Dataset Integrity Check for the Nonalcoholic Steatohepatitis (NASH) Nonalcoholic fatty liver disease (NAFLD) Pediatric Data Files

Prepared by Jane Wang
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton MD 20705
April 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

Nonalcoholic fatty liver disease (NAFLD), which affects 10%-30% of the general U.S. population, is the most prevalent liver disease in American children, and can progress to significant fibrosis and cirrhosis. The Nonalcoholic Steatohepatitis Clinical Research Network (NASH CRN) was initiated by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) in 2002 to conduct multicenter, collaborative studies on the etiology, contributing factors, natural history, complications, and treatment of NASH.

The aim of this study was to determine the associations of readily available demographic, clinical, and laboratory variables with the diagnosis of NASH and its key histological features, and determine the ability of these variables to predict the severity of nonalcoholic fatty liver disease (NAFLD) in pediatric cases.

Note that this data package only contains the analysis datasets that were used for the Lavine et al [1] publication.

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 “peds” dataset. This dataset was created by the DCC from the forms datasets, which are not included.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Lavine et al [1] JAMA, April 27, 2011—Vol 305, No. 16. To verify the integrity of the "peds" dataset, descriptive statistics of characteristics were computed.

5 Results

Table 1 in the publication [1] Sample Characteristics for Children Enrolled in NASH CRN Eligible for Study Inclusion. Table B compares the results calculated from the archived data file to the results published in Table 1. The results of the replication are the same as the published results.

Table 2 in the publication [1] Predictors of NAFLD Pattern (Excluding Borderline Zone 1). Table C compares the results calculated from the archived data file to the results published in Table 2. The results of the replication are the same as the published results.

Table 3 in the publication [1] Predictors of Borderline Zone 1 Versus Definite NASH Pattern in Children With Some Fibrosis Present. Table D compares the results calculated from the archived data file to the results published in Table 3. The results of the replication are similar to the published results, within rounding error.

Table 4 in the publication [1] Predictors of Fibrosis Stage. Table E compares the results calculated from the archived data file to the results published in Table 3. The results of the replication are similar to the published results, within rounding error.

Table 5 in the publication [1] Predictors of NAS. Table F compares the results calculated from the archived data file to the results published in Table 3. The results of the replication are similar to the published results, within rounding error.

6 Conclusions

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

7 References

HEATHER M. PATTON, JOEL E. LAVINE, MARK L. VAN NATTA, JEFFREY B. SCHWIMMER, DAVID KLEINER, JEAN MOLLESTON, and the Nonalcoholic Steatohepatitis Clinical Research Network. Clinical Correlates of Histopathology in Pediatric Nonalcoholic Steatohepatitis. *GASTROENTEROLOGY 2008;135:1961–1971*

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

Table Variable	Variables Used in Replication from the "PEDS" Dataset
Acanthosis nigricans	ACANTH
age	AGE
Albumin (g/dL)	ALB
Alkaline phosphatse (U/L)	ALKA
ALT (U/L)	ALT
ANA (% positive)	ANA
ASMA (% positive)	ASMA
AST (U/L)	AST
biopsy length	BLENGTH
BMI (kg/m2)	BMI
BMI percentile	BMIPCT
GGT(U/L)	GGT
Fasting glucose (mg/dL)	GLUC
Hematocrit	HEMA
Hispanic ethnicity (%)	HISPANIC
HOMA-IR	HOMA
Fasting insulin (mU/mL)	INSU
Biopsy <10 mm	LENG10
Demographics Male	MALE
NAFLD pattern	NAFLD
NAS	NAS
Fibrosis score	NFIBRO
QUICKI	QUICKI
Database	STUDY=1
TONIC	STUDY=3
Tanner stage	TANNER
Percentage of body fat	TOTALFAT
Triglycerides (mg/dL)	TRI
UCSD (%)	UCSD
White blood cell count (1000/mm3)	WBC
White race	WHITE
Fibrosis stage	XFIBRO

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

characteristic	N(%) [Manuscript]	N(%) [DSIC]	N(%) [Difference]
TONIC	136(77)	136.0 (77.0)	0.0 (0.0)
Database	40(23)	40.0 (23.0)	0.0 (0.0)
Male	136(77)	136.0 (77.0)	0.0 (0.0)
Age in y, mean SD (range)	12.4;2.6(6-17)	12.4;2.6(6.0-17.0)	0.0;0.0(0.0-0.0)
White race	128(73)	128.0 (73.0)	0.0 (0.0)
Hispanic ethnicity	104(59)	104.0 (59.0)	0.0 (0.0)
Histology Biopsy length in mm, median pseud	15;6.7(5-53)	15.0;6.7(5.0-53.0)	0.0;0.0(0.0-0.0)
Not NASH	36(20)	36.0 (20.0)	0.0 (0.0)
Borderline zone 3 pattern	26(15)	26.0 (15.0)	0.0 (0.0)
Borderline zone 1 pattern	50(28)	50.0 (28.0)	0.0 (0.0)
Definite NASH	64(36)	64.0 (36.0)	0.0 (0.0)
Fibrosis stage None (score = 0)	45(26)	46.0 (26.0)	-1.0 (0.0)
Fibrosis stage Mild zone 3 only (score = 1)	12(7)	12.0 (7.0)	0.0 (0.0)
Fibrosis stage Moderate zone 3 only (score =	8(5)	8.0 (5.0)	0.0 (0.0)
Fibrosis stage Periportal only (score = 1)	57(32)	57.0 (32.0)	0.0 (0.0)
Fibrosis stage Mild/moderate zone 3 and peri	29(16)	29.0 (16.0)	0.0 (0.0)
Fibrosis stage Bridging (score = 3)	24(14)	24.0 (14.0)	0.0 (0.0)
Fibrosis score, mean SD	1.2(1.0)	1.2 (1.0)	0.0 (0.0)

characteristic	N(%) [Manuscript]	N(%) [DSIC]	N(%) [Difference]
NAS 1	1(1)	1.0 (1.0)	0.0 (0.0)
NAS 2	15(9)	15.0 (9.0)	0.0 (0.0)
NAS 3	33(19)	33.0 (19.0)	0.0 (0.0)
NAS 4	43(24)	43.0 (24.0)	0.0 (0.0)
NAS 5	41(23)	41.0 (23.0)	0.0 (0.0)
NAS 6	27(15)	27.0 (15.0)	0.0 (0.0)
NAS 7	16(9)	16.0 (9.0)	0.0 (0.0)
NAS, mean SD	4.4(1.4)	4.4 (1.4)	0.0 (0.0)
BMI in kg/m2, median pseudo-sigmaa (range)	33;5.2(18.2-57.9)	32.9;5.2(18.2-57.9)	0.0;0.0(0.0-0.0)
BMI age-sex percentile, median pseudo-sigmaa	99.1;0.8(89.6-100.0)	99.1;0.8(89.6-100.0	0.0;0.0(0.0-0.0)
Body fat percentage, median pseudo-sigmaa	44;7(28-59)	43.7;6.8(27.5-59.2)	0.0;0.0(1.0-0.0)
Tanner stage, mean SD (range)	2.5;1.4(1-5)	2.5;1.4(1.0-5.0)	0.0;0.0(0.0-0.0)
Acanthosis nigricans	133(76)	133.0 (76.0)	0.0 (0.0)

