.washington.edu

INTRODUCTION

Self-rated health is a robust indicator of general physical health status that predicts morbidity, mortality, subsequent disability and health care utilization.^{1–6} It is also a stronger predictor of mortality than physician-assessed health.⁷ In this study, we examine a long-standing speculation among researchers in the field that immigrant Asian Americans may respond differently when rating their health relative to their US-born counterparts. In particular, immigrants are perceived to possibly underreport extreme categories on health rating scales.

The motivation for the study comes from two observations. First, recent findings suggest that Asians tend to avoid extreme ratings. For example, Chinese and Japanese students were found more likely to select midpoints on general scales compared to U.S. and Canadian students.⁸ Related to health scales, Japanese tend to avoid extreme positive categories when reporting their emotions.^{9,10} Second, cultural environment is considered a major reason for these differences. Thus, tendency to select midpoints of Chinese and Japanese students was directly linked to the students' identification with collectivist cultures.⁸ Collectivist societies tend to encourage self-criticism, understatement of personal virtues and diffidence in individual behavior. Accordingly, individuals in these societies may be more likely to avoid extreme ratings, either positive or negative, in describing their behavior and emotions. On the other hand, the cultural environment of the U.S. rewards self-enhancement¹¹ and individuals are more likely to use the full range of options in rating their own behavior and emotions. Classic assimilation theory suggests that the process of acculturation, or taking on the cultural habits and practices of the mainstream society, progresses throughout the immigrant's own lifetime and subsequent generations.¹² Consonant with the assimilation theory, foreign-born Asian Americans are more likely to be strongly associated with Asian cultures than US-born.

Accordingly, we hypothesize that foreign-born Asian Americans will endorse the extreme categories of the self-rated health scale less often than US-born Asian Americans, even after adjustment for health and other demographic factors. Given that the 5-category health status scale is used in a wide range of surveys across many countries, it is critical to empirically test whether systematic group differences exist in reporting and to distinguish them from possible group differences in health.

Other research has examined general imbalances in using extreme categories in scales based on sets of items.^{13–16} For example, African Americans were found to be more likely than whites to use extreme response categories in responding to Likert-type attitude and personality items.¹³ In that study, the authors considered some fundamental variables such as academic achievement or geographic area as potential confounders but did the analysis in a univariate (considering the effect of the covariates one by one) rather than multivariate (controlling for a distribution of all the covariates) fashion.

A large body of research situated within the framework of Item Response Theory focuses on studying whether particular groups of individuals tend to endorse items differently.¹⁷ The context is usually a multi-item test or a scale which targets assessment of a certain trait. In this context, methods to detect differential item functioning are available for analyzing group differences in responding.¹⁸ In this study, we face an analogous problem of detecting differences in responding but focus specifically on only two separate outcomes, self-rated physical and mental health.

A traditional way to address possible biases in a single outcome is to fit a regression model with the group indicator being the predictor of interest, controlling for other covariates via their inclusion in the model as additional predictors. Several studies reported their findings on differential group responses based on this approach. For example, one study found women rate their health more favorably than men, controlling for more objective health status,¹⁹ whereas

other studies found no gender differences.^{20,21} Similarly, when compared to younger individuals, those 75 years or older were found to be more positive in their health ratings despite reporting more health-related problems;¹⁹ and obese individuals were found to report lower self-ratings of physical health, controlling for morbidity and functional limitation.²

Although common in the social sciences, regression-type approaches for addressing group differences have several caveats. First, multiple regression analyses usually have to compromise between including either a large number of covariates to control for potentially important characteristics or a smaller number of covariates to satisfy modeling requirements, such as avoiding multicollinearity and overfitting. Often, the compromise is reached through using composite measures, for example, the number of illnesses as opposed to indicators for illnesses themselves.^{19,22} In predictive models, avoiding overfitting by keeping the number of covariates low is especially important for the model's predictive ability in new data.

Second, given the goal of estimating group differences, the researchers need to make sure that groups overlap in their background characteristics sufficiently to carry out meaningful comparisons without unwarranted extrapolations. For example, consider a regression model that is fitted with data on 85- and 45-year olds to study the influence of age. If the inference on the group effect is made but ignores the fact that 85-year olds were predominately female and the 45-year olds were predominately male, that inference is flawed. While one can check for an overlap in each single covariate separately, it is a non-trivial task to do in a high-dimensional setting with many covariates.^{23,24} Note that standard regression diagnostics do not include checking for overlap in the multivariate distribution of covariates.

