Dataset Integrity Check for the Nonalcoholic Steatohepatitis (NASH) Nonalcoholic Fatty Liver Disease (NAFLD) Pediatric Data Files

Prepared by Allyson Mateja
IMS Inc.
3901 Calverton Blvd, Suite 200, Calverton MD, 20705
May 27, 2016

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.

3 Archived Datasets

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

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Patton et al in Gastroenterology in 2008 [1]. To verify the integrity of the datasets, descriptive statistics were computed.

5 Results

Note that there is not a one-to-one match between the NAFLD Pediatric analysis datasets and the NAFLD Pediatric form datasets. As a result, this dataset integrity check was performed by comparing the results from subjects that were in both the analysis datasets and the form datasets. The results from the analysis tables are presented in the 'Manuscript' columns in the tables below. Some discrepancies are expected as a result. In addition, discrepancies are expected due to differing cut-off dates between the analysis datasets and the form datasets.

For Table 1 in the publication [1], Sample Characteristics for Children Enrolled in NASH CRN Eligible for Study Inclusion, Table A lists the variables that were used in the replication and Table B compares the results calculated from the archived data files to the results published in Table 1. The results of the replication are almost an exact match to published results.

For Table 2 in the publication [1], Predictors of NAFLD Pattern (Excluding Borderline Zone 1), Table C lists the variables that were used in the replication and Table D compares the results calculated from the archived data files to the results published in Table 2. The results of the replication are almost an exact match to published results.

For Table 3 in the publication [1], Predictors of Borderline Zone 1 Versus Definite NASH Pattern in Children with Some Fibrosis Present, Table E lists the variables that were used in the replication and Table F compares the results calculated from the archived data files to the results published in Table 3. The results of the replication are almost an exact match to published results.

For Table 4 in the publication [1] Predictors of Fibrosis Stage, Table G lists the variables that were used in the replication and Table H compares the results calculated from the archived data files to the results published in Table 4. The results of the replication are almost an exact match to published results.

For Table 5 in the publication [1] Predictors of NAS, Table I lists the variables that were used in the replication and Table J compares the results calculated from the archived data files to the results published in Table 5. The results of the replication are almost an exact match to published results.

6 Conclusions

The NIDDK repository is confident that the NAFLD Pediatric data files to be distributed are a true copy of the manuscript 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: Sample Characteristics for Children Enrolled in NASH CRN Eligible for Study Inclusion

Table Variable	**dataset.variable**
Study	rg.study
Male	rg.rg111
Age in y	rg.rg110
White race	rg.114e, rg.114a
Hispanic ethnicity	rg.rg112
Biopsy length in mm	cr.cr115
NAFLD pattern	cr.cr123
Fibrosis stage	cr.cr121
Fibrosis score	cr.cr121
NAS	cr.cr116a, cr.cr117a, cr.cr118a
BMI	pe.pe208a, pe.pe208b, pe.pe209a, pe.pe209b, pe208c, pe209c
BMI percentile	pe.pe208a, pe.pe208b, pe.pe209a, pe.pe209b, pe208c, pe209c
Body fat percentage*	N/A
Tanner stage	pe.pe237, pe.pe240
Acanthosis nigricans	pe.pe218

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

Characteristic	N (%) Manuscript	N (%) DSIC	Difference
Study			
TONIC	0 (0)	0 (0)	0 (0)
Database	37 (100)	37 (100)	0 (0)
Demographics			
Male	23 (62)	23 (62)	0 (0)
Age in y, mean ± SD (range)	11.6 ± 2.9 (6-16)	11.6 ± 2.9 (6-16)	0 ± 0 (0-0)
White race	19 (51)	19 (5)	0 (0)
Hispanic ethnicity	22 (59)	22 (59)	0 (0)
Histology			
Biopsy length in mm, median ± pseudo-sigma (range)	15 ± 4.4 (7-28)	15 ± 4.4 (7-28)	0 ± 0 (0-0)
NAFLD pattern			
Not NASH	15 (41)	15 (41)	0 (0)
Borderline zone 3 pattern	4 (11)	4 (11)	0 (0)
Borderline zone 1 pattern	11 (30)	11 (30)	0 (0)
Definite NASH	7 (19)	7 (19)	0 (0)
Fibrosis stage			
None (score = 0)	15 (41)	15 (41)	0 (0)
Mild zone 3 only (score = 1)	1 (3)	1 (3)	0 (0)
Moderate zone 3 only (score = 1)	1 (3)	1 (3)	0 (0)
Periportal only (score = 1)	11 (30)	11 (30)	0 (0)
Mild/moderate zone 3 and periportal (score = 2)	4 (11)	4 (11)	0 (0)
Bridging (score = 3)	5 (14)	5 (14)	0 (0)
Fibrosis score, mean ± SD	0.97 ± 1.04	0.97 ± 1.04	0 ± 0
NAS			
1	1 (3)	1 (3)	0 (0)
2	8 (22)	8 (22)	0 (0)
3	12 (32)	12 (32)	0 (0)
4	7 (19)	7 (19)	0 (0)
5	6 (16)	6 (16)	0 (0)
6	2 (5)	2 (5)	0 (0)
7	1 (3)	1 (3)	0 (0)
NAS, mean ± SD	3.51 ± 1.37	3.51 ± 1.37	0 (0)
Anthropometrics			

Characteristic	N (%) Manuscript	N (%) DSIC	Difference
BMI in kg/m, mean ± pseudo-sigma (range)	31.57 ± 5.15 (18.2-47.2)	32.57 ± 4.66 (18.1-47.2)	1 ± 0.49 (0.1-0)
BMI age-sex percentile, median ± pseudo-sigma (range)	99.2 ± 1.1 (92.1-100)	98.5 ± 1.3 (93.6-100)	0.7 ± 0.2 (1.5-0)
Body fat percentage, median ± pseudo-sigma (range)*	N/A	N/A	N/A
Clinical			
Tanner stage, mean ± SD (range)	2.4 ± 1.5 (1-5)	2.4 ± 1.5 (1-5)	0 ± 0 (0-0)
Acanthosis nigricans	28 (76)	28 (76)	0 (0)

Pseudo-sigma is Interquartile range (IQR)/1.349

Table C: Variables used to replicate Table 2: Predictors of NAFLD Pattern (Excluding Borderline Zone 1)

Characteristic	**dataset.variable**
NAFLD Pattern	cr.cr123
Male	rg.rg111
Age	rg.rg110
White race	rg.114e, rg.114a
Hispanic ethnicity	rg.rg112
Clinic site: UCSD (%)**	N/A
ALT	lr.lr329
AST	lr.lr328
Alkaline phosphatase	lr.lr330
GGT	lr.lr331
Fasting glucose	lr.lr339a
Fasting insulin	lr.lr339b
HOMA-IR	lr.lr339a, lr.lr339b
QUICKI	lr.lr339a, lr.lr339b
ANA	ls.ls123
ASMA	ls.ls124
BMI	pe.pe208a, pe.pe208b, pe.pe209a, pe.pe209b, pe208c, pe209c
BMI percentile	pe.pe208a, pe.pe208b, pe.pe209a, pe.pe209b, pe208c, pe209c
Percentage of body fat*	N/A
Tanner stage	pe.pe237, pe.pe240
Median biopsy length	cr.cr115
< 10 mm	cr.cr115
Fibrosis score	cr.cr121

