Integrity Check for The Diabetes Prevention Type 1 (DPT-1) Files

As a partial check of the integrity of the DPT-1 datasets archived in the NIDDK data repository, a set of tabulations was performed to verify that published results from the DPT-1 study can be reproduced using the archived datasets. A small number of analyses were performed to duplicate published results for the data reported by the Type 1 Diabetes Study Group [1] in NEJM in May, 2002 . The results of this integrity check are described below. The full text of the NEJM article can be found in Attachment 1, and the SAS code for our tabulations is included in Attachment 2.

All sample sizes, demographic data, baseline measurements and corresponding p-values were successfully replicated; however, there are discrepancies for the number person-years of follow-up and the number of hypoglycemic episodes. One reason for this may be that published results are from the Parenteral Study, which ended in 2001, but the datasets archived in the NIDDK data repository contain data from the follow-up continued on a subset of the Parenteral subjects. The DCC reviewed the methodology used in the replication and stated that it appears to be correct. The DCC has also stated that all published results were independently confirmed.

Data and Structures. In the data structure of the DPT-1 study, there is no single baseline dataset. The following datasets were used for the integrity check of the baseline table (Table 1. Baseline Characteristics of the 339 Randomized Subjects).

SUBJECT	Subject Table (contains race, gender, treatment assignment, etc.)
TEST	Test Table (contains all test results for DPT-1)
RELATIVE	Relative Table (contains information on subject's probands)
CODETABLE	Code groups for all coded variables

Sample size was obtained by restricting the SUBJECT dataset to observations with a value of 'Intervention' or 'Observation' for the *treatment* variable.

Age in years was calculated as the value of the *age_at_randomization* variable (SUBJECT dataset) divided by 365.25.

First-phase insulin response was obtained by adding together the insulin values at one and three minutes (TEST dataset, *ins1* and *ins3* variables, where the *testname* variable has a value of 'AB-IVGTT' or 'CO-IVGTT'). Insulin values were transformed from character to numeric, with values of the form '<4', '<3', etc. recoded to zero. The definition of first-phase insulin response is documented in the Methods: Intravenous Glucose-Tolerance Test section of the Skyler article.

Impaired glucose tolerance was defined as a glucose level of 140 to 199 mg/dl at 120 minutes (TEST dataset, *glu120* variable, where the *testname* variable has a value of 'MMTT' or 'OGTT'). Glucose values were transformed from character to numeric. The definitions of impaired and indeterminate glucose tolerance are documented in the Methods: Oral Glucose-Tolerance Test section of the Skyler article. Indeterminate glucose tolerance was not calculated, but frequencies for impaired glucose tolerance match the reported frequencies for impaired/indeterminate glucose tolerance. The DCC has since indicated that no single field exists to capture Fasting Plasma Glucose (FPG). This information is captured in the TEST dataset, variables *glum10* (Glucose Result -10 Minutes), *glum4* (Glucose Results -4 Minutes), and *gluc0* (Glucose Result: 0 Minutes) for *testname* values of 'MMTT', 'OGTT', or 'IVGTT'. The FPG was then defined as the maximum value of the available results (depending on the test). Additionally, if the test was a simple FPG, it was captured in the *result* variable field.

Race/Ethnicity and sex variables are found in the SUBJECT dataset (variable names *race* and *gender*).

The DCC states that index proband was not tracked. Basic proband information was collected on the screening form and stored in the RELATIVE dataset. For purposes of this replication, index patient was defined to be the proband with the closest relationship to the randomized subject (ordered as sibling, parent, offspring, second degree relative). The RELATIVE dataset, which contains data for all probands, was merged with the SUBJECT dataset. The *reltype* variable (RELATIVE dataset) was collapsed into categories of sibling, parent, offspring and second degree relative. This method of determining the index patient gave the closest match to published results.

Values for islet-cell antibodies were obtained from the *result* variable of the TEST dataset, where the *testname* variable has a value of 'ICA'. ICA values were transformed from character to numeric.

Values for insulin antibodies were obtained from the *result* variable of the TEST dataset, where the *testname* variable has a value of 'IAA'. IAA values were transformed from character to numeric.

Subjects may have multiple baseline observations. To identify a single baseline record for each subject, all observations with a reported age less than or equal to the *age_at_randomization* variable (SUBJECT dataset) were kept (Note: study timepoints are identified by subject's age in days). Datasets were then sorted by ID and age. The last observation (i.e., the observation occurring closest to the day of randomization but not after randomization) was then kept as the baseline observation for a subject.