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

characteristic	Not NASH (n=36) [Manuscript]	Not NASH (n=36) [DSIC]	Not NASH (n=36) [Difference]
Male (%)	67	67	0
Mean age (y)	13.1	13.1	0
White race (%)	72	72	0
Hispanic ethnicity (%)	58	58	0
Clinic site: UCSD (%)	50	50	0
ALT, U/L	76	77	-1
AST, U/L	44	44	0
Alkaline phosphatase, U/L	188	189	-1
GGT, U/L	32	32	0
Fasting glucose, mg/dL	89	89	0
Fasting insulin, mU/mL	27	27	0
HOMA-IR	5.5	6	-1
QUICKI	0.299	0.299	0
ANA (% positive)	14	14	0
ASMA (% positive)	28	28	0
Anthropometric, median values BMI (kg/m2)	33	33	0
BMI percentile	98.9	99	0
Percentage of body fat	41	41	0
Tanner stage (mean)	3.3	3.3	0
Median biopsy length (mm)	14	14	0
<10 mm (%)	16	19	-3
Fibrosis score (mean)	0.4	0.4	0

characteristic	Borderline zone 3, adult type (n=26) [Manuscript]	Borderline zone 3, adult type (n=26) [DSIC]	Borderline zone 3, adult type (n=26) [Difference]
Male (%)	73	73	0
Mean age (y)	12.4	12.4	0
White race (%)	69	69	0
Hispanic ethnicity (%)	62	62	0
Clinic site: UCSD (%)	42	42	0
ALT, U/L	90	90	0
AST, U/L	57	57	0
Alkaline phosphatase, U/L	225	225	0
GGT, U/L	36	36	0
Fasting glucose, mg/dL	89	89	0
Fasting insulin, mU/mL	27	27	0
HOMA-IR	6.1	6.1	0
QUICKI	0.295	0.295	0
ANA (% positive)	12	12	0
ASMA (% positive)	24	24	0
Anthropometric, median values BMI (kg/m2)	33	33	0
BMI percentile	99.2	99.2	0
Percentage of body fat	42	42	0
Tanner stage (mean)	2.5	2.5	0
Median biopsy length (mm)	13	13	0
<10 mm (%)	27	27	0
Fibrosis score (mean)	1	1	0

characteristic	Definite NASH (n=64) [Manuscript]	Definite NASH (n=64) [DSIC]	Definite NASH (n=64) [Difference]
Male (%)	78	78	0
Mean age (y)	13.1	13.1	0
White race (%)	77	77	0
Hispanic ethnicity (%)	50	50	0
Clinic site: UCSD (%)	44	44	0
ALT, U/L	102	102	0
AST, U/L	64	64	0
Alkaline phosphatase, U/L	214	215	-1
GGT, U/L	49	49	0
Fasting glucose, mg/dL	89	89	0
Fasting insulin, mU/mL	36	36	0
HOMA-IR	8	8	0
QUICKI	0.285	0.285	0
ANA (% positive)	22	22	0
ASMA (% positive)	38	38	0
Anthropometric, median values BMI (kg/m2)	33	33	0
BMI percentile	99	99	0
Percentage of body fat	44	44	0
Tanner stage (mean)	2.8	2.8	0
Median biopsy length (mm)	15	15	0
<10 mm (%)	9	9	0
Fibrosis score (mean)	1.6	1.6	0

Table D: Comparison of values computed in integrity check to reference article Table 3 values.

characteristic	Borderline zone 1, Pediatric type (n=49) [Manuscript]	Borderline zone 1, Pediatric type (n=49) [DSIC]	Borderline zone 1, Pediatric type (n=49) [Difference]	Definite NASH (n=54) [Manuscript]	Definite NASH (n=54) [DSIC]	Definite NASH (n=54) [Difference]
Male (%)	86	86	0	76	76	0
Mean age (y)	11.1	11.1	0	13	13	0
White race (%)	73	73	0	74	74	0
Hispanic ethnicity (%)	69	69	0	46	46	0
Clinic site: UCSD (%)	61	61	0	41	41	0
ALT (U/L)	84	84	0	101	101	0
AST (U/L)	53	53	0	64	64	0
Alkaline phosphatase (U/L)	267	267	0	216	216	0
GGT (U/L)	39	39	0	47	47	0
Triglycerides (mg/dL)	109	109	0	142	143	-1
Fasting glucose (mg/dL)	88	88	0	88	89	-1
Fasting insulin (mg/dL)	23	23	0	36	36	0
HOMA-IR	4.8	4.8	0	7.7	7.7	0
QUICKI	0.304	0.304	0	0.286	0.286	0
ANA (% positive)	18	18	0	22	22	0
ASMA (% positive)	31	31	0	37	37	0
values BMI (kg/m2)	31	31	0	33	33	0
BMI percentile	99.3	99.3	0	99.1	99.1	0
Percentage of body fat	46	46	0	44	44	0
Tanner stage (mean)	1.8	1.8	0	2.7	2.7	0
Median biopsy length (mm)	16	16	0	15	15	0
<10 mm (%)	4	4	0	11	11	0
Fibrosis score (mean)	1.3	1.3	0	1.9	1.9	0

Table E: Comparison of values computed in integrity check to reference article Table 4 values.