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

Characteristic	Not NASH Manuscript (n=15)	Not NASH DSIC (n=15)	Difference (n=0)	Borderline zone 3, adult type Manuscript (n=4)	Borderline zone 3, adult type DSIC (n=4)	Difference (n=0)
Demographics						
Male (%)	60	60	0	75	75	0
Mean age (y)	12.7	12.7	0	10	10	0
White race (%)	53	53	0	25	25	0
Hispanic ethnicity (%)	60	60	0	25	25	0
Clinic site: UCSD (%)**	50	N/A	N/A	50	N/A	N/A
Laboratory data (median values)						
ALT, U/L	70	70	0	69.5	69.5	0
AST, U/L	38	38	0	40	40	0
Alkaline phosphatase, U/L	202	202	0	294	294	0
GGT, U/L	32	32	0	23.5	23.5	0
Fasting glucose, mg/dL	89	89	0	89	89	0
Fasting insulin, mU/mL	29	29	0	13.9	13.9	0
HOMA-IR	5.5	5.5	0	3.1	3.1	0
QUICKI	0.299	0.341	0.042	0.331	0.485	0.154
ANA (% positive)	13	13	0	0	0	0
ASMA (% positive)	45	45	0	0	0	0
Anthropometric, median values						
BMI (kg/m²)	35	35	0	32	32	0
BMI percentile	99	99.1	0.1	99.4	99.5	0.1
Percentage of body fat*	N/A	N/A	N/A	N/A	N/A	N/A
Clinical						
Tanner stage (mean)	3	3	0	1.25	1.25	0
Histologic						
Median biopsy length (mm)	14	14	0	19	19	0
< 10 mm (%)	13	13	0	0	0	0
Fibrosis score (mean)	0.3	0.3	0	1	1	0

Characteristic	Definite NASH Manuscript (n=7)	Definite NASH DSIC (n=7)	Difference (n=0)
Demographics			
Male (%)	29	29	0
Mean age (y)	13.3	13.3	0
White race (%)	71	71	0
Hispanic ethnicity (%)	43	43	0
Clinic site: UCSD (%)**	29	N/A	N/A
Laboratory data (median values)			
ALT, U/L	69	69	0
AST, U/L	49	49	0
Alkaline phosphatase, U/L	165	165	0
GGT, U/L	57	57	0
Fasting glucose, mg/dL	91	91	0
Fasting insulin, mU/mL	39	39	0
HOMA-IR	8.8	8.8	0
QUICKI	0.282	0.314	0.032
ANA (% positive)	14	14	0
ASMA (% positive)	67	67	0
Anthropometric, median values			
BMI (kg/m^2)	30	30	0
BMI percentile	98.1	98.1	0
Percentage of body fat*	N/A	N/A	N/A
Clinical			
Tanner stage (mean)	3.4	3.4	0
Histologic			
Median biopsy length (mm)	13	13	0
< 10 mm (%)	0	0	0
Fibrosis score (mean)	1.7	1.7	0

Table E: Variables used to replicate Table 3: Predictors of Borderline Zone 1 Versus Definite NASH Pattern in Children with Some Fibrosis Present

Characteristic	**dataset.variable**
NAFLD Pattern	cr.cr123
Male	rg.rg111
Age	rg.rg110
White race	rg.114e, rg.114a
Hispanic ethnicity	rg.rg112
Clinic site: UCSD (%)**	N/A
ALT	lr.lr329
AST	lr.lr328
Alkaline phosphatase	lr.lr330
GGT	lr.lr331
Triglycerides	lr.lr338a
Fasting glucose	lr.lr339a
Fasting insulin	lr.lr339b
HOMA-IR	lr.lr339a, lr.lr339b
QUICKI	lr.lr339a, lr.lr339b
ANA	ls.ls123
ASMA	ls.ls124
BMI	pe.pe208a, pe.pe208b, pe.pe209a, pe.pe209b, pe208c, pe209c
BMI percentile	pe.pe208a, pe.pe208b, pe.pe209a, pe.pe209b, pe208c, pe209c
Percentage of body fat*	N/A
Tanner stage	pe.pe237, pe.pe240
Median biopsy length	cr.cr115
< 10 mm	cr.cr115
Fibrosis score	cr.cr121

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

Characteristic	Borderline zone 1, pediatric type Manuscript (n= 11)	Borderline zone 1, pediatric type DSIC (n=11)	Difference (n=0)	Definite NASH Manuscript (n=6)	Definite NASH DSIC (n=6)	Difference (n=0)
Demographics						
Male (%)	82	82	0	33	33	0
Mean age (y)	9.5	9.5	0	13.5	13.5	0
White race (%)	45	45	0	67	67	0
Hispanic ethnicity (%)	82	82	0	50	50	0
Clinic site: UCSD (%)**	82	N/A	N/A	33	N/A	N/A
Laboratory data, median values						
ALT (U/L)	55	55	0	74	74	0
AST (U/L)	38	38	0	61	61	0
Alkaline phosphatase (U/L)	295	295	0	152.5	152.5	0
GGT (U/L)	28	28	0	58.5	58.5	0
Triglycerides (mg/dL)	136	136	0	198.5	198.5	0
Fasting glucose (mg/dL)	84	84	0	95.5	95.5	0
Fasting insulin (mg/dL)	5.2	5.7	0.5	46	46	0
HOMA-IR	1.1	1.3	0.2	9.7	9.7	0
QUICKI	0.377	0.661	0.284	0.278	0.302	0.024
ANA (% positive)	0	0	0	17	17	0
ASMA (% positive)	29	29	0	80	80	0
Anthropometric, median values						
BMI (kg/m^2)	33.5	33.5	0	30.8	30.8	0
BMI percentile	99.5	99.5	0	98.3	98.3	0
Percentage of body fat*	N/A	N/A	N/A	N/A	N/A	N/A
Clinical						
Tanner stage (mean)	1.5	1.5	0	3.7	3.7	0
Histologic						
Median biopsy length (mm)	15	15	0	16.5	16.5	0
<10 mm (%)	9	9	0	0	0	0
Fibrosis score (mean)	1.5	1.5	0	2	2	0

Table G: Variables used to replicate Table 4: Predictors of Fibrosis Stage

Characteristic	**dataset.variable**
Fibrosis Stage	cr.cr121
Male	rg.rg111
Age	rg.rg110
White race	rg.114e, rg.114a
Hispanic ethnicity	rg.rg112
Clinic site: UCSD (%)**	N/A
ALT	lr.lr329
AST	lr.lr328
Alkaline phosphatase	lr.lr330
GGT	lr.lr331
Albumin	lr.lr333
White blood cell count	lr.lr311
Hematocrit	lr.lr310
Fasting insulin	lr.lr339b
HOMA-IR	lr.lr339a, lr.lr339b
QUICKI	lr.lr339a, lr.lr339b
ANA	ls.ls123
ASMA	ls.ls124
BMI	pe.pe208a, pe.pe208b, pe.pe209a, pe.pe209b, pe208c, pe209c
BMI percentile	pe.pe208a, pe.pe208b, pe.pe209a, pe.pe209b, pe208c, pe209c
Percentage of body fat*	N/A
Tanner stage	pe.pe237, pe.pe240
Median biopsy length	cr.cr115
< 10 mm	cr.cr115
Definite NASH	cr.cr123
NAS	cr.cr116a, cr.cr117a, cr.cr118a