Table 1: Comparison of Baseline Table Values Computed in Integrity Check to Reference Article Values

Variable	Intervention Group			Observation Group		
	Skyler et al (2002)	Integrity check	Difference	Skyler et al (2002)	Integrity check	Difference
Sample Size	169	169	0	170	170	0
Age (yr):						
Median	11.9	11.9	0	12.1	12.1	0
Interquartile range	7.1 – 16.7	8.5 – 18.1	1.4 – 1.4	7.6 – 16.6	8.7 – 17.8	1.1 – 1.2
First-phase insulin response (µU/ml):						
Mean ± SD	70.8 ± 40.4	70.8 ± 40.4	0	72.8 ± 37.1	72.8 ± 37.1	0
Impaired/Indeterminate glucose tolerance:						
N (%)	58 (34.3)	58 (34.3)	0	54 (31.8)	54 (31.8)	0
Race or ethnic group - N (%):						
Non-hispanic white	159 (94.1)	159 (94.1)	0	158 (92.9)	158 (92.9)	0
Non-hispanic black	1 (0.6)	1 (0.6)	0	1 (0.6)	1 (0.6)	0
Hispanic	4 (2.4)	4 (2.4)	0	5 (2.9)	5 (2.9)	0
Other	5 (3.0)	5 (3.0)	0	6 (3.5)	6 (3.5)	0
Sex - N (%):						
Male	87 (51.5)	87 (51.5)	0	89 (52.4)	89 (52.4)	0
Female	82 (48.5)	82 (48.5)	0	81 (47.6)	81 (47.7)	0 (0.1)
Relationship to patient w/ diabetes – N (%):						
Sibling	100 (59.2)	100 (59.2)	0	112 (65.9)	113 (66.7)	1 (0.8)
Parent	47 (27.8)	47 (27.8)	0	40 (23.5)	39 (22.9)	1 (0.6)
Offspring	8 (4.7)	8 (4.7)	0	5 (2.9)	5 (2.9)	0
Second degree relative	14 (8.3)	14 (8.3)	0	13 (7.6)	13 (7.7)	0 (0.1)
Islet cell antibodies – JDF units:						
Median	160	160	0	160	160	0
Interquartile range	40 – 320	40 – 320	0	40 – 320	40 – 320	0
Insulin antibodies (nU/ml):						
Mean ± SD	365 ± 641	365 ± 640	0 ± 1	295 ± 552	295 ± 552	0

The following datasets were used for the integrity check of the follow-up table (Table 2. Hypoglycemic Episodes During Follow-Up).

SUBJECT	Subject Table (contains race, gender, treatment assignment, etc.)
DN	Death Notice
LF	Lost to Follow-up
GP	Home Glucose Monitoring Form
AE	Adverse Reporting Form
HY	Hypoglycemia Reporting Form
MH	Medical History
IF	Insulin Infusion Flow-Sheet
CODETABLE	Code groups for all coded variables

Sample size was obtained by restricting the SUBJECT dataset to observations with a value of 'Intervention' or 'Observation' for the *treatment* variable.

Number of person-years of follow-up was calculated as the sum person-years of follow-up across all subjects. A subject's last date on study was taken to be the earliest date among date of death (DN dataset, *age_at_death* variable), date of diagnosis of diabetes (primary study end point; (SUBJECT dataset, *age_at_iddm_diagnosis* variable)), date lost to follow-up (LF dataset, *age_at_last_contact* variable), and date of study end (SUBJECT dataset, *age_at_last_contact_questionnair, age_at_last_contact_from_forms, age_at_last_dpt1_test* variables). The subject's age at randomization (SUBJECT dataset, *age_at_randomization* variable) was then subtracted from the subject's age on this last date of study; all divided by 365.25.

Number of hypoglycemic episodes detected by quarterly glucose profiles was calculated from the GP dataset, variables *bldglu1, bldglu2, bldglu3, bldglu4* and *bldglu5*. If any of the 5 blood glucose measurements were less than 50 mg/dl, the subject was counted as having a hypoglycemic episode. Throughout the Skyler article, blood glucose measurements less than 50 mg/dl are defined as hypoglycemic.

Number of spontaneously reported hypoglycemic episodes (presumed & definite) was examined using the AE dataset (variables *event1, event2, event3, event4, event5, event6, event7, age_at_onset1, age_at_onset2, age_at_onset3, age_at_onset4, age_at_onset5, age_at_onset6,* and *age_at_onset7*); the HY dataset (variables *bldsg* and *symsweet*); and the MH dataset (variables *numhypo, age_hypoglycemia1, age_hypoglycemia2, age_hypoglycemia3, age_hypoglycemia4, age_hypoglycemia5, and age_hypoglycemia6*).

In the AE data, an episode was counted as 'presumed' if the event code had a value of 30. A value of 40 was counted as a 'definite' episode. These values are found in the CODEBOOK dataset.