Third, all model-based approaches assume a particular functional form for a relationship between the outcome variable and the covariates. While modeling assumptions such as linearity or log-linearity may be less important when there is a sufficient overlap on the covariates, the conclusions can highly depend on a specific form of the model when there is a lack of overlap.^{23,25}

Fourth, model-based approaches in observational studies often require repeated analyses that simultaneously involve outcomes, covariates, and group indicators. It may be difficult for researchers to be objective when findings from different analyses disagree.

On the other hand, for propensity score approaches, multitudes of covariates and overfitting in general are irrelevant concerns because the goal is to develop a sample-specific adjustment and not a predictive model. The propensity score itself provides a simple check for overlap in the multivariate distribution of baseline covariates. Finally, propensity score approaches allow one to treat observational studies in a manner analogous to randomized experiments: without including the outcome variables in the design phase.²⁶

In this paper, using propensity score methodology, we first examine the extent of overlap in the multivariate distributions of background characteristics between US- and foreign-born Asian Americans. Given the lack of overlap between the two groups in the original sample, we derive groups of US- and foreign-born subjects that are comparable in background characteristics. We then assess differences between US- and foreign-born Asian Americans in self-rating their physical and mental health.

METHODS

Data

The data in this study come from a nationally-representative, household survey of Latino and Asian Americans, the National Latino and Asian American Study (NLAAS).²⁷ Trained

interviewers administered the NLAAS questionnaire in the participant's preferred language in a face-to-face interview unless the respondent specifically requested a telephone interview. A total of 2,095 interviews were completed on respondents of Asian descent, who are the primary focus of the present paper. Detailed descriptions of the methods used in NLAAS can be found elsewhere.^{28,29}

Study Design

The likelihood of individuals endorsing extreme categories in a self-rated health scale naturally depends on their objective health. It may also depend on a number of demographic characteristics such as age, education, and income.¹⁵ Thus, natural differences in group composition may result in systematic differences in responses.

Using the propensity score, a scalar summary of multivariate baseline characteristics, yields a simple check for overlap in the multivariate distribution of those characteristics.²⁶ Formally, the propensity score is defined as the conditional probability of being assigned to the treatment group, given background covariates. We defined the propensity score as the conditional probability of being US-born, given background characteristics. The background characteristics we chose to include were demographic indicators associated with self-rated physical and/or mental health, measures of objective health status, and variables that reflected interview conditions.

Measures

The two outcome variables were 5-category subjective measures of *self-rated physical* and *self-rated mental health*. Specifically, an NLAAS interviewer asked separately for physical and mental health, "How would you rate your overall (mental) health – excellent, very good, good, fair or poor?" and the respondent answered verbally.

Background covariates for the propensity score model included demographic and health characteristics and interview conditions (Table 1). Demographic characteristics were the respondent's *age*; highest level of *education* achieved; *gender*; *family size*; *household income*; and *marital status* indicators at the time of the interview. All health covariates available in the survey were self-reported. For our analysis, we retained health indicators that we deemed to be more objective. These were *body mass index (BMI)*; *functional ability* ("How many days in the past 30 were you limited at all in carrying out your normal daily activities because of problems with your physical health, mental health, or substance use?"); *mobility limitation* ("Was there ever a time in the past 30 days when health-related problems caused you difficulties with mobility?"); *presence of a mental disorder* (if the respondent was diagnosed in their lifetime with a mental disorder from a list of 14 mood, anxiety and substance disorders according to DSM-IV criteria); *smoking status*; and binary responses to questions regarding *chronic conditions* such as "Have you ever had chronic back pain or neck problems?" and "Did the doctor ever tell you that you had heart disease?". We excluded 3 health indicators that asked respondents for a difficulty rating in standing or moving to avoid a circular reasoning in the case of a possible general tendency of immigrant Asian Americans to underreport extreme categories. Since presence of other individuals during the interview might have influenced reporting, we included an indicator of whether *others were present* during the interview. We were not able to include the variable that captured English or native language use during the interview because close to 100% of US-born Asian Americans used English.

Missing household income data were imputed for 268 individuals (12.8% of the total) using hot deck imputation. After the imputation, only 2% of the individuals (10 US-born and 33 foreign-born) still had missing records on other covariates. Those individuals were excluded, reducing the total to N=2050.

Propensity Score Model

Table 1 provides univariate comparisons of background characteristics, unadjusted for survey weights. Note that while these values reflect differences between foreign- and US-born Asians for the survey sample only, we found them to be representative of the differences in the overall population obtained from the analogous analysis adjusted for survey weights.

The groups differed significantly on marital status: the foreign-born were more likely to be married. On average, the foreign-born individuals were older, less educated, had lower BMI scores, and lower prevalence of asthma. US-born Asian Americans were more likely to be limited in functional ability, to have mental disorders, allergies, headaches, and back and neck pain, whereas the foreign-born were more likely to have arthritis and tuberculosis.