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

Characteristic	None Manuscript (n=15)	None DSIC (n=15)	Difference (n=0)	Mild Manuscript (n=13)	Mild DSIC (n=13)	Difference (n=0)
Demographics						
Male (%)	53	53	0	77	77	0
Mean age (y)	11.7	11.7	0	11.2	11.2	0
White race (%)	53	53	0	38	38	0
Hispanic ethnicity (%)	47	47	0	77	77	0
Clinic site: UCSD (%)**	53	N/A	N/A	69	N/A	N/A
Laboratory data, median values						
ALT (U/L)	70	70	0	48	48	0
AST (U/L)	39	39	0	32	32	0
Alkaline phosphatase (U/L)	251	251	0	242	242	0
GGT (U/L)	24	24	0	25	25	0
Albumin (g/dL)	4.2	4.2	0	4.3	4.3	0
White blood cell count (1000/mm^3)	7.5	7.5	0	8.1	8.1	0
Hematocrit (%)	41.7	41.7	0	40	40	0
Fasting insulin (mU/mL)	23	23	0	10.5	15.2	4.7
HOMA-IR	5.2	5.2	0	2	2.7	0.7
QUICKI	0.301	0.367	0.066	0.348	0.423	0.075
ANA (% positive)	13	13	0	0	0	0
ASMA (% positive)	30	30	0	40	40	0
Anthropometric, median values						
BMI (kg/m^2)	33.8	33.8	0	33.5	33.5	0
BMI percentile	99	99.1	0.1	99.4	99.4	0
Percentage of body fat*	N/A	N/A	N/A	N/A	N/A	N/A
Clinical						
Tanner stage (mean)	2.5	2.5	0	2.3	2.3	0
Histologic						
Median biopsy length (mm)	15	15	0	13	13	0
<10 mm (%)	13	13	0	0	0	0
Definite NASH (%)	7	7	0	15	15	0
NAS (mean)	3.1	3.1	0	3.3	3.3	0

Characteristic	Moderate Manuscript (n=4)	Moderate DSIC (n=4)	Difference (n=0)	Bridging Manuscript (n=5)	Bridging DSIC (n=5)	Difference (n=0)
Demographics						
Male (%)	50	50	0	60	60	0
Mean age (y)	13	13	0	10.8	10.8	0
White race (%)	75	75	0	60	60	0
Hispanic ethnicity (%)	50	50	0	60	60	0
Clinic site: UCSD (%)**	50	N/A	N/A	60	N/A	N/A
Laboratory data, median values						
ALT (U/L)	96	96	0	104	104	0
AST (U/L)	57.5	57.5	0	66	66	0
Alkaline phosphatase (U/L)	168.5	168.5	0	337	337	0
GGT (U/L)	55.5	55.5	0	60	60	0
Albumin (g/dL)	4.1	4.1	0	4.6	4.6	0
White blood cell count (1000/mm^3)	6.8	6.8	0	7.9	7.9	0
Hematocrit (%)	37.6	37.6	0	40.8	40.8	0
Fasting insulin (mU/mL)	47.8	47.8	0	31	31	0
HOMA-IR	10	10	0	7.2	7.2	0
QUICKI	0.277	0.299	0.022	0.289	0.335	0.046
ANA (% positive)	25	25	0	0	0	0
ASMA (% positive)	50	50	0	67	67	0
Anthropometric, median values						
BMI (kg/m^2)	35.4	35.4	0	27.4	27.4	0
BMI percentile	99.1	99.2	0.1	97.1	97.3	0.2
Percentage of body fat*	N/A	N/A	N/A	N/A	N/A	N/A
Clinical						
Tanner stage (mean)	2	2	0	2.4	2.4	0
Histologic						
Median biopsy length (mm)	13.5	13.5	0	16	16	0
<10 mm (%)	0	0	0	20	20	0
Definite NASH (%)	50	50	0	40	40	0
NAS (mean)	4.5	4.5	0	4.6	4.6	0

Table I: Variables used to replicate Table 5: Predictors of NAS

Characteristic	**dataset.variable**
NAS	cr.cr116a, cr.cr117a, cr.cr118a
Male	rg.rg111
Age	rg.rg110
White race	rg.114e, rg.114a
Hispanic ethnicity	rg.rg112
Clinic site: UCSD (%)**	N/A
ALT	lr.lr329
AST	lr.lr328
Alkaline phosphatase	lr.lr330
GGT	lr.lr331
Fasting glucose	lr.lr339a
Fasting insulin	lr.lr339b
HOMA-IR	lr.lr339a, lr.lr339b
QUICKI	lr.lr339a, lr.lr339b
ANA	ls.ls123
ASMA	ls.ls124
Median biopsy length	cr.cr115
< 10 mm	cr.cr115
Definite NASH	cr.cr123
Fibrosis score	cr.cr121

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

Characteristic	NAS 1-3 Manuscript (n=21)	NAS 1-3 DSIC (n=21)	Difference (n=0)	NAS 4-5 Manuscript (n=13)	NAS 4-5 DSIC (n=13)	Difference (n=0)
Demographics						
Male (%)	62	62	0	69	69	0
Mean age (y)	11.4	11.4	0	11.5	11.5	0
White race (%)	52	52	0	46	46	0
Hispanic ethnicity (%)	62	62	0	62	62	0
Clinic site: UCSD (%)**	57	N/A	N/A	62	N/A	N/A
Laboratory data, median values						
ALT (U/L)	57	57	0	69	69	0
AST (U/L)	38	38	0	41	41	0
Alkaline phosphatase (U/L)	242	242	0	254	254	0
GGT (U/L)	24	24	0	26	26	0
Fasting glucose (mg/dL)	87	87	0	90	90	0
Fasting insulin (mU/mL)	23	24.5	1.5	16.9	16.9	0
HOMA-IR	5.2	5.2	0	3.6	3.6	0
QUICKI	0.301	0.36	0.059	0.318	0.414	0.096
ANA (% positive)	10	10	0	0	0	0
ASMA (% positive)	25	25	0	50	50	0
Histologic						
Median biopsy length (mm)	14	14	0	16	16	0
<10 mm (%)	10	10	0	8	8	0
Definite NASH (%)	5	5	0	23	23	0
Fibrosis score (mean)	0.7	0.7	0	1.2	1.2	0

Characteristic	NAS 6-7 Manuscript (n=3)	NAS 6-7 DSIC (n=3)	Difference (n=0)
Demographics			
Male (%)	33	33	0
Mean age (y)	12.7	12.7	0
White race (%)	67	67	0
Hispanic ethnicity (%)	33	33	0
Clinic site: UCSD (%)**	67	N/A	N/A
Laboratory data, median values			

Characteristic	NAS 6-7 Manuscript (n=3)	NAS 6-7 DSIC (n=3)	Difference (n=0)
ALT (U/L)	110	110	0
AST (U/L)	73	73	0
Alkaline phosphatase (U/L)	172	172	0
GGT (U/L)	60	60	0
Fasting glucose (mg/dL)	91	91	0
Fasting insulin (mU/mL)	39	39	0
HOMA-IR	8.8	8.8	0
QUICKI	0.282	0.314	0.032
ANA (% positive)	33	33	0
ASMA (% positive)	100	100	0
Histologic			
Median biopsy length (mm)	20	20	0
<10 mm (%)	0	0	0
Definite NASH (%)	100	100	0
Fibrosis score (mean)	2.3	2.3	0

*Body fat percentage was only measured in TONIC participants, and was not measured in participants who participated in only the NAFLD Pediatric Database.

