Dataset Integrity Check for the TODAY Data Files

Prepared by Corey Del Vecchio
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton MD 20705
January 7, 2014

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

The TODAY study tested the hypothesis that an aggressive approach to reducing insulin resistance early in the course of T2DM would prolong glycemic control and improve associated risk factors. The trial consisted of three treatment arms: metformin alone, metformin with rosiglitazone, and metformin with an intensive lifestyle intervention program (called the TODAY Lifestyle Program or TLP). Individuals between 10 and 17 years old with a diagnosis of T2DM for less than two years, a body mass index (BMI) ≥ 85th percentile, and an absence of diabetes-related autoimmunity were eligible for the study. Upon enrollment, participants entered a run-in period of 2 to 6 months, with the goals of weaning them from non-study diabetes medications, initiating treatment with metformin, providing standard diabetes education, and documenting adherence to the medication regimen. Participants were required to attain glycemic control, defined as a glycated hemoglobin level of less than 8% and measured monthly for at least 2 months, with metformin alone before proceeding. Following successful completion of the run-in period, participants were randomized to treatment with metformin and either placebo or rosiglitazone as appropriate. For those assigned to treatment with the TLP, the program was delivered in a series of in-person visits and focused on weight loss through family-based changes in eating and activity behaviors. The primary outcome measure compared between groups was time to treatment failure, defined as a persistently elevated glycated hemoglobin level (≥8%) over a period of 6 months or persistent metabolic decompensation (defined as either the inability to wean the participant from insulin within 3 months after its initiation for decompensation or the occurrence of a second episode of decompensation within 3 months after discontinuation of insulin).

Note that the data package includes data from the Beck Depression Inventory form. As the Beck Depression Inventory form is copyrighted, the form is not included in the data package.

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 the different form datasets.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by the TODAY Study Group [1] in the New England Journal of Medicine. To verify the integrity of the datasets, descriptive statistics from the baseline and M48 visits were computed, by treatment group.

5 Results

Figure 2 in the publication [1], Overall Primary Outcome Results. 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 Figure 2. The results of the replication match the published results.

Appendix C in the publication [1]: Baseline Characteristics of 699 Participants, Overall and by Treatment Group. Table C lists the variables that were used in the replication and Table D compares the results calculated from the archived data file to the results published in Appendix C. The results of the replication are very similar to the published results. Note that certain variables in the table were redacted to protect the subjects' confidentiality. These variables are marked as "N/A" in Table C and were not included in the replication.

Appendix D in the publication [1]: Reason for Failure and Median Time to Failure by Treatment Group. Table E lists the variables that were used in the replication and Table F compares the results calculated from the archived data file to the results published in Appendix D. The results of the replication match the published results.

Appendix H in the publication [1]: Secondary Outcomes at Baseline and 24 Months by Treatment Group. Table G lists the variables that were used in the replication and Table H compares the results calculated from the archived data file to the results published in Appendix H. The results of the replication are very similar to the published results.

Replications of Table 1 "Major Coexisting Conditions at Baseline and during the Study, According to Treatment Group" and Table 2 "Adverse Events and Clinical and Biochemical Assessments According to Treatment Group" were not performed. For Table 1, the study group applied algorithms with adjudication to make the diagnoses which were not included in the data package. For Table 2, the adverse event data were not included in the data package to protect the subjects' confidentiality.

6 Conclusions

The NIDDK repository is confident that the TODAY data files to be distributed are a copy of the manuscript data.

7 References

[1] TODAY Study Group "A Clinical Trial to Maintain Glycemic Control in Youth with Type 2 Diabetes." N Engl J Med 2012; 366:2247-2256 June 14, 2012

Table A: Variables used to replicate Figure 2: Overall Primary Outcome Results.

Table Variable	Variables Used in Replication (Dataset)
Failure rate	outcome (primout.sas7bdat)
Treatment group	tx (primout.sas7bdat)

Table B: Comparison of values computed in integrity check to reference article Figure 2 values

Failure Rates:	Manuscript	DSIC	Difference
Metformin alone	51.7%	51.7%	0
Metformin-rosiglitazone	38.6%	38.6%	0
Metformin-lifestyle	46.6%	46.6%	0

Table C: Variables used to replicate Appendix C: Baseline Characteristics of 699 Participants, Overall and by Treatment Group.

