HHS Public Access

Author manuscript

Psychometrika. Author manuscript; available in PMC 2019 December 01.

Published in final edited form as:

Psychometrika. 2018 December ; 83(4): 991–1006. doi:10.1007/s11336-018-9612-2.

NONLINEAR PREDICTIVE MODELS FOR MULTIPLE MEDIATION ANALYSIS: WITH AN APPLICATION TO EXPLORE ETHNIC DISPARITIES IN ANXIETY AND DEPRESSION AMONG CANCER SURVIVORS

Qingzhao Yu,
LOUISIANA STATE UNIVERSITY HEALTH SCIENCES CENTER

Kaelen L. Medeiros,
AMERICAN COLLEGE OF SURGEON

Xiaocheng Wu, and
LOUISIANA TUMOR REGISTRY

Roxanne E. Jensen
LOMBARDI COMPREHENSIVE CANCER CENTER

Abstract

Mediation analysis allows the examination of effects of a third variable (mediator/confounder) in the causal pathway between an exposure and an outcome. The general multiple mediation analysis method (MMA), proposed by Yu et al., improves traditional methods (e.g., estimation of natural and controlled direct effects) to enable consideration of multiple mediators/confounders simultaneously and the use of linear and nonlinear predictive models for estimating mediation/confounding effects. Previous studies find that compared with non-Hispanic cancer survivors, Hispanic survivors are more likely to endure anxiety and depression after cancer diagnoses. In this paper, we applied MMA on MY-Health study to identify mediators/confounders and quantify the indirect effect of each identified mediator/confounder in explaining ethnic disparities in anxiety and depression among cancer survivors who enrolled in the study. We considered a number of socio-demographic variables, tumor characteristics, and treatment factors as potential mediators/confounders and found that most of the ethnic differences in anxiety or depression between Hispanic and non-Hispanic white cancer survivors were explained by younger diagnosis age, lower education level, lower proportions of employment, less likely of being born in the USA, less insurance, and less social support among Hispanic patients.

Correspondence should be made to Qingzhao Yu, Biostatistics Program, School of Public Health, Louisiana State University Health Sciences Center, 3rd Floor, 2020 Gravier Street, New Orleans, LA 70112, USA. qyu@lsuhsc.edu.

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11336-018-9612-2) contains supplementary material, which is available to authorized users.

Keywords

ethnic disparity; health-related quality of life; mediation/confounding analysis; MY-Health study; nonlinear models

1. Introduction

The ultimate goal of health care is to restore or preserve functioning and well-being related to patients' health, known as the health-related quality of life (HRQOL, Rodenberg, 2006). In 2014, the American Cancer Society (ACS) reported that there were as many as 14.5 million cancer survivors living in the USA. The SEER-Medicare Health Outcomes Study showed that patients with lung, colorectal, or prostate cancer had significantly worse HRQOL in terms of mental and physical health outcomes compared with cancer-free control subjects (Reeve et al., 2009, 2012, 2015) due to effects from both the disease and the treatments used to fight cancer. Moreover, there are disparities in HRQOL outcomes among cancer survivors across race and ethnicity. Gallicchio et al. (2013) examined the self-reported HRQOL scores among breast cancer survivors by race and found that black women reported significantly more physical functioning limitations than whites. Culver et al. (2002) identified more severe level of "psychological distress" (which includes depression and anxiety) in the Hispanic breast cancer survivors versus matched non-Hispanic white survivors. There is a growing recognition that racial/ethnic differences in HRQOL exist among cancer survivors. Overall, compared with non-Hispanic whites (NHW), blacks and Hispanics whites (HW) exhibit poorer HRQOL outcomes of general physical, mental health, and role functioning. Our goal is to identify and differentiate the effect of modifiable risk factors that contribute to the racial/ethnic disparities in HRQOL in the USA in order to inform targeted precision medicine efforts. For this purpose, we demonstrate a novel multiple mediation analysis method that enables exploring mechanisms that underlie these disparities and provides information to guide screening methods and interventions that optimally reduce them. In this paper, we focus the research on exploring the ethnic disparities in anxiety and depression among cancer survivors using data collected from the Measuring Your Health (MY-Health) study that was funded by the National Health Institute (NIH). To avoid the confounding effects from race (most Hispanics are whites), this research is stratified to compare between Hispanic and non-Hispanic whites only.

The MY-Health study collected information on HRQOL measures developed by the Patient-Reported Outcomes Measurement Information System (PROMIS) among cancer survivors enrolled in the study. Four population-based cancer registries of the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program participated in the study. The MY-Health study database was linked with the cancer registry data on tumor characteristics and cancer treatments, based on which we conduct research to identify and order factors that contribute to the health disparities.

We apply a novel mediation analytic method to differentiate the relative effects from different risk factors on ethnic disparities in anxiety and depression. A third variable can intervene in the relationship between an explanatory variable and a response variable

through many forms. In this paper, we focus on the two forms: mediation or confounding. Although they are conceptually distinctive, MacKinnon (2000) claimed that these effects are statistically similar in the sense that all of them measure the change of association between the exposure (X, e.g., ethnicity) and response (Y, e.g., anxiety or depression score) variables when considering a third variable (denoted as M). Therefore, the statistical methods developed for mediation framework can be used for confounding effect analysis, although the scientific interpretations of the analysis might be different. In general, indirect effect (IE) refers to effect through X to M, and then to Y, and direct effect (DE) refers to effect from X directly to Y after adjusting for all related factors. The relationship among the variables can be shown by the graphical model in Fig. 1. Note that $X\,M_i\,Y$ is a path between X and Y, where M_i could denote one mediator/confounder, or a vector−of mediators/confounders−within which there are complicated associations. Mediation analysis refers to the statistical techniques attempting to make inferences on mediation/confounding effects (effects from X to Y through different paths). In general, there are two groups of mediation analytic methods, linear structural equations and counterfactual frameworks, that can be applied to differentiating risk factors that contribute to the ethnic disparities in anxiety/depression (Alwin & Hauser, 1975; Judd & Kenny, 1981; Robins & Greenland, 1992; Mackinnon & Dwyer, 1993; Have et al., 2007; Albert, 2008; Vanderweele & Vansteelandt, 2009; VanderWeele, 2009). Exploring ethnic disparities in HRQOL outcomes requires a statistical mediation method that meets the following challenges: (1) the response, exposure, and mediators/confounders can take any formats: continuous, binary, or multi-categorical; (2) transformations among exposure, mediators/confounders, and other covariates should be allowed to account for potential nonlinear associations and interactions among variables; (3) if we can differentiate individual mediation/confounding effect from each mediator/confounder, then we gain knowledge on what are the most important mediators/confounders that explain the exposure–outcome relationship; (4) joint effects from highly correlated variables can be calculated, and (5) can make inferences on estimates of mediation/confounding effects (calculate variances and confidence intervals).

We apply a novel mediation analysis method with multiple mediators/confounders to the MY-Health data to explore ethnic disparities in depression and anxiety among cancer survivors in this paper. The rest of the paper is organized as follows: Sect. 2 introduces the MY-health study and data measurements. Section 3 reviews mediation analysis methods and the novel MMA method and its algorithms for statistical inferences. We also introduce the R package, *mma*, for data analysis. Results for the exploration of ethnic disparities in HRQOL are discussed in Sect. 4. And finally, we make conclusions, discuss limitations of the research, and point out future research directions in Sect. 5.

