Using Allostatic Load to explore the Latino Health Puzzle for Hispanics in the United States

Alexis R. Santos-Lozada, Pennsylvania State University

Jeffrey T. Howard, University of Texas at San Antonio

Introduction

Self-rated health (SRH) is commonly assessed in health surveys (Manor, Matthews, & Power, 2000). A recent line of research has started to test whether biological risk profiles (measured as Allostatic Load) correspond with subjective health ratings (Santos-Lozada & Howard, 2018). This line of research has revealed that SRH does not corresponds similarly to biological risk profiles for non-Hispanic blacks and Hispanics. We argue that the reduced link between AL and SRH for Hispanics is rooted in the language that Hispanics answer surveys.

Why? Numerous articles have found that those who answer health surveys in Spanish to be more likely to report poor/fair SRH (Bzostek, Goldman, & Pebley, 2007; Lorraine, Hammock, & Blanton, 2005; Shetterly, Baxter, Mason, & Hamman, 1996; Viruell-Fuentes, Morenoff, Williams, & House, 2011). Differences in the meaning of the health categories can influence those answering in Spanish to answer "Regular" as an average or medium category (Figure 1) and to be categorized as having bad health.

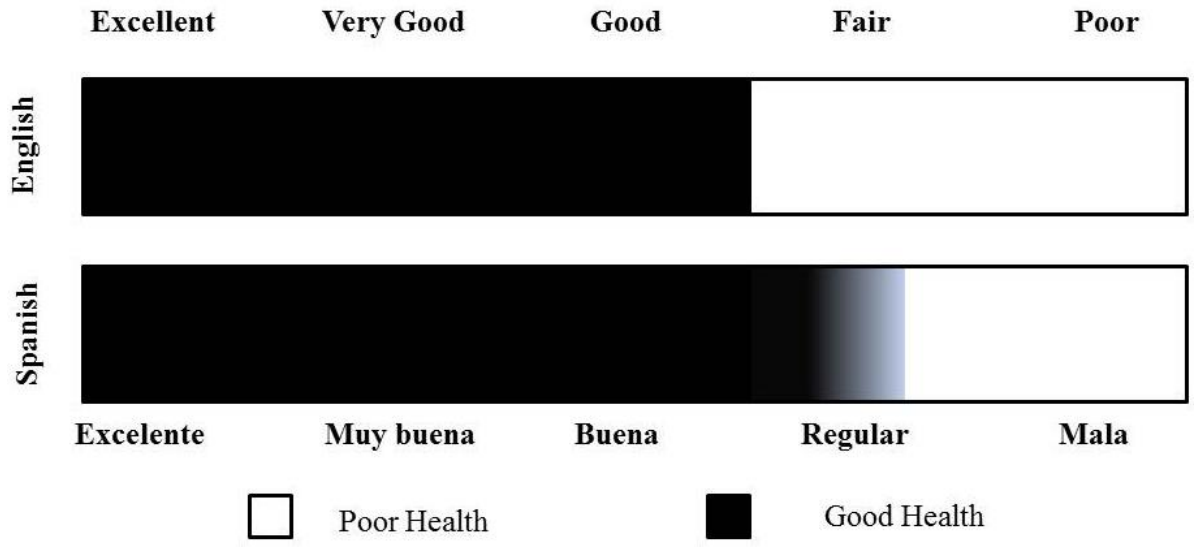

Figure 1 Conceptualization of health categories by language of interview

Figure 1 presents a conceptualization of the meaning of each category of the health by language of interview, which is supported by previous literature (Brewer et al., 2013; Bzostek et al., 2007; Dubard & Gizlice, 2008; Viruell-Fuentes et al., 2011). The usual dichotomization of poor/fair is

valid for individuals interviewed in English. For individuals who answer in Spanish the regular category operates as a gradient ranging from medium-good, medium, or medium-bad. In essence regular is a way of indicating the person is not doing well, without implying poor health.

The primary goal of this study is to determine whether or not self-rated health is valid for multilanguage comparisons based on strength of association with AL score by focusing on four specific questions. First, do differences in AL score explain poor/fair self-rated health for Hispanic adults? Second, do language differences persist even when AL scores and other characteristics are incorporate to the model? Third, does the predictive power of AL score, differ by language in which the survey is conducted? And do these differences in association indicate validity issues for those who answer the survey in Spanish?

The research questions that guide this study, we test five hypotheses. First, we hypothesize that odds of poor/fair SRH will be greater for individuals with higher AL scores, even with controls for potential covariates. Second, given that previous research has shown Spanish respondents to be more likely to report poor/fair SRH than those who answered the survey in English, we expect significant differences in the odds of reporting poor/fair SRH between English and Spanish respondents, with Spanish respondents experiencing an increased risk of reporting the outcome. Third, we hypothesize that the association of AL scores differs by language of interview, with AL scores being less predictive of poor/fair SRH for those who answered the survey in Spanish. Forth, we hypothesize that AL score will be associated with increased odds of poor/fair SRH for English respondents. Lastly, we hypothesize that AL scores are not associated with increased odds of poor/fair SRH for Spanish respondents.

Data and Methods

Sample

Data for this study come from the National Health and Nutrition Examination Survey (NHANES) 2009-2010. The NHANES is designed to assess the health and nutritional status of adults and children in the United States conducted by the Center for Diseases Control and Prevention (Johnson et al., 2013; Zipf et al., 2013). It combines interviews and physical examinations. It includes demographic, socioeconomic, dietary, and health related questions. The examination component includes information of medical, dental, physiological measurements, as well as laboratory tests administered by medical personnel. The initial sample size was 1,581 individuals who reported being Hispanics, 25 years or older, and who were fasting at the moment of being measured for biomarkers that are drawn from the blood sample. List-wise deletion was employed when dealing with missing values for covariates included in this analysis (Allison, 2002). This means only observations with information for all covariates and the outcome are included in the analytic sample, this is not applied for income level. Income was the only variable where we created a category of "Refused/Don't know". We base this decision on previous findings, which indicate that respondents with missing income information may not be a random subset of survey

participants (Kim, Egerter, Cubbin, Takahashi, & Braveman, 2007). The manner in which we dealt with missing values for biomarkers is discussed in the *Measures* section. The final size of the analytical sample was 1,352 observations; this represents a reduction of 14.84% of the initial sample.