Table Variable	Variables Used in Replication (Dataset)
Age (years)	N/A
Tanner stage 4 or 5	tanner (pe.sas7bdat dataset)
BMI Z-score	N/A
Percent overweight	N/A
Duration of diabetes (months)	N/A
Female sex	sex (pat.sas7bdat dataset)
Race/ethnicity	race (pat.sas7bdat dataset)
Household income	houseinc (pat.sas7bdat dataset)
Parent/guardian highest level of education	houseedu (pat.sas7bdat dataset)
Nuclear family history of diabetes	momdbnow, daddbnow, fulldiab, and halfdiab (pat.sas7bdat dataset)
Nuclear family + grandparents history of diabetes	momdbnow, daddbnow, fulldiab, halfdiab, and grandiab (pat.sas7bdat dataset)
Treatment group	tx (primout.sas7bdat)

Table D: Comparison of values computed in integrity check to reference article Appendix C values

Characteristics	Overall (n=699) Manuscript	Overall (n=699) DSIC	Difference
Age (years)	14.0 (2.0)	N/A	N/A
Tanner stage 4 or 5	88.0%	88.5	0.5
BMI Z-score	2.23 (0.47)	N/A	N/A
Percent overweight	78.9 (37.3)	N/A	N/A
Duration of diabetes (months)	7.8 (5.8)	N/A	N/A
Female sex	64.7%	64.7%	0
Race/ethnicity			
White Non-Hispanic	20.3%	20.3%	0
Black Non-Hispanic	32.5%	32.5%	0
Hispanic	39.7%	39.7%	0
American Indian	5.9%	N/A	N/A
Asian	1.6%	N/A	N/A
Household income			
<$25,000	41.5%	41.5%	0
$25,000-49,999	33.7%	33.7%	0
>$49,999	24.8%	24.8%	0
Parent/guardian highest level of education			
12th grade or less	26.5%	26.5%	0
High school graduate/GED/business/technical	25.1%	25.1%	0
Some college/associate degree	31.8%	31.8%	0
Bachelors degree or higher	16.6%	16.6%	0
Nuclear family history of diabetes	59.6%	59.6	0
Nuclear family + grandparents history of diabetes	89.5%	89.3	0.2

Characteristics	M (n=232) Manuscript	M (n=232) DSIC	Difference
Age (years)	14.1 (1.9)	N/A	N/A
Tanner stage 4 or 5	88.8%	90.0%	1.2
BMI Z-score	2.27 (0.45)	N/A	N/A
Percent overweight	82.1 (38.3)	N/A	N/A
Duration of diabetes (months)	7.8 (6.0)	N/A	N/A
Female sex	62.9%	62.9%	0
Race/ethnicity			
White Non-Hispanic	21.1%	21.1%	0
Black Non-Hispanic	33.2%	33.2%	0
Hispanic	39.2%	39.2%	0
American Indian	5.2%	N/A	N/A
Asian	1.3%	N/A	N/A
Household income			
<$25,000	38.9%	38.9%	0
$25,000-49,999	39.9%	39.9%	0
>$49,999	21.2%	21.2%	0
Parent/guardian highest level of education			
12th grade or less	26.2%	26.2%	0
High school graduate/GED/business/technical	24.9%	24.9%	0
Some college/associate degree	33.6%	33.6%	0
Bachelors degree or higher	15.3%	15.3%	0
Nuclear family history of diabetes	57.5%	57.5%	0
Nuclear family + grandparents history of diabetes	92.5%	92.5%	0

