Dataset Integrity Check for the TrialNet(08) Data Files

Prepared by Michael Spriggs
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton MD 20705
March 12, 2015

Table of Contents

1 Standard Disclaimer

The intent of this DSIC is to provide confidence that the data distributed by the NIDDK repository is a true copy of the study data. Our intent is not to assess the integrity of the statistical analyses reported by study investigators. As with all statistical analyses of complex datasets, complete replication of a set of statistical results should not be expected in secondary analysis. This occurs for a number of reasons including differences in the handling of missing data, restrictions on cases included in samples for a particular analysis, software coding used to define complex variables, etc. Experience suggests that most discrepancies can ordinarily be resolved by consultation with the study data coordinating center (DCC), however this process is labor-intensive for both DCC and Repository staff. It is thus not our policy to resolve every discrepancy that is observed in an integrity check. Specifically, we do not attempt to resolve minor or inconsequential discrepancies with published results or discrepancies that involve complex analyses, unless NIDDK Repository staff suspect that the observed discrepancy suggests that the dataset may have been corrupted in storage, transmission, or processing by repository staff. We do, however, document in footnotes to the integrity check those instances in which our secondary analyses produced results that were not fully consistent with those reported in the target publication.

2 Study Background

This multicenter, double-masked, randomized, controlled trial evaluated whether treatment of with a target antigen (glutamic acid decarboxylase) could modulate aggressive autoimmunity effects in subjects recently diagnosed with type 1 diabetes mellitus. Subjects were randomized via a 1:1:1 ratio into one of three treatment groups: three injections of 20 µg GAD-alum, two injections of GAD-alum and one of aluminum hydroxide alone (placebo), or three injections of aluminum hydroxide. These subcutaneous injections were completed at baseline, four weeks, and twelve weeks. All subjects received intensive diabetes management with the goal of achieving excellent glycemic control. Stimulated C-peptide levels were measured at the year 1 visit via a mixed meal tolerance test (MMTT). Subjects continued follow up for a second year including the performance of a MMTT every 6 months.

3 Archived Datasets

All SAS data files, as provided by the Data Coordinating Center (DCC), are located in the "data" folder in the data package. For this replication, variables were taken from different analysis datasets created by the DCC.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Wherrett, et al [1] in Lancet on 2011, July 23. To verify the integrity of the datasets, descriptive statistics of baseline characteristics were computed, by treatment group (Table B).

5 Results

Table 1 in the publication [1], <u>Baseline Characteristics of Subjects at Entry.</u> Table A lists the variables that were used in the replication and Table B compares the results calculated from the archived data file to the results published in Table 1. The results of the replication are similar to the published results.

Table 2 in the publication [1]:<u> Adverse Events</u>. The adverse event data that was included in the data package includes adverse events that were not captured prior to the publication. For this reason, the adverse event data was not compared to the data in the publication.

6 Conclusions

The NIDDK repository is confident that the TrialNet (08) data files to be distributed are a true copy of the study data.

7 References

1. Wherrett DK, Bundy B, Becker DJ, et al. Antigen-based therapy with glutamic acid decarboxylase (GAD) vaccine in patients with recent-onset type 1 diabetes: a randomised double-blind trial. Lancet 2011;378:319-27.

Table A: Variables used to replicate Table 1: <u>Baseline Characteristics of Subjects at Entry</u>

Table Variable	Variables Used in Replication from the Trialnet(08) Dataset
Age- yr	Demographics.age
Female sex- no. of patients (%)	Demographics.sex
Race - no. of patients* (%)	Demographics.race_white
Ethnicity - no. of patients (%)	Demographics.ethnicity
No. of days from diagnosis to first infusion	screening_medical_history. DateOfT1DdiagnosisMonth screening_medical_history. DateOfT1DdiagnosisDay screening_medical_history. DateOfT1DdiagnosisYear study_drug_administration.Date_of_ Visit
Weight (kg)	physical_exam.WEIGHTKG
Body Mass Index (kg/m2)	physical_exam.WEIGHTKG physical_exam.HEIGHTCM
Total daily insulin dose at baseline* – (U/kg)	diab.AverageUnitsOfItermediateInsul diab.AvgUnitsShortActingInsulin physical_exam.WEIGHTKG
Ketoacidosis at diagnosis - no. of patients (%)	Screening_medical_history.DKAatDiagnosis
Diabetes associated HLA alleles present - no. of patients* (%) Glycated hemoglobin at baseline (HbA1c -%) AUC Mean for C-peptide (pmol/ml) No. of autoantibodies** - no. of patients (%) GAD Antibody Titre (index units)	research_labs.VISIT research_labs.SPEC_NAME research_labs.TEST_NAME research_labs.RESULT research_labs.OUTCOME