Next, we compared the multivariate distributions of the background characteristics. We obtained propensity scores by fitting a logistic regression model to predict nativity status based on the 29 variables from Table 1 plus 19 additional variables. These additional variables included two quadratic terms ($education^2$ and BMI^2), an *education-BMI* interaction, an *education-asthma* interaction, and interactions between each of the selected demographic characteristics (*income*, *age*, *female*, *married* and *never married*) and *education*, *BMI*, and *asthma*. We selected those interaction terms that we judged as potentially having different effects in the US- and foreign-born groups of Asian Americans. From the logistic regression, we obtained propensity scores for each individual as the predicted probabilities of being US-born based on individual background characteristics. Note that the two outcome variables, self-rated physical and mental health, were not included in the model for the propensity score.

To assess the overlap in the background characteristics, we followed guidelines by Rubin²⁶ which suggest comparing the means and variances of the propensity scores, and the residual variability in the covariates. Specifically, for a regression adjustment to be trustworthy: (i) the difference between the means of the propensity scores in the two groups standardized by its standard deviation must be less than $\frac{1}{2}$; (ii) the ratio of the variances of the propensity score in the two groups must be close to one, and (iii) for each of the covariates, the ratio of the variances of the residuals after adjusting for the propensity score (obtained by regressing each of the original covariates on the propensity score) must be close to one.²⁶ We found that a standard regression adjustment would not be appropriate for the original survey sample of US- and foreign-born Asian Americans since conditions (i) and (ii) above were clearly violated (Table 2). Table-2-about-here. To derive groups of US- and foreign-born that share similar demographic and health characteristics, we used propensity score matching.³⁰ We chose matching as opposed to stratification by propensity scores³¹ to facilitate contingency table analyses for the multi-category outcome measures. In this case, stratification was a less desirable alternative to matching because it would drastically increase the number of small counts in the contingency tables and small counts pose inferential concerns in discrete data analysis.^{32,33}

We attempted to match each US-born Asian American to a foreign-born Asian American of the same ethnicity (Chinese, Filipino, Vietnamese or Other) with a similar propensity score.

We used logit transformation of the propensity score throughout the matching process. For each US-born, we identified foreign-born as potential matches if their propensity scores fell within the selected caliper of the US-born's propensity. We set the caliper to be one quarter of the standard deviation of the propensity score on the logit scale. This process of selecting matches close to each other in propensity scores corresponds to randomization in designed experiments. To select from a pool of potential matches within the caliper, we calculated the Mahalanobis distance based on the logit of the propensity score and selected important covariates (*age*, *education*, *family size*, *household income*, and *chronic conditions*). Compared

with the Euclidean metric, the Mahalanobis metric automatically accounts for differences in scales between variables (i.e., is scale-invariant) and for correlations between variables. The use of metric matching in this step corresponds to blocking in designed randomized experiments. Analogously to blocking, the selection of covariates for the metric may include key substantive variables as well as variables which are still somewhat out of balance between the two groups. Overall, this matching procedure is known as the nearest available Mahalanobis metric matching within a caliper defined by the propensity score.³⁰

The matching algorithm employed the following steps. First, the US born individuals were randomly ordered, and the first individual was chosen. Then, all foreign born individuals with propensity scores within the caliper of the selected US-born's estimated propensity score were selected as potential matches. Next, Mahalanobis distances were calculated between the US-born individual and all potential matches. An individual closest to the US-born in the Mahalanobis distance was determined to be a match. The matched pair was then removed from the pool and the process was repeated until matches for all of the US-born individuals were chosen. When the pool of potential within the caliper matches for a selected US-born individual was empty, the Mahalanobis metric was calculated for all foreign-born individuals still available for matching and the closest match was selected.

Following the algorithm described, for the Vietnamese group (18 US-born, 492 foreign-born), all of the US-born individuals matched to foreign-born within the caliper. For the Filipino (154 US-born, 341 foreign-born), Chinese (122 US-born, 464 foreign-born) and Other Asian (150 US-born, 309 foreign-born) groups, 68.4%, 77.5%, and 74.3% of US-born individuals matched within the caliper, respectively. All individuals matched within the caliper from all ethnic groups were then recombined to constitute a caliper matched sample of N=335, with one US-born and one foreign-born forming each pair. All matched individuals (not necessarily within the caliper) were also recombined to constitute a full matched sample of N=444.