Characteristics	M+R (n=233) Manuscript	M+R (n=233) DSIC	Difference
Age (years)	14.1(2.1)	N/A	N/A
Tanner stage 4 or 5	88.8%	88.6	0.2
BMI Z-score	2.22 (0.49)	N/A	N/A
Percent overweight	79.1 (38.1)	N/A	N/A
Duration of diabetes (months)	8.0 (5.7)	N/A	N/A
Female sex	65.2%	65.2%	0
Race/ethnicity			
White Non-Hispanic	20.2%	20.2%	0
Black Non-Hispanic	27.5%	27.5%	0
Hispanic	43.3%	43.3%	0
American Indian	6.9%	N/A	N/A
Asian	2.1%	N/A	N/A
Household income			
<$25,000	41.9%	41.9%	0
$25,000-49,999	29.8%	29.8%	0
>$49,999	28.3%	28.3%	0
Parent/guardian highest level of education			
12th grade or less	26.3%	26.3%	0
High school graduate/GED/business/technical	21.5%	21.5%	0
Some college/associate degree	34.2%	34.2%	0
Bachelors degree or higher	18.0%	18.0%	0
Nuclear family history of diabetes	58.8%	58.8%	0
Nuclear family + grandparents history of diabetes	88.5%	88.2%	0.3

Characteristics	M+L (n=233) Manuscript	M+L (n=233) DSIC	Difference
Age (years)	13.8 (2.0)	N/A	N/A
Tanner stage 4 or 5	86.3%	86.8	0.5
BMI Z-score	2.18 (0.46)	N/A	N/A
Percent overweight	75.6 (35.3)	N/A	N/A
Duration of diabetes (months)	7.6 (5.8)	N/A	N/A
Female sex	65.8%	65.8%	0
Race/ethnicity			
White Non-Hispanic	19.7%	19.7%	0
Black Non-Hispanic	36.7%	36.7%	0
Hispanic	36.7%	36.7%	0
American Indian	5.6%	N/A	N/A
Asian	1.3%	N/A	N/A
Household income			
<$25,000	43.6%	43.6%	0
$25,000-49,999	31.3%	31.3%	0
>$49,999	25.1%	25.1%	0
Parent/guardian highest level of education			
12th grade or less	27.1%	27.1%	0
High school graduate/GED/business/technical	28.8%	28.8%	0
Some college/associate degree	27.5%	27.5%	0
Bachelors degree or higher	16.6%	16.6%	0
Nuclear family history of diabetes	62.5%	62.5%	0
Nuclear family + grandparents history of diabetes	87.3%	87.3	0

Table E: Variables used to replicate Appendix D: Reason for Failure and Median Time to Failure by Treatment Group.

Table Variable	Variables Used in Replication (Dataset)
Reason for failure (%)	reason (primout.sas7bdat)
Median time to failure (months)	daystocensor, daystopo_s, reason (pe.sas7bdat dataset)
Treatment group	tx (primout.sas7bdat)

Table F: Comparison of values computed in integrity check to reference article Appendix D values

	Metformin Alone (manuscript)	Metformin Alone (DSIC)	Difference
Reason for failure (%)			
Persistent elevation of A1c	84.2	84.2	0
Metabolic decompensation	15.8	15.8	0
Median time to failure (months)	10.3	10.3	0

	Metformin + rosiglitazone (manuscript)	Metformin + rosiglitazone (DSIC)	Difference
Reason for failure (%)			
Persistent elevation of A1c	75.6	75.6	0
Metabolic decompensation	24.4	24.4	0
Median time to failure (months)	12.0	12.0	0

	Metformin + lifestyle (manuscript)	Metformin + lifestyle (DSIC)	Difference
Reason for failure (%)			
Persistent elevation of A1c	78.9	78.9	0
Metabolic decompensation	21.1	21.1	0
Median time to failure (months)	11.8	11.8	0

Table G: Variables used to replicate Appendix H: Secondary Outcomes at Baseline and 24 Months by Treatment Group.

Table Variable	Variables Used in Replication (Dataset)
Total cholesterol (mg/dL)	chol (cbl.sas7bdat)
LDL (mg/dL)	ldl (cbl.sas7bdat)
HDL (mg/dL)	hdl (cbl.sas7bdat)
Triglycerides (mg/dL)	trig (cbl.sas7bdat)
Urine albumin/creatinine (mg/g)	ualb and ucreat (cbl.sas7bdat)
1/fasting insulin (mL/μU)	ins (cbl.sas7bdat)
Insulinogenic index	ins30min, ins0min, glu30min, and glu0min (cbl.sas7bdat)
Systolic blood pressure	sbp (baseline.sas7bdat and baseline.sas7bdat)
Diastolic blood pressure	dbp (baseline.sas7bdat and baseline.sas7bdat)
DEXA fat mass (kg)	wb_tot_fat_p (dexa.sas7bdat)
DEXA lean mass	wb_tot_fat_p (dexa.sas7bdat)
Treatment group	tx (primout.sas7bdat)

