**Program 2.4** Evaluating Balance Produced by Propensity Score

/\* This section of code evaluates the balance produced by the

propensity score by 1) summarizing the distribution of the propensity

scores via box plots, 2) running two-way models to compare the

balance of covariates before and after adjustment, and 3) computing

standardized treatment differences for each covariate before and

after adjustment. \*/

/\*1. assessing balance between covariates by treatment and quintiles

of propensity scores by box plots\*/

PROC FORMAT;

VALUE BPF **1** = 'Q1-A'

**2** = 'Q1-B'

**3** = 'Q2-A'

**4** = 'Q2-B'

**5** = 'Q3-A'

**6** = 'Q3-B'

**7** = 'Q4-A'

**8** = 'Q4-B'

**9** = 'Q5-A'

**10** = 'Q5-B';

RUN;

DATA ADOS4;

SET ADOS3;

LABEL BP='QUINTILE-TREATMENT';

FORMAT BP BPF.;

IF TX=**1** AND QUINTILES\_PS=**0** THEN BP=**1**;

ELSE IF TX=**0** AND QUINTILES\_PS=**0** THEN BP=**2**;

ELSE IF TX=**1** AND QUINTILES\_PS=**1** THEN BP=**3**;

ELSE IF TX=**0** AND QUINTILES\_PS=**1** THEN BP=**4**;

ELSE IF TX=**1** AND QUINTILES\_PS=**2** THEN BP=**5**;

ELSE IF TX=**0** AND QUINTILES\_PS=**2** THEN BP=**6**;

ELSE IF TX=**1** AND QUINTILES\_PS=**3** THEN BP=**7**;

ELSE IF TX=**0** AND QUINTILES\_PS=**3** THEN BP=**8**;

ELSE IF TX=**1** AND QUINTILES\_PS=**4** THEN BP=**9**;

ELSE IF TX=**0** AND QUINTILES\_PS=**4** THEN BP=**10**;

RUN;

PROC SORT DATA=ADOS4;

BY BP;

RUN;

TITLE 'Distribution of propensity scores by quintiles and treatment';

PROC BOXPLOT DATA=ADOS4;

PLOT PS\*BP;

RUN;

TITLE 'Distribution of Baseline PHQ1 by quintiles and treatment';

PROC BOXPLOT DATA=ADOS4;

PLOT PHQ1\*BP;

RUN;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* MACRO GEN1 assesses balance produced by a propensity \*;

\* stratification adjustment via a two-way model approach \*;

\* (Rosenbaum and Rubin, 1984). A data set with the test \*;

\* statistics and *p*-values for the treatment effect and the\*;

\* treatment by ps strata is produced. \*;

\* INPUT VARIABLES: \*;

\* VAR - covariate to be evaluated \*;

\* DST - NOR for normal, BIN for binary variables \*;

\* LNK - ID for normal, LOGIT for binary variables \*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**%MACRO** GEN1(VAR,DST,LNK);

\* Run main effect and ps-adjusted models using GENMOD,

output parameter estimates to data sets for compilation\*;

PROC GENMOD DATA = ADOS3 DESCENDING;

CLASS TX;

MODEL &VAR = TX / DIST = &DST LINK = &LNK TYPE3;

ODS OUTPUT TYPE3 = TEST1;

TITLE2 'TESTING FOR COVARIATE BALANCE: WITHOUT PS';

TITLE3 "VAR: &VAR"; RUN;

PROC GENMOD DATA = ADOS3 DESCENDING;

CLASS TX QUINTILES\_PS;

MODEL &VAR = TX QUINTILES\_PS TX\*QUINTILES\_PS / DIST =

&DST LINK = &LNK TYPE3;

LSMEANS TX / DIFF;

ODS OUTPUT TYPE3 = TEST2;

ODS OUTPUT LSMEANS = TESTL1;

TITLE2 'TESTING FOR COVARIATE BALANCE: WITH PS';

TITLE3 "VAR: &VAR"; RUN;

DATA TEST1;

SET TEST1;

OVAR = "&VAR";

DUM = **1**;