In the HY data, an episode was defined as 'presumed' if typical symptoms resolved with the intake of food. 'Definite' episodes were defined as a blood glucose concentration of less than 50 mg/dl. These definitions of 'presumed' and 'definite' hypoglycemic episodes are taken from the Methods: Follow-up Assessments section of the Skyler article.

Since the AE and HY datasets did not produce identical counts and since the HY form states that all self reporting of presumed/definite hypoglycemia was to be recorded on both the HY and AE forms, the datasets were combined and episodes were counted as follows: If both forms agreed that an episode was 'presumed', it was counted as 'presumed'; if both forms agreed that an episode was 'definite', it was counted as 'definite'; if the forms disagreed, the episode was counted as 'presumed'.

The MH dataset makes no distinction between 'presumed' and 'definite' episodes. Therefore, for purposes of this replication, all episodes were counted as 'presumed'. Additionally, any number of hypoglycemic

episodes can be reported on the MH form, but no more than 6 event dates can be reported. Therefore, it is not clear how to determine if an episode reported on the MH form is the same episode reported on an AE and/or HY form.

Since multiple sources of hypoglycemic episodes were collected, each episode was cross-referenced with the others to eliminate duplicates. While we eliminated duplicates we could identify, it may be that unresolved duplicates (as well as the fact that this replication analysis uses files containing follow-up and corrected data) are responsible for the remaining discrepancies.

Finally, the IF dataset (intrrpt1, intrrpt2, intrrpt3, intrrpt4, intrrpt5, age_at_infusion_date1, age_at_infusion_date2, age_at_infusion_date3, age_at_infusion_date4, and age_at_infusion_date5 variables) was used to identify episodes of hypoglycemia occurring during an intravenous infusion of insulin. These cases were to be excluded from the number of spontaneously reported hypoglycemic episodes (presumed & definite), per table footnote. For cases of hypoglycemia occurring during an infusion, there were no instances where the infusion date matched with a date on the AE, HY or MH forms. Therefore, no cases were excluded for this reason.

Table 2: Comparison of Values for Hypoglycemic Episodes During Follow-Up Computed in Integrity Check to Reference Article Values

Variable	Intervention Group			Observation Group		
	Skyler et al (2002)	Integrity check	Difference	Skyler et al (2002)	Integrity check	Difference:
Sample Size	169	169	0	170	170	0
No. of person-yrs of follow-up:	438.7	460.5	21.8	428.1	428.8	0.7
Hypoglycemia detected by quarterly glucose profiles (# episodes):	32	20	12	32	15	17
Spontaneously reported hypoglycemia:						
Presumed (# episodes):	591	522	69	243	193	50
Definite (# episodes):	59	61	2	11	15	4

Note: Rates per100 person-years are not presented as part of the integrity check. Since the sum total for person years of follow-up does not match, the rates will not match either. Likewise, p-values are not presented since the counts differ.

Notes

1. Analysis was done using the copy of the database provided to the NIDDK Data Repository. The SAS datasets used were created on July 27, 2005. 11 of the 31 datasets are examined in this analysis. Of the 20 datasets not examined, 5 of them are related to additional reviews of study data and/or an ancillary study.

2. The SAS datasets provided to the NIDDK Data Repository are in an archival format. In order to use SAS Viewer, limit CPU resources and increase performance when using these datasets, they must be converted back to an un-archived state. One method to do this is via PROC MIGRATE, as follows:

   ```
   /* Location of Archived DPT-1 SAS Data Files */
   LIBNAME OLD 'W:\Databases\DPT-1 data\DPT-1_7-27-2005\SAS Data Files';

   /* Location for Un-archived DPT-1 SAS Data Files */
   LIBNAME NEW 'W:\Project data files\DPT-1';

   /* Migrate the datasets */
   PROC MIGRATE IN=OLD OUT=NEW; RUN;
   ```

   Un-archived versions of all the archived datasets in the 'OLD' location will then be created in the 'NEW' location.

Reference

1. Diabetes Prevention Trial – Type 1 Diabetes Study Group (Skyler et al). Effects of insulin in relatives of patients with type 1 diabetes mellitus. **New England Journal of Medicine** 2002; 346: 1685-1691.

ATTACHMENT 1

"The full text of the article referenced will be provided to approved requestors along with the data archive."

Diabetes Prevention Trial – Type 1 Diabetes Study Group. Effects Of Insulin in Relatives of Patients with Type 1 Diabetes Mellitus. New England Journal of Medicine, 2002 May 30;346(22):1685-1691b.