**Clinic site was redacted from the public use datasets for de-identification purposes.

Attachment A: SAS Code

```
**** NASH NAFLD Pediatric DSIC;
**** Programmer: Allyson Mateja;
**** Date: April 25, 2016;

title '/prj/niddk/ims_analysis/NAFLD/prog_initial_analysis/nafld_pediatric_dsic_summary_statistics.sas';
title2 ' ';

options nofmterr;

proc format;
       value $studyf 1 = 'Database'
                     3 = 'TONIC';
       value $nafldf  '0'  = 'Not NASH'
                      '1a' = 'Borderline Zone 3'
                      '1b' = 'Borderline Zone 1'
                      '2'  = 'Definite NASH';

       value $fibrosisf '0'  = 'None'
                        '1a' = 'Mild zone 3 only'
                        '1b' = 'Moderate zone 3 only'
                        '1c' = 'Periportal only'
                        '2'  = 'Mild/moderate zone 3 and periportal'
                        '3'  = 'Bridging';

libname naflddta '/prj/niddk/ims_analysis/NAFLD/private_orig_data/NASHCRN_Data_Sharing_NAFLDDatabase/datasets/SASDATA/';

libname inlib3 xport
'/prj/niddk/ims_analysis/NASH/private_orig_data/NASHCRN_Data_Sharing_PediatricNAFLDDatabase_Gastroenterology_2008/Datasets/peds.xpt';
proc copy in=inlib3 out=work;
libname inlib4 xport
'/prj/niddk/ims_analysis/NASH/private_orig_data/NASHCRN_Data_Sharing_PediatricNAFLDDatabase_Gastroenterology_2008/Datasets/link_id.xpt
';
proc copy in=inlib4 out=work;

data peds_ids;
       set ids;

data table1_peds;
       set peds;

proc sort data = table1_peds;
       by id;

proc sort data = peds_ids;
```

```
       by id;

data table1_peds;
       merge table1_peds (in=val1)
             peds_ids    (in=val2);
       by id;
       if val1 and val2 then output table1_peds;

data ad;
       set naflddta.ad;

data cr;
       set naflddta.cr;

data dr;
       set naflddta.dr;

data hi;
       set naflddta.hi;

data ir;
       set naflddta.ir;

data ld;
       set naflddta.ld;

data lp;
       set naflddta.lp;

data lq;
       set naflddta.lq;

data lr;
       set naflddta.lr;

data ls;
       set naflddta.ls;

data ma;
       set naflddta.ma;

data mv;
       set naflddta.mv;

data pa;
       set naflddta.pa;

data pe;
       set naflddta.pe;
```

```
data pf;
       set naflddta.pf;

data pq;
       set naflddta.pq;

data pr;
       set naflddta.pr;

data ps;
       set naflddta.ps;

data pt;
       set naflddta.pt;

data pv;
       set naflddta.pv;

data pw;
       set naflddta.pw;

data py;
       set naflddta.py;

data qf;
       set naflddta.qf;

data rg;
       set naflddta.rg;

data ie;
       set naflddta.ie;

data bg;
       set naflddta.bg;

data nash_id;
       set naflddta.nash_id;

proc sort data=ad; by nash;
proc sort data=bg; by nash;
proc sort data=cr; by nash;
proc sort data=dr; by nash;
proc sort data=hi; by nash;
proc sort data=ie; by nash;
proc sort data=ir; by nash;
proc sort data=ld; by nash;
proc sort data=lp; by nash;
```

```
proc sort data=lq; by nash;
proc sort data=lr; by nash;
proc sort data=ls; by nash;
proc sort data=ma; by nash;
proc sort data=mv; by nash;
proc sort data=pa; by nash;
proc sort data=pe; by nash;
proc sort data=pf; by nash;
proc sort data=pq; by nash;
proc sort data=pr; by nash;
proc sort data=ps; by nash;
proc sort data=pt; by nash;
proc sort data=pv; by nash;
proc sort data=pw; by nash;
proc sort data=py; by nash;
proc sort data=qf; by nash;
proc sort data=rg; by nash;

proc contents data = rg;
proc contents data = table1_peds;

proc freq data = rg;
       tables rg110 rg123;

proc sort data = table1_peds;
       by nash;

proc print data = table1_peds (obs=20);
       var nash;

proc sort data = rg;
       by nash;

proc print data = rg (obs=20);
       var nash;

proc freq data = cr;
       tables visit;

data cr;
       set cr;
       if visit in ('s1', 's2');

proc sort data = cr;
       by nash visit;

data cr;
       set cr;
       by nash;
```

```
       if first.nash then output;

data in_rg_and_peds in_rg_only in_peds_only;
       merge rg          (in=val1 keep=nash rg110 rg123 rg130a formdate)
             table1_peds (in=val2)
             cr          (in=val3 rename=(formdate = cr_formdate));
       by nash;
       time = abs(formdate - cr108);
       if val1 and val2 and val3 and time <= 182.5 and (exitdt = . or exitdt >= 1000) then output in_rg_and_peds;
       else if val1 and not val2 then output in_rg_only;
       else if val2 and not val1 then output in_peds_only;

proc freq data = in_rg_and_peds;
       tables study;
       format study $studyf.;
       title3 'Manuscript Table 1 - Study';

proc freq data = in_rg_and_peds;
       tables male;
       title3 'Manuscript Table 1 - Male';

proc means data = in_rg_and_peds n mean std min max;
       var age;
       title3 'Manuscript Table 1 - Age';

proc freq data = in_rg_and_peds;
       tables white;
       title3 'Manuscript Table 1 - White';

proc freq data = in_rg_and_peds;
       tables hispanic;
       title3 'Manuscript Table 1 - Hispanic ethnicity';

proc means data = in_rg_and_peds n median qrange min max;
       var blength;
       title3 'Manuscript Table 1 - Biopsy length';

proc freq data = in_rg_and_peds;
       tables nafld;
       title3 'Manuscript Table 1 - NAFLD pattern';

proc freq data = in_rg_and_peds;
       tables xfibro;
       title3 'Manuscript Table 1 - Fibrosis stage';

proc means data = in_rg_and_peds n mean std;
       var nfibro;
       title3 'Manuscript Table 1 - Fibrosis score';
```

```
proc freq data = in_rg_and_peds;
       tables nas;
       title3 'Manuscript Table 1 - NAS';

proc means data = in_rg_and_peds n mean std;
       var nas;

proc means data        = in_rg_and_peds n median qrange min max;
       var bmi;
       title3 'Manuscript Table 1 - BMI';

proc means data = in_rg_and_peds n median qrange min max;
       var bmipct;
       title3 'Manuscript Table 1 - BMI age-sex percentile';

proc means data = in_rg_and_peds n median qrange min max;
       var totalfat;
       title3 'Manuscript Table 1 - Body fat percentage';

proc means data = in_rg_and_peds n mean std min max;
       var tanner;
       title3 'Manuscript Table 1 - Tanner stage';

proc freq data = in_rg_and_peds;
       tables acanth;
       title3 'Manuscript Table 1 - Acanthosis nigricans';

data pe;
       set pe;
       if visit = 's1';

data lr;
       set lr;
       if visit = 's2';

data dsic_table1_values;
       merge in_rg_and_peds (in=val1 keep=nash)
             rg
             bg (in=val2)
        cr (in=val3 drop=registdt)
        pe (drop = registdt)
        lr (drop = registdt)
        ls (drop = registdt);
       by nash;
       age = input(rg110, 8.);
       blength = input(cr115, 8.);
       if blength < 10 then leng10 = 1;
       else leng10 = 0;
       if cr121 ne 'm' then fibro_score = input(substr(cr121, 1, 1), 8.);
```

```
nas = cr116a+cr117a+cr118a;
if pe208c = 1 then do;
       pe208a = 2.54*pe208a;
       pe208b = 2.54*pe208b;
end;
if pe208a ne . and pe208b ne . then height = (pe208a+pe208b)/2;
else if pe208a ne . and pe208b = . then height = pe208a;
else if pe208a = . and pe208b ne . then height = pe208b;
if pe209c = 1 then do;
       pe209a = 0.453592*pe209a;
       pe209b = 0.453592*pe209b;
end;
if pe209a ne . and pe209b ne . then weight = (pe209a+pe209b)/2;
else if pe209a ne . and pe209b = . then weight = pe209a;
else if pe209a = . and pe209b ne . then weight = pe209b;
bmi = weight/((height/100)**2);
if rg111 = '2' and age = 2 then do;
       L=-0.98660853;
