**Example SAS CODE FOR CONDUCTING SEGMENTED REGRESSION on aggregated data
 (Steps 1-4) or on individual patient data (Step 5);**

**\*STEP 1. Create variables for segmented regression;**

 \*1. create a time variable (here, MONTHS) equal to 1 for 1st time period, and sequentially increasing
 through both periods;

 **data** monthdata2; set monthdata;

 year= admission\_year - **2008**;

 months = (year -**1**)\***12** + admission\_month;

 \*2. create an intervention variable;

 if admission\_year < **2012** then intervention=**0**;

 else if admission\_year >= **2012** then intervention=**1**; \*intervention began 1/2/2012;

 \*3. Create an interaction variable between time variable and intervention variable such that the
 intervention variable above estimates the jump in the outcome at start of intervention;

 \*for our AMET example, intervention starts at month 37, and below code creates variable

 "months\_minus\_36" which is 0 for months < 37 and starts at 1 for month 37, incrementing to the end;

 months\_minus\_36= max(**0**, months-**36**); \*Months after start of intervention 1 (started month 37);

**\*STEP 2. Aggregate data into appropriate time intervals in each period;|**

1. **Continuous outcome**

 \*For example, use proc means to calculate and save the average across patients of the outcome
 [here, HOSP\_LOS (hospital length of stay)] and potential confounding variable by month,
 or relevant time interval;

 months= months from start of study
 intervention = 0 or 1 (pre-intervention or post-intervention period)

%let caselevelvars= HOSP\_LOS age bmi comorbidity1-comorbidity100;

 **proc** **sort** data= fulldata; by intervention months;

 **proc** **means** data= fulldata n mean; by intervention months;

 var &caselevelvars;

 output out=means mean= &caselevelvars;  **run**;

 \*the above creates new variables of the mean of the outcome and baseline covariates for each month,

 and keeps the same names as the individual variables (for simplicity);

 **Merge data from Steps 1 and 2 into work.aggregated\_data**

**B. Binary Outcome

Use code from Step1A above for covariates, but the binary outcome should be summarized for each time
period in a way which allows either logistic regression in the events/trials format or analysis
on the count of events for the time interval.

Outcome is rapid response encounter (yes/no) – “Rapid\_Response**”**.
Calculate # events (**nevents\_RR) **and total N (**nTotal\_RR) **for each month.**

 **proc** **sort** data= fulldata; by intervention months;

 **proc** **means** data= fulldata n sum; by intervention months;

 var **Rapid\_Response**;

 output out=response sum= **nevents\_RR** N= **nTotal\_RR;**

 **run**;

 **Merge this binary response outcome data with the summary data on covariates in Step 1A.
 Merge data from Steps 1 and 2 into work.aggregated\_data**

 **\*STEP 3. Conduct segmented regression;**

 \*Choose an appropriate regression model based on the outcome variable and data structure;

 \*For the AMET data we used logistic regression for the rapid response (RR) variable since it

 is binary. We used linear regression for a sensitivity analysis on hospital length of stay.

1. **Continuous outcome – hospital LOS (sensitivity analysis in AMET study)**\*We used the proc GENMOD procedure with identity link (dist=n), as below.
 This is the same as linear regression;

 %let covs=age bmi Surgical\_Encounter RENLFAIL LIVER TUMOR OBESE WGHTLOSS ANEMDEF DRUG DEPRESS;

 **proc** **genmod** data= **aggregated\_data;**

 model outcome= months intervention months\_minus\_36 &caselevelvars / dist= n;

 ods output ParameterEstimates=beta;
 \*saving the beta coefficients and standard errors for calculations;

 **run**;

 \*calculate 95% CIs for segmente regression parameters of interest;

 **data** beta; set beta;

 lowerCL= estimate - **1.96** \* stderr;

 upperCL= estimate + **1.96** \* stderr;

 pval=probchisq; \*renaming P-value;

 **run**;

 title "Segmented Regression results on aggregated data using linear regression";

 **proc** **print** data=beta;

 **run**;

 **B. Binary Outcome – rapid response (yes/no)**

 We used the LOGISTIC procedure with logit link to conduct logistic regression modeling the
 #events/#trials for each month as the outcome variable, as below.

 %let covs=age bmi Surgical\_Encounter RENLFAIL LIVER TUMOR OBESE WGHTLOSS ANEMDEF DRUG DEPRESS;

 **proc** **logistic** data=ed.means\_RR\_nomiss2 descending;

 model nevents\_rr/totalN\_rr = months intervention months\_minus\_36 &covs /rl lackfit link=logit;

 where months < **80**; \*The study period ended here;

 output out=preds p=pred lower=lower\_CLM upper=upper\_CLM XBETA=XBETA STDXBETA=se\_xBETA
 stdreschi=stdreschi stdresdev=stdresdev;

 ods output ParameterEstimates=beta;

 **run**;
 **data** beta; set beta;

 lowerCL= estimate - **1.96** \* stderr;

 upperCL= estimate + **1.96** \* stderr;

 pval=probchisq; \*renaming P-value;

 **run**;

 title "Segmented Regression results on aggregated data using linear regression";

 **proc** **print** data=beta;

 **run**;

**\*STEP 4. Compare predicted values under intervention to counterfactual values
 assuming pre-intervention trend;**

 \*first create separate dataset for each parameter of interest, then merge together below;

 **data** b0 b1 b2 b3; set beta;

 D=**1**;

 if upcase(parameter)="INTERCEPT" then do;

 b0= estimate;