ATTACHMENT 2

SAS Code for Baseline and Follow-Up Tabulations from DPT-1 Datasets in the NIDDK Repository

```
/**********************************************************************************/
/*
/* Program: R:\Norma\DPT-1\NEJM\getdata.sas
/* Author:  Norma Pugh
/* Date:    14 February 06
/* Purpose: Datasets for NEJM replication.
/*
/**********************************************************************************/
/* Location of NIDDK Repository SAS files */
libname data 'r:\databases\dpt-1 data\dpt-1_7-27-2005\sas data files';

/* Location of SAS files created for this replication */
libname out  'r:\Norma\DPT-1\NEJM';

/* Randomized subjects */
data out.subjects;
 set data.subject;
 if treatment in('Intervention','Observation');
 randage=age_at_randomization/365.25;
 keep maskid treatment race gender age_at_randomization randage
      age_at_iddm_diagnosis
      age_at_last_contact_from_forms age_of_last_contact_questionnair
age_at_last_dpt1_test;
run;

/* Death notices */
data out.dn;
 set data.dn;
 keep maskid age_at_death;
run;

/* Lost to follow-up */
data out.lf;
 set data.lf;
 keep maskid age_at_last_contact;
run;

/* Follow-up */
data out.followup;
 set data.followup;
 keep maskid age_at_date_questionnaire_rcvd;

/* Test datasets: First-phase insulin response, Glucose tolerance, ICA, IAA */
data out.fpir(keep=maskid age_at_draw testname ins1 ins3)
     out.glucose(keep=maskid age_at_draw testname glu30 glu60 glu90 glu120)
     out.ica(keep=maskid age_at_draw testname result outcome)
     out.iaa(keep=maskid age_at_draw testname result outcome);
 set data.test(keep=maskid age_at_draw testname result outcome ins1 ins3 glu30 glu60
glu90 glu120);
 if testname='ICA' then output out.ica;
 if testname='IAA' then output out.iaa;
 if testname in('AB-IVGTT','CO-IVGTT') then output out.fpir;
 if testname in('MMTT','OGTT') then output out.glucose;
run;

/* Home glucose monitoring */
data out.gp;
```

```
 set data.gp;
 keep maskid age_at_form_date age_at_date_call age_when_home_monitoring_began bldglu1
bldglu2 bldglu3 bldglu4 bldglu5;
run;

/* Adverse events */
data out.ae;
 set data.ae;
 keep maskid event1 event2 event3 event4 event5 event6 event7 age_at_onset1 age_at_onset2
age_at_onset3 age_at_onset4
      age_at_onset5 age_at_onset6 age_at_onset7;
run;

/* Insulin infusion flow-sheet */
data out.if;
 set data.if;
 keep maskid age_at_infusion_date1 age_at_infusion_date2 age_at_infusion_date3
age_at_infusion_date4 age_at_infusion_date5
                   intrrpt1 intrrpt2 intrrpt3 intrrpt4 intrrpt5;
run;

/* Hypoglycemia Reporting */
data out.hy;
 set data.hy;
 keep maskid age_at_occurence bldsg symsweet;
run;
```

```
/********************************************************************************/
/*
/* Program: R:\Norma\DPT-1\NEJM\table1.sas
/* Author:  Norma Pugh
/* Date:    14 February 06
/* Purpose: NEJM baseline replication (Table 1).
/*
/********************************************************************************/
/* Location of SAS files created for this replication */
libname data 'r:\norma\dpt-1\nejm';

/* Location of NIDDK Repository SAS files */
libname master 'r:\databases\dpt-1 data\sas data files';


/************************************/
/* Age: median, interquartile range */
/************************************/
proc sort data=data.subjects out=subjects; by treatment; run;
proc univariate data=subjects; by treatment; var randage; title'Age Results'; run;


/**********************************************************************/
/* First phase insulin response - Add insulin values at 1 and 3 minutes */
/**********************************************************************/
proc sort data=data.fpir out=fpir; by maskid age_at_draw; run;
proc sort data=subjects; by maskid; run;

data fpir;
 merge fpir subjects(in=keep);
 by maskid;
 if keep & age_at_draw<=age_at_randomization;      /* Keep all baseline records */
run;

data fpir;
 set fpir;
 by maskid age_at_draw;
 if last.maskid;
                                                   /* Keep final baseline record */
 if substr(ins1,1,1)='<' then I1=O; else I1=ins1+O; /* Convert to numeric. Recode '<'
values to O. */
 if substr(ins3,1,1)='<' then I3=O; else I3=ins3+O;
 fpir=I1+I3;
run;

proc sort data=fpir; by treatment; run;

proc means data=fpir; by treatment; var fpir; title'First-phase insulin response