       M=16.42339664;
       S=0.085451785;
end;
if rg111 = '1' and age = 2 then do;
       L=-2.01118107;
       M=16.57502768;
       S=0.080592465;
end;
if rg111 = '2' and age = 3 then do;
       L=-2.096828937;
       M=15.69924188;
       S=0.078605255;
end;
if rg111 = '1' and age = 3 then do;
       L=-1.419991255;
       M=16.00030401;
       S=0.072634432;
end;
if rg111 = '2' and age = 4 then do;
       L=-3.018521987;
       M=15.29854897;
       S=0.078713325;
end;
if rg111 = '1' and age = 4 then do;
       L=-1.714869347;
       M=15.62817269;
       S=0.071889214;
end;
if rg111 = '2' and age = 5 then do;
       L=-3.35007771;
       M=15.15188405;
```

```
        S=0.084300139;
end;
if rg111 = '1' and age = 5 then do;
        L=-2.61516595;
        M=15.41914163;
        S=0.07599225;
end;
if rg111 = '2' and age = 6 then do;
        L=-3.225606516;
        M=15.21690296;
        S=0.093803033;
end;
if rg111 = '1' and age = 6 then do;
        L=-3.21170511;
        M=15.38353217;
        S=0.083048178;
end;
if rg111 = '2' and age = 7 then do;
        L=-2.926186592;
        M=15.45356545;
        S=0.105325289;
end;
if rg111 = '1' and age = 7 then do;
        L=-3.323188896;
        M=15.51286936;
        S=0.092131305;
end;
if rg111 = '2' and age = 8 then do;
        L=-2.617192204;
        M=15.82699517;
        S=0.117158667;
end;
if rg111 = '1' and age = 8 then do;
        L=-3.18305795;
        M=15.78231007;
        S=0.102091189;
end;
if rg111 = '2' and age = 9 then do;
        L=-2.360920527;
        M=16.30609316;
        S=0.128013515;
end;
if rg111 = '1' and age = 9 then do;
        L=-2.971148225;
        M=16.16712234;
        S=0.111720691;
end;
if rg111 = '2' and age = 10 then do;
        L=-2.171295888;
```

```
        M=16.86231366;
        S=0.137057004;
end;
if rg111 = '1' and age = 10 then do;
        L=-2.765648008;
        M=16.64613844;
        S=0.120112464;
end;
if rg111 = '2' and age = 11 then do;
        L=-2.045235058;
        M=17.46906585;
        S=0.143868341;
end;
if rg111 = '1' and age = 11 then do;
        L=-2.590560148;
        M=17.20088732;
        S=0.126734613;
end;
if rg111 = '2' and age = 12 then do;
        L=-1.975521156;
        M=18.10148804;
        S=0.148361495;
end;
if rg111 = '1' and age = 12 then do;
        L=-2.447426113;
        M=17.81463359;
        S=0.131389042;
end;
if rg111 = '1' and age = 13 then do;
        L=-2.3294571;
        M=18.47179706;
        S=0.13414147;
end;
if rg111 = '2' and age = 13 then do;
        L=-1.954977947;
        M=18.73643338;
        S=0.150705138;
end;
if rg111 = '2' and age = 14 then do;
        L=-1.977073595;
        M=19.35257209;
        S=0.151255713;
end;
if rg111 = '1' and age = 14 then do;
        L=-2.227362173;
        M=19.15758672;
        S=0.135251083;
end;
if rg111 = '2' and age = 15 then do;
```

```
               L=-2.034893091;
               M=19.9305662;
               S=0.150511645;
       end;
       if rg111 = '1' and age = 15 then do;
               L=-2.132344989;
               M=19.85766121;
               S=0.135110159;
       end;
       if rg111 = '2' and age = 16 then do;
               L=-2.119156972;
               M=20.45325617;
               S=0.14909006;
       end;
       if rg111 = '1' and age = 16 then do;
               L=-2.039015385;
               M=20.5576474;
               S=0.134198323;
       end;
       if rg111 = '2' and age = 17 then do;
               L=-2.215737645;
               M=20.90575839;
               S=0.147723315;
       end;
       if rg111 = '1' and age = 17 then do;
               L=-1.949134561;
               M=21.24247982;
               S=0.13305669;
       end;
       z = (((bmi/M)**L)-1)/(L*S);
       percentile = 100*probnorm (z);
       if rg111 = '1' then tanner_stage = pe237;
       else if rg111 = '2' then tanner_stage = pe240;
       if pe218 >= 1 then acanth = 1;
       else acanth = 0;
       homa_ir = (lr339a*lr339b)/405;
       quicki = 1/(log10(lr339b)+log10(lr339b));
       if rg114e = '1' and rg114a  = '' then white = 1;
       else white = 0;
       if cr123 = '2' then defnash = 1;
       else defnash = 0;
       if ls124 = '1' and ls124a < 40 then ls124 = '2';
       if val1 then output dsic_table1_values;

proc freq data = dsic_table1_values;
       tables study /missing;
       title3 'DSIC Table 1 - TONIC';

proc freq data = dsic_table1_values;
```

```
       tables rg111;
       title3 'DSIC Table 1 - Male';

proc means data = dsic_table1_values;
       var age;
       title3 'DSIC Table 1 - Age';

proc freq data = dsic_table1_values;
       tables white /list missing;
       title3 'DSIC Table 1 - White race';

proc freq data = dsic_table1_values;
       tables rg112 /missing;
       title3 'DSIC Table 1 - Hispanic ethnicity';

proc means data = dsic_table1_values n median qrange min max;
       var blength;
       title3 'DSIC Table 1 - Biopsy length';

proc freq data = dsic_table1_values;
       tables cr123 ;
       format cr123 $nafldf.;
       title3 'DSIC Table 1 - NAFLD pattern';

proc freq data = dsic_table1_values;
       tables cr121 ;
       format cr121 $fibrosisf.;
       title3 'DSIC Table 1 - Fibrosis stage';

proc means data = dsic_table1_values n mean std;
       var fibro_score;
       title3 'DSIC Table 1 - Fibrosis Score';

proc freq data = dsic_table1_values;
       tables nas;
       title3 'DSIC Table 1 - NAS';

proc means data = dsic_table1_values n mean std;
       var nas;

proc means data = dsic_table1_values n mean std min max;
       var bmi;
       title3 'DSIC Table 1 - BMI';

proc means data = dsic_table1_values n mean std min max;
       var percentile;
       title3 'DSIC Table 1 - BMI Percentile';

proc means data = dsic_table1_values n mean std min max;
```

```
        var tanner_stage;
        title3 'DSIC Table 1 - Tanner Stage';

proc freq data = dsic_table1_values;
        tables acanth;
        title3 'DSIC Table 1 - Acanthosis nigricans';

data table2_manuscript;
        set in_rg_and_peds;
        if nafld ne '1b' then output;

proc freq data = table2_manuscript;
        tables nafld;
        title3 'Table 2 Manuscript';

proc sort data = table2_manuscript;
        by nafld;

proc freq data = table2_manuscript;
        tables male;
        by nafld;
        title3 'Table 2 Manuscript - Male';

proc means data = table2_manuscript;
        var age;
        class nafld;
        title3 'Table 2 Manuscript - Age';

proc freq data = table2_manuscript;
        tables white;
        by nafld;
        title3 'Table 2 Manuscript - White';

proc freq data = table2_manuscript;
        tables hispanic;
        by nafld;
        title3 'Table 2 Manuscript - Hispanic';

proc freq data = table2_manuscript;
        tables ucsd;
        by nafld;
        title3 'Table 2 Manuscript - UCSD';

proc means data = table2_manuscript n median;
        var alt;
        class nafld;
        title3 'Table 2 Manuscript - ALT';

proc means data = table2_manuscript n median;
```

```
       var ast;
       class nafld;
       title3 'Table 2 Manuscript - AST';

proc means data = table2_manuscript n median;
       var alka;
       class nafld;
       title3 'Table 2 Manuscript - Alkaline phosphatase';

proc means data = table2_manuscript n median;
       var ggt;
       class nafld;
       title3 'Table 2 Manuscript - GGT';

proc means data = table2_manuscript n median;
       var gluc;
       class nafld;