Table B: Comparison of values computed in integrity check to reference article Table 1 values

Characteristic	Al×3 [Manuscript]	Al×3 [DSIC]	Al×3 [Diff]
Age- yr			
Mean	16.6 ±9.23	16.60 ± 9.23	0
Median	14.5	14.50	0
Range	4 – 45	4.00-45.00	0
Female sex- no. of patients (%)	19 (39.6)	19 (39.6)	0
Race - no. of patients* (%)	40 (85.1)	40 (83.3)	0(1.8)
Ethnicity - no. of patients (%)	48 (100.0)	43 (89.6)	5(10.4)
No. of autoantibodies** - no. of patients (%)			
1	1 (2.1)	1 (2.1)	0
2	10 (20.8)	10 (20.8)	0
3	24 (50.0)	24 (50.0)	0
4	13 (27.1)	13 (27.1)	0
GAD Antibody Titre (index units)	0.336±0.307	0.336 ± 0.307	0
No. of days from diagnosis to first infusion	87.1 ± 15.5	87.1 ± 15.4	0+-.1
Range	42 - 106	42-106	0-0
Weight (kg)	56.4 ± 21.3	56.4 ± 21.3	0
Body Mass Index (kg/m2)	20.9 ± 4.22	20.9 ± 4.22	0
AUC Mean for C-peptide (pmol/ml)	0.690 ± 0.278	0.707 ± 0.250	-0.017 ± 0.028
Glycated hemoglobin at baseline (HbA1c -%)	6.39 ± 0.844	6.39 ± 0.835	0± 0.009
Total daily insulin dose at baseline* – (U/kg)	0.395 ± 0.251	0.401 ± 0.253	-0.006±-0.002
Ketoacidosis at diagnosis - no. of patients (%)	6 (13.3)	6 (12.5)	0(.8)
Diabetes associated HLA alleles present - no. of patients* (%)			
DR3 and DR4	14 (29.8)	14 (29.8)	0(0)
DR3 only	8 (17.0)	8 (17.0)	0(0)
DR4 only	20 (42.6)	20 (42.6)	0(0)
Neither	5 (10.6)	5 (10.6)	0(0)

Characteristic	GAD/Al×2 + Al [Manuscript]	GAD/Al×2 + Al [DSIC]	GAD/Al×2 + Al [Diff]
Age- yr			
Mean	14.9±8.72	14.86 ± 8.74	.04+-.02
Median	14	14.00	0
Range	3 - 45	3.00-45.00	0
Female sex- no. of patients (%)	31 (63.3)	31 (63.3)	0
Race - no. of patients* (%)	45 (93.8)	47 (95.9)	-2(-2.1)
Ethnicity - no. of patients (%)	49 (100.0)	44 (89.8)	5(10.2)
No. of autoantibodies** - no. of patients (%)			
1	0 (0.0)	0(0)	0
2	10 (20.4)	10 (20.4)	0
3	23 (46.9)	23 (46.9)	0
4	16 (32.7)	16 (32.7)	0
GAD Antibody Titre (index units)	0.349 ± 0.381	0.349 ± 0.381	0
No. of days from diagnosis to first infusion	87.0 ± 14.1	87.3 ± 14.1	.3+-0
Range	47 - 104	47-104	0-0
Weight (kg)	48.7±18.4	48.7 ± 18.4	0
Body Mass Index (kg/m2)	20.0 ± 3.66	19.97 ± 3.67	.03+--.01
AUC Mean for C-peptide (pmol/ml)	0.655 ± 0.374	0.699 ± 0.367	-0.044 ± 0.007
Glycated hemoglobin at baseline (HbA1c -%)	6.63 ± 0.965	6.63 ± 0.965	0
Total daily insulin dose at baseline* – (U/kg)	0.331 ± 0.219	0.341 ± 0.212	-.01+-.003
Ketoacidosis at diagnosis - no. of patients (%)	11(25.0)	12 (25.0)	1(0)
Diabetes associated HLA alleles present - no. of patients* (%)			
DR3 and DR4	7 (14.3)	7 (14.3)	0(0)
DR3 only	13 (26.5)	13 (26.5)	0(0)
DR4 only	21 (42.9)	21 (42.9)	0(0)
Neither	8 (16.3)	8 (16.3)	0(0)

