Dataset Integrity Check for the Nonalcoholic Steatohepatitis (NASH) The Farnesoid X Receptor Ligand Obeticholic Acid in NASH Treatment (FLINT) Trial Data Files

Prepared by Jane Wang
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton MD 20705
August 8, 2020

Contents

1 Standard Disclaimer 3

2 Study Background 3

3 Archived Datasets 3

4 Statistical Methods 4

5 Results 4

6 Conclusions 4

7 References 4

Table A: Variables used to replicate Table 1 in the publication. 5

Table B.1: Comparison of values computed in integrity check to reference article Table 1 values (Mean and Standard Deviation (SD) Data) 7

Table B.2: Comparison of values computed in integrity check to reference article Table 1 values (N (count) and percentage (%) data) 10

Attachment A: SAS Code 11

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 Farnesoid X Receptor (FXR) Ligand Obeticholic Acid in Nonalcoholic Steatohepatitis (NASH) Treatment (FLINT) trial was a multicenter, randomized, placebo-controlled clinical trial conducted by the NASH CRN to investigate the safety and efficacy of treatment of obeticholic acid compared to placebo in adults with nonalcoholic fatty liver disease. Individuals with histologic evidence of NASH, based on a liver biopsy and a disease activity score (NAS) of at least 4 (with a score of at least 1 in each component of the NAS 8-point scale), were enrolled and randomized to treatment with either obeticholic acid (25 mg) or placebo for 72 weeks. Each participant received a baseline oral glucose tolerance test and a baseline MRI scan to determine liver fat content. Additionally, information was collected on demographics, alcohol consumption, and medication use. The primary outcome measure was improvement in the NAS score by at least 2 points with no worsening of liver fibrosis and was assessed by liver biopsy at 72 weeks. Secondary outcome measures included change in NASH diagnosis (from determinate or indeterminate to not-NASH), fibrosis score, hepatocellular ballooning score, hepatic fat fraction, serum aminotransferase and gamma-glutamyl transpeptidase (GGT) levels, fasting glucose levels, and health related quality of life scores.

3 Archived Datasets

All SAS data files, as provided by the Data Coordinating Center (DCC), are located in the “Datasets” folder in the data package of NASHCRN_Data_Sharing_FLINT_LANCET_2015. For this replication, variables were taken from the datasets listed below in Table A below.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Brent A Neuschwander-Tetri et al. Lancet 2015; 385: 956–65 [1]. To verify the integrity of the “Flint Lancet 2015” dataset, descriptive statistics of characteristics were computed. Baseline characteristics were computed by different treatment groups.

5 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 Table 1 to Table 4.The results of this replication are almost an exact match to published data within expected variation.

6 Conclusions

The NIDDK repository is confident that the NASH Flint Lancet 2015 data files to be distributed are a true copy of the study data.

7 References

[1] Brent A Neuschwander-Tetri, Rohit Loomba, Arun J Sanyal, Joel E Lavine, Mark L Van Natta, Manal F Abdelmalek, Naga Chalasani, Srinivasan Dasarathy, Anna Mae Diehl, Bilal Hameed, Kris V Kowdley, Arthur McCullough, Norah Terrault, Jeanne M Clark, James Tonascia, Elizabeth M Brunt, David E Kleiner, Edward Doo, for the NASH Clinical Research Network. Farnesoid X nuclear receptor ligand obeticholic acid fornon-cirrhotic, non-alcoholic steatohepatitis (FLINT): a multicentre, randomised, placebo-controlled trial. Lancet 2015; 385: 956–65

Table A: Variables used to replicate Table 1 in the publication.

Table 1 variables	Variable used to replicate Table 1 (dataset.variable)
Age (years)	Table1.age
Male	Table1.MALE
Race	Table1.Race
Hispanic	Table1.HISP
Physical component summary	Table1.SF36PHYS
Mental component summary	Table1.SF36MENT
Alanine aminotransferase (U/L)	Table1.ALT
Aspartate aminotransferase (U/L)	Table1.AST
Alkaline phosphatase (U/L)	Table1.ALKA
glutamyl transpeptidase (U/L)	Table1.GGT
Total bilirubin (mol/L)	Table1.BILIT_SI
Total cholesterol (mmol/L)	Table1.CHOL_SI
HDL cholesterol (mmol/L)	Table1.HDL_SI
LDL cholesterol (mmol/L)	Table1.LDL_SI
Trigylcerides (mmol/L)	Table1.TRIG_SI
Haemoglobin (g/L)	Table1.HEMO_SI
Haematocrit (proportion of 1.0)	Table1.HEMA_SI
Mean corpuscular volume (fL)	Table1.MCV
White blood cell count (*10. per L)	Table1.WBC_SI
Platelet count (*10. per L)	Table1.PLATELE0
Bicarbonate (mmol/L)	Table1.BICARB_S
Calcium (mmol/L)	Table1.CAL_SI
Phosphate (mmol/L)	Table1.PHOS_SI
Creatinine (mol/L)	Table1.CR_SI
Uric acid (mol/L)	Table1.URIC_SI
Albumin (g/L)	Table1.ALB_SI
Total protein (g/L)	Table1.PROT_SI
Prothrombin time (s)	Table1.PT
International normalised ratio	Table1.INR
Fasting serum glucose (mmol/L)	Table1.GLUCB_SI
Insulin (pmol/L)	Table1.INSUB_SI
HOMA-IR (glucose [mmol/Linsulin	Table1.HOMA_SI
Glycated haemoglobin A1c (mmol/mol)	Table1.HBA1C_SI
Weight (kg)	Table1.WEIGHT
Body-mass index (kg/m2)	Table1.BMI
Waist circumference (cm)	Table1.WAISTCM
Waist-to-hip ratio	Table1.RATIO
Systolic blood pressure (mm Hg)	Table1.SBP