       title3 'Table 2 Manuscript - Glucose';

proc means data = table2_manuscript n median;
       var insu;
       class nafld;
       title3 'Table 2 Manuscript - Insulin';

proc means data = table2_manuscript n median;
       var homa;
       class nafld;
       title3 'Table 2 Manuscript - HOMA-IR';

proc means data = table2_manuscript n median;
       var quicki;
       class nafld;
       title3 'Table 2 Manuscript - QUICKI';

proc freq data = table2_manuscript;
       tables ana;
       by nafld;
       title3 'Table 2 Manuscript - ANA';

proc freq data = table2_manuscript;
       tables asma;
       by nafld;
       title3 'Table 2 Manuscript - ASMA';

proc means data = table2_manuscript n median;
       var bmi;
       class nafld;
       title3 'Table 2 Manuscript - BMI';
```

```
proc means data = table2_manuscript n median;
       var bmipct;
       class nafld;
       title3 'Table 2 Manuscript - BMI percentile';

proc means data = table2_manuscript n median;
       var totalfat;
       class nafld;
       title3 'Table 2 Manuscript - Percentage of body fat';

proc means data = table2_manuscript;
       var tanner;
       class nafld;
       title3 'Table 2 Manuscript - Tanner stage';

proc means data = table2_manuscript n median;
       var blength;
       class nafld;
       title3 'Table 2 Manuscript - Biopsy length';

proc freq data = table2_manuscript;
       tables leng10;
       by nafld;
       title3 'Table 2 Manuscript - < 10 mm';

proc means data = table2_manuscript;
       var nfibro;
       class nafld;
       title3 'Table 2 Manuscript - Fibrosis score';

data table2_dsic;
       set dsic_table1_values;
       if cr123 ne '1b';

proc freq data = table2_dsic;
       tables cr123;
       format cr123 $nafldf.;
       title3 'Table 2 DSIC';

proc sort data = table2_dsic;
       by cr123;

proc freq data = table2_dsic;
       tables rg111;
       by cr123;
       title3 'DSIC Table 2 - Male';

proc means data = table2_dsic;
       var age;
```

```
       class cr123;
       title3 'DSIC Table 2 - Age';

proc freq data = table2_dsic;
       tables white /missing;
       by cr123;
       title3 'DSIC Table 2 - White race';

proc freq data = table2_dsic;
       tables rg112 /missing;
       by cr123;
       title3 'DSIC Table 2 - Hispanic ethnicity';

proc means data = table2_dsic n median;
       var lr329;
       class cr123;
       title3 'DSIC Table 2 - ALT';

proc means data = table2_dsic n median;
       var lr328;
       class cr123;
       title3 'DSIC Table 2 - AST';

proc means data = table2_dsic n median;
       var lr330;
       class cr123;
       title3 'DSIC Table 2 - Alkaline Phosphatase';

proc means data = table2_dsic n median;
       var lr331;
       class cr123;
       title3 'DSIC Table 2 - GGT';

proc means data = table2_dsic n median;
       var lr339a;
       class cr123;
       title3 'DSIC Table 2 - Fasting glucose';

proc means data = table2_dsic n median;
       var lr339b;
       class cr123;
       title3 'DSIC Table 2 - Fasting insulin';

proc means data = table2_dsic n median;
       var homa_ir;
       class cr123;
       title3 'DSIC Table 2 - HOMA-IR';

proc means data = table2_dsic n median;
```

```
       var quicki;
       class cr123;
       title3 'DSIC Table 2 - QUICKI';

proc freq data = table2_dsic;
       tables ls123;
       by cr123;
       title3 'DSIC Table 2 - ANA';

proc freq data = table2_dsic;
       tables ls124;
       by cr123;
       where ls124 in ('1', '2');
       title3 'DSIC Table 2 - ASMA';

proc means data = table2_dsic n median;
       var bmi;
       class cr123;
       title3 'DSIC Table 2 - BMI';

proc means data = table2_dsic n median;
       var percentile;
       class cr123;
       title3 'DSIC Table 2 - BMI Percentile';

proc means data = table2_dsic;
       var tanner_stage;
       class cr123;
       title3 'DSIC Table 2 - Tanner Stage';

proc means data = table2_dsic n median;
       var blength;
       class cr123;
       title3 'DSIC Table 2 - Biopsy length';

proc freq data = table2_dsic;
       tables leng10;
       by cr123;
       title3 'DSIC Table 2 - < 10 mm';

proc means data = table2_dsic;
       var fibro_score;
       class cr123;
       title3 'DSIC Table 2 - Fibrosis Score';

data table3_manuscript;
       set in_rg_and_peds;
       if xfibro ne 0 and nafld in ('1b', '2');
```

```
proc freq data = table3_manuscript;
       tables nafld;
       title3 'Table 3 Manuscript';

proc sort data = table3_manuscript;
       by nafld;

proc freq data = table3_manuscript;
       tables male;
       by nafld;
       title3 'Table 3 Manuscript - Male';

proc means data = table3_manuscript;
       var age;
       class nafld;
       title3 'Table 3 Manuscript - Age';

proc freq data = table3_manuscript;
       tables white;
       by nafld;
       title3 'Table 3 Manuscript - White';

proc freq data = table3_manuscript;
       tables hispanic;
       by nafld;
       title3 'Table 3 Manuscript - Hispanic';

proc freq data = table3_manuscript;
       tables ucsd;
       by nafld;
       title3 'Table 3 Manuscript - UCSD';

proc means data = table3_manuscript n median;
       var alt;
       class nafld;
       title3 'Table 3 Manuscript - ALT';

proc means data = table3_manuscript n median;
       var ast;
       class nafld;
       title3 'Table 3 Manuscript - AST';

proc means data = table3_manuscript n median;
       var alka;
       class nafld;
       title3 'Table 3 Manuscript - Alkaline phosphatase';

proc means data = table3_manuscript n median;
       var ggt;
```

```
        class nafld;
        title3 'Table 3 Manuscript - GGT';

proc means data = table3_manuscript n median;
        var tri;
        class nafld;
        title3 'Table 3 Manuscript - triglycerides';

proc means data = table3_manuscript n median;
        var gluc;
        class nafld;
        title3 'Table 3 Manuscript - Glucose';

proc means data = table3_manuscript n median;
        var insu;
        class nafld;
        title3 'Table 3 Manuscript - Insulin';

proc means data = table3_manuscript n median;
        var homa;
        class nafld;
        title3 'Table 3 Manuscript - HOMA-IR';

proc means data = table3_manuscript n median;
        var quicki;
        class nafld;
        title3 'Table 3 Manuscript - QUICKI';

proc freq data = table3_manuscript;
        tables ana;
        by nafld;
        title3 'Table 3 Manuscript - ANA';

proc freq data = table3_manuscript;
        tables asma;
        by nafld;
        title3 'Table 3 Manuscript - ASMA';

proc means data = table3_manuscript n median;
        var bmi;
        class nafld;
        title3 'Table 3 Manuscript - BMI';

proc means data = table3_manuscript n median;
        var bmipct;
        class nafld;
        title3 'Table 3 Manuscript - BMI percentile';

proc means data = table3_manuscript n median;
```

```
       var totalfat;
       class nafld;
       title3 'Table 3 Manuscript - Percentage of body fat';

proc means data = table3_manuscript;
       var tanner;
       class nafld;
       title3 'Table 3 Manuscript - Tanner stage';

proc means data = table3_manuscript n median;
       var blength;
       class nafld;
       title3 'Table 3 Manuscript - Biopsy length';

proc freq data = table3_manuscript;
       tables leng10;
       by nafld;
       title3 'Table 3 Manuscript - < 10 mm';

proc means data = table3_manuscript;
       var nfibro;
       class nafld;
       title3 'Table 3 Manuscript - Fibrosis score';