2. PROMIS and the MY-Health Study

2.1. PROMIS

Health-related quality of life is defined in different ways. The most commonly used one is by Patrick and Erickson (1993): “Health related quality of life is the value assigned to duration of life as modified by the impairments, functional states, perceptions, and social opportunities that are influenced by disease, injury, treatment, or policy.” The Patient-

Reported Outcomes Measurement Information System (PROMIS) is an NIH-funded initiative designed to provide a standard set of patient-reported outcome measures using modern psychometric theory that provides scores referenced against the general US population. PROMIS uses IRT-calibrated item banks across many symptom and functional domains (Jensen et al., 2015).

PROMIS has created many item banks that cover a range of function and symptom issues prominent in cancer survivorship such as physical function, pain interference, fatigue, sleep disturbance, depression, anxiety, ability to participate in social roles, and cognitive function (NIH, 2017b). For cancer patients, PROMIS measures can provide insight into symptom and functional issues during initial treatment and follow-up care.

PROMIS measures use a *t*-score metric for all domains, with a mean of 50 and a standard deviation of 10. Physical function, pain interference, fatigue, sleep disturbance, depression, and anxiety are calibrated, so a score of 50 reflects the average response in the US general population, and 10 points reflect a standard deviation for the US population. Reference values specific to cancer survivors are also available. The meaning of a high score depends on the domain it measures. For example, a higher fatigue score means higher fatigue severity, but a higher physical function score indicates better functions. The NIH's Assessment Center (http://www.assessmentcenter.net/) provides the PROMIS instrument that can automatically calculate domain scores when survey questions are answered online. If researchers choose not to use the Assessment Center instrument, the Web site also provides manuals on how to measure the domains. The manuals provide instructions on how to calculate raw scores from the questions answered and tables that convert raw scores to scaled *t*-scores. The manuals also give instructions on calculating standard errors of the estimated *t*-scores (NIH, 2017a, c).

2.2. MY-Health Study

The Measuring Your Health (MY-Health) study was supported by NIH and includes investigators from the Georgetown University, the Fred Hutchinson Cancer Research Center, and three SEER cancer registries (California, New Jersey, and Louisiana). The main goal of this study was to validate eight PROMIS domains in a broad, diverse community-based cancer population. Patients were identified and contacted 6–13months postdiagnosis and completed a 6-month follow-up.

The study group chose seven cancer sites (i.e., breast, cervical, uterine, prostate, colorectal, non-Hodgkins lymphoma, and nonsmall cell lung). Survey participants were randomly selected using a stratified sampling method. The strata are designed based on four racial/ethnic subgroups: black, non-Hispanic white, Asian/Pacific Islander (API), and Hispanic, and three age-based subgroups: age 21–49, 50–64, 65–84. The stratified random sampling method was used to ensure diversity in the surveyed population. The participants were recruited through the three participating cancer registries (Jensen et al., 2016, 2017).

Measurement equivalence across the ethnic groups (Hispanic whites and non-Hispanic whites) is an essential prerequisite for exploring the ethnic disparities in the PROMIS measures. Reeve and Terest (2016) reviewed papers that evaluated the measurement

equivalence of the PROMIS and have provided strong evidence supporting through differential item functioning (DIF) the measurement equivalence of the PROMIS in ethnically, socio-demographically diverse groups. Specifically for the measurements of anxiety and depression in the MY-Health study, Teresi et al. (2016a,b) have shown that none survey item had a high magnitude of DIF, therefore supporting the use of depression and anxiety measures in the MY-Health study across ethnically diverse groups.

MY-Health survey population consists of 5506 cancer patients who completed the baseline survey. Among them, 2928 individuals are either NHW (2160) or HW (667). Ethnicity is used as the exposure variable in our study, with the standardized anxiety PROMIS score and depression PROMIS score as the outcomes for two separate analyses. The data cleaning process is described briefly in the supplementary material. We first check whether the PROMIS scores are different among different ethnic groups. ANOVA was used for the tests, and the results are shown in Table 1.

Yost et al. (2011) stated that a difference of around 3.0 in PROMIS scores between groups is considered as important. The difference in the average score between NHW and HW is 3.58 for anxiety and 2.90 for depression. We conclude that the differences in the anxiety and depression PROMIS scores between NHW and HW are clinically meaningful. Therefore, we performed mediation analyses on the dataset to identify potential specific risk factors that may be used to explain this identified difference.

3. Methods

3.1. General Definitions of Mediation/Confounding Effects for Multiple Mediators/Confounders

Yost et al. (2014) proposed general definitions of mediation/confounding effects. These definitions are related to conventional mediation analysis (e.g., Mackinnon & Dwyer, 1993; Mackinnon, 2008), but are more general in that they are consistent for different types of predictors or outcomes. The definitions are based on three basic assumptions (VanderWeele & Robinson, 2014; VanderWeele et al., 2014) that were widely used in defining mediation/confounding effects:

Assumption 1 No unmeasured confounder for the exposure–outcome relationship;

Assumption 2 No unmeasured confounder for the mediator–outcome relationship;

Assumption 3 Mediator M_i is not causally prior to other mediators/confounders M_{-i}.

With these assumptions, given other factor(s) **Z**, Yu et al. (2014) define the mediation/confounding effects through the rate of change in Y when X changes from x by a u^* unit, where u^* is the smallest unit of X, such that $u^*+x \in \text{domain}(X)$, for any $x \in \text{domain}(X)$. Based on the notations, the average total effect is defined as

$$\text{ATE}_{\mathbf{Z}} = E_{x*}\left[\lim_{u \to u*} \frac{E(Y(x^*+u)|\mathbf{Z}) - E(Y(x^*)|\mathbf{Z})}{u}\right],$$

and the average direct effect not through M_i (the ith mediator) is

$$\mathrm{ADE}_{\backslash M_i}|\mathbf{Z} = E_{x*}E_{M_i}\left[\lim_{u \to u*} \frac{E\left(Y\left(x*+u,\, M_i, \mathbf{M}_{-i}(x*+u)\right)\middle|\mathbf{Z}\right) - E\left(Y\left(x*,\, M_i, \mathbf{M}_{-i}(x*)\right)\middle|\mathbf{Z}\right)}{u}\right],$$

where $\mathbf{M}_{-i}$ denotes the vector of mediators excluding M_i. The definition of direct effect not from M_i is analogous to the natural direct effect [defined as $E\{Y(x^*, M(x))\, Y(x, M(x))$ at two different levels of the exposure variable x and x^* (Robins & Greenland, − 1992; Pearl,}2001)] that allows for natural variation in the levels of the mediator between subjects. Instead of allowing M_i to vary conditionally on a fixed level of the predictor, we allow it to vary at its marginal distribution. The average direct effect measures the average changing rate in the potential outcome with the modification of X, where the distribution of M_i does not change with X (marginal distribution), while distributions of all other mediators/ confounders change with X (conditional distributions). Therefore, we call the rate of change in Y the direct effect of X on Y not from M_i, which includes the direct effect of X on Y and the indirect effects through mediators/confounders other than M_i. The definition of indirect effect through M_i is defined straightforwardly as $\mathrm{AIE}(M_i|\mathbf{Z}) = \mathrm{AIE}_{\mathbf{Z}} - \mathrm{ADE}_{\backslash Mi|\mathbf{Z}}$. In summary, the general indirect effect a mediator defined as the changing rate in the outcome when the relationship from the exposure to the mediator/confounder is deactivated. With the general definition, the cross-world assumption, which requires that no measured or unmeasured effect of the exposure that confounds the mediator-outcome relationship (Avin et al., 2005), is not needed to identify the mediation/confounding effect. As an example, Fan (2012) discussed the situation when there are exposure–mediator interaction effects on the outcome for binary and continuous exposures. We include the result in the supplementary material.