Table 1 variables	Variable used to replicate Table 1 (dataset.variable)
Diastolic blood pressure (mm Hg)	Table1.DBP
Hyperlipidaemia*	Table1.HYPERLIP
Hypertension	Table1.HTN
Cardiovascular disease	Table1.CVD
Diabetes	Table1.DIAB
Antilipidaemic	Table1.ANTILIPI
Cardiovascular	Table1.CVHTNRX
Antidiabetic	Table1.DIABRX
Metformin	Table1.METFORM
Pioglitazone	Table1.PIO
Vitamin E	Table1.VITE
Thiazolidinedione	Table1.TZD
Aspirin (81 mg)	Table1.ASPIRIN
De.nite steatohepatitis	Table1.BLDEFNAS
Fibrosis stage?	Table1.BLFIBRO
Total NAFLD activity score?	Table1.BLNAS
Hepatocellular ballooning score	Table1.BLBALL
Steatosis score	Table1.BLSTEATO
Lobular in.ammation score	Table1.BLINFLAM
Portal in.ammation score?	Table1.BLPORTAL
Biopsy length (mm)	Table1.BLLENGTH

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

Variables	Obeticholic (n=141) [Manuscript] Mean (SD)	Obeticholic [DSIC]	Obeticholic [Difference]	Placebo (n=142) [Manuscript] Mean (SD)	Placebo [DSIC]	Placebo [Difference]
Age in years	52(11)	52(11)	0(0)	51(12)	51(12)	0(0)
Physical component summary	45(11)	45(11)	0(0)	44(11)	44(11)	0(0)
Mental component summary	48(12)	48(12)	0(0)	48(12)	48(12)	0(0)
Alanine aminotransferase (U/L)	83(49)	83(49)	0(0)	82(51)	82(51)	0(0)
Aspartate aminotransferase (U/L)	64(38)	64(38)	0(0)	58(34)	58(34)	0(0)
Alkaline phosphatase (U/L)	82(29)	82(29)	0(0)	81(25)	81(25)	0(0)
glutamyl transpeptidase (U/L)	78(85)	78(85)	0(0)	76(97)	76(97)	0(0)
Total bilirubin in mol/L, Mean (SD)	11.5(5.9)	11.5(5.9)	0.0(0.0)	11.3(7.5)	11.3(7.5)	0.0(0.0)
Total cholesterol (mmol/L)	4.9(1.2)	4.9(1.2)	0.0(0.0)	4.8(1.2)	4.8(1.2)	0.0(0.0)
HDL cholesterol (mmol/L)	1.1(0.3)	1.1(0.3)	0.0(0.0)	1.1(0.4)	1.1(0.4)	0.0(0.0)
LDL cholesterol (mmol/L)	2.9(1.0)	2.9(1.0)	0.0(0.0)	2.9(1.1)	2.9(1.1)	0.0(0.0)
Trigylcerides (mmol/L)	2.2(1.5)	2.2(1.5)	0.0(0.0)	2.0(1.7)	2.0(1.7)	0.0(0.0)
Haemoglobin (g/L)	140(15)	140(15)	0(0)	140(14)	140(14)	0(0)
Haematocrit (proportion of 1)	0.41(0.04)	0.41(0.04)	0.00(0.00)	0.41(0.04)	0.41(0.04)	0.00(0.00)
Mean corpuscular volume (fL)	88.7(4.8)	88.7(4.8)	0.0(0.0)	89.0(5.3)	89.0(5.3)	0.0(0.0)
White blood cell count (per L)	7.3(1.9)	7.3(1.9)	0.0(0.0)	6.9(2.3)	6.9(2.3)	0.0(0.0)

Variables	Obeticholic (n=141) [Manuscript] Mean (SD)	Obeticholic [DSIC]	Obeticholic [Difference]	Placebo (n=142) [Manuscript] Mean (SD)	Placebo [DSIC]	Placebo [Difference]
Platelet count (10 per L)	237(59)	237(59)	0(0)	237(65)	237(65)	0(0)
Bicarbonate (mmol/L)	25.9(2.5)	25.9(2.5)	0.0(0.0)	26.2(2.6)	26.2(2.6)	0.0(0.0)
Calcium (mmol/L)	2.4(0.1)	2.4(0.1)	0.0(0.0)	2.4(0.1)	2.4(0.1)	0.0(0.0)
Phosphate (mmol/L)	1.1(0.2)	1.1(0.2)	0.0(0.0)	1.1(0.2)	1.1(0.2)	0.0(0.0)
Creatinine (mol/L)	71(18)	71(18)	0(0)	70(16)	70(16)	0(0)
Uric acid (mol/L)	375(89)	375(89)	0(0)	366(86)	365(86)	1(0)
Albumin (g/L)	43(4)	43(4)	0(0)	43(4)	43(4)	0(0)
Total protein (g/L)	73(5)	73(5)	0(0)	74(5)	74(5)	0(0)
Prothrombin time (s)	11.7(2.1)	11.7(2.1)	0.0(0.0)	11.7(2.2)	11.7(2.2)	0.0(0.0)
International normalized ratio	1.01(0.08)	1.01(0.08)	0.00(0.00)	1.00(0.07)	1.00(0.07)	0.00(0.00)
Fasting serum glucose (mmol/L)	6.5(1.8)	6.5(1.8)	0.0(0.0)	6.4(2.2)	6.4(2.2)	0.0(0.0)
Insulin (pmol/L)	201(226)	201(226)	0(0)	138(129)	138(129)	0(0)
HOMA-IR (glucose [mmol/Linsulin	61(74)	61(74)	0(0)	40(42)	40(42)	0(0)