data table3_dsic;
       set dsic_table1_values;
       if     fibro_score ne 0 and cr123 in ('1b', '2');

proc freq data = table3_dsic;
       tables cr123;
       title3 'Table 3 DSIC';

proc sort data = table3_dsic;
       by cr123;

proc freq data = table3_dsic;
       tables rg111;
       by cr123;
       title3 'DSIC Table 3 - Male';

proc means data = table3_dsic;
       var age;
       class cr123;
       title3 'DSIC Table 3 - Age';

proc freq data = table3_dsic;
       tables white /missing;
       by cr123;
       title3 'DSIC Table 3 - White race';
```

```
proc freq data = table3_dsic;
       tables rg112 /missing;
       by cr123;
       title3 'DSIC Table 3 - Hispanic ethnicity';

proc means data = table3_dsic n median;
       var lr329;
       class cr123;
       title3 'DSIC Table 3 - ALT';

proc means data = table3_dsic n median;
       var lr328;
       class cr123;
       title3 'DSIC Table 3 - AST';

proc means data = table3_dsic n median;
       var lr330;
       class cr123;
       title3 'DSIC Table 3 - Alkaline Phosphatase';

proc means data = table3_dsic n median;
       var lr331;
       class cr123;
       title3 'DSIC Table 3 - GGT';

proc means data = table3_dsic n median;
       var lr338a;
       class cr123;
       title3 'DSIC Table 3 - Triglycerides';

proc means data = table3_dsic n median;
       var lr339a;
       class cr123;
       title3 'DSIC Table 3 - Fasting glucose';

proc means data = table3_dsic n median;
       var lr339b;
       class cr123;
       title3 'DSIC Table 3 - Fasting insulin';

proc means data = table3_dsic n median;
       var homa_ir;
       class cr123;
       title3 'DSIC Table 3 - HOMA-IR';

proc means data = table3_dsic n median;
       var quicki;
       class cr123;
```

```
        title3 'DSIC Table 3 - QUICKI';

proc freq data = table3_dsic;
        tables ls123;
        by cr123;
        title3 'DSIC Table 3 - ANA';

proc freq data = table3_dsic;
        tables ls124;
        by cr123;
        where ls124 in ('1', '2');
        title3 'DSIC Table 3 - ASMA';

proc means data = table3_dsic n median;
        var bmi;
        class cr123;
        title3 'DSIC Table 3 - BMI';

proc means data = table3_dsic n median;
        var percentile;
        class cr123;
        title3 'DSIC Table 3 - BMI Percentile';

proc means data = table3_dsic;
        var tanner_stage;
        class cr123;
        title3 'DSIC Table 3 - Tanner Stage';

proc means data = table3_dsic n median;
        var blength;
        class cr123;
        title3 'DSIC Table 3 - Biopsy length';

proc freq data = table3_dsic;
        tables leng10;
        by cr123;
        title3 'DSIC Table 3 - < 10 mm';

proc means data = table3_dsic;
        var fibro_score;
        class cr123;
        title3 'DSIC Table 3 - Fibrosis Score';

proc sort data = in_rg_and_peds;
        by nfibro;

proc freq data = in_rg_and_peds;
        tables nfibro;
        title3 'Table 4 Manuscript';
```

```
proc freq data = in_rg_and_peds;
       tables male;
       by nfibro;
       title3 'Table 4 Manuscript - Male';

proc means data = in_rg_and_peds;
       var age;
       class nfibro;
       title3 'Table 4 Manuscript - Age';

proc freq data = in_rg_and_peds;
       tables white;
       by nfibro;
       title3 'Table 4 Manuscript - White';

proc freq data = in_rg_and_peds;
       tables hispanic;
       by nfibro;
       title3 'Table 4 Manuscript - Hispanic';

proc freq data = in_rg_and_peds;
       tables ucsd;
       by nfibro;
       title3 'Table 4 Manuscript - UCSD';

proc means data = in_rg_and_peds n median;
       var alt;
       class nfibro;
       title3 'Table 4 Manuscript - ALT';

proc means data = in_rg_and_peds n median;
       var ast;
       class nfibro;
       title3 'Table 4 Manuscript - AST';

proc means data = in_rg_and_peds n median;
       var alka;
       class nfibro;
       title3 'Table 4 Manuscript - Alkaline phosphatase';

proc means data = in_rg_and_peds n median;
       var ggt;
       class nfibro;
       title3 'Table 4 Manuscript - GGT';

proc means data = in_rg_and_peds n median;
       var alb;
       class nfibro;
```

```
       title3 'Table 4 Manuscript - Albumin';

proc means data = in_rg_and_peds n median;
       var wbc;
       class nfibro;
       title3 'Table 4 Manuscript - White blood cell count';

proc means data = in_rg_and_peds n median;
       var hema;
       class nfibro;
       title3 'Table 4 Manuscript - Hematocrit';

proc means data = in_rg_and_peds n median;
       var insu;
       class nfibro;
       title3 'Table 4 Manuscript - Insulin';

proc means data = in_rg_and_peds n median;
       var homa;
       class nfibro;
       title3 'Table 4 Manuscript - HOMA-IR';

proc means data = in_rg_and_peds n median;
       var quicki;
       class nfibro;
       title3 'Table 4 Manuscript - QUICKI';

proc freq data = in_rg_and_peds;
       tables ana;
       by nfibro;
       title3 'Table 4 Manuscript - ANA';

proc freq data = in_rg_and_peds;
       tables asma;
       by nfibro;
       title3 'Table 4 Manuscript - ASMA';

proc means data = in_rg_and_peds n median;
       var bmi;
       class nfibro;
       title3 'Table 4 Manuscript - BMI';

proc means data = in_rg_and_peds n median;
       var bmipct;
       class nfibro;
       title3 'Table 4 Manuscript - BMI percentile';

proc means data = in_rg_and_peds n median;
       var totalfat;
```

```
       class nfibro;
       title3 'Table 4 Manuscript - Percentage of body fat';

proc means data = in_rg_and_peds;
       var tanner;
       class nfibro;
       title3 'Table 4 Manuscript - Tanner stage';

proc means data = in_rg_and_peds n median;
       var blength;
       class nfibro;
       title3 'Table 4 Manuscript - Biopsy length';

proc freq data = in_rg_and_peds;
       tables leng10;
       by nfibro;
       title3 'Table 4 Manuscript - < 10 mm';

proc freq data = in_rg_and_peds;
       tables dnash;
       by nfibro;
       title3 'Table 4 Manuscript - Definite NASH';

proc means data = in_rg_and_peds;
       var nas;
       class nfibro;
       title3 'Table 4 Manuscript - NAS';

proc freq data = dsic_table1_values;
       tables fibro_score;
       title3 'DSIC Table 4';

proc sort data = dsic_table1_values;
       by fibro_score;

proc freq data = dsic_table1_values;
       tables rg111;
       by fibro_score;
       title3 'DSIC Table 4 - Male';

proc means data = dsic_table1_values;
       var age;
       class fibro_score;
       title3 'DSIC Table 4 - Age';

proc freq data = dsic_table1_values;
       tables white /missing;
       by fibro_score;
       title3 'DSIC Table 4 - White race';
```

```
proc freq data = dsic_table1_values;
       tables rg112 /missing;
       by fibro_score;
       title3 'DSIC Table 4 - Hispanic ethnicity';

proc means data = dsic_table1_values n median;
       var lr329;
       class fibro_score;
       title3 'DSIC Table 4 - ALT';

proc means data = dsic_table1_values n median;
       var lr328;
       class fibro_score;
       title3 'DSIC Table 4 - AST';

proc means data = dsic_table1_values n median;
       var lr330;
       class fibro_score;
       title3 'DSIC Table 4 - Alkaline Phosphatase';

proc means data = dsic_table1_values n median;
       var lr331;
       class fibro_score;
       title3 'DSIC Table 4 - GGT';

proc means data = dsic_table1_values n median;
       var lr333;
       class fibro_score;
       title3 'DSIC Table 4 - Albumin';