Measures

Self-Rated Health

The outcome of this analysis is poor/fair SRH. This variable was collected by asking respondents to rate their general health. English respondents could select one of five categories: (1) 'Excellent', (2) 'Very good', (3) 'Good', (4) 'Fair' and (5) 'Poor'. Spanish respondents were given the following options: (1) 'Excelente', (2) 'Muy Buena', (3) 'Buena', (4) 'Regular' and (5) 'Mala'. We followed the usual dichotomization found through most SRH literature (Acevedo-Garcia, Bates, Osypuk, & McArdle, 2010; Idler & Benyamini, 1997; Manor et al., 2000; Subramanian, Acevedo-Garcia, & Osypuk, 2005). On it zero represents whose who rate their health as 'Excellent', 'Very Good' or 'Good' ('Excelente', 'Muy Buena', 'Buena' for Spanish respondents, respectively); and one indicates respondents rated their health as poor or fair ('Mala' or 'Regular' for Spanish respondent, respectively).

Language of Interview

According to NHANES documentation, survey participants selected the language of interview or request a certified translator. Respondents were given the choice of completing the survey in English or Spanish. Previous research has indicated that individuals who answer in Spanish are more likely to report poor/fair SRH when compared to those who answer in English (Dubard & Gizlice, 2008; Kandula, Lauderdale, & Baker, 2007; Viruell-Fuentes et al., 2011). Viruell-Fuentes and colleagues (2011) indicate that the translation of "fair" to "regular" induces respondents to report poorer health than they would in English. Another article indicates that translation issues between Spanish and English versions of the SRH question may cause differences in risk of reporting poor health (Bzostek et al., 2007). We operationalize language of interview as a dichotomous variable indicating whether the respondent answered in English (reference group) or Spanish.

Allostatic Load

A set of 10 biomarkers, inclusive of all biomarkers that could be derived from the physical examinations and laboratory components of NHANES, were used to calculate allostatic load for each participant. These included (1) systolic and (2) diastolic blood pressure, (3) body mass index (BMI), (4) glycated hemoglobin, (5) albumin, (6) creatinine clearance, (7) triglycerides, (8) C-reactive protein, (9) homocysteine, and (10) total cholesterol.

Exploratory analysis revealed that a number of individuals had missing values in one or more biomarkers. In these cases, rather than limiting our sample, we decided to perform imputation of these values. We followed the recommendations of previous literature on the subject of missing data which recommends avoiding missing cases and to consider approaching the issue through imputation techniques (van der Heijden, T. Donders, Stijnen, & Moons, 2006). Missing cases for each element of the biomarkers were imputed using multiple regression procedures with the survey population (Allison, 2002; Shao & Wang, 2002; Shrive, Stuart, Quan, & Ghali, 2006). Typically less than 4% of the sample had missing cases for each measure except for triglycerides (mg/dL) which had a 50.96%. Table 1 presents a list of each biomarker and sample sizes, means, and standard errors with and without imputation. We specified a regression formula where each biomarker acted as the dependent variable and a value was predicted using age, sex, marital status, years of education, origin and income level as independent variables. The following formula illustrates the regression approach for each biomarker. In this formula *y* represents the particular biomarker. The intercept and coefficients are represented by β_0 *and* β_i, respectively. Each independent variable for the regression formula used in the multiple imputation algorithms is represented by x_i and the error term is represented by e_i.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{n-1} x_{n-1} + \beta_n x_n + e_i \qquad (1)$$

Comparisons of the imputed and non-imputed means and standard errors for each element suggest that imputation had a small effect on the distribution for each biomarker. Our regression based approach has been used in previous research (Howard & Sparks, 2015, 2016b, 2016a) to deal with missing values in biomarker information also using NHANES. Comparatively averages (*means*) and standard deviations (*s.d.*) are similar between the non-imputed and the imputed descriptive statistics presented in Table 1. Given the empirical nature of the imputation and the similarity in descriptive statistics we feel confident about the imputation mechanism employed in the analysis. The GLM procedure (Hamer, Johnson, & Simpson, 1998) in SAS 9.4 (SAS Institute, 2013) was used to estimate these regression models.

Table 1 also presents clinical incidence of each biomarker for the analytic sample, which was determined based on exceeding clinically determined thresholds which has been detailed and discussed in previous literature (Howard & Sparks, 2016a). Among the 10 biomarkers considered in our study only four exceeded an incidence of 20%. These were: HDL cholesterol (mg/dL), triglycerides (mg/dL), BMI (kg/m^2) and C-reactive protein (mg/dL). Three exceeded incidence of 10%. These were: systolic blood pressure (mmHg), total cholesterol (mg/dL), and Glycohemoglobin (%). Diastolic blood pressure (mmHg), pulse rate at 60 seconds and albumin (g/dL) had an incidence below 10%. Allostatic load was incorporated in our models by assigning individuals the count of biomarkers that exceeded the clinically determined thresholds as discussed in previous research (Howard & Sparks, 2016a). The value of this allostatic load index could range from 0 to a maximum score of 10 (Geronimus, Hicken, Keene, & Bound, 2006).

Covariates

Covariates were included in the models to control for these factors, including: age, sex, education, marital status, family income, place of birth, smoking and drinking habits. Age was measured as (1) 25-40 years (reference group), (2) 41-60 years and (3) 61 years and older; individuals are assigned to each category based on their reported age at the moment of survey. Sex was measured as dichotomous variable with levels for females and males (reference group). Education was measured as (1) less than a High School diploma (reference group), (2) High School or Some college, and (3) College degree or higher. Marital status was measured as (1) never married (reference group), (2) married, (3) divorced, separated, widowed, or (4) cohabitating. Family income was measured as (1) less than $20,000 (reference group), (2) $20,000 - $64,999, (3) $65,000 or more and (4) don't know/refused. Place of birth was measured as a dichotomous variable with levels for foreign born and U.S. born (reference). Smoking was measured as (1) non-smoker (reference group), (2) current smoker, and (3) former smoker. Drinking habits were measured as (1) non-drinker (reference group), (2) 1 drink per week and (3) over 1 drink per week.