Examining propensity scores, we found the quality of overlap in background characteristics to be very good for the caliper match (Table 2); the distributions of the propensity scores in the two groups had similar means and variances (conditions (i) and (ii) above), and only 6% of the covariates violated condition (iii). Although the quality of overlap in the background characteristics also improved for the full matched sample (Table 2), it may not have improved enough to justify a full match. Our visual inspection of the distribution of propensity scores by nativity within each ethnic group revealed that for all groups but Vietnamese there were some US-born individuals with propensity scores at the high end of the distribution for whom a close match to a foreign-born was not possible.

Table 3 provides univariate comparisons of background covariates between the US-born and the foreign-born in the caliper matched sample. Comparing background characteristics of the US-born group with those of the matched foreign-born, we observed that most of the bias was removed. The two groups had similar means for most of the covariates included in the model except possibly *education* (p-value=0.05) and *other chronic pain* (p-value=0.04). Univariate comparisons in the full matched sample (not shown) had six variables for which statistically significant differences still remained; however, just like for the caliper match, those differences did not appear to be clinically important.

RESULTS

Matched Pairs Analysis of Self-rated Physical and Mental Health

We examined two by two cross-classifications of extreme versus non-extreme responses on self-rated physical and mental health for the caliper matched sample (Tables 4 and 5). Non-extreme categories included "fair", "good", and "very good". Extreme categories included

“poor” and “excellent”. For physical health, there were 50 pairs in which a US-born individual chose an extreme response category and a matched foreign-born chose a non-extreme category as well as 52 pairs in which a US-born individual chose an intermediate category and the matched foreign-born chose an extreme category (Table 4). For mental health, the analogous counts were 73 and 71, respectively (Table 5). We found the distribution of entries in these two by two tables to be consistent with the log-linear model of complete symmetry;^{32,33} the p-values for the Pearson’s Chi-square goodness of fit test were 0.84 and 0.87 for physical and mental health, respectively. Therefore, controlling for background covariates, we found no evidence that foreign-born Asian Americans are different in their likelihood of endorsing extreme categories from US-born Asian Americans when self-rating their physical and mental health.

To address the issue of power of this test, we investigated what is the smallest departure from symmetry that our test could detect at the 0.05 significance level, given observed numbers of discordant pairs (102 pairs for physical and 144 for mental health). For physical health, we found that off-diagonal entries of 46 and 56 produced the smallest detectable departure from symmetry (p-value=0.03); for mental health, the corresponding counts were 66 and 78 (p-value=0.01).

Having obtained matched samples also allowed us to examine the overall distributions of self-rated physical and mental health via analysis of the cross-classifications of the matched pairs by the five initial categories from “poor” to “excellent” (Tables 4 and 5). Notice that the agreement in ratings between matched individuals was rather weak. In fact, a strong agreement was not to be expected for two main reasons. First, even if one assumed the existence of a perfect association between the health background covariates in our study and self-rated health, a number of other covariates were included in the model for propensity scores which might have turned out to be influential in the matching process. Second, because matching was done on the scalar obtained from the logistic regression model (i.e., propensity score), there could have been multiple combinations of covariate values which resulted in similar propensity scores. Nonetheless, after controlling for background covariates, if there was a difference in how the two groups report their health, it should have been reflected in the tables through a significant deviation from symmetry.

We found that log-linear models of symmetry fit the entries in five by five contingency tables (Tables 4 and 5) very well; the exact p-values³⁴ for the Pearson’s Chi-square goodness of fit test were 0.72 and 0.49. In other words, the observed symmetric off-diagonal entries are similar up to a random error. It also follows that marginal distributions of self-rated physical and mental health ratings among foreign- and US-born Asian Americans are similar as well.

Sensitivity Analyses

We corrected for covariates that were still imbalanced after matching by fitting logistic regression models to predict extreme responses conditional on those covariates, the logit of the propensity score, and the nativity status. These regression adjustments did not alter our conclusion; there was no significant association between the nativity status and extreme responses. This conclusion was also robust with respect to moderate changes in our choices of the caliper and the set of interactions in the propensity score model. Results from examining cross-classifications of extreme versus non-extreme responses, disaggregated by ethnic group, confirmed the above conclusions obtained for all ethnic categories combined. Finally, results from the examination of self-rated physical and mental health cross-classifications for the full matched sample (N=444 matches) also showed no differential reporting by the nativity status.

DISCUSSION

This study examined self-reported ratings of physical and mental health by foreign- and US-born Asian Americans from the NLAAS data. In particular, we were interested to see whether foreign-born Asian Americans have a tendency to use intermediate categories of the scale more often than the extremes, compared to US-born individuals. Controlling for a number of demographic and health characteristics, we found no support for a differential use of any self-rated physical or mental health category, including the extremes.