Characteristic	GAD/Al×3 [Manuscript]	GAD/Al×3 [DSIC]	GAD/Al×3 [Diff]
Age- yr			
Mean	17.9 ± 10.4	17.83 ± 10.34	.07+-.06
Median	15.5	15.50	0
Range	3 - 44	3.00-44.00	0
Female sex- no. of patients (%)	14 (29.2)	14 (29.2)	0
Race - no. of patients* (%)	43 (91.5)	45 (93.8)	-2(-2.3)
Ethnicity - no. of patients (%)	48 (100.0)	41 (85.4)	7(14.6)
No. of autoantibodies** - no. of patients (%)			
1	3 (6.2)	3 (6.3)	0(-.1)
2	10 (20.8)	11 (22.9)	-1(-2.1)
3	14 (29.2)	14 (29.2)	0
4	21 (43.8)	20 (41.7)	1(2.1)
GAD Antibody Titre (index units)	0.230 ± 0.224	0.228 ± 0.226	0.002 ±-0.002
No. of days from diagnosis to first infusion	83.9 ± 16.8	84.0 ± 16.8	-.1+-0
Range	45 - 105	45-105	0-0
Weight (kg)	61.4 ± 22.9	61.4 ± 22.9	0
Body Mass Index (kg/m2)	22.3 ± 4.69	22.26 ± 4.69	.04+-0
AUC Mean for C-peptide (pmol/ml)	0.710 ± 0.362	0.765 ± 0.333	-0.055 ± 0.029
Glycated hemoglobin at baseline (HbA1c -%)	6.63 ± 1.19	6.63 ± 1.19	0
Total daily insulin dose at baseline* – (U/kg)	0.386 ± 0.229	0.402 ± 0.235	-.016+-.006
Ketoacidosis at diagnosis - no. of patients (%)	9 (21.4)	9 (18.8)	0(2.6)
Diabetes associated HLA alleles present - no. of patients* (%)			
DR3 and DR4	13 (27.1)	13 (27.1)	0(0)
DR3 only	15 (31.2)	15 (31.3)	0(-0.1)
DR4 only	13 (27.1)	13 (27.1)	0(0)
Neither	7 (14.6)	7 (14.6)	0(0)

Attachment A: SAS Code

```
***********************************************************************************************************
***Programmer: Michael Spriggs
***Date Created: 03/04/2015
***Purpose: To replicate the data in the TN08 manuscript
**********************************************************************************************************;

title1 "%sysfunc(getoption(sysin))";
title2 " ";

options nofmterr source2 mprint symbolgen spool;
libname sas_data "/prj/niddk/ims_analysis/TrialNet/private_orig_data/new_data_3_6_15/Data.Extraction.8.Version150305PW/8/sasv9/";
libname ae_data  "/prj/niddk/ims_analysis/TrialNet/private_orig_data/Data.Extraction.8.Version110531/8/sasv9/";

proc format;
  value yesnof
        1="Yes"
        2="No"
        ;
  value $treatf
        '3 injections of Alum only'="Alx3"
        '2 injections of 20?g GAD-alum followed by one injection with Alum alone'="GAD/Al×2 + Al"
        '3 injections with 20?g GAD-alum'='GAD/Al×3'
        ;
  value treatf
        0="Alx3"
        1="GAD/Al×2 + Al"
        2='GAD/Al×3'
        ;
  value datecutf
        .="Missing"
        0-18777="Before 5/31/2011"
        18778="5/31/2011"
        18779-High="After 5/31/2011"
        ;
  value severef
        1="1"
        2-5="2+"
        ;


*** File containing macro for examining each dataset ***;
%include '/prj/niddk/ims_analysis/sas_macros/redaction_data_summary.sas';

%macro freqdata1(order=, invar=, level=, popvar=, totalvl=);

%if &totalvl.=null %then %do;
  proc freq data=table1 noprint;
        tables &invar*treatmentname/out=data1 outpct;
        format _all_;
```

```
       run;

   data data1(keep=LEVEL treatmentname name CHARALL ORDERER);
     set data1(rename=(&invar=LEVEL));
     length name $100 CHARALL $100;
     name=upcase("&invar");
     PCT_DISP=round(PCT_COL,.1);
     CHARALL=compress(put(COUNT,8.))||" ("||compress(put(PCT_DISP,8.1))||")";
     ORDERER=&order;
     if level in &level then output data1;

%end;
%else %do;
  proc freq data=table1 noprint;
       tables &invar*treatmentname/out=data1 outpct;
       format _all_;
       where &popvar. in &totalvl.;
       run;

   data data1(keep=LEVEL treatmentname name COUNT PCT_DISP ORDERER) total1(keep=accumer1 accumer2);
     set data1(rename=(&invar=LEVEL)) end=end1;
     length name $100;
     retain accumer1 accumer2;
     if _n_=1 then do;
       accumer1=0;
       accumer2=0;
     end;
     if treatmentname=1 then accumer1=accumer1+count;
     else if treatmentname=2 then accumer2=accumer2+count;
     name=upcase("&invar");
     PCT_DISP=round(PCT_COL,.1);
     ORDERER=&order;
     if level in &level then output data1;
     if end1 then output total1;

data total1(drop=accumer:);
  set total1;
  length treatmentname total 8.;
  treatmentname=1;
  total=accumer1;
  output;
  treatmentname=2;
  total=accumer2;
  output;

data data1(drop=COUNT PCT_DISP total);
  merge data1 total1;
  by treatmentname;
  length CHARALL $100;
  CHARALL=compress(put(COUNT,8.))||"/"||compress(put(total,8.))||" ("||compress(put(PCT_DISP,8.1))||")";
```

```
%end;
data accumfreq1;
  set accumfreq1 data1;

%mend freqdata1;

%macro meandata1(order=, invar=, roundvar=, digit=);
proc means data=table1 mean stddev noprint;
      var &invar;
      class treatmentname;
      output out=data1 mean=mean stddev=stddev;
      run;

data data1(drop=_TYPE_ _FREQ_ mean stddev);
  set data1;
  length name CHARALL $100;
  name=upcase("&invar");
  mean=round(mean,&roundvar);
  stddev=round(stddev,&roundvar);
  CHARALL=compress(put(mean,8.&digit))||" ± "||compress(put(stddev,8.&digit));
  ORDERER=&order;

data accummean1;
  set accummean1 data1;

%mend meandata1;

%macro mediandata1(order=, invar=, roundvar=, digit=);
proc means data=table1 median p25 p75 min max noprint;
      var &invar;
      class treatmentname;
      output out=data1 median=median p25=p25 p75=p75 min=min max=max;
      run;

data data1(drop=_TYPE_ _FREQ_ median p25 p75 min max);
  set data1;
  length name CHARALL $100;
  name=upcase("&invar");
  median=round(median,&roundvar);
  min=round(min,&roundvar);
  max=round(max,&roundvar);
  ORDERER=&order;
  CHARALL=compress(put(median,8.&digit));
  output;
  ORDERER=ORDERER+.01;
  CHARALL=compress(put(min,8.&digit)||"-"||put(max,8.&digit));
  output;

data accummedian1;
  set accummedian1 data1;
```