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

Variables	Obeticholic (n=141) [Manuscript] Mean (SD)	Obeticholic [DSIC]	Obeticholic [Difference]	Placebo (n=142) [Manuscript] Mean (SD)	Placebo [DSIC]	Placebo [Difference]
Glycated haemoglobin A1c (mmol/mol)	48(12)	48(12)	0(0)	47(11)	47(11)	0(0)
Weight (kg)	100(23)	100(23)	0(0)	96(18)	96(18)	0(0)
Body-mass index (kg/m2)	35(7)	35(7)	0(0)	34(6)	34(6)	0(0)
Waist circumference (cm)	112(15)	113(15)	1(0)	109(14)	109(14)	0(0)
Waist-to-hip ratio	0.96(0.07)	0.96(0.07)	0.00(0.00)	0.95(0.09)	0.95(0.09)	0.00(0.00)
Systolic blood pressure (mm Hg)	132(17)	133(17)	1(0)	132(15)	132(15)	0(0)
Diastolic blood pressure (mm Hg)	77(11)	77(11)	0(0)	78(10)	78(10)	0(0)
Fibrosis stage?	1.9(1.1)	1.9(1.1)	0.0(0.0)	1.8(1.0)	1.8(1.0)	0.0(0.0)
Total NAFLD activity score?	5.3(1.3)	5.3(1.3)	0.0(0.0)	5.1(1.3)	5.1(1.3)	0.0(0.0)
Hepatocellular ballooning score	1.4(0.7)	1.4(0.7)	0.0(0.0)	1.3(0.7)	1.3(0.7)	0.0(0.0)
Steatosis score	2.1(0.8)	2.1(0.8)	0.0(0.0)	2.0(0.8)	2.0(0.8)	0.0(0.0)
Lobular in.ammation score	1.8(0.7)	1.8(0.7)	0.0(0.0)	1.8(0.7)	1.8(0.7)	0.0(0.0)
Portal in.ammation score?	1.2(0.6)	1.2(0.6)	0.0(0.0)	1.1(0.6)	1.1(0.6)	0.0(0.0)
Biopsy length (mm)	21(10)	21(10)	0(0)	21(10)	21(10)	0(0)

Data for Tables B and B.1 are shown in Mean (SD)

Table B.2: Comparison of values computed in integrity check to reference article Table 1 values

Variables	Obeticholic (n=141) [Manuscript] N (%)	Obeticholic [DSIC]	Obeticholic [Difference]	Placebo (n=142) [Manuscript] N (%)	Placebo [DSIC]	Placebo [Difference]
Male	43(30%)	43(30)	0(0)	53(37%)	53(37)	0(0)
Asian	6(4%)	6(4)	0(0)	10(7%)	10(7)	0(0)
Black	2(1%)	2(1)	0(0)	4(3%)	4(3)	0(0)
White	123(87%)	123(87)	0(0)	111(78%)	111(78)	0(0)
Other	10(7%)	10(7)	0(0)	17(12%)	17(12)	0(0)
Hispanic	22(16%)	22(16)	0(0)	21(15%)	21(15)	0(0)
Hyperlipidaemia	87(62%)	87(62)	0(0)	86(61%)	86(61)	0(0)
Hypertension	87(62%)	87(62)	0(0)	85(60%)	85(60)	0(0)
Cardiovascular disease	7(5%)	7(5)	0(0)	8(6%)	8(6)	0(0)
Diabetes	75(53%)	75(53)	0(0)	74(52%)	74(52)	0(0)
Antilipidaemic	72(51%)	72(51)	0(0)	64(45%)	64(45)	0(0)
Cardiovascular	97(69%)	97(69)	0(0)	92(65%)	92(65)	0(0)
Antidiabetic	67(48%)	67(48)	0(0)	73(51%)	73(51)	0(0)
Metformin	55(39%)	55(39)	0(0)	62(44%)	62(44)	0(0)
Pioglitazone	1(1%)	1(1)	0(0)	6(4%)	6(4)	0(0)
Vitamin E	29(21%)	29(21)	0(0)	32(23%)	32(23)	0(0)
Thiazolidinedione	3(2%)	3(2)	0(0)	5(4%)	5(4)	0(0)
Aspirin (81 mg)	37(26%)	37(26)	0(0)	33(23%)	33(23)	0(0)
De.nite steatohepatitis	114(81%)	114(81)	0(0)	111(79%)	111(79)	0(0)

Data for Tables B.2 are shown in N (%)