PVAL\_TX\_UNADJ = PROBCHISQ;

TSTAT\_TX\_UNADJ = CHISQ;

TSTATDF\_TX\_UNADJ = DF;

KEEP DUM OVAR TSTAT\_TX\_UNADJ TSTATDF\_TX\_UNADJ PVAL\_TX\_UNADJ;

DATA TEST2A;

SET TEST2;

IF SOURCE = 'tx';

OVAR = "&VAR";

DUM=**1**;

PVAL\_TX\_ADJ = PROBCHISQ;

TSTAT\_TX\_ADJ = CHISQ;

TSTATDF\_TX\_ADJ = DF;

KEEP DUM OVAR TSTAT\_TX\_ADJ TSTATDF\_TX\_ADJ PVAL\_TX\_ADJ;

DATA TEST2B;

SET TEST2;

IF SOURCE = 'tx\*QUINTILES\_PS';

OVAR = "&VAR";

DUM=**1**;

PVAL\_TXPS\_ADJ = PROBCHISQ;

TSTAT\_TXPS\_ADJ = CHISQ;

TSTATDF\_TXPS\_ADJ = DF;

KEEP DUM OVAR TSTAT\_TXPS\_ADJ TSTATDF\_TXPS\_ADJ PVAL\_TXPS\_ADJ;

PROC SORT DATA = TEST1; BY DUM; RUN;

PROC SORT DATA = TEST2A; BY DUM; RUN;

PROC SORT DATA = TEST2B; BY DUM; RUN;

DATA BPP\_&VAR;

MERGE TEST1 TEST2A TEST2B;

BY DUM;

%MEND GEN1;

\* Call GEN1 macro to assess balance for each covariate and summarize

output in single data set\*;

ODS LISTING CLOSE;

%***GEN1***(GENDER, BIN, LOGIT); RUN;

%***GEN1***(SPOUSE, BIN, LOGIT); RUN;

%***GEN1***(WORK, BIN, LOGIT); RUN;

%***GEN1***(AGE, NOR, ID); RUN;

%***GEN1***(PHQ1, NOR, ID); RUN;

ODS LISTING;

DATA BPP\_ALL;

SET BPP\_GENDER BPP\_SPOUSE BPP\_WORK BPP\_AGE BPP\_PHQ1;

PROC PRINT DATA = BPP\_ALL;

VAR OVAR TSTAT\_TX\_UNADJ PVAL\_TX\_UNADJ TSTAT\_TX\_ADJ

PVAL\_TX\_ADJ PVAL\_TXPS\_ADJ;

TITLE 'PROPENSITY STRAT. BALANCE ASSESSMNT: 2-WAY MODELS';

TITLE2 'TEST STATISTICS (TSTAT) AND PVALUES (PVAL) FOR

MODELS WITHOUT PROPENSITY';

TITLE3 'ADJUSTMENT (UANDJ) AND WITH PROPENSITY ADJUSTMENT

(ADJ)'; RUN;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* MACRO STRATA is called by MACRO GEN2 and computes the \*;

\* standardized differences for a given subgroup (quintile)\*;

\* of the data. \*;

\* Input Variables: \*;

\* DATAIN - analysis data set \*;

\* DATOUT - output data set containing standardized \*;

\* differences \*;

\* STRN - strata number \*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

%MACRO STRAT(DATIN,DATOUT,STRN);

DATA ONE;

SET &DATIN;

IF QUINTILES\_PS = &STRN;

DATA ONE\_A ONE\_B;

SET ONE;

IF TX = **1** THEN OUTPUT ONE\_A;

IF TX = **0** THEN OUTPUT ONE\_B;

DATA ONE\_A;

SET ONE\_A;

MN\_A\_&STRN = MN;

SD\_A\_&STRN = SD;

NUM\_A\_&STRN = NUM;

DUMM = **1**;

KEEP MN\_A\_&STRN SD\_A\_&STRN NUM\_A\_&STRN DUMM;

DATA ONE\_B;

SET ONE\_B;

MN\_B\_&STRN = MN;