9

```
%mend mediandata1;

%macro rangedata1(order=, invar=, roundvar=, digit=);
proc means data=table1 median p25 p75 min max noprint;
        var &invar;
        class treatmentname;
        output out=data1 min=min max=max;
        run;

data data1(drop=_TYPE_ _FREQ_ min max);
  set data1;
  length name CHARALL $100;
  name=upcase("&invar");
  min=round(min,&roundvar);
  max=round(max,&roundvar);
  ORDERER=&order;
  CHARALL=compress(put(min,8.&digit)||"-"||put(max,8.&digit));
  output;

data accummedian1;
  set accummedian1 data1;

%mend rangedata1;

%macro freqdata2(order=, invar=, level=, popvar=, totalvl=);

%if &totalvl.=null %then %do;
  proc freq data=table2 noprint;
        tables &invar*treatmentname/out=data2 outpct;
        format _all_;
        run;

  data data2(keep=LEVEL treatmentname name CHARALL ORDERER);
    set data2(rename=(&invar=LEVEL));
    length name $100 CHARALL $100;
    name=upcase("&invar");
    PCT_DISP=round(PCT_COL,.1);
    CHARALL=compress(put(COUNT,8.))||" ("||compress(put(PCT_DISP,8.1))||")";
    ORDERER=&order;
    if level in &level then output;
%end;
%else %do;
  proc freq data=table2 noprint;
        tables &invar*treatmentname/out=data2 outpct;
        format _all_;
        where &popvar. in &totalvl.;
        run;

  data data2(keep=LEVEL treatmentname name COUNT PCT_DISP ORDERER) total2(keep=accumer1 accumer2);
    set data2(rename=(&invar=LEVEL)) end=end1;
```