Applying the definitions to the special case in our analysis, the exposure variable is the binary ethnicity (NHW, $X = 1$ or HW, $X = 0$). The average total effect is $E(Y|\mathbf{Z}, X=1) E(|\mathbf{Z}, X=0)$, and the average direct effect of X on Y not through Mj is $\mathrm{ADE}\backslash Mj|\mathbf{Z} = \mathrm{DE}\backslash Mj|\mathbf{Z}(0) = Em\, j\{E\mathbf{M}-j|X=1[E(Y|\mathbf{Z}, Mj = m\, j, \mathbf{M}-j, X = 1)] - E\mathbf{M}-j|X=0[E(Y|\mathbf{Z}, Mj = m\, j, \mathbf{M}-j, X = 0)]\}$.

Compared with conventional definitions of the average total effect that look at the average differences in the expected Y between $x = a$ and $x = a^*$ (for example, Mackinnon & Dwyer, 1993; VanderWeele, 2009; Vansteelandt & Daniel, 2017), the definitions by Yu et al. (2014) are based on the rate of change. The motivation is that first the exposure levels a and a^* do not have to be preset, thus generalizing the definitions from binary to multi-categorical or continuous exposures, and second, the effects will not change with the unit of X. Yu et al. (2014) have shown that the proposed method is equivalent to the conventional method in the single continuous mediator case. They also established the relationship between the proposed definitions of direct or indirect effect and the natural direct or indirect effects in the single binary mediator case. The mediation analysis is generalized so that it can handle binary, multi-categorical or continuous exposure, mediator, and response variables. In addition, algorithms were provided by Yu et al. (2014) so that general predictive models including multivariate additive regression trees (MART) and Cox proportional hazard models can be used in addition to generalized linear models for fitting variable relationships.

In this paper, mediation analysis is used to explore the racial disparity in HRQOL measurements. There are controversies in interpreting the race effect since it is impossible to establish causal effect of race. However, "effect of race" can be defined. VanderWeele and Robinson (2014) extensively discussed the challenges and different interpretations of race effect. We use their interpretation where "the effect of race involves the joint effects of race-associated physical phenotype (e.g., skin color), parental physical phenotype, genetic background, and cultural context when such variables are thought to be hypothetically manipulable and if adequate control for confounding were possible" (VanderWeele & Robinson, 2014). Combined with the method proposed above, direct effect of race is interpreted as the remaining racial disparity if distributions of various risk factors across racial groups could be equalized. The indirect effect from a certain risk factor (mediator/confounder) is the change in the health disparity if the distributions of the risk factor can be set as the same across racial groups, while distributions for other risk factors are kept as observed. With this interpretation, the hypothetical manipulation on race is not required. Instead, the interpretation was performed by framing around more manipulable risk factors, such as environmental and healthcare facility variables. For detailed discussion on explaining "race effect," the readers are referred to VanderWeele and Robinson (2014).

3.2. Multiple Mediation Analysis for Non-/Semi-Parametric Predictive Models with Binary Exposure

When (generalized) linear regression is insufficient to describe the relationships among variables, we seek to use the non-/semi-parametric predictive models. Algorithms 1 and 2 that derived directly from the definitions of mediation/confounding effects provide the method to calculate mediation/confounding effects when the exposure variable is binary with the assumption that the sample size at each exposure level is large. More general mediation analysis with any types of exposures is discussed in a separate paper (Yu & Li, 2017). The R package *mma* was built using the algorithms to make inferences on mediation/confounding effects.

The algorithms are based on the assumption that the prediction model of $E(Y)$ has the form

$$E(Y_i) = f(x_i, M_{1i}, \ldots, M_{pi}), \text{ for } i = 1, \ldots, n, \quad (1)$$

In the R package *mma*, MART and generalized linear models (GLM) can be chosen to build f.

Algorithm 1 Estimate the total effect:

1. Randomly draw N vectors of mediators/confounders from the sub-population where $X = 0$, denoting as $(M_1 j_1, \ldots, MTpj1)$, for $j=1, \ldots, N$.
2. Randomly draw N vectors of mediators/confounders= from the sub-population where $X = 1$, denoting as $(M_{1j2}, \ldots, M_{pj2})^T$, for $j = 1, \ldots, N$.
3. $\text{TE}=\frac{1}{N}\left[\sum_{j=1}^{N} f\left(1, M_{1j2,\ldots,} M_{pj2}\right)^T - \sum_{j=1}^{N} f\left(0, M_{1j1,\ldots,} M_{pj1}\right)\right]$.

We can also calculate the total effect using $E(Y|x=1) - E(Y|x=0)$ directly from the observations.

Algorithm 2 Estimate the direct effect not through M_k:

1. Use the samples generated by Steps 1 and 2 of Algorithm 1.
2. Combine the vectors $\{M_{kj1}\}_{j=1}^{N}$ and $\{M_{kj1}\}_{j=1}^{N}$ and randomly permute the combined vector, denote the new vector as $\{\widetilde{M}_{kj1}\}_{j=1}^{2N}$. $\{\widetilde{M}_{kj1}\}_{j=1}^{2N}$ forms a sample of M_k from itsmarginal distribution.
3. $DE_{\setminus Mk}$ is estimated by
$$\frac{1}{N}\Big[\sum_{j=1}^{N} f\left(1, M_{1j2}, \ldots, M_{k-1,j2}, \widetilde{M}_{kj}, M_{k+1,j2}, \ldots, M_{pj2}\right) - \sum_{j=1}^{N} f\left(0, M_{1j1}, \ldots, M_{k-1,j1}, \widetilde{M}_{k,(N+j)}, M_{k+1,j1}, \ldots, M_{pj1}\right)\Big].$$

Due to the randomness brought in by sampling, the two algorithms are repeated, and the average results from the repetitions are estimates of mediation/confounding effects. For the analysis in this paper, we did 20 permutations.

3.3. The R Package mma

The *mma* package, available on the Comprehensive R Archive Network (CRAN) (Yu & Li, 2017), was generated based on the above two algorithms for mediation analysis. It has two sets of functions for multiple mediation analysis. One is the step-by-step process, where the function *data.org* helps identify the mediators/confounders and covariates. Then the organized data sets are read into the function *med* to estimate the mediation effects [indirect effect (IE), total effect (TE), direct effect (DE)]. Finally, the function *boot.med* helps to report summary statistics of the estimated mediation/confounding effects, where the standard deviations and confidence intervals are calculated using the bootstrap method. An alternative process combines the three steps in one function, *mma*. For both sets of functions, linear or nonlinear predictive models can be chosen for mediation analysis. By default, generalized linear models are used to model the associations among variables. If nonlinear method is chosen, MART and smoothing splines are used to model the relationships.

MART, originally proposed by Friedman (2001), is an ensemble technique that aims to improve the performance of a single model by fitting many models and combining them for prediction. MART employs two algorithms: “regression tree” from Classification And Regression Tree (CART) (Breiman et al., 1984) and “boosting” which builds and combines a collection of models, i.e., trees. We choose MART as a modeling method in mediation analysis since first, MART is able to capitalize on the nonlinear relationships between the dependent and independent variables with no need for specifying the basis functions. Second, due to the hierarchical splitting scheme in regression trees, MART is able to capture complex and/or high-order interaction effects. And third, as a tree-based method, MART can handle mixed-type predictors (i.e., quantitative and qualitative covariates) and missing values in covariates. Moreover, MART is able to handle time-to-event (survival) outcomes

(Yu et al., 2009). Smoothing spline is a functional estimate that balances between the "goodness-of-fit" and "smoothness." Smoothing spline is fitted to a set of spline basis functions, typically by least squares with a penalty on the "roughness." Since spline bases are used for model fitting, nonlinear relationships can be fitted among variables. Also, the method can be easily extended to deal with data sets of hierarchical structure (Gu, 2013).