Results'; run;


/*****************************************/
/* Impaired/Indeterminate glucose result */
/*****************************************/
proc sort data=data.glucose out=glucose; by maskid age_at_draw; run;

data glucose;
```

```
 merge glucose subjects(in=keep);
 by maskid;
 if keep & age_at_draw<=age_at_randomization;        /* Keep all baseline records */
run;

data glucose;
 set glucose;
 by maskid age_at_draw;
 if last.maskid;
                                                     /* Keep final baseline record */
 G30=glu30+0;
                                                     /* Convert to numeric */
 G60=glu60+0;
 G90=glu90+0;
 G120=glu120+0;
 if 140<=g120<=199 then impair=1; else impair=0;
run;

proc sort data=glucose; by treatment; run;

proc freq data=glucose; by treatment; tables impair; title'Impaired/Indeterminate glucose
tolerance Results'; run;


/*****************************/
/* Race & Gender frequencies */
/*****************************/
proc sort data=subjects; by treatment; run;
proc freq data=subjects; by treatment; tables race gender; title'Race/Gender Results';
run;


/********************************/
/* Relationship to index patient */
/********************************/
proc sort data=master.relative out=relative(keep=maskid reltype qual); by maskid; run;
proc sort data=subjects; by maskid; run;

proc freq data=relative; tables reltype; title'Check values of reltype';  run;

data relation;
 merge relative subjects(in=keep);
 by maskid;
 if keep;
 if reltype in('B','IT','SIB','SIS','TWIN') then relation='1_sibling';     /* Categorize
relatives */
 else if reltype in('F','M') then relation='2_parent';
 else if reltype in('D','SON') then relation='3_offsprg';
 else if reltype
in('A','C','FC','GA','GF','GGF','GM','GP','GU','MA','MGA','MGM','MU','NE','PA','PGF','PGM
','PU','SC','U') then relation='4_2nd_deg';
run;

proc sort data=relation; by maskid relation; run;
        /* Keep closest proband */
data relation; set relation; by maskid relation; if first.maskid; run;
```

```
proc sort data=relation; by treatment; run;
proc freq data=relation; by treatment; tables relation; title'Relationship to index
patient Results'; run;


/***************/
/* Baseline ICA */
/***************/
proc sort data=data.ica out=ica; by maskid age_at_draw; run;

data ica;
 merge ica subjects(in=keep);
 by maskid;
 if keep & age_at_draw<=age_at_randomization;        /* Keep all baseline records */
run;

data ica;
 set ica;
 by maskid age_at_draw;
 if last.maskid;
                                                     /* Keep final baseline record */
 icavalue=result+0;
                                                     /* Convert to numeric */
run;

proc sort data=ica; by treatment; run;

proc univariate data=ica; by treatment; var icavalue; title'ICA Results'; run;

/***************/
/* Baseline IAA */
/***************/
proc sort data=data.iaa out=iaa; by maskid age_at_draw; run;

data iaa;
 merge iaa subjects(in=keep);
 by maskid;
 if keep & age_at_draw<=age_at_randomization;        /* Keep all baseline records */
run;

data iaa;
 set iaa;
 by maskid age_at_draw;
 if last.maskid;
                                                     /* Keep final baseline record */
 iaavalue=result+0;
                                                     /* Convert to numeric */
run;

proc sort data=iaa; by treatment; run;

proc means data=iaa; by treatment; var iaavalue; title'IAA Results'; run;
```

```
/*****************************************************************************/
/*
/* Program: R:\Norma\DPT-1\NEJM\table2.sas
/* Author:  Norma Pugh
/* Date:    19 February 06
/* Purpose: NEJM baseline replication (Table 2).
/*
/*****************************************************************************/
/* Location of SAS files created for this replication */
libname data 'r:\norma\dpt-1\nejm';


/*****************************/
/* Person-years of follow-up */
/*****************************/
proc sort data=data.subjects out=subjects; by maskid; run;
proc sort data=data.dn out=dn; by maskid; run;
proc sort data=data.lf out=lf; by maskid; run;

data yrs_fu;
 merge subjects(in=keep) dn lf;
 by maskid;
 if keep;
 /* last date = earliest of death, diagnosis, lost to follow-up, end of study */
 last_dt=min(age_at_death, age_at_iddm_diagnosis, age_at_last_contact,