Statistical Procedures

Survey weights, strata and clustering are included in the NHANES data to account for complex sample design and adjust for probability of selection of respondents, subgroup selection, and unit non-response. The survey procedures in SAS 9.4 were employed to incorporate these design aspects into the calculation of standard errors in our statistical models. We followed NHANES guidelines for the incorporation of survey weights (WTMEC2YR), strata (SDMVSTRA) and clustering (SDMVPSU) in our analysis of the subsample who participated in the physical examination component of the study (Johnson et al., 2013). Descriptive statistics are generated for each variable of interest, covariates, and the dependent variable. Bivariate descriptive statistics were generated for all variables, with corresponding statistical comparisons. These comparisons are done using Rao-Scott Chi-Squares (Rao & Scott, 1987) and corresponding p-values to ascertain differences in characteristics for Hispanics who answered the NHANES in English and Spanish. The SURVEYMEANS and SURVEYFREQ procedure (Siller & Tompkins, 2005) in SAS 9.4 were employed to calculate all descriptive statistics and to adjust these estimates for complex survey design (SAS Institute, 2013). All descriptive statistics and corresponding comparisons are presented in Table 2 and/or illustrated in Figures 1 and 2.

Multivariate logistic regression models were used to test specific hypotheses regarding the relationship between odds of reporting poor/fair SRH by allostatic load, language of interview and additional controls. The SURVEYLOGISTIC procedure (An, 2002) also in SAS 9.4 was used fit all regression models presented in this paper (SAS Institute, 2013). The SURVEYLOGISTIC procedure was chosen because it allows to account for complex survey design in the estimation of associations and for the results to be representative of the Hispanic population in the U.S. for 2009-2010 (Hardy, 1993; Lee & Forthofer, 2005). Five logistic regression models were fit in a nested structure to assess the association between AL and language of interview with odds of poor/fair

SRH. The first model only contains AL to provide an unadjusted estimate of the odds ratio (O.R.). Model 2 includes language of survey to study initial associations for both variables. Model 3 includes age, sex, education, marital status, family income, and place of birth in addition to AL and language of interview. Model 4 incorporates smoking and drinking habits in addition to all the variables incorporated in Model 3. All these models are presented in Table 3. Model 5, presented in Table 4, incorporated an interaction effect between AL and language of interview to ascertain whether AL is associated in a different way with poor/fair SRH based on the language in which the survey was answered. Two additional models were fit stratifying by language of interview. Models 6 and 7, in Table 5, presents associations for English and Spanish respondents, respectively. Results from the logistic regression models are presented as odds ratios (O.R.) and 95% confidence intervals (95% C.I.) with corresponding significance levels. Significant levels are detailed in each table. Interaction effects are presented as coefficients (β), standard errors of the mean (*s.e.*) and p-value.

Results

Descriptive Analysis

Figure 2, presents the distribution of allostatic load scores based on the language of interview. No differences were found for the allostatic scores distribution between English and Spanish respondents. A slightly higher proportion of individuals are in the 5-7 categories for the English respondents (6.65%) than for Spanish respondents (5.01%). This indicates that from a biological standpoint Spanish respondents are as healthy as English respondents. If language does not biases the way individuals answer the SRH question no empirical difference should be found between both groups with regards to SRH and language, nor for the effect of allostatic load in predicting poor/fair SRH by language of interview. Figure 3 presents average allostatic load scores for English and Spanish respondents. Individuals who answered in English and indicate their health was good (Excellent, Very Good, Good) had an average allostatic load score of 1.77 while those who answered in Spanish had an average of 1.87. For those who indicated their health was poor/fair in English, average allostatic load score was 2.77. Average score for those who answered in Spanish was 2.20. The difference in average scores for English respondents was 1.00 and 0.33 for Spanish respondents. These differences in average scores indicate that biological profile is closer for Spanish respondents than for English respondents, based on how they rate their health.

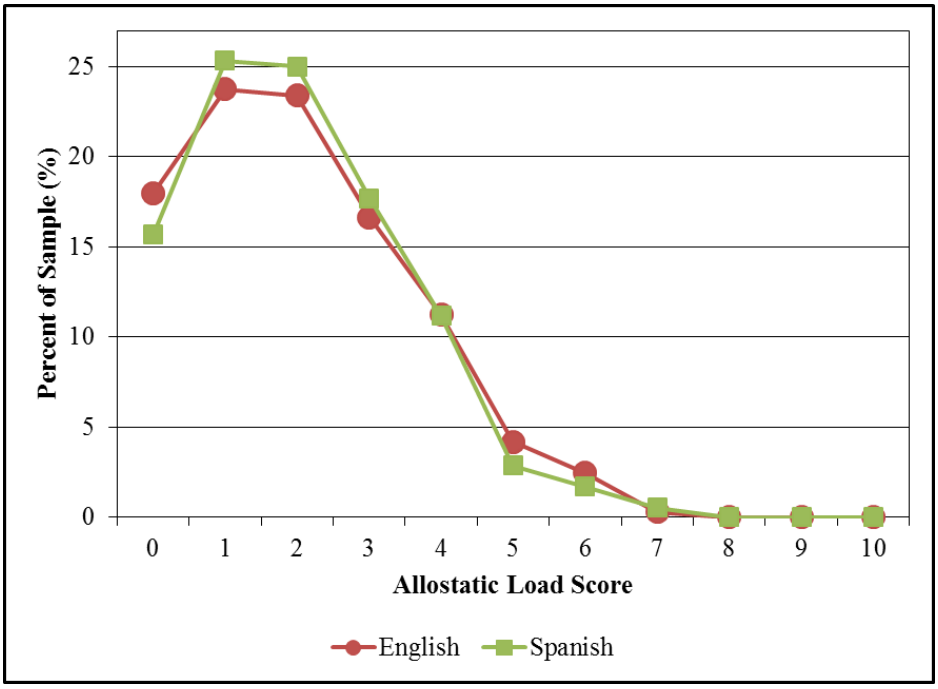

Figure 2: Distribution of allostatic load scores for English and Spanish respondents