The *mma* package provides generic functions to help explain the results from mediation analysis. For our research, any variable that is significantly related to ethnicity and is significantly associated with the outcome when other variables are included in the model is treated as a potential mediator. Variables that significantly associate with the outcome when other variables are adjusted, but not significantly related to the ethnicity are included in the mediation analysis as other covariates. Variables that are not related to the outcome when other variables are considered are excluded for further analysis. The function *data.org* tests associations among variables and identifies mediators/confounders and covariates. Its results can be summarized to show selected variables and the test results (*p* values) for each associations of interests. Moreover, the outputs from the function *boot.med* or *mma* can be summarized to show the inferences results on mediation/confounding effects (estimates, standard deviations, and confidence intervals). The graphic function, *plot*, helps researchers visualize the complicated relationships and explain the ethnic differences in health outcome. The readers are referred to Yu and Li (2017) for details on how to use the *mma* package and Yu et al. (2017) for an example of applying the *mma* package.

4. Results

We explore the ethnic disparity in HRQOL outcomes (anxiety and depression) using the mediation analysis with multiple mediators/confounders (Yu et al., 2014) using the MY-Health study on cancer patients.

4.1. Selection of Potential Mediators/Confounders

Potential mediators/confounders are the variables that can potentially explain the ethnic disparities. Usually, mediators/confounders are significantly related to ethnicity and with the HRQOL outcome (Baron & Kenny, 1986). Our first step of analysis was to clean the MY-Health data set and identify potential mediators/confounders.

The MY-Health data set was well maintained by the research group. We performed the following two steps for data cleaning. Firstly, some variables were created. We created the variable *comorbidities* to indicate the comorbidity conditions of patients using the "Self-Reported Comorbidity" section of the MY-Health survey. The comorbid conditions (e.g., heart attack and asthma) were reported by patients with the possible answers for each condition "Yes," "No," or "Unsure." The variable *comorbidities* record the total number of comorbidities a patient reported to have experienced. Its values range from 0 to 13. We also regrouped some variables into more convenient categories. For example, "surgery of the primary site" was site specific indicating the type of surgery. For example, tumor destruction and resection were code as 10–19 and 20–80, respectively. We combined the values into the variable "surgery," which indicates whether surgery was performed as part of the treatment. Secondly, missing data were evaluated. Variables like "age to USA" and "years lived in

USA," which do not apply to patients born in the USA, tended to have many missings. For such cases, we included the variable "US born" as an indicator of whether the patient was born in the USA and excluded "age to USA" and "years lived in USA" for further analysis. Based on literature reviews and data availability, we included all variables listed in Table 2 other than the exposure variable and outcomes, as potential mediators/confounders. In the table, "social support" is a PROMIS score that measures the companionship, and emotional, information, and instrumental supports one can get. The variable "spirituality" measures one's spiritual support. The variable "income" was recoded to five ordered categories from "less than $10,000" to "$200,000 or more." Education was also ordered to five categories from "lower than high school" to "graduate degree." Since both variables are ordered and MART was used to fit models in which linear relationship assumption is not required, we treated both variables as continuous. The variables, their variable formats and data sources are listed in the table.

To understand the relationships among variables and to identify potential mediators/confounders and covariates, we tested the significance of two associations: (1) between ethnicity and the potential mediator/confounder; and (2) between the potential mediator/confounder and the outcome, when other variables are controlled. For this initial selection process, we set the significance level at 0.1 to reduce the risk of falsely ignoring important variables. To test the first relationship, we use ANOVA for continuous mediators/confounders and Chi-square test for categorical mediators/confounders. For the second relationship, we use the type-3 tests in a linear model with all potential variables. Table 3 shows the test results and identified potential mediators/confounders (with $*$) and covariates (with^{-}) for outcomes *anxiety* and *depression*, respectively. Note that the significant exposure–mediator relationship and mediator–outcome relationship are not prerequisites for variables to be included in the mediation analysis. Since the variables employment, insurance, and income are highly correlated, we forced the three variables into the analysis as potential mediators/confounders and estimated the joint effect from them.

From Table 3, we found that mostly, the mediators/confounders and covariates were similarly selected for both *anxiety* and *depression*. The differences are that "days from diagnosis" was chosen as a potential mediator/confounder for depression, but not for anxiety. Also, "radiation" and "hormonal therapy" were covariates for *depression* but not for *anxiety*, while "grade of cancer" was chosen as a covariate for *anxiety* but not for *depression*.

4.2. Mediation/Confounding Effects Explaining the Ethnic Disparities in HRQOL Outcomes

Tables 4 and 5 show the estimated direct and indirect effects in explaining the ethnic disparity in *anxiety* and *depression*, respectively, using the generalized linear models and nonparametric models. The confidence intervals were calculated based on 1000 bootstrap samples. The relative effect (RE) is defined as the ratio of the indirect or direct effect over the total effect. In the tables, joint effect is the joint effect of *income, employment*, and *insurance*. For both outcomes, more ethnic disparities are explained by nonlinear models since they account for potential nonlinear relationships. To explain the results, we focus on

using the nonlinear models. Figure 2 shows the relative effects for the outcomes from the nonlinear models. Although “days to diagnosis” was selected as a potential mediator/confounder for depression score, it does not have a significant indirect effect in explaining the ethnic difference. Overall, the significant mediators/confounders are the same for both depression and anxiety scores and the order of mediation/confounding effects are roughly the same. In explaining the mediation/confounding effects, we focus on the *anxiety* outcome. Results for the *depression* outcome can be similarly deduced.

Compared with Hispanic whites, non-Hispanic whites have an average lower anxiety PROMIS score (TE = −3.08). If the “US nativity” could be set equivalent among different eth nic groups, the ethnic disparity in *anxiety* would reduce by 22.5%. Other variables such as *social support* (21.6%), *education* (18.5%), *age at diagnosis* (16.7%), and the joint effect of *income, insurance*, and *employment* (5.4%) also significantly explain the ethnic difference. An interesting variable is *spirituality*, which have a positive indirect effect (opposite to the total effect) and a significant negative relative effect (− 19.7%). This is observed since the anxiety score decreases with spirituality support, which was on average higher among Hispanics compared with non-Hispanic whites. Spirituality support is a protective factor. If the spirituality support is kept equivalent between non-Hispanic white and Hispanic whites, the ethnic disparity would increase 19.7%.