Attachment A: SAS Code

```
*****************************************************************************************************************
***
***Program:
***Programmer: Jane Wang
***Date Created: 04/12/2020
***Purpose:
*****************************************************************************************************************
**;

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


libname sasfile '/prj/niddk/ims_analysis/NASH/private_orig_data/NASHCRN_Data_Sharing_FLINT_LANCET_2015/Datasets/';
libname t_1 xport "/prj/niddk/ims_analysis/NASH/private_orig_data/NASHCRN_Data_Sharing_FLINT_LANCET_2015/Datasets/table1.xpt";proc
copy in =t_1 out= work;
libname t_2 xport "/prj/niddk/ims_analysis/NASH/private_orig_data/NASHCRN_Data_Sharing_FLINT_LANCET_2015/Datasets/table2.xpt";proc
copy in =t_2 out= work;
libname t_3 xport "/prj/niddk/ims_analysis/NASH/private_orig_data/NASHCRN_Data_Sharing_FLINT_LANCET_2015/Datasets/table3.xpt";proc
copy in =t_3 out= work;
libname t_4 xport "/prj/niddk/ims_analysis/NASH/private_orig_data/NASHCRN_Data_Sharing_FLINT_LANCET_2015/Datasets/table4.xpt";proc
copy in =t_4 out= work;

*** Data from the Primary outcome paper that was converted to .csv format so that the DSIC data could be easily compared;
FILENAME table1  '/prj/niddk/ims_analysis/NASH/private_created_data/nash_flint_lancet_2015_table1.csv';
FILENAME table2  '/prj/niddk/ims_analysis/NASH/private_created_data/nash_flint_lancet_2015_table2.csv';
FILENAME table3  '/prj/niddk/ims_analysis/NASH/private_created_data/nash_flint_lancet_2015_table3.csv';
FILENAME table4  '/prj/niddk/ims_analysis/NASH/private_created_data/nash_flint_lancet_2015_table4.csv';

*** Output CSV files that will be converted to .xls before being added to the DSIC document;
FILENAME out_t1  '/prj/niddk/ims_analysis/NASH/private_created_data/nash_flint_lancet_2015_table1_dsic.csv';
FILENAME out_t2  '/prj/niddk/ims_analysis/NASH/private_created_data/nash_flint_lancet_2015_table2_dsic.csv';
FILENAME out_t3  '/prj/niddk/ims_analysis/NASH/private_created_data/nash_flint_lancet_2015_table3_dsic.csv';
FILENAME out_t4  '/prj/niddk/ims_analysis/NASH/private_created_data/nash_flint_lancet_2015_table4_dsic.csv';

%macro baseline_freq1(dataset_name,var_name);

            *** Creating a frequency table in the format of Table 1 in the primary outcome paper;
 proc freq data = &dataset_name ;
            table (&var_name.)*tx ;
            title3 "Frequency table of the &var_name. variable in the analysis dataset";

            *** Outputting the frequency data to work.&var_name._cross using the ODS output;
 ods output CrossTabFreqs = work.&var_name._cross;

 data &var_name._cross (keep =  tx Frequency colpercent table_name) ;
  set &var_name._cross;
  if &var_name = 1 and tx ne '';
  length table_name $30.;
  table_name  ="&var_name";
*proc print data = &var_name._cross;

 proc sort data = &var_name._cross;
   by table_name tx;

 data &var_name._cross_1(drop =  tx Frequency colpercent i);
  set &var_name._cross;
  by table_name;
  array temp1(2) stat1 stat2 ;
  array temp2(2) stat3 stat4 ;
  retain stat1-stat4;
```

```
   if first.table_name then do i = 1 to 2;
     temp1(i) = .;
     temp2(i) = .;
   end;
     temp1(_n_) = Frequency;
     temp2(_n_) = colpercent;
   if last.table_name;
*proc print data = &var_name._cross_1;

%mend;
%macro baseline_freq2(dataset_name,var_name);

                *** Creating a frequency table in the format of Table 1 in the primary outcome paper;
  proc freq data = &dataset_name ;
                table (&var_name.)*tx ;
                title3 "Frequency table of the &var_name. variable in the analysis dataset";

                *** Outputting the frequency data to work.&var_name._cross using the ODS output;
  ods output CrossTabFreqs = work.&var_name._cross;
proc print data = &var_name._cross;

  data &var_name._cross (keep =&var_name   tx Frequency colpercent table_name) ;
   set &var_name._cross;
   if &var_name ne '' and tx ne '';
   length table_name $30.;
   table_name  =&var_name;
proc print data = &var_name._cross;

  proc sort data = &var_name._cross;
    by table_name  tx;

  data &var_name._cross;
   set &var_name._cross;
   by table_name  ;
   retain count 0;
   if first.table_name then count = 0;
   count = count+1;

proc print data = &var_name._cross;

  data &var_name._cross_1(drop = &var_name tx Frequency colpercent i count);
   set &var_name._cross;
   by table_name  ;
   array temp1(2) stat1 stat2 ;
   array temp2(2) stat3 stat4 ;
   retain stat1-stat4  ;
   if first.table_name  then do i = 1 to 2;
     temp1(i) = .;
     temp2(i) = .;
   end;
      temp1(count) = Frequency;
     temp2(count) = colpercent;
   if last.table_name ;
proc print data = &var_name._cross_1;

%mend;

%macro baseline_means(dataset_name,var_name);

*** Creating a frequency table in the format of Table 1 in the primary outcome paper;
  proc means data = &dataset_name mean Std ;
                var &var_name.;
                by tx;
                title3 "Frequency table of the &var_name. variable in the analysis dataset";

                *** Outputting the frequency data to work.&var_name._cross using the ODS output;
```