SD\_B\_&STRN = SD;

NUM\_B\_&STRN = NUM;

DUMM = **1**;

KEEP MN\_B\_&STRN SD\_B\_&STRN NUM\_B\_&STRN DUMM;

\* This step merges the summary stats for each treatment and

computes the pooled variances and then the standardized

difference. For binary data variances a percentage value

between .05 and .95 is used to avoid infinite values. \*;

DATA &DATOUT;

MERGE ONE\_A ONE\_B;

BY DUMM;

MN\_DIFF\_&STRN = MN\_A\_&STRN - MN\_B\_&STRN;

MN2\_A\_&STRN = MAX(MN\_A\_&STRN,**.05**); MN2\_A\_&STRN =

MIN(MN2\_A\_&STRN,**.95**);

MN2\_B\_&STRN = MAX(MN\_B\_&STRN,**.05**); MN2\_B\_&STRN =

MIN(MN2\_B\_&STRN,**.95**);

IF &BNRY = **0** THEN SD\_DIFF\_&STRN = SQRT( **0.5**\*(

SD\_A\_&STRN\*\***2** + SD\_B\_&STRN\*\***2** ));

IF &BNRY = **1** THEN SD\_DIFF\_&STRN = SQRT( (MN2\_A\_&STRN\*(**1**-

MN2\_A\_&STRN) + MN2\_B\_&STRN\*(**1**-MN2\_B\_&STRN)) / **2** );

STDDIFF\_&STRN = MN\_DIFF\_&STRN / SD\_DIFF\_&STRN;

%MEND STRAT;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* MACRO GEN2 computes the standardized differences for a \*;

\* given covariate within each propensity score strata \*;

\* (by calling the MACRO STRAT), unadjusted in the full \*;

\* sample (without propensity scoring), and averaging \*;

\* across the propensity score strata (adjusted) \*;

\* INPUT VARIABLES: \*;

\* VAR - covariate to be evaluated \*;

\* BNRY - enter 1 for binary covariate, 0 for continuous\*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

%MACRO GEN2(VAR,BNRY);

\* Generate summary statistics for entire sample using PROC SUMMARY

and then compute the standardized difference for the unadjusted

full sample \*;

PROC SUMMARY DATA = ADOS3;

CLASS TX;

VAR &VAR;

OUTPUT OUT=SSTAT MEAN=MN STD=SD N=NUM;

DATA SSTAT1;

SET SSTAT;

IF TX = **1**;

MEAN\_A = MN;

SD\_A = SD;

N\_A = NUM;

DUMM = **1**;

DATA SSTAT2;

SET SSTAT;

IF TX = **0**;

MEAN\_B = MN;

SD\_B = SD;

N\_B = NUM;

DUMM = **1**;

PROC SORT DATA = SSTAT1; BY DUMM; RUN;

PROC SORT DATA = SSTAT2; BY DUMM; RUN;

DATA SSTATF;

MERGE SSTAT1 SSTAT2;

BY DUMM;

MN\_DIFF = MEAN\_A - MEAN\_B;

SDP = SQRT( ( (SD\_A\*\***2**) + (SD\_B\*\***2**) ) / **2** );

MEAN\_A2 = MAX(MEAN\_A,**.05**); MEAN\_A2 = MIN(MEAN\_A2,**.95**);

MEAN\_B2 = MAX(MEAN\_B,**.05**); MEAN\_B2 = MIN(MEAN\_B2,**.95**);

IF &BNRY = **1** THEN SDP = SQRT( (MEAN\_A2\*(**1**-MEAN\_A2) +

MEAN\_B2\*(**1**-MEAN\_B2)) / **2** );

STDDIFF\_UNADJ = MN\_DIFF / SDP;

OVAR = "&VAR";

KEEP OVAR DUMM MN\_DIFF SDP STDDIFF\_UNADJ;

\* Generate summary statistics for each propensity strata

using PROC SUMMARY and then compute the standardized

difference for each strata using STRAT macro \*;

PROC SORT DATA = ADOS3; BY QUINTILES\_PS; RUN;