The *mma* R package provides visual aids to understand the relationships among variables. For example, Fig. 3 describes how *age* mediated the ethnic disparity in *anxiety*. The top plot shows the relationship between age and anxiety score. The line was fitted using the Multiple Additive Regression Trees (MART). We can see that the average anxiety was the highest when the patients are diagnosed with cancer at an age younger than forty. After that, the anxiety score decreased sharply with age, until about the age 78, at which the decreasing rate became level. The lower two plots show the age distribution among non-Hispanic (middle plot) and Hispanic (lower plot) population. We found that non-Hispanic whites were diagnosed with cancer at a relatively older age than were Hispanic whites. Therefore, non-Hispanic whites have an average lower anxiety score. If the age distributions could be set as equivalent across the ethnic groups, the ethnic disparity would decrease by 16.7%. Similarly, Fig. 4 shows how US nativity helps explain the ethnic disparity in anxiety score. In conclusion, US-born patients have an average lower anxiety score than foreign-born patients, and there are more foreign-born Hispanic whites (58.25%) than non-Hispanic whites (5.83%). Therefore, the variable US born explained part of the ethnic disparity. The supplementary material provides the graphs for mediators/confounders to explain how each of them explains the ethnic disparity in anxiety score. Overall, having social support, being employed, male, having private insurance, no chemotherapy, high income, high education, old age, US born, and few comorbidity are related to low anxiety. And in general, there are higher proportion of non-Hispanic whites with such properties.

After accounting for all the mediators/confounders, there is still 27% of the ethnic disparity in anxiety score that cannot be explained (the direct effect). This could be due to variables that can explain the difference was not all included in this study, such as the environmental factors.

5. Conclusions and Future Research

Mediation analysis is used most often in health sciences and psychological research, with different behavioral and socio-demographic variables used as the mediators/confounders. Its appli cation in a study such as this one is to identify which factors could be important in explaining the relationship between a predictor and an outcome from a large pool of candidate mediators/confounders. Both the MY-Health study in general and our assessment using the data aim at taking a broader approach to study cancer survivors' experiences in terms of HRQOL. MY-Health did not limit their study by cancer site, instead included seven different cancers (female breast, uterine and cervical cancer, prostate cancer, and male and female colorectal cancer, non-Hodgkin's lymphoma, and nonsmall cell lung cancer), at three different age stratification groups (21–49, 50–64, 65–84). A stratified random sampling scheme was adapted in order to ensure a racially and ethnically diverse survivor population (Hispanic, non-Hispanic white, black, and Asian/Pacific Islander). The results of this study have allowed us to shed light on the factors that mediate the ethnic disparities in two very serious mental symptoms, anxiety and depression. The number of cancer survivors in the USA is growing rapidly as treatments for cancers improve. There will be a growing number of survivors, and our study helps show how we can care for their psychological needs along ethnic lines. Clinicians should focus on screening those survivors with less education, who were diagnosed at a younger age, with no insurance or having only government insurance, with multiple comorbidities, less socially supported, and having lower levels of spirituality. Researchers should focus more on the modifiable factors as they develop interventions and long-term care strategies. If an intervention is chosen to improve the HRQOL of cancer patients, the proposed method can also be used to evaluate the intervention. The two comparison populations will then be the groups with and without intervention separately. The effect of the intervention is measured by the total effect between the two populations.

However, there remained disparities in the anxiety and depression scores after accounting for all the mediators/confounders we collected. It is likely due to that we unknowingly omitted some important variables or the variables were not collected for the MY-Health study. This is considered to be the most common cause of specification error by Judd and Kenny, and they note that this problem is often not easy to fix (Judd & Kenny, 1981). There are many variables in the MY-Health data set that we omitted since either we did not consider them relevant when designing our analysis (e.g., type of radiation treatment administered, SEER region, or the different MY-Health exclusive scores such as financial well-being), or we found through exploratory analysis that did not seem to be relevant (survey mode: by mail or phone). It is possible any one of these variables may actually be a crucial mediator/confounder, and needs further exploration. Moreover, we believe geographic variables about build (e.g., walkability) and social environment (e.g., census tract poverty level) may have an influence on anxiety and depression, but were not collected for this study. Our next step is to build the environmental risk factor data set and link it with the MY-Health study. We will develop a multilevel mediation analysis to account for mediators/confounders from different levels and sources.

Acknowledgments

Funding This study was partially funded by the NIMHD award # 1R15MD012387 and by the Louisiana State University Health Sciences Center Pilot Fund.

References

Albert JM (2008). Mediation analysis via potential outcomes models. Statistics in Medicine, 27(8), 1282–1304. [PubMed: 17691077]

Alwin DF, & Hauser RM (1975). The decomposition of effects in path analysis. American Sociological Review,40(1), 37.

Avin C, Shpitser I, & Pearl J (2005). Identifiability of path-specific effects In van Zaanen M, Oates T, Paliouras G, & de la Higuera C (Eds.), Proceedings of the international joint conference on artificial intelligence IJCAI-05. Edinburgh: Morgan-Kaufmann Publishers.

Baron R, & Kenny D (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51(6), 1173–1182. [PubMed: 3806354]

Breiman L, Friedman J, Olshen RA, & Stone C (1984). Classification and regression trees. Pacific Grove: Wadsworth.

Culver JL, Arena PL, Antoni MH, & Carver CS (2002). Coping and distress among women under treatment for early stage breast cancer: Comparing African Americans, Hispanics and non-Hispanic Whites. Psycho-Oncology, 11(6), 495–504. [PubMed: 12476431]

Fan Y (2012). Multiple mediation analysis with general predictive models, Dissertation. New Orleans, LA: Louisiana State University Health Sciences Center.

Friedman J (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29, 1189–1232.

Gallicchio L, Calhoun C, & Helzlsouer KJ (2013). Association between race and physical functioning limitations among breast cancer survivors. Supportive Care in Cancer, 22(4), 1081–1088. [PubMed: 24292017]

Gu C (2013). Smoothing spline ANOVA models. New York: Springer.

Have TRT, Joffe MM, Lynch KG, Brown GK, Maisto SA, & Beck AT (2007). Causal mediation analyses with rank preserving models. Biometrics, 63(3), 926–934. [PubMed: 17825022]

Jensen RE, Moinpour CM, Keegan THM, Cress RD, Wu X-C, & Paddock L (2017). The measuring your health study: Leveraging community-based cancer registry recruitment to establish a large, diverse cohort of cancer survivors for analyses of measurement equivalence and validity of the patient reported outcomes measurement information system (promis) short form items. Psychological Test and Assessment Modeling, 58(1), 99–117.

Jensen RE, Moinpour CM, Potosky AL, Lobo T, Hahn EA, Hays RD, et al. (2016). Responsiveness of 8 patient-reported outcomes measurement information system (PROMIS) measures in a large, community-based cancer study cohort. Cancer, 123(2), 327–335. [PubMed: 27696377]

Jensen RE, Potosky AL, Reeve BB, Hahn E, Cella D, Fries J, et al. (2015). Validation of the PROMIS physical function measures in a diverse US population-based cohort of cancer patients. Quality of Life Research, 24(10), 2333–2344. [PubMed: 25935353]

Judd CM, & Kenny DA (1981). Process analysis: Estimating mediation in treatment evaluations. Evaluation Review,5(5), 602–619.

MacKinnon DP (2000). Equivalence of the mediation, confounding and suppression effect. Prevention Science, 1(4),173–181. [PubMed: 11523746]

Mackinnon DP (2008). Introduction to statistical mediation analysis. New York: Lawrence Erlbaum Associates.

Mackinnon DP, & Dwyer JH (1993). Estimating mediated effects in prevention studies. Evaluation Review, 17(2),144–158.

NIH. Calibration studies testing. http://www.nihpromis.org/science/calibrationtesting. Accessed 09 Feb 2017.

NIH. Promis domain frameworks. http://www.nihpromis.org/measures/domainframework. Accessed 09 Feb 2017.

NIH. Promis methodology. http://www.nihpromis.org/science/methodology. Accessed 09 Feb 2017.

Patrick D, & Erickson P (1993). Health policy, quality of life: Health care evaluation and resource allocation. NewYork, NY: Oxford University Press.