characteristic	None (n=46) [Manuscript]	None (n=46) [DSIC]	None (n=46) [Difference]	Mild (n=77) [Manuscript]	Mild (n=77) [DSIC]	Mild (n=77) [Difference]
Male (%)	72	72	0	79	79	0
Mean age (y)	13.1	13.1	0	12.1	12.1	0
White race (%)	74	74	0	71	71	0
Hispanic ethnicity (%)	63	63	0	69	69	0
Clinic site: UCSD (%)	52	52	0	53	53	0
ALT (U/L)	84	85	-1	82	82	0
AST (U/L)	48	48	0	52	52	0
Alkaline phosphatase (U/L)	202	202	0	250	250	0
GGT (U/L)	38	38	0	32	32	0
Albumin (g/dL)	4.5	4.5	0	4.3	4.3	0
White blood cell count (1000/mm3)	7.2	7.2	0	7.6	7.6	0
Hematocrit (%)	42	42	0	41	41	0
Fasting insulin (mU/mL)	30	30	0	24	24	0
HOMA-IR	6.7	6.7	0	5.2	5.2	0
QUICKI	0.291	0.291	0	0.301	0.301	0
ANA (% positive)	20	20	0	19	19	0
ASMA (% positive)	27	27	0	32	32	0
BMI (kg/m2)	34	34	0	32	32	0
BMI percentile	99	99	0	99.1	99.1	0
Percentage of body fat	41	41	0	44	44	0
Tanner stage (mean)	3	3	0	2.3	2.3	0
Median biopsy length (mm)	14	15	-1	15	15	0
<10 mm (%)	17	17	0	6	6	0
Definite NASH (%)	22	22	0	26	26	0
NAS (mean)	4.1	4.1	0	4.2	4.2	0

characteristic	Moderate (n=29) [Manuscript]	Moderate (n=29) [DSIC]	Moderate (n=29) [Difference]	Bridging (n=24) [Manuscript]	Bridging (n=24) [DSIC]	Bridging (n=24) [Difference]
Male (%)	72	72	0	88	88	0
Mean age (y)	13	13	0	11.5	11.5	0
White race (%)	76	76	0	74	74	0
Hispanic ethnicity (%)	31	31	0	54	54	0
Clinic site: UCSD (%)	38	38	0	46	46	0
ALT (U/L)	96	96	0	126	126	0
AST (U/L)	56	56	0	76	76	0
Alkaline phosphatase (U/L)	218	218	0	263	263	0
GGT (U/L)	52	52	0	54	55	-1
Albumin (g/dL)	4.2	4.2	0	4.4	4.4	0
White blood cell count (1000/mm3)	7.8	7.8	0	8.3	8.3	0
Hematocrit (%)	40	40	0	40	40	0
Fasting insulin (mU/mL)	33	33	0	33	33	0
HOMA-IR	7.5	7.5	0	7.6	7.6	0
QUICKI	0.287	0.287	0	0.287	0.287	0
ANA (% positive)	24	24	0	4	4	0
ASMA (% positive)	37	37	0	32	32	0
BMI (kg/m2)	34	34	0	32	32	0
BMI percentile	99	99	0	99.2	99.2	0
Percentage of body fat	45	45	0	44	44	0
Tanner stage (mean)	2.7	2.7	0	2	2	0
Median biopsy length (mm)	13	13	0	15	15	0
<10 mm (%)	17	17	0	17	17	0
Definite NASH (%)	72	72	0	54	54	0
NAS (mean)	5	5	0	5.1	5.1	0

Table F: Comparison of values computed in integrity check to reference article Table 5 values.

	NASH 1-3 (n=49) [Manuscript]	**NASH 1-3 (n=49) [DSIC]**	**NASH 1-3 (n=49) [Diff]**	**NASH 4-5 (n=84) [Manuscript]**	**NASH 4-5 (n=84) [DSIC]**	**NASH 4-5 (n=84) [Diff]**	**NASH 6-7 (n=43) [Manuscript]**	**NASH 6-7 (n=43) [DSIC]**	**NASH 6-7 (n=43) [Diff]**
Demographic Male (%)	73	73	0	75	75	0	86	86	0
Mean age (y)	12.4	12.4	0	12.4	12.4	0	12.5	12.5	0
White race (%)	67	67	0	75	75	0	77	77	0
Hispanic ethnicity (%)	55	55	0	62	62	0	58	58	0
UCSD (%)	45	45	0	49	49	0	56	56	0
ALT (U/L)	74	74	0	87	87	0	114	114	0
AST (U/L)	42	42	0	56	56	0	66	66	0
Alkaline phosphatse (U/L)	239	239	0	236	237	-1	257	257	0
GGT(U/L)	29	29	0	36	36	0	53	53	0
Fasting glucose (mg/dL)	87	87	0	88	88	0	89	89	0
Fasting insulin (mU/mL)	23	23	0	30	30	0	37	37	0
HOMA-IR	5	5	0	6.1	6.1	0	8.5	8.5	0
QUICKI	0.303	0.303	0	0.295	0.295	0	0.293	0.283	0.01
ANA (% positive)	16	16	0	17	17	0	23	23	0
ASMA (% positive)	20	20	0	31	31	0	45	45	0
Median biopsy length	14	14	0	15	15	0	15	15	0
<10 mm (%)	14	14	0	14	14	0	7	7	0
Definite NASH (%)	6	6	0	31	31	0	81	81	0
Fibrosis score (mean)	0.8	0.8	0	1.2	1.3	-0.1	1.4	1.4	0