Table H: Comparison of values computed in integrity check to reference article Appendix H values

Metformin only	N Manuscript	50 (25, 75) Manuscript	N DSIC	50 (25, 75) DSIC	N Diff	50 (25, 75) Diff
Total cholesterol (mg/dL)						
Baseline	231	141 (124, 163)	231	141 (124, 163)	0	0 (0, 0)
Month 24	193	153 (135, 178)	194	153 (134, 178)	1	0 (1, 0)
LDL (mg/dL)						
Baseline	230	83 (68, 101)	230	83 (68, 101)	0	0 (0, 0)
Month 24	193	88 (73, 110)	194	88 (73, 110)	1	0 (0, 0)
HDL (mg/dL)						
Baseline	230	36 (31, 43)	230	36 (31, 43)	0	0 (0, 0)
Month 24	193	39 (33, 46)	194	39 (33, 46)	1	0 (0, 0)
Triglycerides (mg/dL)						
Baseline	231	99 (70, 147)	231	99 (70, 147)	0	0 (0, 0)
Month 24	193	102 (72, 158)	194	101 (72, 158)	1	1 (0, 0)
Urine albumin/creatinine (mg/g)						
Baseline	231	7 (4, 15)	231	7 (4, 15)	0	0 (0, 0)
Month 24	194	7 (4, 15)	195	7 (4, 15)	1	0 (0, 0)
1/fasting insulin (mL/μU)						
Baseline	228	0.036 (0.025, 0.050)	230	0.035 (0.024, 0.048)	2	0.001 (0.001, 0.002)
Month 24	167	0.037 (0.023, 0.061)	193	0.037 (0.023, 0.060)	26	0 (0, 0.001)
Insulinogenic index						
Baseline	220	1.02 (0.47, 1.98)	219	1.03 (0.47, 1.99)	1	0.01 (0, 0.01)
Month 24	150	0.75 (0.33, 1.39)	151	0.75 (0.32, 1.39)	1	0 (0.01, 0)
Systolic blood pressure						
Baseline	232	114 (106.5, 122)	232	113.25 (106, 120.5)	0	0.75 (0.5, 1.5)
Month 24	200	114 (107.5, 123.5)	201	114 (108, 124)	1	0 (0.5, 0.5)
Diastolic blood pressure						
Baseline	232	66.5 (60.75, 72.25)	232	66.5 (60.75, 72.25)	0	0 (0, 0)
Month 24	200	68 (62.75, 73.75)	201	68 (63, 74)	1	0 (0.25, 0.25)
DEXA fat mass (kg)						
Baseline	157	34.2 (26.9, 39.4)	157	34.2 (26.9, 39.4)	0	0 (0, 0)
Month 24	130	35.2 (29.5, 42.1)	130	35.2 (29.5, 42.1)	0	0 (0, 0)
DEXA lean mass (kg)						
Baseline	157	55.6 (46.8, 63.3)	157	55.6 (46.8, 63.3)	0	0 (0, 0)
Month 24	130	55.2 (48.9, 66.0)	130	55.2 (48.9, 66.0)	0	0 (0, 0)