Pearl J (2001). Direct and indirect effects In Proceedings of the seventeenth conference on uncertainty and artificial intelligence. Morgan Kaufmann.

Reeve BB, Potosky AL, Smith AW, Han PK, Hays RD, Davis WW, et al. (2009). Impact of cancer on health-related quality of life of older americans. JNCI Journal of the National Cancer Institute, 101(12), 860–868. [PubMed: 19509357]

Reeve BB, Stover AM, Jensen RE, Chen RC, Taylor KL, Clauser SB, et al. (2012). Impact of diagnosis and treatment of clinically localized prostate cancer on health-related quality of life for older Americans. Cancer, 118(22), 5679–5687. [PubMed: 22544633]

Reeve BB, & Terest JA (2016). Overview to the two-part series: Measurement equivalence of the patient reported outcomes measurement information system (promis) short forms. Psychological Test and Assessment Modeling, 58(1), 31–35.

Reeve BB, Thissen D, DeWalt DA, Huang I-C, Liu Y, Magnus B, et al. (2015). Linkage between the PROMIS® pediatric and adult emotional distress measures. Quality of Life Research, 25(4), 823–833. [PubMed: 26424169]

Robins JM, & Greenland S (1992). Identifiability and exchangeability for direct and indirect effects. Epidemiology,3(2), 143–155. [PubMed: 1576220]

Rodenberg CA (2006). A review of: Assessing quality of life in clinical trials: Theory and methods, second edition, by Fayers P and Hays R (eds.). Journal of Biopharmaceutical Statistics, 16(5), 761–763.

Teresi JA, Ocepek-Welikson K, Kleinman M, Ramirez M, & Kim G (2016a). Measurement equivalence of the patient reported outcomes measurement information system (promis) anxiety short forms in ethnically diverse groups. Psychological Test and Assessment Modeling, 58(1), 183–219. [PubMed: 28649483]

Teresi JA, Ocepek-Welikson K, Kleinman M, Ramirez M, & Kim G (2016b). Psychometric properties and performance of the patient reported outcomes measurement information system (promis) depression short forms in ethnically diverse groups. Psychological Test and Assessment Modeling, 58(1), 141–181. [PubMed: 28553573]

VanderWeele TJ (2009). Marginal structural models for the estimation of direct and indirect effects. Epidemiology,20(1), 18–26. [PubMed: 19234398]

VanderWeele TJ, & Robinson WR (2014). On the causal interpretation of race in regressions adjusting for confounding and mediating variables. Epidemiology, 25(4), 473–484. [PubMed: 24887159]

Vanderweele TJ, & Vansteelandt S (2009). Conceptual issues concerning mediation, interventions and composition.Statistics and Its Interface, 2(4), 457–468.

VanderWeele T, Vansteelandt S, & Daniel RM (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology, 25(2), 300–306. [PubMed: 24487213]

Vansteelandt S, & Daniel RM (2017). Interventional effects for mediation analysis with multiple mediators. Epidemiology, 28(2), 258–265. [PubMed: 27922534]

Yost K, Eton D, Garcia S, & Cella D (2011). Minimally important differences were estimated for six promis cancer scales in advanced-stage cancer patients. Journal of Clinical Epidemiology, 64(5), 507–516. [PubMed: 21447427]

Yu Q, Fan Y, & Wu X (2014). General multiple mediation analysis with an application to explore racial disparities in breast cancer survival. Journal of Biometrics & Biostatistics, 5(189), 189.

Yu Q, & Li B (2017). mma: An r package for multiple mediation analysis. Journal of Open Research Software, 5, 11.

Yu Q, Li B, & Scribner RA (2009). Hierarchical additive modeling of nonlinear association with spatial correlations—An application to relate alcohol outlet density and neighborhood assault rates. Statistics in Medicine, 28(14), 1896–1912. [PubMed: 19402025]

Yu Q, Scribner R, Leonardi C, Zhang L, Park C, Chen L, et al. (2017). Exploring racial disparity in obesity: A mediation analysis considering geo-coded environmental factors. Spatial and Spatio-temporal Epidemiology, 21, 13–23. [PubMed: 28552184]

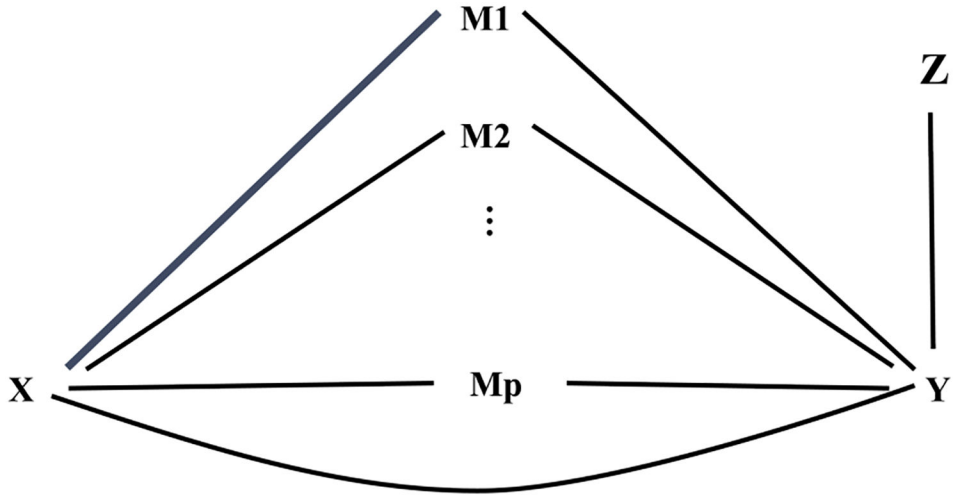

Figure 1.
Graphical model of mediation/confounding effects. Note that there can be lines that connect between mediators, M, and Z.

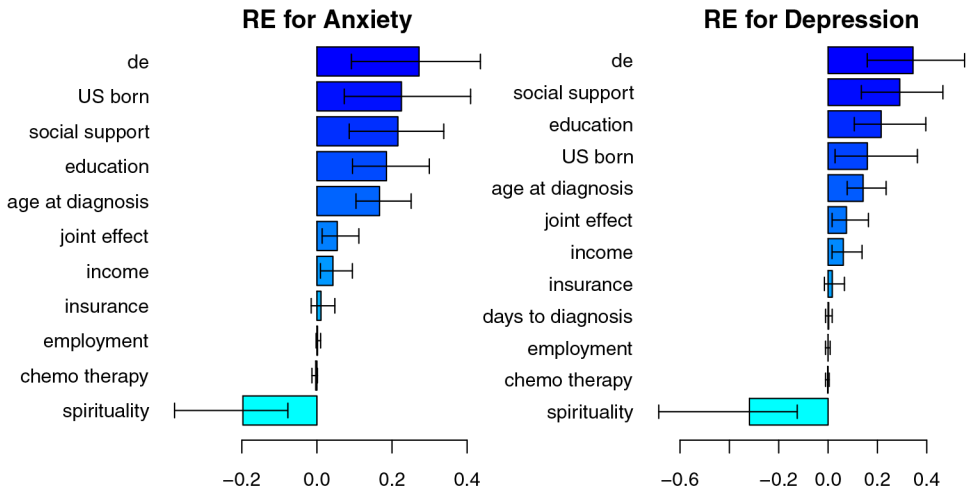

Figure 2.