We based our analytic approach on a propensity score method. Compared to regression-type adjustments, this method has a number of advantages which include: intuitive appeal and persuasiveness to non-technical audiences, objectivity of the analysis since similar groups can be formed before even looking at the outcome variables,^{25,35} more straightforward diagnostics as compared to those for regression analysis,²⁵ and lower standard errors of treatment effect as compared to those from regression models.³⁶

The propensity score methodology originated from comparing treatment and control groups in observational studies.^{24,30,35,36} There are only a few examples that use propensity score methodology to draw descriptive comparisons between natural groups such as gender.²⁴ We should emphasize that comparisons between social or demographic groups provide no basis for causal interpretations but only for descriptive comparisons.

Theoretically, under certain regularity conditions which include sufficient overlap in the distribution of covariates and a close adherence to the regression-specified functional form, results from multiple regression and propensity score analyses lead to the same conclusions.^{24,35,26} In certain circumstances, such as when one desires to describe not only the group effect but also the effects of other covariates, regression-type approaches should be preferred over propensity score methods.²⁵ In our case, a typical regression approach was not appropriate because of the lack of overlap in the distributions of background covariates between the two groups.

The ultimate goal of our analysis was to examine potential differences in reporting self-rated physical and mental health between US- and foreign-born Asian Americans, controlling for background characteristics. Propensity score matching allowed us to address this question by using log-linear models for cross-classifications of matched pairs. There are a number of log-linear models, including symmetry and quasi-symmetry,³² which are applicable to matched pairs data analysis. In our case, symmetry models fit the data very well.

Since our data come from a survey, we had to address the question of survey weights. Although survey weights' adjustments in a propensity score analysis by subclassification have been suggested elsewhere,²⁵ we argue that adjusting for weights in propensity score matching is not necessary due to the nature of the matching process which directly depends only on the multivariate distribution of covariates.

In this study, we excluded about 2% of individuals because of missing data on the covariates. A very small percentage of missing data in our case made the case-wise deletion to be the method of choice, even though methods to deal with missing data within a propensity score analysis framework exist.³¹ Another limitation of our study is that we were not able to control for language variables because of a high confounding with nativity status.

In conclusion, we found no systematic differences in using categories of the self-rated health scales between foreign- and US-born Asian Americans. Our results only pertain to verbal responses on the 5-category self-rated physical and mental health scales with the categories ranging from "excellent" to "poor," and not to other scales or other types of interviews.

Moreover, our study is limited to detecting differences between nativity groups. We recognize that nativity may not fully capture the cultural differences in reporting styles among respondents. There may be other more meaningful measures to capture the cultural differences in reporting styles among Asian Americans such as native language ability. However, the propensity score methodology demonstrated in the manuscript can be broadly applicable for detecting differences in reporting styles between other social or demographic groups especially when there is a reason that their reporting styles may differ.

Acknowledgments

This project is supported by NIMH grant #U01 MH62207 and MH62209 with additional support from SAMHSA and OBSSR.