```
 ods output Summary = work.&var_name._means;
 run;
*  proc print data = &var_name._means;

 data &var_name._means;
  set &var_name._means;
  length table_name $30.;
  table_name  ="&var_name";

  proc sort data = &var_name._means;
   by table_name tx;

 data &var_name._means;
  set &var_name._means;
  by table_name  ;
  retain count 0;
  if first.table_name then count = 0;
  count = count+1;

 data &var_name._means_1(drop = tx &var_name._Mean &var_name._StdDev i  count);
  set &var_name._means;
  by table_name;
  array temp1(2) stat1 stat2 ;
  array temp2(2) stat3 stat4 ;
  retain stat1-stat4;
  if first.table_name then do i = 1 to 2;
   temp1(i) = .;
   temp2(i) = .;
  end;
   temp1(count) = &var_name._Mean;
   temp2(count) = &var_name._StdDev;

  if last.table_name;

*  proc print data = &var_name._means_1;

%mend;
*******************************************************************************************;
*************** Check Table 1 ********************************************************;
*******************************************************************************************;

*** Running the baseline_freq on the  categorical variables in the Table 1 manuscript file;
proc sort data = table1;
 by tx;


%baseline_means(table1, age         );
%baseline_means(table1, SF36PHYS     );
%baseline_means(table1, SF36MENT     );
%baseline_means(table1, ALT         );
%baseline_means(table1, AST         );
%baseline_means(table1, ALKA        );
%baseline_means(table1, GGT         );
%baseline_means(table1, BILIT_SI     );
%baseline_means(table1, CHOL_SI      );
%baseline_means(table1, HDL_SI       );
%baseline_means(table1, LDL_SI       );
%baseline_means(table1, TRIG_SI      );
%baseline_means(table1, HEMO_SI      );
%baseline_means(table1, HEMA_SI      );
%baseline_means(table1, MCV         );
%baseline_means(table1, WBC_SI       );
%baseline_means(table1, PLATELE0     );
%baseline_means(table1, BICARB_S     );
%baseline_means(table1, CAL_SI       );
%baseline_means(table1, PHOS_SI      );
```

```
%baseline_means(table1, CR_SI       );
%baseline_means(table1, URIC_SI     );
%baseline_means(table1, ALB_SI      );
%baseline_means(table1, PROT_SI     );
%baseline_means(table1, PT          );
%baseline_means(table1, INR         );
%baseline_means(table1, GLUCB_SI    );
%baseline_means(table1, INSUB_SI    );
%baseline_means(table1, HOMA_SI     );
%baseline_means(table1, HBA1C_SI    );
%baseline_means(table1, WEIGHT      );
%baseline_means(table1, BMI         );
%baseline_means(table1, WAISTCM     );
%baseline_means(table1, RATIO       );
%baseline_means(table1, SBP         );
%baseline_means(table1, DBP         );
%baseline_means(table1, BLFIBRO     );
%baseline_means(table1, BLNAS       );
%baseline_means(table1, BLBALL      );
%baseline_means(table1, BLSTEATO    );
%baseline_means(table1, BLINFLAM    );
%baseline_means(table1, BLPORTAL    );
%baseline_means(table1, BLLENGTH              );


%baseline_freq1(table1,MALE        );
%baseline_freq2(table1,RACE        );
%baseline_freq1(table1,HISP        );
%baseline_freq1(table1,HYPERLIP    );
%baseline_freq1(table1,HTN         );
%baseline_freq1(table1,CVD         );
%baseline_freq1(table1,DIAB        );
%baseline_freq1(table1,ANTILIPI    );
%baseline_freq1(table1,CVHTNRX     );
%baseline_freq1(table1,DIABRX      );
%baseline_freq1(table1,METFORM     );
%baseline_freq1(table1,PIO         );
%baseline_freq1(table1,VITE        );
%baseline_freq1(table1,TZD         );
%baseline_freq1(table1,ASPIRIN     );
%baseline_freq1(table1,BLDEFNAS    );

data table1_compare;
set
age_means_1
MALE_cross_1
RACE_cross_1
HISP_cross_1
SF36PHYS_means_1
SF36MENT_means_1
ALT_means_1
AST_means_1
ALKA_means_1
GGT_means_1
BILIT_SI_means_1
CHOL_SI_means_1
HDL_SI_means_1
LDL_SI_means_1
TRIG_SI_means_1
HEMO_SI_means_1
HEMA_SI_means_1
MCV_means_1
WBC_SI_means_1
PLATELE0_means_1
BICARB_S_means_1
CAL_SI_means_1
```

```
PHOS_SI_means_1
CR_SI_means_1
URIC_SI_means_1
ALB_SI_means_1
PROT_SI_means_1
PT_means_1
INR_means_1
GLUCB_SI_means_1
INSUB_SI_means_1
HOMA_SI_means_1
HBA1C_SI_means_1
WEIGHT_means_1
BMI_means_1
WAISTCM_means_1
RATIO_means_1
SBP_means_1
DBP_means_1

HYPERLIP_cross_1
HTN_cross_1
CVD_cross_1
DIAB_cross_1
ANTILIPI_cross_1
CVHTNRX_cross_1
DIABRX_cross_1
METFORM_cross_1
PIO_cross_1
VITE_cross_1
TZD_cross_1
ASPIRIN_cross_1
BLDEFNAS_cross_1
BLFIBRO_means_1
BLNAS_means_1
BLBALL_means_1
BLSTEATO_means_1
BLINFLAM_means_1
BLPORTAL_means_1
BLLENGTH_means_1
;

proc print data = table1_compare;
*** Importing the Table 1 Data taken from the primary outcome paper;
data table1_data(drop = stat2_tnum2_);
  infile table1 delimiter = ',' MISSOVER DSD  firstobs=2 ls=1080;
  length character $100. table_name $ 30. stat1_char_ stat2_char_ $ 12.;;
  input character         $  table_name $ stat1_char_ $ stat2_char_ $ ;
  ordernum = _n_;
  stat1_char_ = compress(stat1_char_) ;
  stat2_char_ = compress(stat2_char_);
  stat1_ = input(substr(stat1_char_,1,index(stat1_char_, '(')-1),8.);
  stat2_ =input( substr(stat2_char_,1,index(stat2_char_, '(')-1),8.);
  if index (stat1_char_, '%') >0 then do;
   stat3_ = input(substr(stat1_char_,index(stat1_char_, '(')+1, length(stat1_char_)-index(stat1_char_, '(')-2),8.);