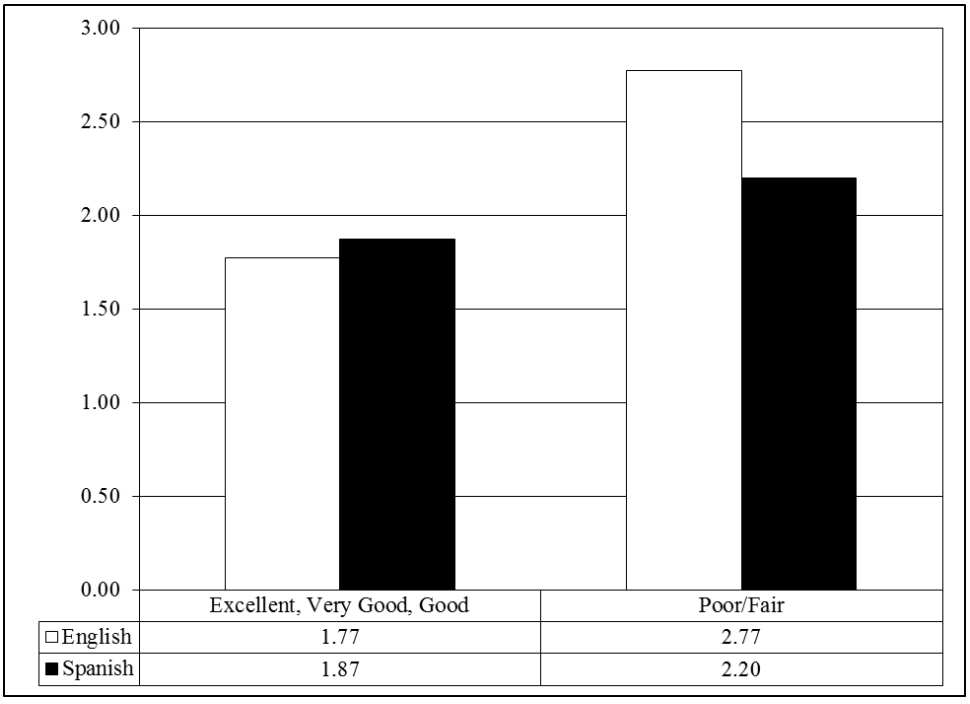

Figure 3: Mean allostatic load score by self-rated health and language of interview

Table 2 presents weighted descriptive statistics for the overall sample and by language of interview. Overall, 35.75% of Hispanic adults report poor/fair SRH. On average they had an allostatic load of 2.02 on the 10 points scale. In terms of age distribution 45.86% of the respondents

are between 25-40 years old, and the rest are 41 years or older. Among the study population 54.26% answered the survey in Spanish. Male-Female distribution is 53.59%-46.41%. The educational profile of the population was: 47.51% with less than a High School diploma, 40.91% with a High School diploma, and 11.58% had a college degree or higher. Income was predominantly in the $20,000 and $64,999 category, followed by those who earn less than $20,000 with 49.86% and 24.73%, respectively. 33.42% of respondents were born in the United States and 66.58% were born abroad. 40% were either smokers or former smokers. With regards to drinking habits 67.98% of respondents had 1 drink or more per week, while the rest indicated they were non-drinkers.

Next, we calculated descriptive statistics by language of survey. Results from the stratified analysis indicate that Spanish respondents reported poor/fair SRH at a higher proportion than English respondents (44.40% and 25.50%, respectively). From a descriptive perspective support is found for hypothesis 1. Allostatic load scores measured in the 0-10 count were found to be different with average scores being higher for English respondents. The demographic profile for both groups does not differ significantly; these include sex ratios, age profile, and marital status. Hispanics who answered in English had higher educational attainment with 26.62% of the population that have not completed a high school diploma; which contrasts with 68.50% of the Spanish respondents who are in that same educational attainment category.

Income level, which is closely associated with educational attainment profile, also differs between English and Spanish respondents. English respondents had 77.14% of the sample concentrated in the mid- and high-income categories, in comparison to 58.51% of Spanish respondents. Differences in smoking habits of both groups were significant but only at a marginally significant level ($p \leq 0.10$). Drinking habits differ significantly for English respondents with 74.09% having 1 drink or more compared to 62.82% for Spanish respondents.

Empirical Analysis

Table 3 presents nested regression models where we study initial effects of allostatic load score and language; and also while controlling for socioeconomic status, place of birth and health behaviors. Model 1 presents the initial effect of AL score for poor/fair SRH; a unit increase in allostatic load results in 28% higher odds of reporting the outcome. Model 2 incorporates language of interview to Model 1. In this model, a unit increase in allostatic load score is translated in 30% higher odds of reporting the outcome. Spanish respondents were found to be at 143% higher odds (O.R. =2.43, 95% C.I. =1.60-3.68) of reporting poor/fair SRH than English respondents.

Model 3 incorporates demographic and socioeconomic characteristics, and place of birth to the previous model. In this model, a unit increase in allostatic load score is translated in 23% higher odds of reporting poor/fair SRH. Spanish respondents were found to be at 63% higher odds of reporting poor/fair SRH when they are compared to English respondents. We found a significant positive trend for age groups, with individuals in older age groups having higher odds of reporting

the outcome when compared to the younger age group (25-40 years). Females were found to be at higher odds of reporting the outcome when compared to males (O.R. =1.44, 95% C.I. =1.14-1.82). Higher educational attainment is found to reduce odds of reporting poor/fair SRH. Individuals with a High School Diploma or some college and college or more were found to be at lower odds of reporting poor/fair SRH (O.R. =0.45 and O.R.=0.20, respectively). No difference was found for marital status. Similar to education, higher income level was associated with lower probability of reporting the outcome. Individuals who indicated earning $20,000-$64,999 or $65,000 or more were found to have lower odds of reporting poor/fair SRH in comparison to those who earn less than $20,000. No difference was found between those who expressed not knowing their income level or refused to disclose it in comparison to the reference group. No difference was found for those born outside of the U.S. compared to those who were.