References

1. Farmer MM, Ferraro KF. Are Racial Disparities in Health Conditional on Socioeconomic Status? *Soc Sci Med* 2005;60:191–204. [PubMed: 15482878]
2. Ferraro KF, Yu Y. Body Weight and Self-ratings of Health. *J Health Soc Behav* 1995;36:274–284. [PubMed: 7594359]
3. Gomez SL, Kelsey JL, Glaser SL, et al. Immigration and Acculturation in Relation to Health and Health-Related Risk Factors Among Specific Asian Subgroups in a Health Maintenance Organization. *Am J Public Health* 2004;94(11):1977–1984. [PubMed: 15514240]
4. Idler EL, Benyamini Y. Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies. *J Health Soc Behav* 1997;28(1):21–37. [PubMed: 9097506]
5. Mutchler JE, Burr JA. Racial Differences in Health and Health Care Service Utilization in Later Life: The Effect of Socioeconomic Status. *J Health Soc Behav* 1991;32(4):342–356. [PubMed: 1765625]
6. Wickrama KAS, Conger RD, Wallace LE, et al. Linking Early Social Risks to Impaired Physical Health during the Transition to Adulthood. *J Health Soc Behav* 2003;44(1):61–74. [PubMed: 12751311]
7. Mossey JM, Shapiro E. Self-rated Health: A Predictor of Mortality Among the Elderly. *Am J Public Health* 1982;72:800–808. [PubMed: 7091475]
8. Chen C, Lee SY, Stevenson HW. Response Style and Cross-Cultural Comparisons of Rating Scales among East Asian and North American Students. *Psychol Sci* 1995;6(3):170–175.
9. Iwata N, Roberts CR, Kawakami N. Japan–U.S. Comparison of Responses to Depression Scale Items among Adult Workers. *Psychiatry Res* 1995;58:237–245. [PubMed: 8570779]
10. Iwata N, Higuchi HR. Responses of Japanese and American University Students to the STAI Items That Assess the Presence or Absence of Anxiety. *J Pers Assess* 2000;74(1):48–62. [PubMed: 10779932]
11. Kitayama S, Markus HR, Matsumoto H, Norasakkunkit V. Individual and Collective Processes in the Construction of the Self-Enhancement in the United States and Self-Criticism in Japan. *J Pers Soc Psychol* 1997;72:1245–1267. [PubMed: 9177018]
12. Alba, R.; Nee, V. *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. Cambridge: Harvard University Press; 2003.
13. Bachman JG, O'Malley PM. Yea-saying, Nay-saying, and Going to Extremes: Black-White Differences in Response Style. *Public Opin Q* 1984;48(2):491–509.
14. Crandall JE. Social Interest, Extreme Response Style, and Implications for Adjustment. *J Res Pers* 1982;16:82–89.
15. Greenleaf EA. Measuring Extreme Response Style. *Public Opin Q* 1992;56(3):328–351.
16. Hui C, Triandis H. Effects of Culture and Response Format on Extreme Response Style. *J Cross Cult Psychol* 1989;20:296–309.
17. van der Linden, WJ.; Hambleton, RK. *Handbook of Modern Item Response Theory*. New York: Springer-Verlag; 1996.
18. Santor DA, Ramsay JO. Progress in the Technology of Measurement: Applications of Item Response Models. *Psychol Assess* 1998;10(4):345–359.

19. Ferraro KF. Self-Ratings of Health among the Old and the Old-Old. *J Health Soc Behav* 1980;21:377–383. [PubMed: 7204931]
20. Anson O, Paran E, Neumann L, et al. Gender Differences in Health Perceptions and Their Predictors. *Soc Sci Med* 1993;36:419–427. [PubMed: 8434267]
21. Jylha M, Guralnik JM, Ferrucci L, et al. Is Self-Rated Health Comparable Across Cultures and Genders. *J Gerontol* 1998;55B(3):144–152.
22. Harrell, F, Jr. *Regression Modeling Strategies*. New York: Springer; 2001.
23. Rubin DB. Estimating Causal Effects from Large Data Sets Using Propensity Scores. *Ann Intern Med* 1997;127(8S):757–763. [PubMed: 9382394]
24. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 1983;70(1):41–55.
25. Zanutto EL. A Comparison of Propensity Score and Linear Regression Analysis of Complex Survey Data. *Journal of Data Science* 2006;4:67–91.
26. Rubin DB. Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. *Health Services & Outcomes Research Methodology* 2001;2:169–188.
27. Alegria M, Takeuchi D, Canino G, et al. Considering Context, Place and Culture: The National Latino and Asian American Study. *Int J Methods Psychiatr Res* 2004;13(4):208–220. [PubMed: 15719529]
28. Heeringa, SG. *Technical Sampling Design Documentation: 2002–2003 National Latino and Asian American Study (NLAAS)*. Ann Arbor: Institute for Social Research; 2004.
29. Pennell B-E, Bowers A, Carr D, et al. The Development and Implementation of the National Comorbidity Survey Replication, the National Survey of American Life, and the National Latino and Asian American Survey. *Int J Methods Psychiatr Res* 2004;13(4):241–269. [PubMed: 15719531]
30. Rosenbaum PR, Rubin DB. Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *Am Stat* 1985;39(1):33–38.
31. D’Agostino RB, Rubin DB. Estimating and Using Propensity Scores with Incomplete Data. *J Am Stat Assoc* 2000;95:749–759.
32. Agresti, A. *Categorical Data Analysis*. New Jersey: John Wiley and Sons, Inc; 2002.
33. Bishop, YMM.; Fienberg, SE.; Holland, PW. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: MIT Press; 1975.
34. Good, PI. *Resampling Methods: A Practical Guide to Data Analysis*. Boston: Birkhauser; 2001.
35. D’Agostino RB. Tutorial in Biostatistics: Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group. *Stat Med* 1998;17:2265–2281. [PubMed: 9802183]
36. Smith HL. Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies. *Sociol Methodol* 1997;27(1):325–353.