10

```
      length name $100;
      retain accumer1 accumer2;
      if _n_=1 then do;
        accumer1=0;
        accumer2=0;
      end;
      if treatmentname=1 then accumer1=accumer1+count;
      else if treatmentname=2 then accumer2=accumer2+count;
      name=upcase("&invar");
      PCT_DISP=round(PCT_COL,.1);
      ORDERER=&order;
      if level in &level then output data2;
      if end1 then output total2;

data total2(drop=accumer:);
  set total2;
  length treatmentname total 8.;
  treatmentname=1;
  total=accumer1;
  output;
  treatmentname=2;
  total=accumer2;
  output;

data data2(drop=COUNT PCT_DISP total);
  merge data2 total2;
  by treatmentname;
  length CHARALL $100;
  CHARALL=compress(put(COUNT,8.))||"/"||compress(put(total,8.))||" ("||compress(put(PCT_DISP,8.1))||")";


%end;
data accumfreq2;
  set accumfreq2 data2;

%mend freqdata2;

%macro meandata2(order=, invar=, roundvar=, digit=);
proc means data=table2 mean stddev noprint;
       var &invar;
       class treatmentname;
       output out=data2 mean=mean stddev=stddev;
       run;

data data2(drop=_TYPE_ _FREQ_ mean stddev);
  set data2;
  length name CHARALL $100;
  name=upcase("&invar");
  mean=round(mean,&roundvar);
  stddev=round(stddev,&roundvar);
  CHARALL=compress(put(mean,8.&digit))||" ± "||compress(put(stddev,8.&digit));
```

11

```
   ORDERER=&order;

data accummean2;
   set accummean2 data2;

%mend meandata2;

%macro mediandata2(order=, invar=, roundvar=, digit=);
proc means data=table2 median p25 p75 noprint;
       var &invar;
       class treatmentname;
       output out=data2 median=median p25=p25 p75=p75;
       run;

data data2(drop=_TYPE_ _FREQ_ median p25 p75);
   set data2;
   length name CHARALL $100;
   name=upcase("&invar");
   median=round(median,&roundvar);
   p25=round(p25,&roundvar);
   p75=round(p75,&roundvar);
   ORDERER=&order;
   CHARALL=compress(put(median,8.));
   output;
   ORDERER=ORDERER+.01;
   CHARALL=compress(put(p25,8.)||"-"||put(p75,8.));
   output;

data accummedian2;
   set accummedian2 data2;

%mend mediandata2;


data accumfreq1 accummean1 accummedian1
     accumfreq2 accummean2 accummedian2;
   set _null_;

data demographics;
   set SAS_DATA.demographics;

data treatment_table;
   set SAS_DATA.treatment_table;

data physical;
   set SAS_DATA.physical_exam;
   where visit="Screening";
   BMI=WEIGHTKG/(HEIGHTCM*HEIGHTCM/10000);

data medhist(keep=MASKID DXDATE DKAatDiagnosis);
   set SAS_DATA.screening_medical_history;
```

```
   where visit="Screening";
       if DateOfT1DdiagnosisMonth="Jan" then DateOfT1DdiagnosisMonthn=1;
   else if DateOfT1DdiagnosisMonth="Feb" then DateOfT1DdiagnosisMonthn=2;
   else if DateOfT1DdiagnosisMonth="Mar" then DateOfT1DdiagnosisMonthn=3;
   else if DateOfT1DdiagnosisMonth="Apr" then DateOfT1DdiagnosisMonthn=4;
   else if DateOfT1DdiagnosisMonth="May" then DateOfT1DdiagnosisMonthn=5;
   else if DateOfT1DdiagnosisMonth="Jun" then DateOfT1DdiagnosisMonthn=6;
   else if DateOfT1DdiagnosisMonth="Jul" then DateOfT1DdiagnosisMonthn=7;
   else if DateOfT1DdiagnosisMonth="Aug" then DateOfT1DdiagnosisMonthn=8;
   else if DateOfT1DdiagnosisMonth="Sep" then DateOfT1DdiagnosisMonthn=9;
   else if DateOfT1DdiagnosisMonth="Oct" then DateOfT1DdiagnosisMonthn=10;
   else if DateOfT1DdiagnosisMonth="Nov" then DateOfT1DdiagnosisMonthn=11;
   else if DateOfT1DdiagnosisMonth="Dec" then DateOfT1DdiagnosisMonthn=12;
   DXDATE=MDY(DateOfT1DdiagnosisMonthn,DATEOFT1DDiagnosisDay,DATEOFT1DDiagnosisYear);

data cbcwdiff;
  set SAS_DATA.cbc_wdiff_results;
  where visit="Screening";

data diab;
  set SAS_DATA.diabetes_management;
  where visit="Baseline";

data drugadmin(keep=MASKID DATE_OF_VISIT);
  set SAS_DATA.study_drug_administration;
  where visit="Baseline";

data research_labs;
  set SAS_DATA.research_labs;

proc contents data=research_labs;

data adverseevent_14;
  set sas_data.adverseevent;

data adverseevent(drop=MASKID rename=(NEW_MASKID=MASKID));
  set ae_data.adverseevent;
  length NEW_MASKID 8.;
  NEW_MASKID=put(MASKID,8.);

proc contents data=adverseevent_14;
proc contents data=adverseevent;

data adverseevent_14;
  set adverseevent_14(rename=(serious=serious_14));

proc sort data=adverseevent_14;
  by InitialAEID AEID MASKID;

proc sort data=adverseevent;
  by InitialAEID AEID MASKID;
```