PROC SUMMARY DATA = ADOS3;

BY QUINTILES\_PS;

CLASS TX;

VAR &VAR;

OUTPUT OUT=PSSTAT MEAN=MN STD=SD N=NUM;

DATA PSSTAT;

SET PSSTAT;

IF TX = ' ' THEN DELETE;

%*STRAT*(PSSTAT,SD0,0); RUN;

%*STRAT*(PSSTAT,SD1,1); RUN;

%*STRAT*(PSSTAT,SD2,2); RUN;

%*STRAT*(PSSTAT,SD3,3); RUN;

%*STRAT*(PSSTAT,SD4,4); RUN;

DATA MRG;

MERGE SD0 SD1 SD2 SD3 SD4;

BY DUMM;

ADJ\_DIFF = (MN\_DIFF\_0 + MN\_DIFF\_1 + MN\_DIFF\_2 + MN\_DIFF\_3

+ MN\_DIFF\_4) / **5**;

\* Create final data set with standardized differences from

unadjusted, adjusted, and within each quintile approach.

The unadjusted SD is used here rather than a pooled

within SD across strata to provide a direct comparison

with the unadjusted standardized difference. \*;

DATA FINAL\_&VAR;

MERGE MRG SSTATF;

BY DUMM;

STDDIFF\_ADJ = ADJ\_DIFF / SDP;

KEEP OVAR STDDIFF\_UNADJ STDDIFF\_ADJ STDDIFF\_0 STDDIFF\_1

STDDIFF\_2 STDDIFF\_3 STDDIFF\_4 ;

%MEND GEN2;

\* Compute the standardized difference for each covariate by

running GEN2 macro and then compile results into a single data

set for summarizing. \*;

%*GEN2*(GENDER,1); RUN;

%*GEN2*(SPOUSE,1); RUN;

%*GEN2*(WORK,1); RUN;

%*GEN2*(AGE,0); RUN;

%*GEN2*(PHQ1,0); RUN;

DATA FINAL;

SET FINAL\_GENDER FINAL\_SPOUSE FINAL\_WORK FINAL\_AGE

FINAL\_PHQ1;

PROC PRINT DATA = FINAL;

VAR OVAR STDDIFF\_UNADJ STDDIFF\_ADJ STDDIFF\_0 STDDIFF\_1

STDDIFF\_2 STDDIFF\_3 STDDIFF\_4;

TITLE 'STANDARDIZED DIFFERENCES BEFORE PS ADJUSTMENT

(STAND\_DIFF\_UNADJ), AFTER PS ';

TITLE2 ' ADJUSTMENT AVERAGING ACROSS STRATA

(STAND\_DIFF\_ADJ), AND WITHIN EACH PS';

TITLE3 ' QUINTILE (STDDIFF\_0 ... STDIFF\_4)'; RUN;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\* MACRO GEN3 assesses the balance produced by a propensity\*;

\* scoring for a propensity score regression analysis. \*;

\* Weighted standardized differences (Austin, 2007) are \*;

\* produced for a given covariate. \*;

\* INPUT VARIABLES: \*;

\* DVAR - covariate to be evaluated \*;

\* BNR - enter 1 for binary variable, 0 for continuous \*;

\* DST - NOR for normal, BIN for binary variables \*;

\* LNK - ID for normal, LOGIT for binary variables \*;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

%MACRO GEN3(DVAR,BNR,DST,LNK);

\* Run the two-way model and output parameter estimates \*;

PROC GENMOD DATA = ADOS3;

CLASS TX;

MODEL &DVAR = TX PS TX\*PS / DIST = &DST LINK = &LNK TYPE3;

LSMEANS TX / DIFF;

ODS OUTPUT PARAMETERESTIMATES = TEST11;

ODS OUTPUT MODELFIT = TEST111;

TITLE2 'TESTING FOR COVARIATE BALANCE: WITH PS'; RUN;

DATA TRT\_EST (KEEP = DUM TRT0\_EST) PS\_EST (KEEP = DUM

PS\_EST) TRTPS\_EST (KEEP = DUM TRT0PS\_EST) INTRCPT\_EST

(KEEP = DUM INTRCPT\_EST);