TABLE 1

Comparison of US- and Foreign-Born Asian Americans Before Matching

Variable	US-Born (N=444) ^a	Foreign-Born (N=1606) ^a	p ^b	Standardized difference ^c
Demographic				
Age (years)	36.7	42.5	<.001	2.6
Education (years)	14.3	13.5	<.001	-8.1
Family Size (# of individuals)	3.7	4.3	<.001	10.9
Gender (female=1, male=0)	50.2%	53.0%	0.301	11.0
Household Income (thousands of dollars)	84.9	85.1	0.963	0.0
Married (yes=1, no=0)	39.4%	73.0%	<.001	153.1
Never Married (yes=1, no=0)	46.9%	18.3%	<.001	-139.3
Interview Characteristics				
Others Present (yes=1, no=0)	33.7%	40.3%	0.011	28.3
Language of Interview English ^d	98.2%	54.7%	<.001	-247.6
Health Indicators				
BMI (lbs/in ²)	25.8	23.5	<.001	-9.7
Functional Ability (# of days of activity limitation)	1.7	1.0	0.011	-2.9
Mobility Limitation (yes=1, no=0)	10.4%	7.6%	0.055	-33.8
Presence of Any Mental Disorder (yes=1, no=0)	28.4%	14.9%	<.001	-79.6
Smoking Status (smoker=1, non-smoker=0)	15.9%	12.5%	0.062	-27.4
Difficulty Standing (1=none, 5=cannot do) ^d	2.9	2.6	0.142	-0.2
Difficulty Moving in Home (1=none, 5=cannot do) ^d	2.3	2.2	0.864	-0.0
Difficulty Walking Distance (1=none, 5=cannot do) ^d	2.9	2.9	0.777	0.0
Presence of Chronic Conditions (yes=1, no=0)				
Allergies	37.0%	29.7%	0.003	-33.2
Arthritis	12.1%	17.0%	0.012	38.9
Asthma	19.6%	4.9%	<.0001	-126.4
Back or Neck Pain	24.0%	16.9%	0.001	-44.0
Cancer	2.0%	0.9%	0.059	-69.8
Diabetes	7.7%	7.2%	0.717	-7.2
Epilepsy	1.1%	0.2%	0.014	-108.7
Heart Attack	1.8%	1.0%	0.2089	-50.5
Heart Disease	3.5%	3.6%	0.935	2.3
Headaches	18.1%	12.5%	0.002	-43.1
Hypertension	16.3%	16.3%	1.000	0.0
Lung Disease	0.4%	0.6%	0.672	32.1
Other Chronic Pain	7.5%	7.3%	0.904	-2.5
Stroke	1.1%	1.2%	0.918	5.4
Tuberculosis	1.1%	2.3%	0.105	68.5
Ulcer	5.5%	6.8%	0.312	22.7

^aNotes: Means are reported for continuous variables and percentages for binary categorical variables^bUnivariate comparisons were made using Chi-Square tests for binary categorical variables and two-sample T-tests for continuous variables.^cThe standardized difference is the mean difference as a percentage of the average standard deviation: $100(x_{US}-x_f)/\sqrt{\{(s_{US}^2+s_f^2)/2\}}$, where for each covariate x_{US} and x_f are the sample means or percentages in the US-born and foreign-born groups, respectively, and s_{US}^2 and s_f^2 are the corresponding sample variances.^dNot included in propensity score calculation.

Comparison of Multivariate Overlap in Background Characteristics between US- and Foreign-Born Asian Americans Based on the Propensity Scores

TABLE 2

	Bias ^a	Ratio ^b	Number of covariates with specified ratio of the variances of the residuals obtained after a linear adjustment for the propensity score			
			≤ 0.5	> 0.5 and ≤ 0.8	> 0.8 and ≤ 1.25	> 1.25 and ≤ 2
Before Match	-0.78	0.65	2	4	18	15
Full Match	-0.32	2.06	0	1	28	14
Selected Match	-0.10	1.07	0	0	28	17
						9
						5
						3

^aNotes: Standardized difference between the means of the propensity scores in the two groups $((x_{us}-x_f)/\sqrt{((s_{us}^2+s_f^2)/2)})$, where for each covariate x_{us} and x_f are the sample means or percentages in the US-born and foreign-born groups, respectively, and s_{us}^2 and s_f^2 are the corresponding sample variances.

^bRatio of the variances of the propensity scores in the two groups.