13

```
data adverseevent_cc;
  merge adverseevent(keep=InitialAEID AEID MASKID Severity1 serious Category) adverseevent_14(keep=InitialAEID AEID MASKID
severity_index serious_14);
  by InitialAEID AEID MASKID;

proc sort data=adverseevent_cc;
  by MASKID;

proc sort data=treatment_table;
  by MASKID;
/*
data adverseevent_cc;
  merge adverseevent_cc treatment_table(keep=maskid treatmentname);
  by maskid;
*/
proc freq data=adverseevent_cc;
  tables severity1*severity_index*serious*serious_14/missing list;
  Title3 'Severity cross check';

proc sort data=adverseevent;
  by Maskid;

proc sort data=demographics;
  by MASKID;

proc sort data=drugadmin;
  by MASKID;

proc sort data=treatment_table;
  by MASKID;

proc sort data=physical;
  by MASKID;

proc sort data=medhist;
  by MASKID;

proc sort data=cbcwdiff;
  by MASKID;

proc sort data=diab;
  by MASKID;

proc sort data=adverseevent;
  by MASKID;

data research_aa(drop=test_name result rename=(s_test_name=test_name s_result=result));
  set research_labs;
  length s_test_name s_result $10. result_n 8.;
  s_test_name=strip(test_name);
```

```
    s_result=strip(result);
    if spec_name="Serum - Autoantibodies" and visit="Screening" and test_name ne " " then do;
      if s_result='<.05' then result_n=0.025;
      else if s_result ne " " then result_n=put(s_result,8.2);
      output;
    end;

data research_peptide(drop=test_name result rename=(s_test_name=test_name s_result=result));
  set research_labs;
  length s_test_name s_result $10. result_n 8.;
  s_test_name=strip(test_name);
  s_result=strip(result);
  if index(spec_name,"peptide")>0 then do;
    if s_result='<.05' then result_n=0.025;
    else if s_result ne " " then result_n=put(s_result,8.2);
    output;
  end;

data research_hemo(drop=test_name result rename=(s_test_name=test_name s_result=result));
  set research_labs;
  length s_test_name s_result $10. Glyc_HEMO 8.;
  s_test_name=strip(test_name);
  s_result=strip(result);
  if index(spec_name,"HbA1c")>0 and Visit="Baseline" then do;
    if s_result ne " " then Glyc_HEMO =put(s_result,8.2);
    output;
  end;

data research_DR(drop=test_name result rename=(s_test_name=test_name s_result=result));
  set research_labs;
  length s_test_name s_result $10.;
  s_test_name=strip(test_name);
  s_result=strip(result);
  if index(test_name,"DR")>0 then  output;

data research_GAD(drop=test_name result rename=(s_test_name=test_name s_result=result));
  set research_labs;
  length s_test_name s_result $10. GAD_NUM 8.;
  if index(test_name,"GAD65")>0 and index(test_name,"GAD65H")=0 and Visit="Screening" then do;
    s_test_name=strip(test_name);
    s_result=strip(result);
    GAD_NUM=put(s_result,8.3);
    output;
  end;

data research_DR_short(keep=MaskID DR_NET);
  set research_DR;
  by MaskID;
  length DR_N 8. DR_NET $10;
  retain DR_N;
  if first.MASKID then DR_N=0;
```

```
    if strip(test_name)="DR3" and strip(result)="ABSENT" then DR_N=DR_N+2;
    if strip(test_name)="DR4" and strip(result)="ABSENT" then DR_N=DR_N+1;
    if last.MASKID then do;
      DR_NET=strip(put(DR_N,8.));
      output;
    end;

data research_short(keep=MASKID ANTIBODY_CHAR);
  set research_aa;
  by MASKID;
  length ANTIBODY_COUNT 8. ANTIBODY_CHAR $8.;
  retain ANTIBODY_COUNT;
  if first.MASKID then ANTIBODY_COUNT=0;
  if test_name="GAD65" and result_n>.032 then ANTIBODY_COUNT=ANTIBODY_COUNT+1;
  if test_name="GAD65H" and result_n>20 then ANTIBODY_COUNT=ANTIBODY_COUNT+1;
  if test_name="ICA512" and result_n>.049 then ANTIBODY_COUNT=ANTIBODY_COUNT+1;
  if test_name="ICA" and result_n>=10 then ANTIBODY_COUNT=ANTIBODY_COUNT+1;
  if test_name="IA-2H" and result_n>5 then ANTIBODY_COUNT=ANTIBODY_COUNT+1;
  if test_name="MIAA" and result_n>.01 then ANTIBODY_COUNT=ANTIBODY_COUNT+1;
  if test_name="ZnT8" and result_n>.02 then ANTIBODY_COUNT=ANTIBODY_COUNT+1;
  if last.MASKID then do;
    ANTIBODY_CHAR=strip(put(ANTIBODY_COUNT,8.));
    output;
  end;

proc freq data=research_short;
  tables ANTIBODY_CHAR/missing list;
  title3 'Antibody check';

data research_auc(drop=test_name result rename=(s_test_name=test_name s_result=result));
  set research_labs;
  length s_test_name s_result $10.;
  s_test_name=strip(test_name);
  s_result=strip(result);
  if visit="Screening" and index(test_name,"PEP")>0 then output;

proc freq data=research_auc;
  tables test_name/missing list;

data research_auc_proc(keep=MaskID result_:);
  set research_auc;
      by MaskID;
  length resultn
        result_m10
        result_0
        result_15
        result_30
        result_60
        result_90
        result_120
        result_150
```