age_of_last_contact_questionnair, age_at_last_contact_from_forms, age_at_last_dpt1_test);
 last_yrs=(last_dt-age_at_randomization)/365.25;
run;

proc sort data=yrs_fu; by treatment; run;
proc univariate data=yrs_fu; by treatment; var last_yrs; title'Person-years of follow-up
Results'; run;


/*****************************/
/* Quarterly glucose profiles */
/*****************************/
proc sort data=data.gp out=gp; by maskid; run;

data quarterly;
 merge subjects(in=keep keep=maskid treatment age_at_randomization) gp;
 by maskid;
 if keep;
 if bldglu1=. & bldglu2=. & bldglu3=. & bldglu4=. & bldglu5=. then hypo=.;
  else if .<bldglu1<50 or .<bldglu2<50 or .<bldglu3<50 or .<bldglu4<50 or .<bldglu5<50
then hypo=1;
  else hypo=0;
run;

proc freq data=quarterly; tables treatment*hypo / chisq; title'Hypoglycemia: quarterly
glucose profiles'; run;


/*************************************************************/
/* Spontaneous Hypoglycemia: Presumed or Definite (from AE form) */
/*************************************************************/
proc sort data=data.ae out=ae; by maskid; run;
```

```
/* Get cases */
%macro ae(num);
 data ae&num(keep=maskid hypo hypoDT);
  set ae;
  if event&num in(30,40) then do;
   output ae&num;
   rename event&num=hypo age_at_onset&num=hypoDT;
  end;
 run;
%mend ae;

%ae(1); %ae(2); %ae(3); %ae(4); %ae(5); %ae(6); %ae(7);

data aes;
 set ae1 ae2 ae3 ae4 ae5 ae6 ae7;
 by maskid;
run;

data aes;
 merge subjects(in=keep) aes;
 by maskid;
 if keep;
 presumed='n'; definite='n';
 if hypo=30 then presumed='y';
 if hypo=40 then definite='y';
run;

proc freq data=aes; tables treatment*presumed / chisq; title'Hypoglycemia: Presumed(30) -
AE Form only'; run;
proc freq data=aes; tables treatment*definite / chisq; title'Hypoglycemia: Definite(40) -
AE Form only'; run;


/****************************************************************/
/* Spontaneous Hypoglycemia: Presumed or Definite (from HY form) */
/****************************************************************/
proc sort data=data.hy out=hy; by maskid; run;

data hy;
 merge subjects(in=keep) hy;
 by maskid;
 if keep;
 presumed='n'; definite='n';
 if .<bldsg<50 then definite='y';
 if not(.<bldsg<50) & symsweet=2 then presumed='y';
run;

proc freq data=hy; tables treatment*presumed / chisq; title'Hypoglycemia: Presumed - HY
Form only'; run;
proc freq data=hy; tables treatment*definite / chisq; title'Hypoglycemia: Definite - HY
Form only'; run;


/*********************************************************************/
/* Spontaneous Hypoglycemia: Presumed or Definite (from AE & HY forms) */
/*********************************************************************/
```

```
data aes(keep=maskid treatment presumed_ae definite_ae hypoDt); set aes(keep=maskid
treatment presumed definite hypoDt); rename presumed=presumed_ae definite=definite_ae;
run;
data hy(keep=maskid treatment presumed_hy definite_hy hypoDt);  set hy(keep=maskid
treatment presumed definite age_at_occurence);    rename presumed=presumed_hy
definite=definite_hy age_at_occurence=hypoDt; run;

proc sort data=aes; by maskid treatment hypoDt; run;
proc sort data=hy; by maskid treatment hypoDt; run;

data all;
 merge aes hy;
 by maskid treatment hypoDt;
 presumed='n'; definite='n';
 /* Concordant */
 if presumed_ae='y' & presumed_hy='y' then presumed='y';
 if definite_ae='y' & definite_hy='y' then definite='y';
 /* Discordant */
 if (presumed_ae='y' & presumed_hy='n') or
    (presumed_ae='n' & presumed_hy='y') or
    (definite_ae='y' & definite_hy='n') or
    (definite_ae='n' & definite_hy='y') then presumed='y';
run;

proc freq data=all; tables treatment*presumed / chisq; title'Hypoglycemia: Presumed - AE
& HY Forms'; run;
proc freq data=all; tables treatment*definite / chisq; title'Hypoglycemia: Definite - AE
& HY Forms'; run;


/***********************************************************************************/
/* Isolate hypoglycemic episodes occuring during an infusion - these will be deleted */
/***********************************************************************************/
proc sort data=data.if out=if; by maskid; run;

%macro todelete(num);
 data if&num._todelete(keep=maskid hypoDT);
  set if;
  if intrrpt&num=1 then do;
   output if&num._todelete;
   rename age_at_infusion_date&num=hypoDT;