TABLE 3

Comparison of US- and Foreign-Born Asian Americans After Matching

Variable	US-Born (N=335) ^a	Foreign-Born (N=335) ^a	p ^b	Standardized difference ^c	Reduction in standardized difference after matching
Demographic					
Age (years)	37.5	36.8	0.525	0.3	2.2
Education (years)	14.4	14.7	0.052	-7.1	1.1
Family Size (# of individuals)	3.9	4.0	0.434	-2.7	8.3
Gender (female=1, male=0)	49.9%	49.9%	1.000	0.0	11.0
Household Income (thousands of dollars)	88.6	84.2	0.377	0.1	-0.1
Married (yes=1, no=0)	46.0%	48.1%	0.588	-8.4	144.7
Never Married (yes=1, no=0)	40.3%	40.3%	1.000	0.0	139.3
Interview Characteristics					
Others Present (yes=1, no=0)	33.7%	37.9%	0.259	-18.2	10.1
Language of Interview English ^d	97.6%	81.2%	<.001	149.9	97.7
Health Indicators					
BMI (lbs/in ²)	24.7	24.5	0.569	1.0	8.7
Functional Ability (# of days of activity limitation)	1.2	1.0	0.517	1.2	1.7
Mobility Limitation (yes=1, no=0)	8.1%	8.7%	0.780	-7.8	26.0
Presence of Any Mental Disorder (yes=1, no=0)	22.1%	19.7%	0.447	14.4	65.2
Smoking Status (smoker=1, non-smoker=0)	14.3%	14.3%	1.000	0.0	27.4
Difficulty Standing (1=none, 5=cannot do) ^d	2.8	2.6	0.361	25.9	-25.7
Difficulty Moving in Home (1=none, 5=cannot do)	2.1	2.0	0.683	11.1	-11.1
Difficulty Walking Distance (1=none, 5=cannot do) ^d	2.7	2.6	0.770	6.6	-6.5
Presence of Chronic Conditions (yes=1, no=0)					
Allergies	34.6%	31.0%	0.324	16.3	16.9
Arthritis	11.3%	10.8%	0.805	6.0	33.0
Asthma	9.9%	6.0%	0.063	52.1	74.4
Back or Neck Pain	20.6%	18.5%	0.495	13.3	30.7
Cancer	1.5%	0.9%	0.477	48.4	21.3
Diabetes	6.6%	4.8%	0.316	33.1	-25.8
Epilepsy	0.9%	0.0%	0.083	143.0	-34.2
Heart Attack	0.6%	0.6%	1.000	0.0	50.5
Heart Disease	2.7%	18%	0.433	40.3	-38.0
Headaches	15.5%	13.1%	0.378	19.4	23.7
Hypertension	15.2%	14.0%	0.662	9.5	-9.5
Lung Disease	0.3%	0.3%	1.000	0.0	32.1
Other Chronic Pain	7.2%	3.6%	0.039	67.4	-64.9
Stroke	0.6%	0.3%	0.563	63.7	-58.4
Tuberculosis	0.6%	0.0%	0.157	142.6	-74.0
Ulcer	5.7%	3.0%	0.088	62.1	-39.4

^aNotes: Means are reported for continuous variables and percentages for binary categorical variables

^bUnivariate comparisons were made using Chi-Square tests for binary categorical variables and two-sample T-tests for continuous variables.

^cThe standardized difference is the mean difference as a percentage of the average standard deviation: $100(x_{US}-x_F)/\sqrt{((s_{US}^2+s_F^2)/2)}$, where for each covariate x_{US} and x_F are the sample means or percentages in the US-born and foreign-born groups, respectively, and s_{US}^2 and s_F^2 are the corresponding sample variances.

d_i Not included in propensity score calculation.

Cross-Tabulation of Responses on Self-Rated Physical Health for Foreign- and US-Born Asian Americans (N=335)

TABLE 4

	Excellent	Very Good	Foreign-Born		Fair	Poor	Total
			Good	Fair			
US-Born	Excellent	10	21	22	5	0	58
	Very Good	24	53	43	15	3	138
	Good	21	43	34	11	0	109
	Fair	3	11	8	4	1	27
	Poor	1	1	1	0	0	3
Total	59	129	108	35		4	335
	Extreme		Non-extreme		Total		
US-Born	Extreme	11	50				61
	Non-extreme	52	222				274
Total	63	272					335

TABLE 5
Cross-Tabulation of Responses on Self-Rated Mental Health for Foreign- and US-Born Asian Americans (N=334)^a

		Excellent	Very Good	Foreign-Born Good	Fair	Poor	Total
US-Born	Excellent	45	39	27	6	1	118
	Very Good	44	55	31	10	0	140
	Good	23	23	17	2	1	66
	Fair	3	3	2	1	0	9
	Poor	0	0	1	0	0	1
Total		115	120	78	19	2	334
		Extreme		Non-extreme		Total	
US-Born	Extreme	46		73		119	
	Non-extreme	71		144		215	
	Total	117		217		334	

^aNotes: One individual that was included in the analysis of self-rated physical health had missing data on self-rated mental health, making the total N=334.