Estimated relative effects (RE) on the ethnic disparities in anxiety (left) and depression (right) by nonlinear models. Note that “de” refers to the direct effect, the ethnic disparities cannot be explained by mediators/confounders, and “joint effect” is the joint indirect effect from employment, insurance, and income.

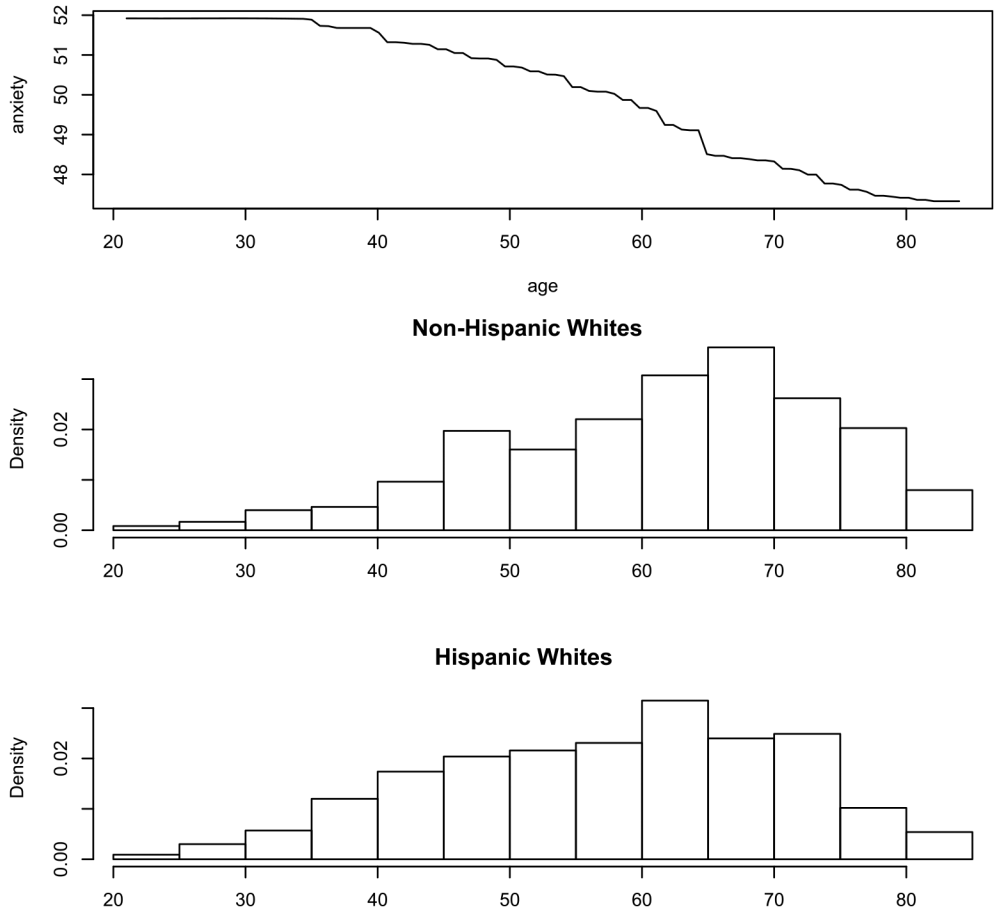

Figure 3.
Indirect effect of *age* (of diagnosis with cancer) on *anxiety*.

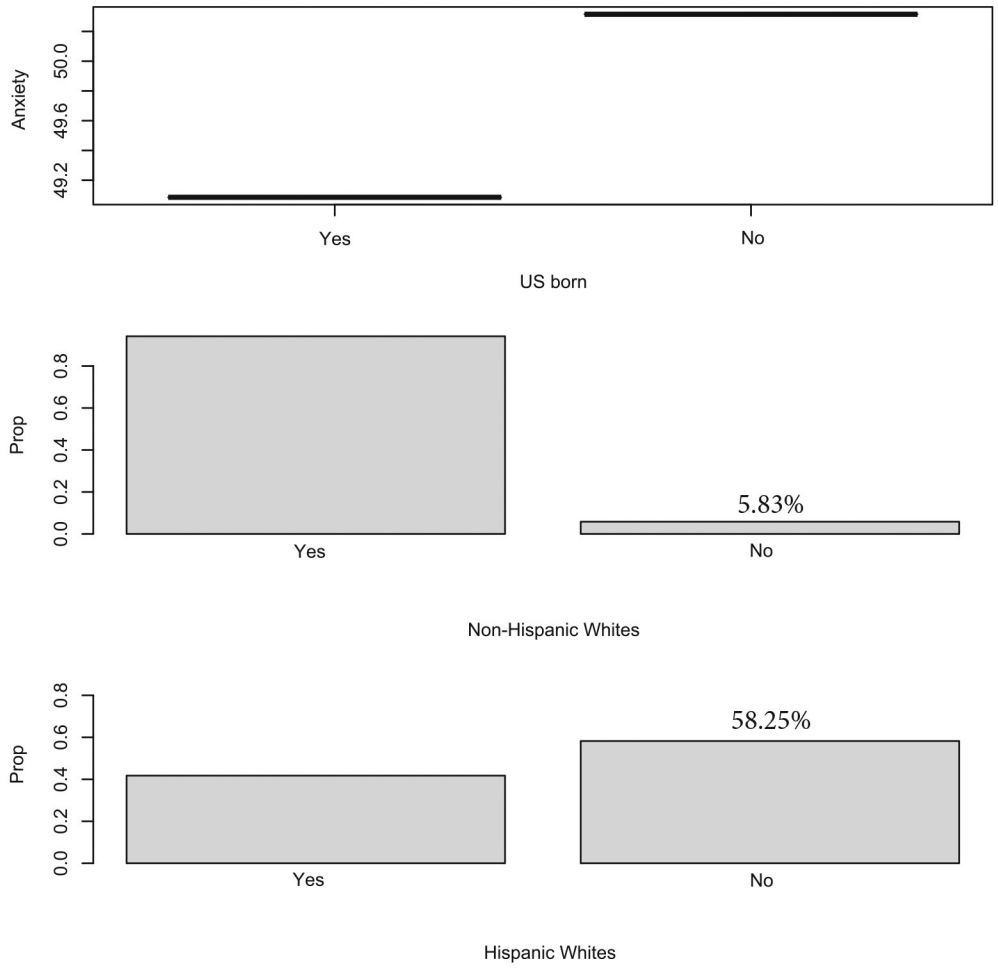

Figure 4.
Indirect effect of *US born* (whether the patient was born in USA or not) on *anxiety*.

Table 1.

ANOVA to compare anxiety and depression PROMIS scores by ethnicity.

Score	HW mean (SD)	NHW mean (SD)	*p* value
Anxiety	52.05 (11.77)	48.47 (10.57)	1.548e-13
Depression	50.74(11.54)	47.84 (10.45)	1.316e-09

Table 2.

Variables, formats, and data sources.

Variable groups	Variable (formats)	Data sources
Outcome	Anxiety score (continuous) Depression score (continuous)	MY-Health Survey
Individual Information	Ethnicity (HW; NHW) Married (no; yes) Employment (no; yes) Sex (M; F) Kids live at home (no; yes) US born (no; yes) Insurance (no; public; private) Education (continuous) Income (continuous) Social support (continuous) Spirituality (continuous) Comorbidities (continuous) Age at diagnosis (continuous) Days of diagnosis (continuous)	Medical record Cancer registries MY-Health Survey
Tumor Characteristics	AJCC stage (I, II, III, IV) Primary site (categorical) Tumor grade (categorical)	Cancer Registries
Treatment Information	Chemotherapy (no; yes) Radiation (no; yes) Surgery (no; yes) Hormonal therapy (no; yes)	Cancer Registries

Table 3.