  end;
  else stat3_ = input(substr(stat1_char_,index(stat1_char_, '(')+1, length(stat1_char_)-index(stat1_char_, '(')-1),8.);
  if index (stat2_char_, '%') >0 then do;
   stat2_tnum2_ = substr(stat2_char_,index(stat2_char_, '(')+1, length(stat2_char_)-index(stat2_char_, '(')-2);
  end;
  else stat2_tnum2_ = substr(stat2_char_,index(stat2_char_, '(')+1, length(stat2_char_)-index(stat2_char_, '(')-1);
;
  stat2_tnum2_ = translate (stat2_tnum2_, '',')%');
  stat4_ = input(stat2_tnum2_,8.);
  if stat1_char_ ne '';


proc print data = table1_data;
```

```
proc sort data = table1_data;
  by table_name ;
proc sort data = table1_compare;
  by table_name ;

data table1_combine;
  merge  table1_data (in = in2) table1_compare (in = in1);
  by table_name;
  if in1 and in2;

data table1_combine;
  set table1_combine;

  if table_name in ('HEMA_SI' 'INR' 'RATIO') then do;
    stat1_char = compress(put(stat1,8.2) || '(' || put(stat3,8.2) || ')');
    stat2_char = compress(put(stat2,8.2) || '(' || put(stat4,8.2) || ')');
    stat1_diff = compress(put((stat1-stat1_),8.2) || '(' || put((stat3-stat3_),8.2) || ')');
    stat2_diff = compress(put((stat2-stat2_),8.2) || '(' || put((stat4-stat4_),8.2) || ')');
  end;
  else if table_name in ('BILIT_SI' 'CHOL_SI' 'HDL_SI' 'LDL_SI' 'TRIG_SI' 'MCV' 'WBC_SI'  'BICARB_S' 'CAL_SI' 'PHOS_SI' 'PT' 'GLUCB_SI' 'BLFIBRO'
'BLNAS' 'BLBALL' 'BLSTEATO'
                'BLINFLAM' 'BLPORTAL') then do;
    stat1_char = compress(put(stat1,8.1) || '(' || put(stat3,8.1) || ')');
    stat2_char = compress(put(stat2,8.1) || '(' || put(stat4,8.1) || ')');
    stat1_diff = compress(put((stat1-stat1_),8.1) || '(' || put((stat3-stat3_),8.1) || ')');
    stat2_diff = compress(put((stat2-stat2_),8.1) || '(' || put((stat4-stat4_),8.1) || ')');
  end;
  else do;
    stat1_char = compress(put(stat1,8.) || '(' || put(stat3,8.) || ')');
    stat2_char = compress(put(stat2,8.) || '(' || put(stat4,8.) || ')');
    stat1_diff = compress(put((stat1-stat1_),8.) || '(' || put((stat3-stat3_),8.) || ')');
    stat2_diff = compress(put((stat2-stat2_),8.) || '(' || put((stat4-stat4_),8.) || ')');
  end;
  label
   character          = "Reasons for staying in TEDDY"
   stat1_char_        = "Obeticholic [Manuscript]"
   stat1_char         = "Obeticholic [DSIC]      "
   stat1_diff         = "Obeticholic [Difference]"
   stat2_char_        = "Placebo [Manuscript]"
   stat2_char         = "Placebo [DSIC]      "
   stat2_diff         = "Placebo [Difference]"
  ;
proc sort data = table1_combine;
  by ordernum;


*******************************************************************************************;
*************** Check Table 2 *******************************************************;
*******************************************************************************************;

*** Running the baseline_freq on the  categorical variables in the Table 1 manuscript file;
proc sort data = table2;
  by tx;

%baseline_freq1(table2,IMPROVE_        );
%baseline_freq1(table2,RESOLVE         );
%baseline_freq1(table2,IFIBRO          );
%baseline_freq1(table2,IBALL           );
%baseline_freq1(table2,ISTEATO         );
%baseline_freq1(table2,IINFLAM         );
%baseline_freq1(table2,IPORTAL      );

%baseline_means(table2, DFIBRO          );
%baseline_means(table2, DNAS            );
%baseline_means(table2, DBALL           );
%baseline_means(table2, DSTEATO         );
```

```
%baseline_means(table2, DINFLAM         );
%baseline_means(table2, DPORTAL         );

data table2_compare;
set
IMPROVE__CROSS_1
RESOLVE_CROSS_1
IFIBRO_CROSS_1
DFIBRO_means_1
DNAS_means_1

IBALL_CROSS_1
DBALL_means_1

ISTEATO_CROSS_1
DSTEATO_means_1

IINFLAM_CROSS_1
DINFLAM_means_1

IPORTAL_CROSS_1
DPORTAL_means_1
;

proc print data = table2_compare;

*** Importing the Table 2 Data taken from the primary outcome paper;
data table2_data(drop = stat2_tnum2_);
  infile table2 delimiter = ',' MISSOVER DSD  firstobs=1 ls=1080;
  length character $100.  stat1_char_ stat2_char_ $ 12. table_name $ 30.;
  input character         $ table_name $  stat1_char_ $ stat2_char_ $ ;
  ordernum = _n_;
  stat1_char_ = compress(stat1_char_) ;
  stat2_char_ = compress(stat2_char_);
    stat1_ = input(substr(stat1_char_,1,index(stat1_char_, '(')-1),8.);
    stat2_ =input( substr(stat2_char_,1,index(stat2_char_, '(')-1),8.);
    if index (stat1_char_, '%') >0 then do;
     stat3_ = input(substr(stat1_char_,index(stat1_char_, '(')+1, length(stat1_char_)-index(stat1_char_, '(')-2),8.);
    end;
    else stat3_ = input(substr(stat1_char_,index(stat1_char_, '(')+1, length(stat1_char_)-index(stat1_char_, '(')-1),8.);
    if index (stat2_char_, '%') >0 then stat2_tnum2_ = substr(stat2_char_,index(stat2_char_, '(')+1, length(stat2_char_)-index(stat2_char_, '(')-2);
    else  stat2_tnum2_ = substr(stat2_char_,index(stat2_char_, '(')+1, length(stat2_char_)-index(stat2_char_, '(')-1);

    stat2_tnum2_ = translate (stat2_tnum2_, '',')%');
    stat4_ = input(stat2_tnum2_,8.);
    if missing(stat2_num1_) then do;
      stat2_num1_ = stat2_num2_;
      stat2_num2_ = .;
      stat1_num2_ = .;
    end;
proc print data = table2_data;

proc sort data = table2_data;