16

```
        result_180
        result_210
        result_240
8.;
  retain result_m10
         result_0
         result_15
         result_30
         result_60
         result_90
         result_120
         result_150
         result_180
         result_210
         result_240
         ;

  if result="<.05" then resultn=.025;
  else resultn=input(result,8.);
  resultn=resultn*.33;

  if first.maskid then do;
    result_m10=.;
    result_0  =.;
    result_15 =.;
    result_30 =.;
    result_60 =.;
    result_90 =.;
    result_120=.;
    result_150=.;
    result_180=.;
    result_210=.;
    result_240=.;
  end;

  if test_name="PEP-10" then result_m10  =resultn;
  if test_name="PEP0"   then result_0    =resultn;
  if test_name="PEP120" then result_120  =resultn;
  if test_name="PEP15"  then result_15   =resultn;
  if test_name="PEP150" then result_150  =resultn;
  if test_name="PEP180" then result_180  =resultn;
  if test_name="PEP210" then result_210  =resultn;
  if test_name="PEP240" then result_240  =resultn;
  if test_name="PEP30"  then result_30   =resultn;
  if test_name="PEP60"  then result_60   =resultn;
  if test_name="PEP90"  then result_90   =resultn;

  if last.maskid then output;

data research_auc_proc_b;
  set research_auc_proc;
```

17

```
   length result_string $50;
   if result_m10>. then result_string=strip(result_string)||"-10,";
   if result_0  >. then result_string=strip(result_string)||"0,"  ;
   if result_15 >. then result_string=strip(result_string)||"15," ;
   if result_30 >. then result_string=strip(result_string)||"30," ;
   if result_60 >. then result_string=strip(result_string)||"60," ;
   if result_90 >. then result_string=strip(result_string)||"90," ;
   if result_120>. then result_string=strip(result_string)||"120,";
   if result_150>. then result_string=strip(result_string)||"150,";
   if result_180>. then result_string=strip(result_string)||"180,";
   if result_210>. then result_string=strip(result_string)||"210,";
   if result_240>. then result_string=strip(result_string)||"240,";

   if result_string in("-10,0,15,30,60,90,120," "-10,0,15,30,60,90,120,150,180,210,240," "-10,0,15,30,60,90,120,150,180," "-
10,0,15,30,60,90,120,150,210,240,") then do;
      AUC_RESULT=(7.5*result_0+15*result_15+22.5*result_30+30*result_60+30*result_90+15*result_120)/120;
   end;
   else if result_string in("-10,15,30,60,90,120,150,180,210,240,") then do;
      AUC_RESULT=(7.5*result_m10+15*result_15+22.5*result_30+30*result_60+30*result_90+15*result_120)/120;
   end;

data table1;
   merge demographics(in=in_demo)
         treatment_table
         physical
         medhist
         cbcwdiff
         diab
         research_short
         research_hemo
         research_DR_short(in=in_DR)
         research_GAD
         drugadmin
         research_auc_proc_b
         ;
   by maskid;
   length c_race_white $8.;
   c_race_white=strip(put(race_white,8.));
   total_insulin_kg=(AverageUnitsOfItermediateInsul+AvgUnitsShortActingInsulin)/weightkg;
   demo_ok=in_demo;
   dr_ok=in_dr;
   INFUSE_DAYS=DATE_OF_VISIT-DXDATE;
   if in_demo and treatmentname ne " " then output;

proc freq data=table1;
   tables ANTIBODY_CHAR*treatmentname/missing;
   title3 'Antibody check';

data table2;
   merge treatment_table
         adverseevent_cc;
```

