Supplemental Material

March 24, 2018

1 Data Cleaning Process for the MY-Health Data

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.

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* records 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 was evaluated. Variables like "age to US" and "years lived in US," which do not apply to patients born in the United States, 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 United States, and exclude "age to US" and "years lived in US" for further analysis. Based on literature reviews and data availability, we included all variables listed in Table 1 as potential mediators. 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 variables, their variable formats and data sources are listed in the table.

Table 1: 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 statge (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

2 Exposure-Mediator Interactions

We discuss the situation when there exists a exposure-mediator interaction effect on the outcome. The underlying models for the associations among the variables are assumed to be:

$$\begin{aligned} logit(Pr(M=1)) &= a_{01} + a_1 X \\ Y &= b_0 + b_1 M + b_2 XM + cX \end{aligned}$$

Lemma With the above assumptions, for binary X, the average indirect effect through M is $(b_1 + 0.5b_2)\left(\frac{e^{a_{01}+a_1}}{1+e^{a_{01}+a_1}} - \frac{e^{a_{01}}}{1+e^{a_{01}}}\right)$. For continuous X, the average indirect effect through M is $E_X[(b_1 + b_2 x)a_1 Pr(M=1|X=x)(1-Pr(M=1|X=x))]$, **?**.

If X has no effect on M, ie., $a_1 = 0$, the average indirect effect through M is 0 as expected.

3 Summary of Mediation Effect Estimations for *depression* PROMIS Score

Table 2: Summary of Mediation Effect Estimations for *depression* PROMIS Score.

Mediator	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)	

Note: "joint effect" refers to the joint indirect effect from employment, insurance and income.

4 R Code for Data Analysis

```
library(mma)
#anxiety
x2<-data.frame(employment=as.factor(mh1$Employment),
               married=as.factor(mh1$Married), sex=as.factor(mh1$Sex),
 kids=as.factor(mh1$kids_lt_18),
               usborn=as.factor(mh1$born_in_US), insurance=as.factor(mh1$insurance),
 stage=as.factor(mh1$DAJCC_Stage),chemo=as.factor(mh1$Survey_chemo),
```

```
               rad=as.factor(mh1$Survey_rad), horm=as.factor(mh1$Survey_horm),
               surg=as.factor(mh1$surgery), grade=as.factor(mh1$grade),
 site=as.factor(mh1$Cancer_site),days=mh1$dx_to_comp_days,
               income=mh1$income, edu=mh1$edu, age=mh1$Age_at_diagnosis,
               socialsupport=mh1$social_promis_score_AS, cmrb=mh1$cmrb,
 spiritual=mh1$spirituality_score)
y2<-mh1$anxiety_promis_score_AS

#data.org
data2<-data.org(x2, y2, pred=mh1$Ethnicity, predref=1,mediator=1:20, alpha = 0.1,
                alpha2 = 0.1,jointm = list(n=1,j1=c("employment","insurance","income")))
summary(data2)

med2.linear<-med(data=data2,n=100,seed=1)
med2.mart<-med(data=data2,n=100,seed=1,nonlinear=T,df=4)
med2.linear
med2.mart

#depression
x3<-data.frame(employment=as.factor(mh1$Employment),
               married=as.factor(mh1$Married), sex=as.factor(mh1$Sex),
 kids=as.factor(mh1$kids_lt_18), usborn=as.factor(mh1$born_in_US),
 insurance=as.factor(mh1$insurance), stage=as.factor(mh1$DAJCC_Stage),
               chemo=as.factor(mh1$Survey_chemo), rad=as.factor(mh1$Survey_rad),
 horm=as.factor(mh1$Survey_horm), surg=as.factor(mh1$surgery),
 grade=as.factor(mh1$grade), site=as.factor(mh1$Cancer_site),
               days=mh1$dx_to_comp_days, income=mh1$income, edu=mh1$edu,
 age=mh1$Age_at_diagnosis,socialsupport=mh1$social_promis_score_AS,
 cmrb=mh1$cmrb, spiritual=mh1$spirituality_score)
y3<-mh1$depression_promis_score_AS

#data.org
data3<-data.org(x3, y3, pred=mh1$Ethnicity, predref=1,mediator=1:20, alpha = 0.1,
                alpha2 = 0.1,jointm = list(n=1,j1=c("employment","insurance",
"income")))
summary(data3)

med3.linear<-med(data=data3,n=100,seed=1)
med3.mart<-med(data=data3,n=100,seed=1,nonlinear=T,df=4)
med3.linear
med3.mart

#bootstrap
#for anxiety
mma2.boot.linear<-mma(x2, y2, pred=mh1$Ethnicity, predref=1,mediator=1:20, alpha = 0.1,
                      alpha2 = 0.1,jointm = list(n=1,j1=c("employment","insurance",
```

```
"income")),n=20,seed=1,n2=1000)

pdf("boot2_glm_mma.pdf")
summary(mma2.boot.linear)
mnames2<-names(mma2.boot.linear$data$x[,-mma2.boot.linear$data$dirx])
for (i in mnames2)
  plot(mma2.boot.linear,vari=i)
dev.off()

mma2.boot.mart<-mma(x2, y2, pred=mh1$Ethnicity, predref=1,mediator=1:20, alpha = 0.1,
                    alpha2 = 0.1,jointm = list(n=1,j1=c("employment","insurance",
"income")),n=20,seed=1,n2=1000, nonlinear=T,df=4)

boot2.linear.summary<-summary(mma2.boot.linear)
boot2.mart.summary<-summary(mma2.boot.mart)

pdf("boot2_mart_mma.pdf")
summary(mma2.boot.mart)
mnames2<-names(mma2.boot.mart$data$x[,-mma2.boot.mart$data$dirx])
for (i in mnames2)
  plot(mma2.boot.mart,vari=i)
dev.off()

#for depression
mma3.boot.linear<-mma(x3, y3, pred=mh1$Ethnicity, predref=1,mediator=1:20, alpha = 0.1,
                      alpha2 = 0.1,jointm = list(n=1,j1=c("employment","insurance",
"income")),n=20,seed=1,n2=1000)
pdf("boot3_glm_mma.pdf")
summary(mma3.boot.linear)
mnames3<-names(mma3.boot.linear$data$x[,-mma3.boot.linear$data$dirx])
for (i in mnames3)
  plot(mma3.boot.linear,vari=i)
dev.off()

mma3.boot.mart<-mma(x3, y3, pred="eth", predref=1,mediator=2:21, alpha = 0.1,
                    alpha2 = 0.1,n2=1000, nonlinear=T,df=4,n=20,seed=1,
                    jointm = list(n=1,j1=c("employment","insurance","income")))

boot3.linear.summary<-summary(mma3.boot.linear)
boot3.mart.summary<-summary(mma3.boot.mart)

pdf("boot3_mart_mma.pdf")
summary(mma3.boot.mart)
mnames3<-names(mma3.boot.mart$data$x[,-mma3.boot.mart$data$dirx])
for (i in mnames3)
  plot(mma3.boot.mart,vari=i)
```

```
dev.off()
```

5 Graphs to Explore Ethnic disparity in Anxiety Score