  end;
 run;

 proc sort data=if&num._todelete; by maskid hypoDT; run;
%mend todelete;

%todelete(1); %todelete(2); %todelete(3); %todelete(4); %todelete(5);

data if_todelete;
 set if1_todelete if2_todelete if3_todelete if4_todelete if5_todelete;
 by maskid hypoDT;
run;

/* See .log file - check # observations after deletions */
proc sort data=aes; by maskid hypoDT; run;
proc sort data=hy; by maskid hypoDT; run;
```

```
proc sort data=if_todelete; by maskid hypoDT; run;

data tryae; merge aes(in=x1) if_todelete(in=x2); by maskid hypoDT; if x2 then delete;
run;
data tryhy; merge hy(in=x1) if_todelete(in=x2); by maskid hypoDT; if x2 then delete; run;
/* No deletions happened... */


/*********************************************************************************/
/*
/* Program: R:\Norma\DPT-1\NEJM\table2_update.sas
/* Author:  Norma Pugh
/* Date:    30 August 06
/* Purpose: NEJM replication (Table 2 - Hypoglycemic Episodes).
/*
/*********************************************************************************/
/* Location of NIDDK Repository SAS files */
libname master 'r:\03_Data_And_Tools\Database\databases\dpt-1 data\dpt-1_7-27-2005\sas
data files';

/* Location of SAS files created for this replication */
libname data   'r:\05_Users\Norma\DPT-1\NEJM';


/************/
/* Subjects */
/************/
proc sort data=data.subjects out=subjects; by maskid; run;


/*****************************************************************/
/* Spontaneous Hypoglycemia: Presumed or Definite (from MH form) */
/*****************************************************************/
proc sort data=master.mh out=mh; by maskid; run;

data medhx;
 merge subjects(in=keep) mh;
 by maskid;
 if keep;
 if numhypo>0;
 keep maskid treatment numhypo;
run;

proc transpose data=medhx out=tr_medhx(drop=_name_ _label_);
 by maskid;
 var numhypo;
 copy treatment;
run;

/* MH dataset with # episodes only */
data mh_counts;
 set tr_medhx;
 mh_presumed=sum(of col1-col6);
 if mh_presumed>0;
 keep maskid treatment mh_presumed;
run;

/* Get dates */
```

```
%macro mh(num);
 data mhdate&num(keep=maskid treatment hypoDT);
  merge subjects(in=keep) mh;
  by maskid;
  if keep;
  if numhypo>0;
  rename Age_Hypoglycemia&num = hypoDT;
 run;
%mend mh;

%mh(1); %mh(2); %mh(3); %mh(4); %mh(5); %mh(6);

data mhdates;
 set mhdate1 mhdate2 mhdate3 mhdate4 mhdate5 mhdate6;
 by maskid;
 if hypodt=. then delete;
run;

/* MH dataset with # episodes and dates */
data mh_counts_dates;
 merge mh_counts mhdates;
 by maskid;
run;


/****************************************************************/
/* Spontaneous Hypoglycemia: Presumed or Definite (from AE form) */
/****************************************************************/
proc sort data=data.ae out=ae; by maskid; run;

/* Get cases */
%macro ae(num);
 data ae&num(keep=maskid hypo hypoDT);
  set ae;
  if event&num in(30,40) then do;
   output ae&num;
   rename event&num=hypo age_at_onset&num=hypoDT;
  end;
 run;
%mend ae;

%ae(1); %ae(2); %ae(3); %ae(4); %ae(5); %ae(6); %ae(7);

data aes;
 set ae1 ae2 ae3 ae4 ae5 ae6 ae7;
 by maskid;
run;

data aes;
 merge subjects(in=keep) aes;
 by maskid;
 if keep;
 presumed='n'; definite='n';
 if hypo=30 then presumed='y';
 if hypo=40 then definite='y';
run;
```

```
/*****************************************************************/
/* Spontaneous Hypoglycemia: Presumed or Definite (from HY form) */
/*****************************************************************/
proc sort data=data.hy out=hy; by maskid; run;

data hy;
 merge subjects(in=keep) hy;
 by maskid;
 if keep;
 presumed='n'; definite='n';
 if .<bldsg<50 then definite='y';
 if not(.<bldsg<50) & symsweet=2 then presumed='y';
run;


/*********************************************************************/
/* Spontaneous Hypoglycemia: Presumed or Definite (from AE & HY forms) */
/*********************************************************************/
data aes(keep=maskid treatment presumed_ae definite_ae hypoDt); set aes(keep=maskid
treatment presumed definite hypoDt); rename presumed=presumed_ae definite=definite_ae;
run;
data hy(keep=maskid treatment presumed_hy definite_hy hypoDt);  set hy(keep=maskid
treatment presumed definite age_at_occurence); rename presumed=presumed_hy
definite=definite_hy age_at_occurence=hypoDt; run;

proc sort data=aes; by maskid treatment hypoDt; run;