Metformin + rosiglitazone	N Manuscript	50 (25, 75) Manuscript	N DSIC	50 (25, 75) DSIC	N Diff	50 (25, 75) Diff
Total cholesterol (mg/dL)						
Baseline	233	146 (125, 167)	233	146 (125, 167)	0	0 (0, 0)
Month 24	187	154 (126, 179)	187	154 (126, 179)	0	0 (0, 0)
LDL (mg/dL)						
Baseline	231	80 (68, 106)	231	80 (68, 106)	0	0 (0, 0)
Month 24	186	84.5 (67, 108)	186	84.5 (67, 108)	0	0 (0, 0)
HDL (mg/dL)						
Baseline	231	38 (33, 44)	231	38 (33, 44)	0	0 (0, 0)
Month 24	187	40 (34, 49)	187	40 (34, 49)	0	0 (0,0)
Triglycerides (mg/dL)						
Baseline	233	98 (67, 138)	233	98 (67, 138)	0	0 (0, 0)
Month 24	187	104 (68, 168)	187	104 (68, 168)	0	0 (0, 0)
Urine albumin/creatinine (mg/g)						
Baseline	230	6 (4, 13)	230	6 (4, 13)	0	0 (0, 0)
Month 24	183	7 (4, 19)	183	7 (4, 19)	0	0 (0, 0)
1/fasting insulin (mL/μU)						
Baseline	228	0.040 (0.028, 0.064)	232	0.040 (0.027, 0.061)	4	0 (0.001, 0.003)
Month 24	157	0.049 (0.031, 0.068)	185	0.047 (0.030, 0.067)	28	0.002 (0.001, 0.001)
Insulinogenic index						
Baseline	223	0.92 (.47, 1.58)	223	0.92 (0.47, 1.58)	0	0 (0, 0)
Month 24	147	0.83 (.28, 1.38)	145	0.85 (0.30, 1.38)	2	0.02 (0.02, 0)
Systolic blood pressure						
Baseline	233	113 (106, 121)	233	112.5 (106.5, 120)	0	0.5 (0.5, 1)
Month 24	192	115.25 (109, 122.5)	192	115.5 (109, 123)	0	0.25 (0, 0.5)
Diastolic blood pressure						
Baseline	233	67 (61.5, 72)	233	67 (61.5, 72)	0	0 (0, 0)
Month 24	192	69.75 (64.5, 74.5)	192	70 (65, 75)	0	0.25 (0.5, 0.5)
DEXA fat mass (kg)						
Baseline	160	32.3 (26.1, 40.6)	160	32.3 (26.1, 40.6)	0	0 (0, 0)
Month 24	123	38.3 (29.4, 46.4)	123	38.3 (29.4, 46.4)	0	0 (0, 0)
DEXA lean mass (kg)						
Baseline	160	55.1 (45.3, 63.0)	160	55.1 (45.3, 63.0)	0	0 (0, 0)
Month 24	123	57.5 (49.3, 64.8)	123	57.5 (49.3, 64.8)	0	0 (0, 0)