18

```
   by maskid;
   if treatmentname=" " then delete;

proc freq data=table1;
   tables treatmentname/missing list;
   format treatmentname $treatf.;
   title3 'Table 1 N';

%meandata1(order=1,   invar=age               , roundvar=.01, digit=2);
%mediandata1(order=2, invar=age, roundvar=.01, digit=2);
%freqdata1(order=4,   invar=sex               , level=("Female"),popvar=, totalvl=null);
%freqdata1(order=5,   invar=c_race_white      , level=("1"),popvar=, totalvl=null);
%freqdata1(order=6,   invar=ethnicity         , level=("Not Hispanic,Latino or Spanish origin"),popvar=, totalvl=null);
%freqdata1(order=7,   invar=ANTIBODY_CHAR     , level=("1"),popvar=, totalvl=null);
%freqdata1(order=8,   invar=ANTIBODY_CHAR     , level=("2"),popvar=, totalvl=null);
%freqdata1(order=9,   invar=ANTIBODY_CHAR     , level=("3"),popvar=, totalvl=null);
%freqdata1(order=10,  invar=ANTIBODY_CHAR     , level=("4"),popvar=, totalvl=null);
%meandata1(order=10.5, invar=GAD_NUM          , roundvar=.001, digit=3);
%meandata1(order=11,  invar=INFUSE_DAYS       , roundvar=.1, digit=1);
%rangedata1(order=11.5, invar=INFUSE_DAYS, roundvar=1, digit=0);
%meandata1(order=12,  invar=weightkg          , roundvar=.1, digit=1);
%meandata1(order=13,  invar=bmi               , roundvar=.01, digit=2);
%meandata1(order=13.5, invar=AUC_RESULT       , roundvar=.001, digit=3);
%meandata1(order=14,  invar=GLYC_HEMO , roundvar=.001, digit=3);
%meandata1(order=15,  invar=total_insulin_kg      , roundvar=.001, digit=3);
%freqdata1(order=17,  invar=DKAatDiagnosis , level=("Yes"),popvar=, totalvl=null);
%freqdata1(order=18,  invar=DR_NET   , level=("0"),popvar=, totalvl=null);
%freqdata1(order=19,  invar=DR_NET   , level=("1"),popvar=, totalvl=null);
%freqdata1(order=20,  invar=DR_NET   , level=("2"),popvar=, totalvl=null);
%freqdata1(order=21,  invar=DR_NET   , level=("3"),popvar=, totalvl=null);


data accumtab1;
   set accumfreq1 accummean1 accummedian1;
   if treatmentname=" " then delete;

proc sort data=accumtab1;
   by treatmentname orderer;

proc print data=accumtab1 noobs;
   by treatmentname;
   format treatmentname $treatf.;
   where treatmentname='3 injections of Alum only';
   title3 'Table 1 stats (list)';

proc print data=accumtab1 noobs;
   by treatmentname;
   format treatmentname $treatf.;
   where treatmentname='2 injections of 20?g GAD-alum followed by one injection with Alum alone';
   title3 'Table 1 stats (list)';
```

```
proc print data=accumtab1 noobs;
  by treatmentname;
  format treatmentname $treatf.;
  where treatmentname='3 injections with 20?g GAD-alum';
  title3 'Table 1 stats (list)';

data table2;
  set table2;
  length treatmentnum 8.;
  if treatmentname='3 injections of Alum only' then treatmentnum=0;
  else if treatmentname='2 injections of 20?g GAD-alum followed by one injection with Alum alone' then treatmentnum=1;
  else if treatmentname='3 injections with 20?g GAD-alum' then treatmentnum=2;
  else abort;
  if Category=" " then delete;

proc sort data=table2;
  by maskid severity_index;

data table2_short;
  set table2;
  by maskid severity_index;
  if severity_index>1 then output;

proc freq data=table2_short;
  tables Category*treatmentnum/missing norow nocum nocol;
  title3 'Table 2 part 2: Events where severity>1.';
  format treatmentnum treatf.;

data table2_person;
  set table2;
  by maskid severity_index;
  if last.maskid;

proc freq data=table2_person;
  tables Category*treatmentnum/missing norow nocum nocol;
  title3 'Table 2 part 2: one record per maskid';
  format treatmentnum treatf.;
```

20