Health behaviors are incorporated in Model 4, which will be referred to as the fully specified model. Higher allostatic load scores are associated with increased odds of reporting the outcome. Age group differences are weakened by the inclusion of health behaviors. In this model, only the older age group (61 years and older) is found to be at higher odds of reporting the outcome, but this difference is marginally significant ($p \leq 0.10$). Language differences are also found in the fully specified model. In model 4, individuals who answered in Spanish were found to be at 60% higher odds of reporting poor/fair SRH, which is a comparable effect to that presented in Model 3. No differences in the effects were found for sex, education, marital status, family income, or place of birth. No difference in odds of reporting the outcome was found based in smoking behaviors or individuals. Those who reported having more than one drink per week were found to be at lower odds of reporting poor/fair SRH in comparison to non-drinkers.

In order to test whether allostatic load operates differently based on language of interview for poor/fair SRH we calculated interaction effects for language and allostatic load. Table 3 presents coefficients, standard errors, and p-values for interactions based on logistic regression models. We found that AL score is associated at a lower degree with poor/fair SRH for Spanish respondents, when they are compared to those who answered the survey in English. This means that English respondents assess their SRH closer to their biological profile, assessed through the aforementioned biomarkers.

Based on the results presented in Table 4, we calculated stratified models based on language of interview. Results from the stratified models are presented in Table 5. Fully specified models were fit to ascertain differences in the effects of allostatic load in poor/fair SRH for individuals who answered the survey in different languages. Model 5 presents associations for individuals who answered in English and Model 6 does so for Spanish respondents. A unit increase in allostatic load score is associated with 45% higher odds of reporting poor/fair SRH. No difference was found for age groups. Females were found to be at 49% higher odds of reporting poor/fair SRH in comparison to male respondents (O.R. =1.49, 95% C.I.=1.04-2.16). Higher education and higher income were found to reduce odds of reporting poor/fair SRH when compared to the lower levels of each variable. Significant differences were found for all marital status, at a marginally

significant level. Married, Divorced/Separated/Widowed, and Cohabitating individuals were found to be at lower odds of reporting poor/fair SRH in comparison to never married individuals. No differences were found for smoking or drinking habits for English respondents.

Finally, Model 6 presents associations for Spanish respondents. Allostatic load was not associated poor/fair SRH (O.R. =1.10, 95% C.I. =0.97-1.26). No differences were found for age group, marital status. Females have higher odds of reporting poor/fair SRH when compared to males (O.R. =1.29, 95% C.I. =0.91-1.67), but at a marginally significant level. Higher educational attainment was associated with significantly lower odds of reporting poor/fair SRH. For income level only individuals who reported earning $65,000 or more were at significantly lower odds of reporting the outcome when compared to those in the lower income level. No difference was found for smoking patterns with regards to the outcome. Individuals who had more than one drink per week were found to be at 51% lower odds of reporting poor/fair SRH, when compared to not-drinkers.

Limitations

We acknowledge the limitations of this study. First, this analysis was based on cross-sectional data which means causality cannot be inferred. Longitudinal methods could be used to study changes in health ratings and allostatic load measurements to better understand factors associated with SRH, even among those who take the survey in different language. Half of the analytic sample contained missing values in the Triglycerides (mg/DL); we employed imputation to allow the analysis to be performed in a representative sample of Hispanics, rather than on a restricted one. We will explain how imputation affects our conclusions in the *Sensitivity Analysis* section of this paper. Finally, the NHANES was not designed to capture Hispanic population by their language of preference. Despite these limitations, we feel that the complex survey design provides us with reliable and appropriate information to ascertain significance levels in the associations noted above.

Sensitivity Analysis

Missing Values

In the *Measures* section we described the methodology employed to impute values in cases where one of the biomarkers is missing. Descriptive statistics for each biomarker with and without imputation were presented in Table 1. No marked differences are found between these values. A sensitivity analysis was deemed appropriate to measure the impact of the imputation in the associations that guided this analysis. We performed the analysis using the sample that only contained valid information for all the biomarkers, this approach has been described previously and allows for the comparison of associations or effects between samples with complete information (restricted sample) and the analytic sample (Sinharay, Stern, & Russell, 2001). First, we calculated descriptive statistics for the restricted sample. The population in the restricted sample is of 649 individuals which represent 48% of the analytic sample. Among the restricted

sample, 35.33% of respondents indicated they had poor/fair SRH, and 57.76% answered the survey in Spanish. Average AL score for respondents was 1.84 (95% C.I. =1.63-2.05); this confidence interval includes the average score of AL score for the analytic sample which is indicative of non-significant differences in AL scores between both samples. Our descriptive analysis does not indicate the restricted sample to be different from the analytic sample.

We fit three additional regression models to determine whether allostatic load and language of interview are associated with poor/fair SRH using the restricted sample. For these models we only discuss the associations of interest based on the research questions that guided our analysis. Model 8 included AL score and Model 9 only included language of interview. According to Model 8 an increase of one point in AL score is associated with 34% higher odds of reporting the outcome (O.R. =1.34, 95% C.I. =1.20-1.49). Model 9, indicated that individuals who answered in Spanish were at 113% higher odds of reporting poor/fair SRH (O.R. =2.13, 95% C.I. =1.17-3.88). Model 10, was fit to study the interaction effect; for this model we only controlled for the variables included in the regression models used for imputation (see *Data and Methods)*. In Model 10, we found allostatic load to be less predictive of poor/fair SRH for those who answered in Spanish when compared to those who answered in English at a marginally significant level (β=-0.16, s.e.=0.10, p-value=0.10). All associations are consistent with those found in the analysis of the analytic sample. These results indicate that the imputation of missing values did not bias the principal associations of our analysis.

Discussion

The data does not illustrate clear differences in the distribution of allostatic load scores for those who answered in English when compared to those interviewed in Spanish. Despite similarities in biological profiles, Spanish respondents were found to have higher percentages of respondents answering their health was poor/fair (mala/regular) as well as odds of reporting the outcome in all the empirical models. However, the allostatic load score only increased the odds of reporting the outcome for those who answered in English. This is evidence that self-reported health does not corresponds similarly to biological risk profiles for Hispanics who answered the NHANES in different languages.

Acknowledgments

We acknowledge assistance provided by the Population Research Institute at Penn State University, which is supported by an infrastructure grant by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (P2CHD041025).