Attachment A: SAS Code

```
/*************************************************************************************************
**********************
***Program:  /prj/niddk/ims_analysis/NASH/prog_initial_analysis/nash_pediatric_integrity_check.sas;
***Programmer: Jane Wang
***Date Created: 2/21/2014
***        The numbers in Tables 1,2,3,4 and 5 of the primary outcome paper will compared to the
NASH data received;
*************************************************************************************************
**********************/

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

options nofmterr;

*** Reading in the analysis datasets used for the DSIC;
libname inlib3 xport
"/prj/niddk/ims_analysis/NASH/private_orig_data/NASHCRN_Data_Sharing_PediatricNAFLDDatabase_Gastr
oenterology_2008/Datasets/peds.xpt";
proc copy in =inlib3  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_pediatric_table1.csv';
FILENAME table2  '/prj/niddk/ims_analysis/NASH/private_created_data/nash_pediatric_table2.csv';
FILENAME table3  '/prj/niddk/ims_analysis/NASH/private_created_data/nash_pediatric_table3.csv';
FILENAME table4  '/prj/niddk/ims_analysis/NASH/private_created_data/nash_pediatric_table4.csv';
FILENAME table5  '/prj/niddk/ims_analysis/NASH/private_created_data/nash_pediatric_table5.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_pediatric_table1_dsic.csv';
FILENAME out_t2
'/prj/niddk/ims_analysis/NASH/private_created_data/nash_pediatric_table2_dsic.csv';
FILENAME out_t3
'/prj/niddk/ims_analysis/NASH/private_created_data/nash_pediatric_table3_dsic.csv';
FILENAME out_t4
'/prj/niddk/ims_analysis/NASH/private_created_data/nash_pediatric_table4_dsic.csv';
FILENAME out_t5
'/prj/niddk/ims_analysis/NASH/private_created_data/nash_pediatric_table5_dsic.csv';


%macro baseline_freq1way(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.;
       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 OneWayFreqs = work.&var_name._cross;
  run;

  data &var_name._cross(drop  =Frequency cumFrequency Percent cumPercent );
    set &var_name._cross;
    length table_name $ 30.;
         table_name =  upcase("&var_name.") ;
    first_stat = round(Frequency,1);
    second_stat = round(Percent,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 MEDIAN  Min Max QRANGE ;;
       var &var_name.;
       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;

  data &var_name._means;
    set &var_name._means;
    length table_name $ 30.;
         table_name =  upcase("&var_name.") ;
    third_stat =  round(&var_name._min,0.1);
    fourth_stat = round(&var_name._max,0.1);
    first_mean_stat    =  round(&var_name._Mean,0.1);
    first_median_stat  =  round(&var_name._Median,0.1);
    second_sd_stat     =  round(&var_name._StdDev,0.1);
    second_pseudo_stat =  round(&var_name._QRange/1.349,0.1);

%mend;

%macro baseline_freq(dataset_name,var_name,by_var);

       *** Creating a frequency table in the format of Table 1 in the primary outcome paper;
  proc freq data = &dataset_name;
       table (&var_name.)*&by_var;
       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(drop = table &var_name. _TYPE_    _TABLE_    Frequency    Percent
RowPercent Missing);
    set &var_name._cross;
    if &var_name. = 1 and not missing(&by_var);
    length table_name $ 30.;
         table_name =  upcase("&var_name.") ;

%mend;

%macro baseline_median(dataset_name,var_name,by_var);


       *** Creating a means table in the format of Table 1 in the primary outcome paper that
contain the median 25th percentile and 75th percentile;
  proc means data = &dataset_name mean MEDIAN  ;
    var &var_name.;
    by &by_var;

       *** Outputting the statisitics to the work.&var_name._summary dataset using the ODS
output;
  ods output Summary = work.&var_name._means;

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

  proc transpose data=&var_name._means out=&var_name._transpose  ;
    by  &by_var;
  data &var_name._transpose (keep =  &by_var      table_name      _LABEL_            COL1) ;
    set &var_name._transpose  ;
    length table_name $ 30.;
         table_name =  upcase("&var_name.") ;

%mend;

*******************************************************************************;
*************** Check Table 1 ************************************************;
*******************************************************************************;

%baseline_freq1way(peds, study    );
%baseline_freq1way(peds, male    );
%baseline_freq1way(peds, white    );
```

```
%baseline_freq1way(peds, hispanic    );
%baseline_freq1way(peds, NAFLD    );
%baseline_freq1way(peds, XFIBRO    );
%baseline_freq1way(peds, nas    );
%baseline_freq1way(peds, acanth    );

data study_cross ;  set study_cross ;  table_name = compress(table_name) || compress(study);
data NAFLD_cross ;  set NAFLD_cross ;  table_name = compress(table_name) || compress(NAFLD);
data nas_cross   ;  set nas_cross   ;  table_name = compress(table_name) || compress(nas);
data XFIBRO_cross;  set XFIBRO_cross;  table_name = compress(table_name) || compress(XFIBRO);


%baseline_means(peds, age     );
%baseline_means(peds, TANNER  );
%baseline_means(peds, nas     );
%baseline_means(peds, blength );
%baseline_means(peds, nfibro  );
%baseline_means(peds, bmi     );
%baseline_means(peds, bmipct  );
%baseline_means(peds, totalfat);

data table1_compare(keep = table_name first_stat second_stat third_stat fourth_stat);
  set age_means      (keep = table_name first_mean_stat   second_sd_stat     third_stat
fourth_stat rename = (first_mean_stat  =first_stat second_sd_stat    =second_stat     ))
      TANNER_means   (keep = table_name first_mean_stat   second_sd_stat     third_stat
fourth_stat rename = (first_mean_stat  =first_stat second_sd_stat    =second_stat     ))
      nfibro_means   (keep = table_name first_mean_stat   second_sd_stat
rename = (first_mean_stat=first_stat second_sd_stat=second_stat))
      nas_means      (keep = table_name first_mean_stat   second_sd_stat
rename = (first_mean_stat  =first_stat second_sd_stat    =second_stat     ))
      blength_means  (keep = table_name first_median_stat second_pseudo_stat third_stat
fourth_stat rename = (first_median_stat=first_stat second_pseudo_stat=second_stat))