Metformin + lifestyle	N Manuscript	50 (25, 75) Manuscript	N DSIC	50 (25, 75) DSIC	N Diff	50 (25, 75) Diff
Total cholesterol (mg/dL)						
Baseline	234	145 (127, 162)	234	145 (127, 162)	0	0 (0, 0)
Month 24	203	153 (130, 171)	203	153 (130, 171)	0	0 (0, 0)
LDL (mg/dL)						
Baseline	233	84 (69, 100)	233	84 (69, 100)	0	0 (0, 0)
Month 24	203	88 (69, 104)	203	88 (69, 104)	0	0 (0, 0)
HDL (mg/dL)						
Baseline	233	38 (34, 44)	233	38 (34, 44)	0	0 (0, 0)
Month 24	203	40 (35, 46)	203	40 (35, 46)	0	0 (0, 0)
Triglycerides (mg/dL)						
Baseline	234	87.5 (63, 128)	234	87.5 (63, 128)	0	0 (0, 0)
Month 24	203	92 (64, 148)	203	92 (64, 148)	0	0 (0, 0)
Urine albumin/creatinine (mg/g)						
Baseline	231	6 (4, 10)	231	6 (4, 10)	0	0 (0, 0)
Month 24	200	6 (4, 14)	200	7 (4, 14)	0	0 (0, 0)
1/fasting insulin (mL/μU)						
Baseline	231	0.040 (0.027, 0.067)	229	0.041 (0.028, 0.070)	2	0.001 (0.001, 0.003)
Month 24	167	0.039 (0.027, 0.064)	203	0.04 (0.025, 0.061)	36	0.001 (0.002, 0.003)
Insulinogenic index						
Baseline	221	0.87 (0.47, 1.89)	219	0.89 (0.47, 1.91)	2	0.02 (0, 0.02)
Month 24	155	0.71 (0.26, 1.69)	154	0.72 (0.28, 1.69)	1	0.01 (0.02, 0)
Systolic blood pressure						
Baseline	234	112 (105, 120)	234	111.5 (105, 120)	0	0.5 (0, 0)
Month 24	209	114 (108.5, 121)	209	114 (109, 121)	0	0 (0.5, 0)
Diastolic blood pressure						
Baseline	234	65 (61, 71)	234	65 (61, 71)	0	0 (0, 0)
Month 24	209	68 (62, 74)	209	68 (62, 74)	0	0 (0, 0)
DEXA fat mass (kg)						
Baseline	170	31.5 (25.5, 38.3)	170	31.5 (25.5, 38.3)	0	0 (0, 0)
Month 24	134	33.2 (25.3, 39.0)	134	33.2 (25.3, 39.0)	0	0 (0, 0)
DEXA lean mass (kg)						
Baseline	170	53.1 (44.6, 61.6)	170	53.1 (44.6, 61.6)	0	0 (0, 0)
Month 24	134	53.7 (47.1, 61.5)	134	53.7 (47.1, 61.5)	0	0 (0, 0)

```
************************************************************************************************************************
***Program: /prj/niddk/ims_analysis/TODAY/prog_initial_analysis/today.data.integrity.check.sas
***Programmer: Corey Del Vecchio
***Date Created: 11/14/2014
***Purpose: To replicate the results in the "A Clinical Trial to Maintain Glycemic Control in Youth with Type 2 Diabetes" publication.
***
************************************************************************************************************************;

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

options mprint symbolgen;

proc format;
  value tx_form 1 = "Metformin only"
                2 = "Metformin + rosiglitazone"
                3 = "Metformin + lifestyle";

  value racef   1 = "Black, Non-Hispanc"
                2 = "Hispanic"
                3 = "White, Non-Hispanic"
                4 = "Other";

  value sexf    1 = "Female"
                2 = "Male";

  value HOUSEINCf 1 = "<$24,999"
                  2 = "$25,000 - $49,999"
                  3 = "> $50,000";

  value HOUSEEDUf 1 = "Less than HS degree"
                  2 = "HS, GED, business or technical school"
                  3 = "Some college but no degree"
                  4 = "At least some college degree";


*** This macro will generate the information needed for Appendix H rounding to two decimal places;

%macro percentile(dataset_name=, var_name=);

proc sort data = &dataset_name.;
       by tx;

proc means data = &dataset_name. N P50 P25 P75 maxdec=2;
       var &var_name.;
       by tx;
       where mvisit = "M00";
       format tx tx_form.;
       title3 "Appendix H Replication of Dataset: &dataset_name. Variable: &var_name. Visit: M00";

run;

proc means data = &dataset_name. N P50 P25 P75 maxdec=2;
       var &var_name.;
       by tx;
       where mvisit = "M24";
       format tx tx_form.;
       title3 "Appendix H Replication of Dataset: &dataset_name. Variable: &var_name. Visit: M24";

run;
```

```
%mend percentile;

run;

*** This macro will generate the information needed for Appendix H rounding to three decimal places;

%macro percentile3(dataset_name=, var_name=);

proc sort data = &dataset_name.;
       by tx;

proc means data = &dataset_name. N P50 P25 P75 maxdec=3;
       var &var_name.;
       by tx;
       where mvisit = "M00";
       format tx tx_form.;