References

Acevedo-Garcia, D., Bates, L. M., Osypuk, T. L., & McArdle, N. (2010). The effect of immigrant generation and duration on self-rated health among US adults 2003-2007. *Social Science & Medicine, 71*(6), 1161–1172. https://doi.org/10.1016/j.socscimed.2010.05.034

Allison, P. D. (2002). *Missing Data*. *Quantitative Applications in the Social Sciences* (Vol. 136).

An, A. (2002). Performing logistic regression on survey data with the new SURVEYLOGISTIC procedure. *Proceedings of SAS Users Group International*, (1987).

Brewer, J. V. V, Miyasato, G. S., Gates, M. A., Curto, T. M., Hall, S. A., & McKinlay, J. B. (2013). Contributors to self-reported health in a racially and ethnically diverse population: focus on Hispanics. *Annals of Epidemiology*, *23*(1), 19–24. https://doi.org/10.1016/j.annepidem.2012.09.013

Bzostek, S., Goldman, N., & Pebley, A. (2007). Why do Hispanics in the USA report poor health? *Social Science & Medicine*, *65*(5), 990–1003. https://doi.org/10.1016/j.socscimed.2007.04.028

Dubard, C. A., & Gizlice, Z. (2008). Language spoken and differences in health status, access to care, and receipt of preventive services among US hispanics. *American Journal of Public Health*, *98*(11), 2021–2028. https://doi.org/10.2105/AJPH.2007.119008

Geronimus, A. T., Hicken, M., Keene, D., & Bound, J. (2006). "Weathering" and age patterns of allostatic load scores among blacks and whites in the United States. *American Journal of Public Health*, *96*(5), 826–833. https://doi.org/10.2105/AJPH.2004.060749

Hamer, R. M., Johnson, U. R. W., & Simpson, P. M. (1998). An Introduction to the Analysis of Repeated Measures for Continuous Response Data using PROC GLM and PROC MIXED. *Proceedings of the Twenty-Third Annual SAS Users Group International Conference*, Paper 58-23.

Hardy, M. A. (1993). Regression with dummy variables. *Sage University Paper Series on Quantitative Applications in the Social Sciences*, *7*(93). https://doi.org/10.1080/00401706.1994.10485828

Howard, J. T., & Sparks, P. J. (2015). The Role of Education in Explaining Racial/Ethnic Allostatic Load Differentials in the United States. *Biodemography and Social Biology*, *61*(1), 18–39. https://doi.org/10.1080/19485565.2014.937000

Howard, J. T., & Sparks, P. J. (2016a). Does allostatic load calculation method matter? Evaluation of different methods and individual biomarkers functioning by race/ethnicity and educational level. *American Journal of Human Biology*, *28*(5), 627–635. https://doi.org/10.1002/ajhb.22843

Howard, J. T., & Sparks, P. J. (2016b). The Effects of Allostatic Load on Racial/Ethnic Mortality Differences in the United States. *Population Research and Policy Review*, *35*, 421–443.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior*, *38*(1), 21–37. https://doi.org/10.2307/2955359

Johnson, C. L., Paulose-Ram, R., Ogden, C. L., Carroll, M. D., Kruszon-Moran, D., Dohrmann, S. M., & Curtin, L. R. (2013). National health and nutrition examination survey: analytic guidelines, 1999-2010. *Vital and Health Statistics. Series 2, Data Evaluation and Methods*

Research, (161), 1–24.

Kandula, N. R., Lauderdale, D. S., & Baker, D. W. (2007). Differences in Self-Reported Health Among Asians, Latinos, and Non-Hispanic Whites: The Role of Language and Nativity. *Annals of Epidemiology*, *17*, 191–198. https://doi.org/10.1016/j.annepidem.2006.10.005

Kim, S., Egerter, S., Cubbin, C., Takahashi, E. R., & Braveman, P. (2007). Potential implications of missing income data in population-based surveys: an example from a postpartum survey in California. *Public Health Reports (Washington, D.C. : 1974)*, *122*(6), 753–763.

Lee, E., & Forthofer, R. (2005). *Analyzing complex survey data*. SAGE Publications.

Lorraine, P. J., Hammock, R. L., & Blanton, J. M. (2005). Predictors of self-rated health status among Texas residents. *Preventing Chronic Disease*, *2*, A12.

Manor, O., Matthews, S., & Power, C. (2000). Dichotomous or categorical response? Analysing self-rated health and lifetime social class. *International Journal of Epidemiology*, *29*(1), 149–157. https://doi.org/10.1093/ije/29.1.149

Rao, J. N. K., & Scott, A. J. (1987). On simple adjustments to chi-square tests with sample survey data. *The Annals of Statistics*, *15*(1), 385–397. https://doi.org/10.1214/aos/1176348654

Santos-Lozada, A. R., & Howard, J. T. (2018). Using Allostatic Load to Validate Self-rated Health for Racial/Ethnic Groups in the United States. *Biodemography and Social Biology*, *64*(1), 1–14. https://doi.org/10.1080/19485565.2018.1429891

SAS Institute. (2013). SAS/STAT Version 9.4. Cary, NC: SAS Institute, Inc.

Shao, J., & Wang, H. (2002). Sample Correlation Coefficients Based on Survey Data Under Regression Imputation. *Journal of the American Statistical Association*, *97*(458), 544–552. https://doi.org/doi:10.1198/016214502760047078

Shetterly, S. M., Baxter, J., Mason, L. D., & Hamman, R. F. (1996). Self-rated health among Hispanic vs non-Hispanic White adults: The San Luis Valley health and aging study. *American Journal of Public Health*, *86*(12), 1798–1801. https://doi.org/10.2105/AJPH.86.12.1798

Shrive, F. M., Stuart, H., Quan, H., & Ghali, W. a. (2006). Dealing with missing data in a multi-question depression scale: a comparison of imputation methods. *BMC Medical Research Methodology*, *6*, 57. https://doi.org/10.1186/1471-2288-6-57

Siller, A. B., & Tompkins, L. (2005). The Big Four: Analyzing Complex Sample Survey Data Using SAS, SPSS, STATA, and SUDAAN. *18th Annual Conference of the NorthEast SAS Users Group*.

Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods*, *6*(4), 317–329. https://doi.org/10.1037/1082-989X.6.4.317

Subramanian, S. V, Acevedo-Garcia, D., & Osypuk, T. L. (2005). Racial residential segregation and geographic heterogeneity in black/white disparity in poor self-rated health in the US: a multilevel statistical analysis. *Social Science & Medicine (1982)*, *60*(8), 1667–1679. https://doi.org/10.1016/j.socscimed.2004.08.040

van der Heijden, G. J. M. G., T. Donders, A. R., Stijnen, T., & Moons, K. G. M. (2006). Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example. *Journal of Clinical Epidemiology*, *59*(10), 1102–1109. https://doi.org/10.1016/j.jclinepi.2006.01.015

Viruell-Fuentes, E. a, Morenoff, J. D., Williams, D. R., & House, J. S. (2011). Language of interview, self-rated health, and the other Latino health puzzle. *American Journal of Public Health*, *101*(7), 1306–1313. https://doi.org/10.2105/AJPH.2009.175455

Zipf, G., Chiappa, M., Porter, K. S., Ostchega, Y., Lewis, B. G., & Dostal, J. (2013). National health and nutrition examination survey: plan and operations, 1999-2010. *Vital and Health Statistics. Ser. 1, Programs and Collection Procedures*, (56), 1–37. https://doi.org/312'.0723s 80-607914

Table 1: Weighted descriptive statistics for 10 allostatic load biomarkers with and without imputation for U.S. Hispanic Adults, NHANES 2009-2010

	Without imputed values			With imputed values			Clinically based AL	
Biomarker	n	Mean	S.E.	n	Mean	S.E.	Threshold	Incidence (%)
Cardiovascular markers								
Diastolic blood pressure (mmHg)	1,335	70.16	0.82	1,352	70.17	0.81	≥ 90	5.10
Systolic blood pressure (mmHg)	1,335	120.98	1.12	1,352	120.94	1.11	≥ 140	12.30
Pulse rate at 60 seconds	1,336	72.11	0.42	1,352	72.12	0.42	≥ 90	8.15
Metabolic markers								
Total cholesterol (mg/dL)	1,306	199.16	2.11	1,352	199.16	2.04	≥ 240	14.20
HDL cholesterol (mg/dL)	1,307	49.82	0.76	1,352	49.79	0.73	< 40	24.71
Triglycerides (mg/dL)	663	151.49	6.97	1,352	150.49	3.66	≥ 150	44.57
Glycohemoglobin (%)	1,311	5.85	0.07	1,352	5.84	0.07	≥ 6.4	12.10
Body Mass Index (kg/m^2)	1,349	29.54	0.27	1,352	29.54	0.28	≥ 30	40.14
Inflammation markers								
Albumin (g/dL)	1,306	4.26	0.01	1,352	4.26	0.01	< 3.8	4.69
C-reactive protein (mg/dL)	1,311	0.38	0.02	1,352	0.39	0.02	≥ 0.03	36.30

Survey Design: Sampling Unit=SDMVPSU, Stratum=SDMSTRA Weight=WTMEC2YR

Table 2 Weighted descriptive statistics for Poor/Fair Self-Rated Health, Allostatic Load Scores, Language, Sociodemographic Characteristics and Health Behaviors for U.S. Hispanic Adults, NHANES 2009-2010 (n=1,352)

			Language of Interview					
	Total		English		Spanish		Rao-Scott	
Variable	Mean	S.E.	Mean	S.E.	Mean	S.E.	Chi-Square	p-value
Poor/Fair Self-Rated Health	35.75	0.02	25.50	3.93	44.40	3.06	12.09	0.0005
Allostatic Load Score[1]	2.02	0.07	2.03	0.06	2.01	0.05	1,297.50	< 0.0001
Age Group								
25-40 years	45.86	0.02	48.61	3.13	43.53	3.35	2.30	0.32
41-60 years	39.44	0.01	36.47	1.94	41.95	2.04		
61 years and older	14.70	0.01	14.92	1.57	14.52	2.08		
Language of Interview								
English	45.74	0.03	--	--	--	--	--	--
Spanish	54.26	0.03	--	--	--	--	--	--
Sex								
Male	53.59	0.01	54.08	1.55	53.18	1.13	0.20	0.65
Female	46.41	0.01	45.92	1.55	46.82	1.13		
Education								
Less than High School	47.51	0.02	22.62	2.37	68.50	3.49	101.42	< 0.0001
High School/Some College	40.91	0.01	58.57	1.44	26.02	2.28		
College Degree or higher	11.58	0.01	18.81	1.78	5.48	1.71		
Marital Status								
Never Married	13.53	0.02	16.19	2.78	11.29	2.32	3.43	0.33
Married	54.39	0.03	52.51	3.44	55.98	3.43		
Divorced/Separated/Widowed	19.79	0.02	20.31	2.31	19.34	1.73		
Cohabitating	12.28	0.01	10.99	1.59	13.38	1.76		
Family income								
Less than $20,000	24.73	0.02	19.36	2.79	29.25	2.52	77.28	<0.0001
$20,000-$64,999	49.86	0.02	45.20	3.47	53.78	2.11		
$65,000 or more	17.18	0.02	31.94	4.10	4.73	1.02		
Don't know/Refused	8.23	0.01	3.50	1.20	12.23	2.16		
Place of Birth								
U.S. Born	33.42	0.05	66.68	5.51	5.38	1.46	9,358,152.00	< 0.0001
Foreign Born	66.58	0.05	33.32	5.51	94.62	1.46		
Smoking								
Non-smoker	61.62	0.02	58.24	2.44	64.57	2.11	4.60	0.10
Current smoker	18.48	0.01	20.95	1.65	16.40	1.60		
Former smoker	19.91	0.01	20.82	2.01	19.14	1.47		
Drinking Habits								
Non-drinker	32.02	0.02	25.91	2.00	37.17	2.26	28.49	<0.0001
1 drink per week	50.07	0.01	50.72	1.62	49.51	1.99		
Over 1 drink per week	17.91	0.01	23.37	1.60	13.31	1.26		
Unweighted n	1,352		607		745			