proc means data = dsic_table1_values n median;
       var lr311;
       class fibro_score;
       title3 'DSIC Table 4 - White blood cell count';

proc means data = dsic_table1_values n median;
       var lr310;
       class fibro_score;
       title3 'DSIC Table 4 - Hematocrit';

proc means data = dsic_table1_values n median;
       var lr339b;
       class fibro_score;
       title3 'DSIC Table 4 - Fasting insulin';

proc means data = dsic_table1_values n median;
       var homa_ir;
       class fibro_score;
```

```
       title3 'DSIC Table 4 - HOMA-IR';

proc means data = dsic_table1_values n median;
       var quicki;
       class fibro_score;
       title3 'DSIC Table 4 - QUICKI';

proc freq data = dsic_table1_values;
       tables ls123;
       by fibro_score;
       title3 'DSIC Table 4 - ANA';

proc freq data = dsic_table1_values;
       tables ls124;
       by fibro_score;
       where ls124 in ('1', '2');
       title3 'DSIC Table 4 - ASMA';

proc means data = dsic_table1_values n median;
       var bmi;
       class fibro_score;
       title3 'DSIC Table 4 - BMI';

proc means data = dsic_table1_values n median;
       var percentile;
       class fibro_score;
       title3 'DSIC Table 4 - BMI Percentile';

proc means data = dsic_table1_values;
       var tanner_stage;
       class fibro_score;
       title3 'DSIC Table 4 - Tanner Stage';

proc means data = dsic_table1_values n median;
       var blength;
       class fibro_score;
       title3 'DSIC Table 4 - Biopsy length';

proc freq data = dsic_table1_values;
       tables leng10;
       by fibro_score;
       title3 'DSIC Table 4 - < 10 mm';

proc freq data = dsic_table1_values;
       tables defnash;
       by fibro_score;
       title3 'DSIC Table 4 - Definite NASH';

proc means data = dsic_table1_values;
```

```
        var nas;
        class fibro_score;
        title3 'DSIC Table 4 - NAS';

data table5_manuscript;
        set in_rg_and_peds;
        if nas in (1, 2, 3) then nas_group = 1;
        if nas in (4,5) then nas_group = 2;
        if nas in (6,7) then nas_group = 3;

proc freq data = table5_manuscript;
        tables nas_group;
        title3 'Table 5 Manuscript';

proc sort data = table5_manuscript;
        by nas_group;

proc freq data = table5_manuscript;
        tables male;
        by nas_group;
        title3 'Table 5 Manuscript - Male';

proc means data = table5_manuscript;
        var age;
        class nas_group;
        title3 'Table 5 Manuscript - Age';

proc freq data = table5_manuscript;
        tables white;
        by nas_group;
        title3 'Table 5 Manuscript - White';

proc freq data = table5_manuscript;
        tables hispanic;
        by nas_group;
        title3 'Table 5 Manuscript - Hispanic';

proc freq data = table5_manuscript;
        tables ucsd;
        by nas_group;
        title3 'Table 5 Manuscript - UCSD';

proc means data = table5_manuscript n median;
        var alt;
        class nas_group;
        title3 'Table 5 Manuscript - ALT';

proc means data = table5_manuscript n median;
        var ast;
```

```
       class nas_group;
       title3 'Table 5 Manuscript - AST';

proc means data = table5_manuscript n median;
       var alka;
       class nas_group;
       title3 'Table 5 Manuscript - Alkaline phosphatase';

proc means data = table5_manuscript n median;
       var ggt;
       class nas_group;
       title3 'Table 5 Manuscript - GGT';

proc means data = table5_manuscript n median;
       var gluc;
       class nas_group;
       title3 'Table 5 Manuscript - Glucose';

proc means data = table5_manuscript n median;
       var insu;
       class nas_group;
       title3 'Table 5 Manuscript - Insulin';

proc means data = table5_manuscript n median;
       var homa;
       class nas_group;
       title3 'Table 5 Manuscript - HOMA-IR';

proc means data = table5_manuscript n median;
       var quicki;
       class nas_group;
       title3 'Table 5 Manuscript - QUICKI';

proc freq data = table5_manuscript;
       tables ana;
       by nas_group;
       title3 'Table 5 Manuscript - ANA';

proc freq data = table5_manuscript;
       tables asma;
       by nas_group;
       title3 'Table 5 Manuscript - ASMA';

proc means data = table5_manuscript n median;
       var blength;
       class nas_group;
       title3 'Table 5 Manuscript - Biopsy length';

proc freq data = table5_manuscript;
```

```
        tables leng10;
        by nas_group;
        title3 'Table 5 Manuscript - < 10 mm';

proc freq data = table5_manuscript;
        tables dnash;
        by nas_group;
        title3 'Table 5 Manuscript - Definite NASH';

proc means data = table5_manuscript;
        var nfibro;
        class nas_group;
        title3 'Table 5 Manuscript - Fibrosis score';

data dsic_table1_values;
        set dsic_table1_values;
        if nas in (1,2,3) then nas_group = 1;
        if nas in (4,5) then nas_group = 2;
        if nas in (6,7) then nas_group = 3;

proc freq data = dsic_table1_values;
        tables nas_group;
        title3 'DSIC Table 5';

proc sort data = dsic_table1_values;
        by nas_group;

proc freq data = dsic_table1_values;
        tables rg111;
        by nas_group;
        title3 'DSIC Table 5 - Male';

proc means data = dsic_table1_values;
        var age;
        class nas_group;
        title3 'DSIC Table 5 - Age';

proc freq data = dsic_table1_values;
        tables white /missing;
        by nas_group;
        title3 'DSIC Table 5 - White race';

proc freq data = dsic_table1_values;
        tables rg112 /missing;
        by nas_group;
        title3 'DSIC Table 5 - Hispanic ethnicity';

proc means data = dsic_table1_values n median;
        var lr329;
```

```
       class nas_group;
       title3 'DSIC Table 5 - ALT';

proc means data = dsic_table1_values n median;
       var lr328;
       class nas_group;
       title3 'DSIC Table 5 - AST';

proc means data = dsic_table1_values n median;
       var lr330;
       class nas_group;
       title3 'DSIC Table 5 - Alkaline Phosphatase';

proc means data = dsic_table1_values n median;
       var lr331;
       class nas_group;
       title3 'DSIC Table 5 - GGT';

proc means data = dsic_table1_values n median;
       var lr339a;
       class nas_group;
       title3 'DSIC Table 5 - Fasting glucose';

proc means data = dsic_table1_values n median;
       var lr339b;
       class nas_group;
       title3 'DSIC Table 5 - Fasting insulin';

proc means data = dsic_table1_values n median;
       var homa_ir;
       class nas_group;
       title3 'DSIC Table 5 - HOMA-IR';

proc means data = dsic_table1_values n median;
       var quicki;
       class nas_group;
       title3 'DSIC Table 5 - QUICKI';

proc freq data = dsic_table1_values;
       tables ls123;
       by nas_group;
       title3 'DSIC Table 5 - ANA';

proc freq data = dsic_table1_values;
       tables ls124;
       by nas_group;
       where ls124 in ('1', '2');
       title3 'DSIC Table 5 - ASMA';
```

```
proc means data = dsic_table1_values n median;
       var blength;
       class nas_group;
       title3 'DSIC Table 5 - Biopsy length';

proc freq data = dsic_table1_values;
       tables leng10;
       by nas_group;
       title3 'DSIC Table 5 - < 10 mm';

proc freq data = dsic_table1_values;
       tables defnash;
       by nas_group;
       title3 'DSIC Table 5 - Definite NASH';

proc means data = dsic_table1_values;
       var fibro_score;
       class nas_group;
       title3 'DSIC Table 5 - NAS';
```