       title3 "Appendix H Replication of Dataset: &dataset_name. Variable: &var_name. Visit: M00";

run;

proc means data = &dataset_name. N P50 P25 P75 maxdec=3;
       var &var_name.;
       by tx;
       where mvisit = "M24";
       format tx tx_form.;
       title3 "Appendix H Replication of Dataset: &dataset_name. Variable: &var_name. Visit: M24";

run;

%mend percentile3;

run;


** Location of the TODAY datasets;
libname today "/prj/niddk/ims_analysis/TODAY/private_created_data/sas_data/";

*** Reading in the datasets needed for this replication;

data BASELINE ; set today.BASELINE ;
data CBL      ; set today.CBL      ;
data PRIMOUT  ; set today.PRIMOUT  ;
data VISIT    ; set today.VISIT    ;
data PE       ; set today.PE       ;
data PRIMOUT  ; set today.PRIMOUT  ;
data dexa     ; set today.dexa;

data pat      ; set today.pat;


*** Replicating the percentages that are included in Figure 2;

proc sort data = PRIMOUT;
  by tx;

proc freq data = PRIMOUT;
  tables outcome /list missing;
  by tx;
  format tx tx_form.;
```

```
  title 'Replicating the percentages that are included in Figure 2';

*** Replicating the data that is included in Appendix C";


proc sort data = primout;
       by releaseid;

proc sort data = PE;
       by releaseid;

*** Adding tx (treatment group variable) to the PE dataset;

data PE;
       merge primout (in = in1 keep = releaseid tx)
                           PE     (in = in2);
       by releaseid;
       if in1 and in2 then output PE;
       else if in2 then abort;

proc sort data = PE;
       by tx;


proc freq data = PE;
       table MVISIT*tanner / list missing;
       where mvisit = "M00";
       title3 "Appendix C Replication of the tanner variable (Overall)";

proc freq data = PE;
       table MVISIT*tanner / list missing;
       by tx;
       where mvisit = "M00";
       format tx tx_form.;
       title3 "Appendix C Replication of the tanner variable (by treatment group)";


*** Creating the variables have_nuc_fam (Nuclear family history of diabetes) and have_nuc_fam_grand (Nuclear family + grandparents history of diabetes) that
are included in Appendix C;

data pat;
       set pat;

       if MOMDBNOW = 1 or DADDBNOW = 1 or FULLDIAB = 1 or HALFDIAB = 1 then have_nuc_fam = 1;
       else if MOMDBNOW = 0 or DADDBNOW = 0 or FULLDIAB = 0 or HALFDIAB = 0 then have_nuc_fam = 0;
       else if MOMDBNOW = . and DADDBNOW = . and FULLDIAB = . and HALFDIAB = . then have_nuc_fam = .;
       else abort;

       if MOMDBNOW = 1 or DADDBNOW = 1 or FULLDIAB = 1 or HALFDIAB = 1 or GRANDIAB = 1 then have_nuc_fam_grand = 1;
       else if MOMDBNOW = 0 or DADDBNOW = 0 or FULLDIAB = 0 or HALFDIAB = 0 or GRANDIAB = 0 then have_nuc_fam_grand = 0;
       else if MOMDBNOW = . and DADDBNOW = . and FULLDIAB = . and HALFDIAB = . and GRANDIAB = . then have_nuc_fam_grand = .;
       else abort;

*** Adding the tx variable to the pat dataset;

data pat;
       merge primout (in = in1 keep = releaseid tx)
                           pat     (in = in2);
       by releaseid;
       if in1 and in2 then output pat;
```

```
        else if in2 then abort;

proc sort data = pat;
        by tx;

proc freq data = pat;
        table sex race HOUSEINC HOUSEEDU have_nuc_fam have_nuc_fam_grand / list;
        format race racef. sex sexf. HOUSEINC HOUSEINCf. HOUSEEDU HOUSEEDUf.;
        title3 "Appendix C Replication of the multiple variables (Overall)";

proc freq data = pat;
        table sex race HOUSEINC HOUSEEDU have_nuc_fam have_nuc_fam_grand / list;
        by tx;
        format tx tx_form. race racef. sex sexf. HOUSEINC HOUSEINCf. HOUSEEDU HOUSEEDUf.;
        title3 "Appendix C Replication of the multiple variables (by treatment group)";


*** Replicating the data in Appendix D";

proc sort data = primout;
        by tx;

run;

data primout;
        set primout;

        *** Creating the reason_group (Reason for Failure) variable ;

        if reason = 1 then reason_group = 1;
        else if reason in(2,3) then reason_group = 2;
        else reason_group = .;

        *** Creating the censor_time (Median time to failure (months)) variable;