      bmi_means      (keep = table_name first_median_stat second_pseudo_stat third_stat
fourth_stat rename = (first_median_stat=first_stat second_pseudo_stat=second_stat))
      bmipct_means   (keep = table_name first_median_stat second_pseudo_stat third_stat
fourth_stat rename = (first_median_stat=first_stat second_pseudo_stat=second_stat))
      totalfat_means (keep = table_name first_median_stat second_pseudo_stat third_stat
fourth_stat rename = (first_median_stat=first_stat second_pseudo_stat=second_stat))
      study_cross
      male_cross (where=( MALE=1))
      white_cross (where=( white=1))
      hispanic_cross (where=( hispanic=1))
      NAFLD_cross
      XFIBRO_cross
      nas_cross
      acanth_cross (where=( acanth=1))
;
  table_name = upcase(table_name);

proc sort data = table1_compare nodupkey;  by table_name;

*** Importing the Table 1 Data taken from the primary outcome paper;
data table1_data;
  infile table1 delimiter = ',' MISSOVER DSD  firstobs=2 ls=1080;
  length characteristic $45 char_stat $23 table_name $ 30.;
  input
  characteristic $ table_name $  stat1       stat2   stat3   stat4 char_stat $
;
  if lengthn(characteristic) NE 0 then output table1_data;

data table1_data;
  set table1_data;
  sort_order = _n_;
  table_name = upcase(table_name);
  char_stat = compress(char_stat);

proc sort data = table1_data nodupkey;  by table_name;

data combined_table1_dataset;
  merge table1_compare
        table1_data    ;
```

```
  by table_name;
  diff1 = round((stat1 - first_stat),  1);
  diff2 = round((stat2 - second_stat), 1);
  if third_stat ne . then diff3 = round((stat3 - third_stat),  1);
  if fourth_stat ne . then diff4 = round((stat4 - fourth_stat), 1);

  if third_stat= . and fourth_stat = . then compare_char = strip(put(first_stat,8.1)) || " (" ||
strip(put(second_stat,8.1)) || ")";
  else compare_char = strip(put(first_stat,8.1)) || ";" || strip(put(second_stat,8.1)) || "(" ||
strip(put(third_stat,8.1)) || "-" || strip(put(fourth_stat,8.1)) || ")";

  if diff3= . and diff4 = . then diff_char = strip(put(diff1,8.1)) || " (" ||
strip(put(diff2,8.1)) || ")";
  else diff_char = strip(put(diff1,8.1)) || ";" || strip(put(diff2,8.1)) || "(" ||
strip(put(diff3,8.1)) || "-" || strip(put(diff4,8.1)) || ")";
    label
     char_stat          = "N(%) [Manuscript]"
     compare_char       = "N(%) [DSIC]      "
     diff_char          = "N(%) [Difference]"
;


proc freq data = combined_table1_dataset;
  tables diff1-diff4/list missing;

*** Outputting the dataset to a csv file to be added to the DSIC;

proc sort data = combined_table1_dataset; by sort_order;
ods csv file = out_t1;

run;

proc print data = combined_table1_dataset NOOBS label;
       var characteristic table_name  char_stat compare_char diff_char;
       title3 "DSIC Check of Table 1";

run;

ods csv close;




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

data peds_table2;
  set peds;
  if NAFLD ne '1b';

proc sort data = peds_table2;
  by NAFLD;

%baseline_freq(peds_table2, male,NAFLD     );
%baseline_freq(peds_table2, white,NAFLD     );
%baseline_freq(peds_table2, hispanic,NAFLD     );
%baseline_freq(peds_table2, ucsd,NAFLD     );
%baseline_freq(peds_table2, ana,NAFLD     );
%baseline_freq(peds_table2, asma,NAFLD     );
%baseline_freq(peds_table2, leng10,NAFLD     );

data table2_freq(rename = (ColPercent = stat));
  set
 male_cross
 white_cross
 hispanic_cross
 ucsd_cross
 ana_cross
 asma_cross
 leng10_cross
```

```
;

%baseline_median(peds_table2, age      ,NAFLD     );
%baseline_median(peds_table2, tanner   ,NAFLD     );
%baseline_median(peds_table2, nfibro   ,NAFLD     );
%baseline_median(peds_table2, alt      ,NAFLD     );
%baseline_median(peds_table2, ast      ,NAFLD     );
%baseline_median(peds_table2, alka     ,NAFLD     );
%baseline_median(peds_table2, ggt      ,NAFLD     );
%baseline_median(peds_table2, gluc     ,NAFLD     );
%baseline_median(peds_table2, insu     ,NAFLD     );
%baseline_median(peds_table2, homa     ,NAFLD     );
%baseline_median(peds_table2, quicki   ,NAFLD     );
%baseline_median(peds_table2, bmi      ,NAFLD     );
%baseline_median(peds_table2, bmipct   ,NAFLD     );
%baseline_median(peds_table2, totalfat ,NAFLD     );
%baseline_median(peds_table2, blength  ,NAFLD     );

data table2_means(drop = _LABEL_ rename = (COL1 = stat));
  set
age_transpose     (where = (_label_ = 'Mean')     )
tanner_transpose  (where = (_label_ = 'Mean')     )
nfibro_transpose  (where = (_label_ = 'Mean')     )
alt_transpose     (where = (_label_ = 'Median')   )
ast_transpose     (where = (_label_ = 'Median')   )
alka_transpose    (where = (_label_ = 'Median')   )
ggt_transpose     (where = (_label_ = 'Median')   )
gluc_transpose    (where = (_label_ = 'Median')   )
insu_transpose    (where = (_label_ = 'Median')   )
homa_transpose    (where = (_label_ = 'Median')   )
quicki_transpose  (where = (_label_ = 'Median')   )
bmi_transpose     (where = (_label_ = 'Median')   )
bmipct_transpose  (where = (_label_ = 'Median')   )
totalfat_transpose(where = (_label_ = 'Median')   )
blength_transpose (where = (_label_ = 'Median')   )
;

data table2_freq_mean;
  set table2_freq table2_means;

proc sort data = table2_freq_mean;
  by  table_name ;

proc transpose data=table2_freq_mean out=table2_compare ;
   by  table_name ;
data table2_compare (drop =    _NAME_                _LABEL_    rename = (col1 = first_stat col2
= second_stat col3 = third_stat));
  set table2_compare;


*** Importing the Table 2 Data taken from the primary outcome paper;
data table2_data;
  infile table2 delimiter = ',' MISSOVER DSD  firstobs=3 ls=1080;
  length characteristic $45 char_statp $23 table_name $ 30.;
  input
  characteristic $ table_name $  stat1       stat2   stat3    char_statp $
;
  if lengthn(characteristic) NE 0 then output table2_data;

data table2_data;
  set table2_data;
  sort_order = _n_;
  table_name = upcase(table_name);

proc sort data = table2_data nodupkey;  by table_name;

data combined_table2_dataset;
  merge table2_compare
        table2_data    ;
  by table_name;
  if table_name = 'QUICKI' then do;
```

```
    first_stat  = round(first_stat,0.001);
    second_stat = round(second_stat,0.001);