  by table_name ;
proc sort data = table2_compare;
  by table_name ;

data table2_combine;
  merge  table2_data (in = in2) table2_compare (in = in1);
  by table_name;
  if in1 and in2;

data table2_combine;
  set table2_combine;

  if table_name in ('DFIBRO' 'DNAS' 'DBALL' 'DSTEATO' 'DINFLAM' 'DPORTAL' ) then do;
```

```
   stat1_char = compress(put(stat1,8.1) || '(' || put(stat3,8.1) || ')');
   stat2_char = compress(put(stat2,8.1) || '(' || put(stat4,8.1) || ')');
   stat1_diff = compress(put((stat1-stat1_),8.1) || '(' || put((stat3-stat3_),8.1) || ')');
   stat2_diff = compress(put((stat2-stat2_),8.1) || '(' || put((stat4-stat4_),8.1) || ')');
  end;
  else do;
   stat1_char = compress(put(stat1,8.) || '(' || put(stat3,8.) || ')');
   stat2_char = compress(put(stat2,8.) || '(' || put(stat4,8.) || ')');
   stat1_diff = compress(put((stat1-stat1_),8.) || '(' || put((stat3-stat3_),8.) || ')');
   stat2_diff = compress(put((stat2-stat2_),8.) || '(' || put((stat4-stat4_),8.) || ')');
  end;
  label
   character          = "Reasons for staying in TEDDY"
   stat1_char_        = "Obeticholic [Manuscript]"
   stat1_char         = "Obeticholic [DSIC]      "
   stat1_diff         = "Obeticholic [Difference]"
   stat2_char_        = "Placebo [Manuscript]"
   stat2_char         = "Placebo [DSIC]      "
   stat2_diff         = "Placebo [Difference]"
  ;
proc sort data = table2_combine;
  by ordernum;

*******************************************************************************************;
*************** Check Table 3 ******************************************************;
*******************************************************************************************;

proc sort data  = table3;
  by tx;

%baseline_means(table3,ALT_D       );
%baseline_means(table3,AST_D       );
%baseline_means(table3,ALKA_D      );
%baseline_means(table3,GGT_D       );
%baseline_means(table3,BILIT_S1    );
%baseline_means(table3,CHOL_SI0    );
%baseline_means(table3,HDL_SI_D    );
%baseline_means(table3,LDL_SI_D    );
%baseline_means(table3,TRIG_SI0    );
%baseline_means(table3,HEMO_SI0    );
%baseline_means(table3,HEMA_SI0    );
%baseline_means(table3,MCV_D       );
%baseline_means(table3,WBC_D       );
%baseline_means(table3,PLATELE4    );
%baseline_means(table3,BICARB_1    );
%baseline_means(table3,CAL_SI_D    );
%baseline_means(table3,PHOS_SI0    );
%baseline_means(table3,CR_SI_D     );
%baseline_means(table3,URIC_SI0    );
%baseline_means(table3,ALB_SI_D    );
%baseline_means(table3,PROT_SI0    );
%baseline_means(table3,PT_D        );
%baseline_means(table3,INR_D       );
%baseline_means(table3,GLUCB_S1    );
%baseline_means(table3,INSUB_S1    );
%baseline_means(table3,HOMA_SI0    );
%baseline_means(table3,HBA1C_S1    );
%baseline_means(table3,WEIGHT_D    );
%baseline_means(table3,BMI_D       );
%baseline_means(table3,WAISTCM0    );
%baseline_means(table3,RATIO_D     );
%baseline_means(table3,SBP_D       );
%baseline_means(table3,DBP_D       );
%baseline_means(table3,SF36PHY1    );
%baseline_means(table3,SF36MEN1    );
```

```
data table3_compare;
set
ALT_D_means_1
AST_D_means_1
ALKA_D_means_1
GGT_D_means_1
BILIT_S1_means_1
CHOL_SI0_means_1
HDL_SI_D_means_1
LDL_SI_D_means_1
TRIG_SI0_means_1
HEMO_SI0_means_1
HEMA_SI0_means_1
MCV_D_means_1
WBC_D_means_1
PLATELE4_means_1
BICARB_1_means_1
CAL_SI_D_means_1
PHOS_SI0_means_1
CR_SI_D_means_1
URIC_SI0_means_1
ALB_SI_D_means_1
PROT_SI0_means_1
PT_D_means_1
INR_D_means_1
GLUCB_S1_means_1
INSUB_S1_means_1
HOMA_SI0_means_1
HBA1C_S1_means_1
WEIGHT_D_means_1
BMI_D_means_1
WAISTCM0_means_1
RATIO_D_means_1
SBP_D_means_1
DBP_D_means_1
SF36PHY1_means_1
SF36MEN1_means_1
;


proc print data = table3_compare;


*** Importing the Table 3 Data taken from the primary outcome paper;
data table3_data(drop = or_char_  p_ );
  infile table3 delimiter = ',' MISSOVER DSD  firstobs=2 ls=1080;
  length character $100.  stat1_char_ stat2_char_ $ 12. table_name $ 30.;
  input character          $ table_name $  stat1_char_ $ stat2_char_ $ or_char_ $ p_ $;
  ordernum = _n_;
  stat1_char_ = compress(stat1_char_) ;
  stat2_char_ = compress(stat2_char_);
  stat1_ = input(substr(stat1_char_,1,index(stat1_char_, '(')-1),8.);
  stat2_ =input( substr(stat2_char_,1,index(stat2_char_, '(')-1),8.);
  stat3_ = input(substr(stat1_char_,index(stat1_char_, '(')+1, length(stat1_char_)-index(stat1_char_, '(')-1),8.);
  stat4_ = input(substr(stat2_char_,index(stat2_char_, '(')+1, length(stat2_char_)-index(stat2_char_, '(')-1),8.);

proc print data = table3_data;

proc print data = table3_data;

proc sort data = table3_data;
  by table_name ;
proc sort data = table3_compare;
  by table_name ;

data table3_combine;
```

```
  merge  table3_data (in = in2) table3_compare (in = in1);
  by table_name;
  if in1 and in2;

data table3_combine;
  set table3_combine;

  if table_name in ('HDL_SI_D' 'LDL_SI_D' 'TRIG_SI0' 'HEMA_SI0' 'CAL_SI_D' 'PHOS_SI0' 'INR_D' 'RATIO_D' 'CHOL_SI0') then do;
    stat1_char = compress(put(stat1,8.2) || '(' || put(stat3,8.2) || ')');
    stat2_char = compress(put(stat2,8.2) || '(' || put(stat4,8.2) || ')');
    stat1_diff = compress(put((stat1-stat1_),8.2) || '(' || put((stat3-stat3_),8.2) || ')');