proc sort data=hy; by maskid treatment hypoDt; run;

/* Delete cases counted on both forms (AE and HY) */
data nodups;
 merge aes hy;
 by maskid treatment hypodt;
 if last.hypodt;
 if hypodt=. then delete;     /* episode = no */
run;

/* AE/HY dataset with # episodes and dates */
data ae_hy_counts_dates;
 set nodups;
 presumed=0; definite=0;
 /* Concordant */
 if presumed_ae='y' & presumed_hy='y' then presumed=1;
 if definite_ae='y' & definite_hy='y' then definite=1;
 /* Discordant */
 if (presumed_ae='y' & presumed_hy='n') or
    (presumed_ae='n' & presumed_hy='y') or
    (definite_ae='y' & definite_hy='n') or
    (definite_ae='n' & definite_hy='y') then presumed=1;
 /* Missing */
 if (presumed_ae='y' & presumed_hy='') or
    (presumed_ae='' & presumed_hy='y') then presumed=1;
 if (definite_ae='y' & definite_hy='') or
    (definite_ae='' & definite_hy='y') then definite=1;
run;

/* AE/HY datset with # episodes only */
```

```
proc transpose data=ae_hy_counts_dates out=temp1 prefix=presumed;
 by maskid;
 var presumed;
 copy treatment;
run;
proc transpose data=ae_hy_counts_dates out=temp2 prefix=definite;
 by maskid;
 var definite;
 copy treatment;
run;
/*
proc contents data=temp1; title'Check number of variables'; run;
proc contents data=temp2; title'Check number of variables'; run;
*/
data temp1; set temp1; ae_hy_presumed=sum(of presumed1-presumed28); drop _name_
presumed1-presumed28; run;
data temp2; set temp2; ae_hy_definite=sum(of definite1-definite28); drop _name_
definite1-definite28; run;

data ae_hy_counts; merge temp1 temp2; by maskid; if ae_hy_presumed=. & ae_hy_definite=.
then delete; run;


/***********************************/
/* Get duplicate dates: MH vs AE/HY */
/***********************************/
data temp_mh;
 set mh_counts_dates;
 mhdt=hypodt;
 keep maskid hypodt mhdt;
run;

data temp_ae_hy;
 set ae_hy_counts_dates;
 keep maskid hypodt;
run;

proc sort data=temp_mh; by maskid hypodt; run;
proc sort data=temp_ae_hy; by maskid hypodt; run;

data temp_dates;
 merge temp_mh temp_ae_hy;
 by maskid hypodt;
 if hypodt=mhdt then dup=1; else dup=0;
run;

proc transpose data=temp_dates out=dupdates prefix=dupcount;
 by maskid;
 var dup;
run;
/*
proc contents data=dupdates; title'Check number of variables'; run;
*/
data dupdates; set dupdates; numdups=sum(of dupcount1-dupcount30); drop _name_ dupcount1-
dupcount30; run;
```

```
/**************************************/
/* Merge all data and get final counts */
/**************************************/
data all;
 merge mh_counts ae_hy_counts dupdates;
 by maskid;
 if mh_presumed=.    then mh_presumed=0;
 if ae_hy_presumed=. then ae_hy_presumed=0;
 if ae_hy_definite=. then ae_hy_definite=0;
 if numdups=.        then numdups=0;
run;

data all;
 set all;
 presumed=mh_presumed+ae_hy_presumed-numdups;
 definite=ae_hy_definite;
run;

proc means data=all n sum min max;
 class treatment;
 var presumed definite;
 title2'Hypoglycemia: Presumed & Definite - AE/HY/MH Forms';
run;


/*****************************************************************************************/
/* Isolate hypoglycemic episodes occuring during an infusion - these will be deleted */
/*****************************************************************************************/
proc sort data=data.if out=if; by maskid; run;

%macro todelete(num);
 data if&num._todelete(keep=maskid hypoDT);
  set if;
  if intrrpt&num=1 then do;
   output if&num._todelete;
   rename age_at_infusion_date&num=hypoDT;
  end;
 run;

 proc sort data=if&num._todelete; by maskid hypoDT; run;
%mend todelete;

%todelete(1); %todelete(2); %todelete(3); %todelete(4); %todelete(5);

data if_todelete;
 set if1_todelete if2_todelete if3_todelete if4_todelete if5_todelete;
 by maskid hypoDT;
run;

/* See .log file - check # observations after deletions */
proc sort data=aes; by maskid hypoDT; run;
proc sort data=hy; by maskid hypoDT; run;
proc sort data=mhdates; by maskid hypoDT; run;
proc sort data=if_todelete; by maskid hypoDT; run;

data tryae; merge aes(in=x1) if_todelete(in=x2); by maskid hypoDT; if x2 then delete;
run;
```

```
data tryhy; merge hy(in=x1) if_todelete(in=x2); by maskid hypoDT; if x2 then delete; run;
data trymh; merge mhdates(in=x1) if_todelete(in=x2); by maskid hypoDT; if x2 then delete;
run;
/* No deletions happened... */
```