    third_stat  = round(third_stat,0.001);

    diff1 = round((stat1 - first_stat),  0.001);
    diff2 = round((stat2 - second_stat), 0.001);
    diff3 = round((stat3 - third_stat),  0.001);
  end;
  else if table_name in ('AGE' 'TANNER' 'NFIBRO') then do;
    first_stat  = round(first_stat,0.1);
    second_stat = round(second_stat,0.1);
    third_stat  = round(third_stat,0.1);

    diff1 = round((stat1 - first_stat),  0.1);
    diff2 = round((stat2 - second_stat), 0.1);
    diff3 = round((stat3 - third_stat),  0.1);
  end;
  else do;
    first_stat  = round(first_stat,1);
    second_stat = round(second_stat,1);
    third_stat  = round(third_stat,1);

    diff1 = round((stat1 - first_stat),  1);
    diff2 = round((stat2 - second_stat), 1);
    diff3 = round((stat3 - third_stat),  1);
  end;

    label
     stat1          = "Not NASH (n=36) [Manuscript]"
     first_stat     = "Not NASH (n=36) [DSIC]      "
     diff1          = "Not NASH (n=36) [Difference]"
     stat2          = "Borderline zone 3, adult type (n=26) [Manuscript]"
     second_stat    = "Borderline zone 3, adult type (n=26) [DSIC]      "
     diff2          = "Borderline zone 3, adult type (n=26) [Difference]"
     stat3          = "Definite NASH (n=64) [Manuscript]"
     third_stat     = "Definite NASH (n=64) [DSIC]      "
     diff3          = "Definite NASH (n=64) [Difference]"
;


proc freq data = combined_table2_dataset;
  tables diff1-diff3/list missing;

*** Outputting the dataset to a csv file to be added to the DSIC;

proc sort data = combined_table2_dataset; by sort_order;
ods csv file = out_t2;

run;

proc print data = combined_table2_dataset NOOBS label;
       var characteristic table_name stat1 first_stat diff1 stat2 second_stat diff2 stat3
third_stat diff3
;
       title3 "DSIC Check of Table 2";

run;

ods csv close;




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

data peds_table3;
  set peds;
  if NAFLD in ('1b','2') and NFIBRO ne 0;

proc sort data = peds_table3;
```

```
  by NAFLD;

%baseline_freq(peds_table3, male,NAFLD     );
%baseline_freq(peds_table3, white,NAFLD     );
%baseline_freq(peds_table3, hispanic,NAFLD     );
%baseline_freq(peds_table3, ucsd,NAFLD     );
%baseline_freq(peds_table3, ana,NAFLD     );
%baseline_freq(peds_table3, asma,NAFLD     );
%baseline_freq(peds_table3, leng10,NAFLD     );

data table3_freq(rename = (ColPercent = stat));
  set
 male_cross
 white_cross
 hispanic_cross
 ucsd_cross
 ana_cross
 asma_cross
 leng10_cross
;

*proc print data = table3_freq;

%baseline_median(peds_table3, age      ,NAFLD     );
%baseline_median(peds_table3, tanner   ,NAFLD     );
%baseline_median(peds_table3, nfibro   ,NAFLD     );
%baseline_median(peds_table3, alt      ,NAFLD     );
%baseline_median(peds_table3, ast      ,NAFLD     );
%baseline_median(peds_table3, alka     ,NAFLD     );
%baseline_median(peds_table3, ggt      ,NAFLD     );
%baseline_median(peds_table3, gluc     ,NAFLD     );
%baseline_median(peds_table3, insu     ,NAFLD     );
%baseline_median(peds_table3, homa     ,NAFLD     );
%baseline_median(peds_table3, quicki   ,NAFLD     );
%baseline_median(peds_table3, bmi      ,NAFLD     );
%baseline_median(peds_table3, bmipct   ,NAFLD     );
%baseline_median(peds_table3, totalfat ,NAFLD     );
%baseline_median(peds_table3, blength  ,NAFLD     );

%baseline_median(peds_table3, TRI      ,NAFLD     );


data table3_means(drop = _LABEL_ rename = (COL1 = stat));
  set
age_transpose     (where = (_label_ = 'Mean')     )
tanner_transpose  (where = (_label_ = 'Mean')     )
nfibro_transpose  (where = (_label_ = 'Mean')     )
alt_transpose     (where = (_label_ = 'Median')   )
ast_transpose     (where = (_label_ = 'Median')   )
alka_transpose    (where = (_label_ = 'Median')   )
ggt_transpose     (where = (_label_ = 'Median')   )
gluc_transpose    (where = (_label_ = 'Median')   )
insu_transpose    (where = (_label_ = 'Median')   )
homa_transpose    (where = (_label_ = 'Median')   )
quicki_transpose  (where = (_label_ = 'Median')   )
bmi_transpose     (where = (_label_ = 'Median')   )
bmipct_transpose  (where = (_label_ = 'Median')   )
totalfat_transpose(where = (_label_ = 'Median')   )
blength_transpose (where = (_label_ = 'Median')   )
TRI_transpose     (where = (_label_ = 'Median')   )
;

data table3_freq_mean;
  set table3_freq table3_means;

proc sort data = table3_freq_mean;
  by  table_name ;

proc transpose data=table3_freq_mean out=table3_compare ;
   by  table_name ;
```

```
data table3_compare (drop =    _NAME_                _LABEL_    rename = (col1 = first_stat col2
= second_stat));
  set table3_compare;

*** Importing the Table 3 Data taken from the primary outcome paper;
data table3_data;
  infile table3 delimiter = ',' MISSOVER DSD  ls=1080;
  length characteristic $45 char_statp $23 table_name $ 30.;
  input
  characteristic $ table_name $  stat1       stat2    char_statp $
;
  if lengthn(characteristic) NE 0 then output table3_data;

data table3_data;
  set table3_data;
  sort_order = _n_;
  table_name = upcase(table_name);


proc sort data = table3_data nodupkey;  by table_name;

data combined_table3_dataset;
  merge table3_compare
        table3_data    ;
  by table_name;
  if table_name = 'QUICKI' then do;
    first_stat  = round(first_stat,0.001);
    second_stat = round(second_stat,0.001);

    diff1 = round((stat1 - first_stat),  0.001);
    diff2 = round((stat2 - second_stat), 0.001);
  end;
  else if table_name in ('AGE' 'TANNER' 'NFIBRO' 'HOMA' 'BMIPCT') then do;
    first_stat  = round(first_stat,0.1);
    second_stat = round(second_stat,0.1);

    diff1 = round((stat1 - first_stat),  0.1);
    diff2 = round((stat2 - second_stat), 0.1);
  end;
  else do;
    first_stat  = round(first_stat,1);
    second_stat = round(second_stat,1);

    diff1 = round((stat1 - first_stat),  1);
    diff2 = round((stat2 - second_stat), 1);
   end;

    label
     stat1          = "Borderline zone 1, Pediatric type (n=49) [Manuscript]"
     first_stat     = "Borderline zone 1, Pediatric type (n=49) [DSIC]      "
     diff1          = "Borderline zone 1, Pediatric type (n=49) [Difference]"
     stat2          = "Definite NASH (n=54) [Manuscript]"
     second_stat    = "Definite NASH (n=54) [DSIC]      "