    stat2_diff = compress(put((stat2-stat2_),8.2) || '(' || put((stat4-stat4_),8.2) || ')');
  end;

  else if table_name in ('BILIT_S1'  'HEMO_SI0' 'MCV_D' 'WBC_D' 'BICARB_1' 'CR_SI_D' 'ALB_SI_D' 'PROT_SI0' 'PT_D' 'GLUCB_S1' 'HBA1C_S1'
'WEIGHT_D' 'BMI_D' 'WAISTCM0') then do;
    stat1_char = compress(put(stat1,8.1) || '(' || put(stat3,8.1) || ')');
    stat2_char = compress(put(stat2,8.1) || '(' || put(stat4,8.1) || ')');
    stat1_diff = compress(put((stat1-stat1_),8.1) || '(' || put((stat3-stat3_),8.1) || ')');
    stat2_diff = compress(put((stat2-stat2_),8.1) || '(' || put((stat4-stat4_),8.1) || ')');
  end;
  else do;
    stat1_char = compress(put(stat1,8.) || '(' || put(stat3,8.) || ')');
    stat2_char = compress(put(stat2,8.) || '(' || put(stat4,8.) || ')');
    stat1_diff = compress(put((stat1-stat1_),8.) || '(' || put((stat3-stat3_),8.) || ')');
    stat2_diff = compress(put((stat2-stat2_),8.) || '(' || put((stat4-stat4_),8.) || ')');
  end;
  label
   character          = "Reasons for staying in TEDDY"
   stat1_char_        = "Obeticholic [Manuscript]"
   stat1_char         = "Obeticholic [DSIC]      "
   stat1_diff         = "Obeticholic [Difference]"
   stat2_char_        = "Placebo [Manuscript]"
   stat2_char         = "Placebo [DSIC]      "
   stat2_diff         = "Placebo [Difference]"
  ;
proc sort data = table3_combine;
  by ordernum;

*********************************************************************************************;
*************** Check Table 4 ******************************************************;
*********************************************************************************************;
proc freq data = table4;
  tables tx/list missing;

 proc freq data = table4;
   tables tx * AE * AETYPE;
  ods output CrossTabFreqs = table4_cross;
proc print data = table4_cross;

 data table4_cross;
   set table4_cross;
   if ae ne '' and  Frequency ne 0 and AETYPE ne '';;

proc sort data = table4_cross;
  by AE       AETYPE  tx ;

  data table4_cross(keep =  AE       AETYPE  tx Frequency count);
    set table4_cross;
    by AE       AETYPE    ;
    retain count 0;
    if first.AETYPE then count = 0;
    count = count+1;

proc print data = table4_cross;
```

```
 data table4_cross_1(drop = Frequency    count i );
   set table4_cross;
   by AE       AETYPE;
   array temp1(2) stat1 stat2 ;
    retain stat1-stat2;
   if first.AETYPE then do i = 1 to 2;
     temp1(i) = .;
   end;
     temp1(count) = Frequency;

   if last.AETYPE;

data table4_cross_1;
  set table4_cross_1;
  length table_name $ 30.;
  table_name = compress(ae) || '/' || compress(aetype);

 proc print data = table4_cross_1;




proc freq data = table4;
  tables tx * AE * GRADE ;
  where ae = 'Pruritus';
 ods output CrossTabFreqs = Pruritus_cross;

data Pruritus_cross;
  set Pruritus_cross;
  if ae ne '';

proc print data = Pruritus_cross;

proc sort data = Pruritus_cross;
  by grade tx ;

  data Pruritus_cross(keep =  AE       GRADE  tx Frequency count);
    set Pruritus_cross;
    by grade tx ;
    retain count 0;
    if first.grade then count = 0;
    count = count+1;

proc print data = Pruritus_cross;

  data Pruritus_cross_1(drop = Frequency    count i);
    set Pruritus_cross;
    by grade tx;
    array temp1(2) stat1 stat2 ;
     retain stat1-stat2;
    if first.grade then do i = 1 to 2;
      temp1(i) = .;
    end;
      temp1(count) = Frequency;

    if last.grade;

data Pruritus_cross_1;
  set Pruritus_cross_1;
  length table_name $ 30.;
  if grade = . then table_name = 'Pruritus';
  else if grade = 1 then table_name = 'grade 1';
  else if grade = 2 then table_name = 'grade 2';
  else if grade = 3 then table_name = 'grade 3';

  proc print data = Pruritus_cross_1;
```

```
data table4_compare;
  set table4_cross_1 Pruritus_cross_1;

proc print data = table4_compare;

*** Importing the Table 4 Data taken from the primary outcome paper;
data table4_data;
  infile table4 delimiter = ',' MISSOVER DSD  firstobs=1 ls=1080;
  length character $100. table_name $ 30. ;
  input character          $  table_name $ stat1_  stat2_ p $;
  ordernum = _n_;

proc print data = table4_data;

proc sort data = table4_data;
  by table_name ;
proc sort data = table4_compare;
  by table_name ;

data table4_combine;
  merge  table4_data (in = in2) table4_compare (in = in1);
  by table_name;
  if in1 and in2;

data table4_combine;
  set table4_combine;

    stat1_diff = stat1-stat1_;
    stat2_diff = stat2-stat2_;

  label
   character      = "Reasons for staying in TEDDY"
   stat1_         = "Obeticholic [Manuscript]"
   stat1          = "Obeticholic [DSIC]      "
   stat1_diff     = "Obeticholic [Difference]"
   stat2_         = "Placebo [Manuscript]"
   stat2          = "Placebo [DSIC]      "
   stat2_diff     = "Placebo [Difference]"
  ;
proc sort data = table4_combine;
  by ordernum;


*** Outputting the data to a csv format to be added to the DSIC;
ods csv file = out_t1;
run;

proc print data = table1_combine NOOBS label;
                var
character
stat1_char_
stat1_char
stat1_diff
stat2_char_
stat2_char
stat2_diff

;
                title "DSIC Check of Table 1 Baseline characteristics of the study population";
```