SET TEST11;

DUM = **1**;

IF PARAMETER = 'tx' AND LEVEL1 = ‘A’ THEN DO;

TRT0\_EST = ESTIMATE;

OUTPUT TRT\_EST;

END;

IF PARAMETER = 'PS' THEN DO;

PS\_EST = ESTIMATE;

OUTPUT PS\_EST;

END;

IF PARAMETER = 'PS\*tx' AND LEVEL1 = **’A’** THEN DO;

TRT0PS\_EST = ESTIMATE;

OUTPUT TRTPS\_EST;

END;

IF PARAMETER = 'Intercept' THEN DO;

INTRCPT\_EST = ESTIMATE;

OUTPUT INTRCPT\_EST;

END;

DATA TEST111;

SET TEST111;

IF CRITERION = 'Deviance';

SIGHAT = SQRT(VALUEDF);

DUM = **1**;

KEEP DUM SIGHAT;

DATA EST;

MERGE TEST111 TRT\_EST PS\_EST TRTPS\_EST INTRCPT\_EST;

DUM = **1**;

KEEP TRT0\_EST PS\_EST TRT0PS\_EST INTRCPT\_EST SIGHAT DUM;

\* Merge parameter estimates with analysis data to allow computation

of predicted values for each patient. \*;

DATA ADOS3;

SET ADOS3;

DUM = **1**;

PROC SORT DATA = ADOS3; BY DUM; RUN;

PROC SORT DATA = EST; BY DUM; RUN;

DATA ALL;

MERGE ADOS3 EST;

BY DUM;

\* For each observation, compute the predicted value assuming each

treatment group \*;

PRED0 = INTRCPT\_EST + TRT0\_EST + PS\_EST\*PS +

TRT0PS\_EST\*PS;

PRED1 = INTRCPT\_EST + PS\_EST\*PS;

\* Compute the standardized difference for continuous and binary

covariates \*;

IF &BNR = **0** THEN DO;

TRTDIFF = TRT0\_EST + TRT0PS\_EST\*PS;

STDDIFF = ABS(TRT0\_EST + TRT0PS\_EST\*PS) / SIGHAT;

END;

IF &BNR = **1** THEN DO;

PRED0B = EXP(PRED0) / (**1** + EXP(PRED0));

PRED1B = EXP(PRED1) / (**1** + EXP(PRED1));

TRTDIFF = PRED0B - PRED1B;

STDDIFF = ABS( TRTDIFF / SQRT( (PRED0B\*(**1**-PRED0B) +

PRED1B\*(**1**-PRED1B)) / **2** ) );

END;

DATA OUT\_&DVAR;

SET ALL;

STDDIFF\_&DVAR = STDDIFF;

KEEP AGE PHQ1 GENDER SPOUSE WORK PS STDDIFF\_&DVAR;

%MEND GEN3;

\* Call GEN3 macro for each covariate to compute the weighted

standardized differences and then combine the results into a

single data set for reporting. \*;

ODS LISTING CLOSE;

%*GEN3*(GENDER, 1, BIN, LOGIT); RUN;

%*GEN3*(SPOUSE, 1, BIN, LOGIT); RUN;

%*GEN3*(WORK, 1, BIN, LOGIT); RUN;

%*GEN3*(AGE, 0, NOR, ID); RUN;

%*GEN3*(PHQ1, 0, NOR, ID); RUN;

ODS LISTING;

DATA REGSTD;

SET OUT\_GENDER OUT\_SPOUSE OUT\_WORK OUT\_AGE OUT\_PHQ1;

PROC MEANS DATA = REGSTD N MEAN STD MIN MAX;

VAR STDDIFF\_GENDER STDDIFF\_SPOUSE STDDIFF\_WORK

STDDIFF\_AGE STDDIFF\_PHQ1;

TITLE 'Assessing Propensity Score Balance for PS

Regression Analyses';

TITLE2 'Summary of Weighted Standardized Differences for

all covariates'; RUN;