     diff2          = "Definite NASH (n=54) [Difference]"
;

proc freq data = combined_table3_dataset;
  tables diff1-diff2/list missing;

*** Outputting the dataset to a csv file to be added to the DSIC;

proc sort data = combined_table3_dataset; by sort_order;
ods csv file = out_t3;

run;

proc print data = combined_table3_dataset NOOBS label;
       var characteristic table_name stat1 first_stat diff1 stat2 second_stat diff2
;
       title3 "DSIC Check of Table 3";
```

```
run;

ods csv close;




*******************************************************************************;
*************** Check Table 4 ************************************************;

proc sort data = peds;
  by NFIBRO;

%baseline_freq(peds, male,NFIBRO     );
%baseline_freq(peds, white,NFIBRO     );
%baseline_freq(peds, hispanic,NFIBRO     );
%baseline_freq(peds, ucsd,NFIBRO     );
%baseline_freq(peds, ana,NFIBRO     );
%baseline_freq(peds, asma,NFIBRO     );
%baseline_freq(peds, leng10,NFIBRO     );

  proc freq data =peds;
        table NAFLD*NFIBRO ;
        title3 "Frequency table of the NAFLD*NFIBRO  variable in the analysis dataset";

        *** Outputting the frequency data to work.&var_name._cross using the ODS output;
  ods output CrossTabFreqs = NAFLD_cross;
  proc print data = NAFLD_cross;
    title3 "NAFLD_cross";

  data NAFLD_cross(drop = table NAFLD _TYPE_    _TABLE_    Frequency    Percent  RowPercent
Missing);
    set NAFLD_cross;
    if NAFLD = '2' and not missing(NFIBRO);
    length table_name $ 30.;
          table_name =  upcase("NAFLD") ;

data table4_freq(rename = (ColPercent = stat));
  set
 male_cross
 white_cross
 hispanic_cross
 ucsd_cross
 ana_cross
 asma_cross
 leng10_cross
 NAFLD_cross
;


%baseline_median(peds, age      ,NFIBRO     );
%baseline_median(peds, tanner   ,NFIBRO     );

%baseline_median(peds, alt      ,NFIBRO     );
%baseline_median(peds, ast      ,NFIBRO     );
%baseline_median(peds, alka     ,NFIBRO     );
%baseline_median(peds, ggt      ,NFIBRO     );
%baseline_median(peds, insu     ,NFIBRO     );
%baseline_median(peds, homa     ,NFIBRO     );
%baseline_median(peds, quicki   ,NFIBRO     );
%baseline_median(peds, bmi      ,NFIBRO     );
%baseline_median(peds, bmipct   ,NFIBRO     );
%baseline_median(peds, totalfat ,NFIBRO     );
%baseline_median(peds, blength  ,NFIBRO     );

%baseline_median(peds, NAs      ,NFIBRO     );
%baseline_median(peds, ALB     ,NFIBRO     );
%baseline_median(peds, WBC     ,NFIBRO     );
%baseline_median(peds, HEMA     ,NFIBRO     );
```

```
data table4_means(drop = _LABEL_ rename = (COL1 = stat));
  set
age_transpose     (where = (_label_ = 'Mean')     )
tanner_transpose  (where = (_label_ = 'Mean')     )
nas_transpose     (where = (_label_ = 'Mean')     )

alt_transpose     (where = (_label_ = 'Median')   )
ast_transpose     (where = (_label_ = 'Median')   )
alka_transpose    (where = (_label_ = 'Median')   )
ggt_transpose     (where = (_label_ = 'Median')   )
insu_transpose    (where = (_label_ = 'Median')   )
homa_transpose    (where = (_label_ = 'Median')   )
quicki_transpose  (where = (_label_ = 'Median')   )
bmi_transpose     (where = (_label_ = 'Median')   )
bmipct_transpose  (where = (_label_ = 'Median')   )
totalfat_transpose(where = (_label_ = 'Median')   )
blength_transpose (where = (_label_ = 'Median')   )

ALB_transpose     (where = (_label_ = 'Median')   )
WBC_transpose     (where = (_label_ = 'Median')   )
HEMA_transpose     (where = (_label_ = 'Median')   )
;

data table4_freq_mean;
  set table4_freq table4_means;

proc sort data = table4_freq_mean;
  by  table_name ;

proc transpose data=table4_freq_mean out=table4_compare ;
   by  table_name ;


data table4_compare (drop =    _NAME_                _LABEL_    rename = (col1 = first_stat col2
= second_stat col3 = third_stat col4 = fourth_stat));
  set table4_compare;
  if _NAME_ = 'stat';

*** Importing the Table 4 Data taken from the primary outcome paper;
data table4_data;
  infile table4 delimiter = ',' MISSOVER DSD  firstobs=3 ls=1080;
  length characteristic $45 char_statp $23 table_name $ 30.;
  input
  characteristic $ table_name $  stat1       stat2   stat3   stat4 char_statp $
;
  if lengthn(characteristic) NE 0 then output table4_data;

data table4_data;
  set table4_data;
  sort_order = _n_;
  table_name = upcase(table_name);

proc sort data = table4_data nodupkey;  by table_name;

data combined_table4_dataset;
  merge table4_compare
        table4_data    ;
  by table_name;
  if table_name = 'QUICKI' then do;
    first_stat  = round(first_stat,0.001);
    second_stat = round(second_stat,0.001);
    third_stat  = round(third_stat,0.001);
    fourth_stat = round(fourth_stat,0.001);

    diff1 = round((stat1 - first_stat),  0.001);
    diff2 = round((stat2 - second_stat), 0.001);
    diff3 = round((stat3 - third_stat),  0.001);
    diff4 = round((stat4 - fourth_stat), 0.001);
  end;
  else if table_name in ('AGE' 'ALB' 'WBC'  'HOMA' 'BMIPCT'  'TANNER' 'NAS' ) then do;
```

```
    first_stat  = round(first_stat,0.1);
    second_stat = round(second_stat,0.1);
    third_stat  = round(third_stat,0.1);
    fourth_stat = round(fourth_stat,0.1);

    diff1 = round((stat1 - first_stat),  0.1);
    diff2 = round((stat2 - second_stat), 0.1);
    diff3 = round((stat3 - third_stat),  0.1);
    diff4 = round((stat4 - fourth_stat), 0.1);
  end;
  else do;
    first_stat  = round(first_stat,1);
    second_stat = round(second_stat,1);
    third_stat  = round(third_stat,1);
    fourth_stat = round(fourth_stat,1);

    diff1 = round((stat1 - first_stat),  1);
    diff2 = round((stat2 - second_stat), 1);
    diff3 = round((stat3 - third_stat),  1);
    diff4 = round((stat4 - fourth_stat), 1);
   end;

    label
     stat1          = "None (n=46) [Manuscript]"
     first_stat     = "None (n=46) [DSIC]      "
     diff1          = "None (n=46) [Difference]"
     stat2          = "Mild (n=77) [Manuscript]"
     second_stat    = "Mild (n=77) [DSIC]      "
     diff2          = "Mild (n=77) [Difference]"
     stat3          = "Moderate (n=29) [Manuscript]"
     third_stat     = "Moderate (n=29) [DSIC]      "
     diff3          = "Moderate (n=29) [Difference]"
     stat4          = "Bridging (n=24) [Manuscript]"