1. Differences for continous outcomes were analyzed using a F-test using a regression algorithm

Survey Design: Sampling Unit=SDMVPSU, Stratum=SDMSTRA Weight=WTMEC2YR

Table 3 Weighted logistic regression models for Poor/Fair Self-Rated Health and Allostatic Load Scores, Language of Interview, Sociodemographic Characteristics and Health Behaviors for U.S. Adults, NHANES 2009-2010 (n=1,352)

	Model 1 - Allostatic Load		Model 2 - Allostatic Load, Language		Model 3 - Allostatic Load, SES		Model 4 - AL, Language, SES, Health Behaviors	
Variable	O.R.	95 % C.I.	O.R.	95 % C.I.	O.R.	95 % C.I.	O.R.	95 % C.I.
Allostatic Load Score	1.28	(1.21-1.36)***	1.30	(1.23-1.38)***	1.23	(1.14-1.33)***	1.22	(1.13-1.32)***
Age Group								
25-40 years					1.00		1.00	
41-60 years					1.36	(0.99-1.87)†	1.28	(0.91-1.82)
61 years and older					1.53	(1.15-2.02)**	1.33	(0.97-1.82)†
Language of Interview								
English			1.00		1.00		1.00	
Spanish			2.43	(1.60-3.68)***	1.63	(1.15-2.29)**	1.60	(1.13-2.26)**
Sex								
Male					1.00		1.00	
Female					1.44	(1.13-1.84)**	1.36	(1.07-1.73)**
Education								
Less than High School					1.00		1.00	
High School/Some College					0.45	(0.30-0.67)***	0.46	(0.31-0.68)***
College Degree or higher					0.20	(0.12-0.34)***	0.21	(0.12-0.37)***
Marital Status								
Never Married					1.00		1.00	
Married					0.96	(0.72-1.28)	0.97	(0.73-1.28)
Divorced/Separated/Widowed					0.80	(0.59-1.09)	0.82	(0.62-1.09)
Cohabitating					0.71	(0.43-1.20)	0.70	(0.42-1.18)
Family income								
Less than $20,000					1.00		1.00	
$20,000-$64,999					0.68	(0.46-0.99)*	0.70	(0.49-1.01)†
$65,000 or more					0.45	(0.30-0.65)***	0.47	(0.33-0.67)***
Don't know/Refused					0.80	(0.43-1.48)	0.83	(0.47-1.48)
Place of Birth								
U.S. Born					1.00		1.00	
Foreign Born					1.30	(0.92-1.82)	1.27	(0.90-1.81)
Smoking								
Non-smoker							1.00	
Current smoker							1.30	(0.82-2.08)
Former smoker							1.29	(0.89-1.87)
Drinking Habits								
Non-drinker							1.00	
1 drink per week							0.79	(0.51-1.20)
Over 1 drink per week							0.56	(0.32-0.96)*
Intercept	-1.11***		-1.64***		-0.85**		-0.70*	

Significance Level: † $p \leq 0.10$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

Survey Design: Sampling Unit=SDMVPSU, Stratum=SDMSTRA Weight=WTMEC2YR

Table 4: Weighted interaction effects from logistic regression models for Allostatic Load Score and Race/Ethnicity and Poor/Fair Self-Rated for U.S. Hispanic Adults, NHANES 2009-2010 (n=1,352)

Interaction (Model 6)		**Coeff.**	**s.e.**	**p-value**
English * AL Score		--	--	
Spanish * AL Score		-0.26	0.09	0.002

Survey Design: Sampling Unit=SDMVPSU, Stratum=SDMSTRA Weight=WTMEC2YR
Model includes controls for AL score, language, age, sex, education, marital status, income, place of birth, smoking and drinking habits.

Table 5 Weighted logistic regression models for Poor/Fair Self-Rated Health and Allostatic Load Scores, Sociodemographic Characteristics and Health Behaviors stratified by language of interview for U.S. Hispanic Adults, NHANES 2009-2010 (n=1,352)

	Model 6: English (n=607)		**Model 7: Spanish (n=745)**	
Variable	O.R.	95 % C.I.	O.R.	95 % C.I.
Allostatic Load Score	1.44	(1.30-1.59)***	1.10	(0.97-1.26)
Age Group				
25-40 years	1.00		1.00	
41-60 years	1.25	(0.68-2.31)	1.31	(0.87-1.97)
61 years and older	1.43	(0.81-2.53)	1.47	(0.91-2.38)
Sex				
Male	1.00		1.00	
Female	1.49	(1.04-2.16)*	1.29	(1.00-1.67)†
Education				
Less than High School	1.00		1.00	
High School/Some College	0.40	(0.24-0.66)**	0.51	(0.32-0.84)**
College Degree or higher	0.14	(0.07-0.26)***	0.34	(0.19-0.61)***
Marital Status				
Never Married	1.00		1.00	
Married	0.64	(0.39-1.06)†	1.33	(0.88-2.02)
Divorced/Separated/Widowed	0.61	(0.35-1.07)†	1.01	(0.57-1.78)
Cohabitating	0.47	(0.20-1.06)†	0.90	(0.48-1.68)
Family income				
Less than $20,000	1.00		1.00	
$20,000-$64,999	0.50	(0.23-1.10)†	0.82	(0.60-1.12)
$65,000 or more	0.46	(0.26-0.80)**	0.50	(0.24-1.05)†
Don't know/Refused	0.95	(0.39-2.30)	0.86	(0.46-1.60)
Place of Birth				
U.S. Born	1.00		1.00	
Foreign Born	1.25	(0.78-2.00)	1.05	(0.66-1.69)
Smoking				
Non-smoker	1.00		1.00	
Current smoker	1.18	(0.82-1.70)	1.37	(0.67-2.78)
Former smoker	1.39	(0.85-2.29)	1.17	(0.67-2.05)
Drinking Habits				
Non-drinker	1.00		1.00	
1 drink per week	0.78	(0.50-1.20)	0.73	(0.47-1.16)
Over 1 drink per week	0.65	(0.34-1.22)	0.49	(0.29-0.83)**
Intercept	-0.54		-0.32	

Significance Level: † $p \leq 0.10$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

Survey Design: Sampling Unit=SDMVPSU, Stratum=SDMSTRA Weight=WTMEC2YR