        if daystocensor NE . and daystopo_s = . then censor_time = round(daystocensor/30.4, 0.1);
        else if daystocensor = . and daystopo_s NE . then censor_time = round(daystopo_s/30.4, 0.1);
        else abort;

proc freq data = primout;
        table reason_group;
        by tx;
        format tx tx_form.;
        title3 "Replicating the Reason for Failure variable in Appendix D";

proc means data = primout median;
        var censor_time;
        by tx;
        where reason_group in(1,2);
        format tx tx_form.;
        title3 "Replicating the Median time to failure (months) variable in Appendix D";

*** Replicating the data that is included in Appendix H (cbl dataset)";

proc sort data = cbl;
        by releaseid;

proc sort data = primout;
        by releaseid;
```

```
*** Adding the tx variable to the cbl dataset;

data cbl;
	merge primout (in = in1 keep = releaseid tx)
			cbl     (in = in2);
	by releaseid;
	if in1 and in2 then do;

		*** Creating the insulinogenic_index (Insulinogenic index) variable that is included in Appendix H;

		if (ins30min - ins0min) > 0 and (glu30min - glu0min) > 0 then insulinogenic_index = (ins30min - ins0min)/(glu30min - glu0min);
		else insulinogenic_index = .;

		*** Creating the ualbcreat_mg_g (Urine albumin/creatinine (mg/g) variable that is included in Appendix H;

		ualbcreat_mg_g = UALB / (UCREAT*.001);

		*** Creating the recip_INS (1/fasting insulin (mL/µU)) variable that is included in Appendix H;

		recip_INS = 1/INS;

		output cbl;
	end;
	else if in2 then abort;

proc sort data = cbl;
	by tx;

run;

*** Generating the data for Appendix H for data that is included in the cbl dataset;

%percentile(dataset_name= cbl, var_name=chol);
%percentile(dataset_name= cbl, var_name=ldl);
%percentile(dataset_name= cbl, var_name=hdl);
%percentile(dataset_name= cbl, var_name=Trig);
%percentile(dataset_name= cbl, var_name=ualbcreat_mg_g);

*** In the manuscript, the data is rounded to three decimal places;
%percentile3(dataset_name= cbl, var_name=recip_INS);

%percentile(dataset_name= cbl, var_name=insulinogenic_index);


*** Replicating the data that is included in Appendix H (baseline and visit datasets)";

proc sort data = baseline;
	by releaseid;

proc sort data = primout;
	by releaseid;

* Adding the mvisit variable to the baseline dataset;

data baseline;
	set baseline;
	mvisit = "M00";

*** Adding baseline data to the visit dataset to create the total_visits dataset;
```

```
data total_visits;
       set baseline
                      visit;

proc sort data = total_visits;
       by releaseid;

*** Adding tx to the total_visits datset;

data total_visits;
       merge primout      (in = in1 keep = releaseid tx)
                             total_visits (in = in2);
       by releaseid;
       if in1 and in2 then output total_visits;
       else if in2 then abort;

proc sort data = total_visits;
       by tx;

*** Generating the data for Appendix H for data that is included in the baseline and visit datasets;

%percentile(dataset_name= total_visits, var_name=SBP);
%percentile(dataset_name= total_visits, var_name=DBP);


*** Replicating the data that is included in Appendix H (dexa dataset)";

proc sort data = dexa;
       by releaseid;

*** Adding tx to the dexa datset;

data dexa;
       merge primout (in = in1 keep = releaseid tx)
                             dexa    (in = in2);
       by releaseid;
       if in1 and in2 then do;

              *** Creating the WB_TOT_FAT_P_kg (DEXA fat mass (kg)) variable that is included in Appendix H;

              WB_TOT_FAT_P_kg = WB_TOT_FAT_P/1000;

              *** Creating the WB_TOT_LEAN_P_kg (DEXA lean mass (kg)) variable that is included in Appendix H;

              WB_TOT_LEAN_P_kg = WB_TOT_LEAN_P/1000;

              output dexa;
       end;
       else if in2 then abort;

proc sort data = dexa;
       by tx;

*** Generating the data for Appendix H for data that is included in the dexa dataset;

%percentile(dataset_name= dexa, var_name=WB_TOT_FAT_P_kg);
%percentile(dataset_name= dexa, var_name=WB_TOT_LEAN_P_kg);
```