Potential mediators/confounders and covariates.

Variables	Anxiety	Depression	HW (667)	NHW (2160)	*p* value
Married (yes)	0.442	0.386	56.73%	65.29%	0.000
Employment (working)	0.549 *	0.748 *	43.41%	46.18%	0.953
Sex (male)	0.007 -	0.016 -	39.58%	42.31%	0.227
Kids live at home (yes)	0.315	0.496	26.84%	14.35%	0.000
US born (yes)	0.000 *	0.008 *	41.75%	94.17%	0.000
Insurance (private)	0.321 *	0.028 *	55.77%	78.26%	0.000
Education (≤ high school)	0.002 *	0.000 *	37.17%	8.12%	0.000
Income < 10,000	0.088 *	0.098 *	16.70%	5.66%	0.000
AJCC stage (IV)	0.680	0.423	10.79%	13.33%	0.134
Primary site (lung)	0.231	0.316	5.85%	20.46%	0.000
Tumor grade (I)	0.019 -	0.116	13.19%	12.96%	0.969
Chemotherapy (yes)	0.021 *	0.009 *	50.92%	46.73%	0.067
Radiation (yes)	0.293	0.002 -	41.04%	42.00%	0.696
Surgery (yes)	0.410	0.903	26.77%	31.02%	0.041
Hormonal therapy (yes)	0.508	0.004 -	22.27%	21.67%	0.791
Social support	0.000 *	0.000 *	49.72(11.01)	51.20(10.85)	0.003
Spirituality	0.000 *	0.000 *	37.87(8.80)	35.99(9.80)	0.000
Comorbidities	0.000 -	0.000 -	1.97(1.84)	2.10(1.80)	0.118
Age at diagnosis	0.000 *	0.000 *	59(13.10)	64(12.43)	0.000
Days of diagnosis	0.397	0.087 *	288.8(48.16)	294.5(50.59)	0.010

Columns 1 and 2 show the *p* values of Type III tests for the corresponding variables in the full model in predicting anxiety and depression scores, respectively. Columns 3 and 4 show the summary statistics (proportion of a representative category for categorical variable, mean (standard deviation) for continuous variable) for Hispanic whites (HW) and non-Hispanic whites (NHW), respectively. Column 5 shows the *p* value for testing the association between ethnicity and the corresponding variable.

* The variable is identified as a potential mediator/confounder.

– A covariate, for the column variable.

Table 4.

Summary of mediation/confounding effect estimations for ethnic disparity in *anxiety* PROMIS score.

Mediator/confounder	Linear models		Nonparametric models	
	IE (95% CI)	RE(%)	IE (95% CI)	RE (%)
Income	−0.19 (−0.60,0.23)	5.4 (−7.1, 17.9)	−0.13 (−0.25, −0.21)	4.3 (0.0, 8.6)
Education	−0.30 (−0.55, −0.06)	9.1 (1.3, 17.0)	−0.56 (−0.83, −0.39)	18.5 (8.4, 28.7)
Age at diagnosis	−0.48 (−0.70,−0.26)	14.4 (6.6, 22.2)	−0.50 (−0.70, −0.31)	16.7 (9.2, 24.2)
Social support	−0.59 (−1.07, −0.11)	17.1 (4.5, 29.8)	−0.67 (−1.14, −0.20)	21.6 (8.6, 34.5)
Spirituality	0.62 (0.25, 0.98)	−19.2 (−35.6, −2.8)	0.57 (0.28, 0.87)	19.7 (−34.8, −4.6)
Employment	0.03 (−0.10, 0.16)	−0.9 (−5,3.1)	−0.005 (−0.02, 0.01)	0.2 (−0.4, 0.8)
US born	−1.01 (−1.60, −0.43)	90.7(9.8,51.5)	−0.68 (−1.14, −0.21)	22.5 (5.8, 39.3)
Chemotherapy	−0.05 (−0.15, 0.05)	1.6 (−1.6, 4.8)	0.010 (−0.01,0.03)	0.3 (−1.1, 0.4)
Insurance	0.02 (−0.20, 0.23)	−0.6 (−7.2, 6.1)	−0.032 (−0.12, 0.06)	1.1 (−2.0, 4.1)
Joint effect	−0.13 (−0.60, 0.34)	3.6 (−10.8, 18.0)	−0.16 (−0.31, −0.02)	5.4 (1, 10.7)
Direct effect	−1.38 (−2.42, −0.35)	40.3 (14.1,66.5)	−0.84 (−1.47, −0.21)	27.2 (9.4. 45.0)
Total effect	−3.40 (−4.55, −2.26)		−3.08 (−4.08, −2.02)	

IE indirect effect, *RE* relative effect. “joint effect” refers to the joint indirect effect from employment, insurance, and income. Non-Hispanic whites are the reference group.

Table 5.

Summary of mediation/confounding effect estimations for ethnic disparity in *depression* PROMIS score.

Mediator/confounder	Linear models		Nonparametric models	
	IE (95% CI)	RE (%)	IE (95% CI)	RE (%)
Days from diagnosis	0.01 (−0.02, 0.04)	−0.4 (−1.7. 1.0)	−0.01 (−0.04,0.02)	0.2 (−1.0, 1.7)
Income	−0.24 (−0.6, 0.13)	9.1 (−5.5, 23.7)	−0.15 (−0.29, −0.04)	6.2 (1.6, 13.8)
Education	−0.34 (−0.58, −0.10)	13.6(1.7, 25.5)	−0.51 (−0.77, −0.27)	21.5 (10.6, 39.6)
Age at diagnosis	−0.35 (−0.53,−0.18)	14.1 (4.1, 24.2)	−0.34 (−0.50, −0.19)	16.7 (9.2, 24.2)
Social support	−0.52 (−1.02, −0.03)	19.5 (1.4, 37.6)	−0.71 (−1.12, −0.24)	29.1 (13.5, 46.6)
Spirituality	0.74 (0.31, 1.17)	−31.4 (−68, −8.7)	0.72 (0.39, 1.08)	−31.9 (−68.6, −12.5)
Employment	0.04 (−0.09, 0.17)	−1.5 (−7.1, 4.1)	0.00 (−0.02, 0.02)	0.0 (−1.0, 0.9)
US born	−.62 (−1.16, −0.09)	25.3 (−3.3, 54)	−0.38 (−0.76, −0.07)	15.9 (2.8, 36.2)
Chemotherapy	−0.0 (−0.10, 0.06)	0.9 (−2.3, 4.1)	0.00 (−0.01,0.03)	−0.2 (−1.0, 0.4)
Insurance	0.01 (−0.25, 0.23)	0.3 (−9.4, 10.0)	−0.04 (−0.15,0.03)	1.7 (−1.5, 6.6)
Joint effect	−0.21 (−0.66, 0.24)	7.7 (−9.9, 25.3)	−0.18 (−0.34, −0.04)	7.5 (1.6, 16.4)
Direct effect	−1.33 (−2.32, −0.33)	50.2(16.6, 83.9)	−0.84 (−1.46, −0.28)	34.4 (15.9, 55.3)
Total effect	−1.33 (−1.44, −3.76)		−2.45 (−3.40, −1.42)	

IE indirect effect, *RE* relative effect. “joint effect” refers to the joint indirect effect from employment, insurance, and income. Non-Hispanic whites are the reference group.