 KEEP B0 D;

 output b0;

 end;

 if upcase(parameter)="MONTHS" then do;

 b1=estimate;

 b1\_lowercl = lowercl;

 b1\_uppercl = uppercl;

 SE\_B1= StdErr;

 KEEP B1 B1\_LOWERCL B1\_UPPERCL SE\_B1 D;

 ouTput b1;

 end;

 if upcase(parameter)="INTERVENTION" then do;

 b2 =estimate;

 b2\_lowercl = lowercl;

 b2\_uppercl = uppercl;

 SE\_B2= StdErr;

 KEEP B2 B2\_LOWERCL B2\_UPPERCL SE\_B2 D;

 ouTput b2;

 end;

 if upcase(parameter)="MONTHS\_MINUS\_36" then do;

 b3 =estimate;

 b3\_lowercl = lowercl;

 b3\_uppercl = uppercl;

 SE\_B3= StdErr;

 KEEP B3 B3\_LOWERCL B3\_UPPERCL SE\_B3 D;

 ouTput b3;

 end;

 **run**;

 \*Joining estimates for each parameter of interest into a single row for analysis;

 **DATA** beta\_preds; UPDATE b0 b1; BY D;

 **DATA** beta\_preds; UPDATE beta\_preds b2; BY D;

 **DATA** beta\_preds; UPDATE beta\_preds b3; BY D;

 \*SELECT A STARTTIME INDICATING WHEN THE INTERVENTION BEGAN;

 %let starttime=37; \*AMET start month;

 \*SELECT A TIME AT WHIFH TO COMPARE INTERVENTIN AND PRE-INTERVENTION TRENDS;

 %LET PTIME=60; \*COMPARE AFTER 24 MONTHS OF INTERVENTION;

 **data** beta\_preds; length dependent $**30**;

 set beta\_preds;

 ptime= &ptime;

 starttime =&starttime;

 Y\_ptime\_with = b0 + b1 \* &ptime + b2 \***1** + b3 \* (&ptime - &starttime);
 \*Equation 3: predicted y at ptime after starting interventi0no at starttime;

 Y\_ptime\_without = b0 + b1 \* &ptime;
 \*Equation 4: predicted y at ptime without intervention (assumign pre-intercention trend);

 Y\_diff\_ptime = b2 \***1** + b3 \* (&ptime - &starttime);
 \*y at ptime --- predicted intervention effect -- equation 2 minus 3 in manuscript;

 Y\_diff\_var = **1**\*\***2** \*se\_b2 \*\***2** + ( ((&ptime - &starttime)\*\***2**) \* se\_b3\*\***2**) ;
 \*variance of A SUM; \*var (cx + dy) = sum (c2 var(x) d2 var(y));

 Y\_diff\_se = sqrt(y\_diff\_var);

 Y\_diff\_low95 = Y\_diff\_ptime - **1.96** \* Y\_diff\_se;

 Y\_diff\_up95 = Y\_diff\_ptime + **1.96** \* Y\_diff\_se;

 z\_stat\_diff = Y\_diff\_ptime/ Y\_diff\_se;

 P\_val = **2**\*(**1**- probnorm(abs(z\_stat\_diff)));

 **run**;

 TITLE3 "Comparing predicted value of Y at &ptime under intervention versus w/o intervention";

 **proc** **print** dat=beta\_preds;

 var ptime starttime Y\_ptime\_with Y\_ptime\_without Y\_diff\_ptime Y\_diff\_var Y\_diff\_se

 Y\_diff\_low95 Y\_diff\_up95 z\_stat\_diff p\_val;

 **run**;

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

**\*STEP 5: SEGMENTED REGRESSION ON PATIENT-LEVEL DATA;**

 \*For patient level data the actual covariate values are used, if available;

 \*Choose a model that fits the subject level outcome variable (e.g., logistic regression if binary);

 \*If patients represented more than once, adjust for possible within-subject correlation;

 \*Otherwise, methods are the same as for the aggregated data;

title "AMET Segmented regression --RRT yes/no--- patient level";

title2 "ALL OBS per patient -- adjust for correlation";

%let covariates= age bmi comorbidity1-comorbidity100;

**proc** **genmod** data=AMET DESCENDING;

 class MRN;

 model Rapid\_Response = months intervention months\_minus\_36 &covariates / dist=b;
 \*use logistic regression since outcome is binary;

 repeated subject=MRN / type=cs; \*assume exchangeable correlation;
 \*repeated statement uses GEE model adjusting for within-subject correlation;

 ods output GEEEmpPEst =beta; \*same betas and robust variance-covariance estimates from GEE;

**run**;

**data** beta; set beta;

 lowerCL= estimate - **1.96** \* stderr;

 upperCL= estimate + **1.96** \* stderr;

 OR=exp(estimate); \*odds ratio for each parameter of interest;

 OR\_low= exp(estimate - **1.96**\*stderr);

 OR\_up = exp(estimate + **1.96**\*stderr);

**run**;

**SAS Code for Autocorrelation analyses.**

In SAS an autocorrelation-adjusted segmented regression analysis could be done using the AUTOREG procedure in the ETS module, as below.

SAS code for adjusting for autocorrelation in segmented regression using continuous outcome.

 proc autoreg data=a;

 Model y = time intervention time\_after\_intervention / nlag=1 method=ml;

 run;
 The NLAG=1 option is used to estimate the first-order auto-correlation estimation and adjustment.