     fourth_stat    = "Bridging (n=24) [DSIC]      "
     diff4          = "Bridging (n=24) [Difference]"
;


proc freq data = combined_table4_dataset;
  tables diff1-diff4/list missing;

*** Outputting the dataset to a csv file to be added to the DSIC;

proc sort data = combined_table4_dataset; by sort_order;
ods csv file = out_t4;

run;

proc print data = combined_table4_dataset NOOBS label;
       var characteristic table_name stat1 first_stat diff1 stat2 second_stat diff2 stat3
third_stat diff3 stat4 fourth_stat diff4
;
       title3 "DSIC Check of Table 4";

run;

ods csv close;




*******************************************************************************;
*************** Check Table 5 ************************************************;

data peds_table5;
  set peds;
  if nas in (1,2,3) then nas_group = 1;
  else if nas in (4,5) then nas_group = 2;
  else if nas in (6,7) then nas_group = 3;

proc sort data = peds_table5;
```

```
  by nas_group;

%baseline_freq(peds_table5, male,nas_group     );
%baseline_freq(peds_table5, white,nas_group     );
%baseline_freq(peds_table5, hispanic,nas_group     );
%baseline_freq(peds_table5, ucsd,nas_group     );
%baseline_freq(peds_table5, ana,nas_group     );
%baseline_freq(peds_table5, asma,nas_group     );
%baseline_freq(peds_table5, leng10,nas_group     );

  proc freq data =peds_table5;
       table NAFLD*nas_group ;
       title3 "Frequency table of the NAFLD*nas_group  variable in the analysis dataset";

       *** Outputting the frequency data to work.&var_name._cross using the ODS output;
  ods output CrossTabFreqs = NAFLD_cross;
  proc print data = NAFLD_cross;
    title3 "NAFLD_cross";

  data NAFLD_cross(drop = table NAFLD _TYPE_    _TABLE_    Frequency    Percent  RowPercent
Missing);
    set NAFLD_cross;
    if NAFLD = '2' and not missing(nas_group);
    length table_name $ 30.;
         table_name =  upcase("NAFLD") ;

data table5_freq(rename = (ColPercent = stat));
  set
 male_cross
 white_cross
 hispanic_cross
 ucsd_cross
 ana_cross
 asma_cross
 leng10_cross
 NAFLD_cross
;


%baseline_median(peds_table5, age      ,nas_group     );
%baseline_median(peds_table5, alt      ,nas_group     );
%baseline_median(peds_table5, ast      ,nas_group     );
%baseline_median(peds_table5, alka     ,nas_group     );
%baseline_median(peds_table5, ggt      ,nas_group     );
%baseline_median(peds_table5, insu     ,nas_group     );
%baseline_median(peds_table5, homa     ,nas_group     );
%baseline_median(peds_table5, quicki   ,nas_group     );
%baseline_median(peds_table5, blength  ,nas_group     );
%baseline_median(peds_table5, gluc     ,nas_group     );
%baseline_median(peds_table5, nfibro   ,nas_group     );

data table5_means(drop = _LABEL_ rename = (COL1 = stat));
  set
age_transpose     (where = (_label_ = 'Mean')     )

alt_transpose     (where = (_label_ = 'Median')   )
ast_transpose     (where = (_label_ = 'Median')   )
alka_transpose    (where = (_label_ = 'Median')   )
ggt_transpose     (where = (_label_ = 'Median')   )
insu_transpose    (where = (_label_ = 'Median')   )
homa_transpose    (where = (_label_ = 'Median')   )
quicki_transpose  (where = (_label_ = 'Median')   )
blength_transpose (where = (_label_ = 'Median')   )

gluc_transpose     (where = (_label_ = 'Median')   )
nfibro_transpose     (where = (_label_ = 'Mean')   )
;

data table5_freq_mean;
  set table5_freq table5_means;
```

```
proc sort data = table5_freq_mean;
  by  table_name ;
proc transpose data=table5_freq_mean out=table5_compare ;
   by  table_name ;

data table5_compare (drop =    _NAME_                _LABEL_    rename = (col1 = first_stat col2
= second_stat col3 = third_stat ));
  set table5_compare;
  if _NAME_ = 'stat';

*** Importing the Table 5 Data taken from the primary outcome paper;
data table5_data;
  infile table5 delimiter = ',' MISSOVER DSD  ls=1080;
  length characteristic $45 char_statp $23 table_name $ 30.;
  input
  characteristic $ table_name $  stat1       stat2   stat3    char_statp $
;
  if lengthn(characteristic) NE 0 then output table5_data;

data table5_data;
  set table5_data;
  sort_order = _n_;
  table_name = upcase(table_name);

proc sort data = table5_data nodupkey;  by table_name;

data combined_table5_dataset;
  merge table5_compare
        table5_data    ;
  by table_name;
  if table_name = 'QUICKI' then do;
    first_stat  = round(first_stat,0.001);
    second_stat = round(second_stat,0.001);
    third_stat  = round(third_stat,0.001);

    diff1 = round((stat1 - first_stat),  0.001);
    diff2 = round((stat2 - second_stat), 0.001);
    diff3 = round((stat3 - third_stat),  0.001);
  end;
  else if table_name in ('AGE' 'HOMA'  'NFIBRO') then do;
    first_stat  = round(first_stat,0.1);
    second_stat = round(second_stat,0.1);
    third_stat  = round(third_stat,0.1);

    diff1 = round((stat1 - first_stat),  0.1);
    diff2 = round((stat2 - second_stat), 0.1);
    diff3 = round((stat3 - third_stat),  0.1);
  end;
  else do;
    first_stat  = round(first_stat,1);
    second_stat = round(second_stat,1);
    third_stat  = round(third_stat,1);

    diff1 = round((stat1 - first_stat),  1);
    diff2 = round((stat2 - second_stat), 1);
    diff3 = round((stat3 - third_stat),  1);
  end;

    label
     stat1          = "NASH 1-3 (n=49) [Manuscript]"
     first_stat     = "NASH 1-3 (n=49) [DSIC]      "
     diff1          = "NASH 1-3 (n=49) [Difference]"
     stat2          = "NASH 4-5 (n=84) [Manuscript]"
     second_stat    = "NASH 4-5 (n=84) [DSIC]      "
     diff2          = "NASH 4-5 (n=84) [Difference]"
     stat3          = "NASH 6-7 (n=43) [Manuscript]"
     third_stat     = "NASH 6-7 (n=43) [DSIC]      "
     diff3          = "NASH 6-7 (n=43) [Difference]"
;
```

```
proc freq data = combined_table5_dataset;
  tables diff1-diff3/list missing;

*** Outputting the dataset to a csv file to be added to the DSIC;

proc sort data = combined_table5_dataset; by sort_order;
ods csv file = out_t5;

run;

proc print data = combined_table5_dataset NOOBS label;
        var characteristic table_name stat1 first_stat diff1 stat2 second_stat diff2 stat3
third_stat diff3
;
        title3 "DSIC Check of Table 5";

run;

ods csv close;



endsas;
```