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

Prepared by Allyson Mateja
IMS Inc.
3901 Calverton Blvd, Suite 200, Calverton MD, 20705
April 29, 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) affects 10%-30% of the general U.S. population 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 is 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).

3 Archived Datasets

All SAS data files, as provided by the Data Coordinating Center (DCC), are located 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 Neuschwander-Tetri et al in Hepatology in September 2010 [1].

To verify the integrity of the three datasets, descriptive statistics were computed.

5 Results

Note that there is not a one-to-one match between the NAFLD Adult analysis datasets and the NAFLD Adult 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], Characteristics of Adult Patients with NAFLD Enrolled in the NASH CRN Studies, Table A lists the variables that were used in the replication and Table B compares the results calculated from the archived data file to the results published in Table 1. The results of the replication are similar to the published results.

For Table 2 in the publication [1], Characteristics of Adult Patients with NAFLD with Contemporaneous Biopsies and Clinical Factors by Presence of Definite NASH, Table C lists the variables that were used in the replication and Table D compares the results calculated from the archived data file to the results published in Table 2. The results of the replication are similar to the published results.

For Table 3 in the publication [1], Characteristics of Adult Patients with NAFLD with Contemporaneous Biopsies and Clinical Factors by Fibrosis Stage, Table E lists the variables that were used in the replication and Table F compares the results calculated from the archived data file to the results published in Table 3. The results of the replication are similar to the published results.

6 Conclusions

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

7 References

[1] Neuschwander-Tetri, B. A., Clark, J. M., Bass, N. M., Van Natta, M. L., Unalp-Arida, A., Tonascia, J., Zein, C. O., Brunt, E. M., Kleiner, D. E., McCullough, A. J., Sanyal, A. J., Diehl, A. M., Lavine, J. E., Chalasani, N., Kowdley, K. V. and NASH Clinical Research Network (2010), Clinical, laboratory and histological associations in adults with nonalcoholic fatty liver disease. Hepatology, 52: 913–924. doi: 10.1002/hep.23784.

Table A: Variables used to replicate Table 1: Characteristics of Adult Patients with NAFLD Enrolled in the NASH CRN Studies

Table Variable	**dataset.variable**
Proximity of Liver Biopsy to Enrollment	bg.bg320a, cr.cr108
Acanthosis nigricans Positive (%)	pe.pe218
Acanthosis nigricans severity score	pe.pe218
Age - yrs	rg.rg110
Albumin - g/dL	lr.lr333
Alkaline phosphatase - U/L	lr.lr330
Isolated Alkaline phosphatase	lr.lr330, lr.lr330a, lr.lr328a, lr.lr329a
AST - U/L	lr.lr328
Abnormal AST	lr.lr328, lr.lr328a
AMA	ls.ls125
ANA	ls.ls123
ANA and ASMA	ls.ls123, ls.ls124
ASMA	ls.ls124
ALT - U/L	lr.lr329
Abnormal ALT	lr.lr329, lr.lr329a
Ballooning	cr.cr118a
Bilirubin, direct - mg/dL	lr.lr327
Bilirubin, total - mg/dL	lr.lr326
Body Mass Index - kg/m²	pe.pe208a, pe.pe208b, pe.pe209a, pe.pe209b, pe208c, pe209c
Total cholesterol - mg/dL	lr.lr338b
Presence of NASH	cr.cr123
Type 2 diabetes	bg.bg349b
Ferritin ng/mL	ls.ls109c
GGT - U/L	lr.lr331
Globulin - g/dL	lr.lr332, lr.lr333
Fasting serum glucose mg/dL	lr.lr339a
HbA1c - %	lr.lr324
HDL cholesterol - mg/dL	lr.lr338c
Hematocrit - %	lr.lr310
Hispanic ethnicity	rg.rg112
HOMA-IR - mg/dL/uU/mL/405	lr.lr339a, lr.lr339b
Hypertension	bg.bg349af

Table Variable	**dataset.variable**
Fibrosis score	cr.cr121
Steatosis	cr.cr116a
Lobular inflammation	cr.cr117a
NAFLD Activity Score	cr.cr116a, cr.cr117a, cr.cr118a
International normalized ratio	lr.lr335
Fasting serum insulin uU/mL	lr.lr339b
Portal inflammation	cr.cr117d
LDL cholesterol - mg/dL	lr.lr338b
Biopsy length	cr.cr115
Male gender	rg.rg111
Mallory bodies	cr.cr119
Metabolic syndrome	rg.rg111, pe.pe210a, pe.pe210b, pe.pe210c, lr.lr338a, lr.lr338c, pe.pe214a, pe.pe214b, lr.lr339a
Platelets - 1000/mm3	lr.lr312
AST/ALT ratio	lr.lr328, lr.lr329
Triglycerides - mg/dL	lr.lr338a
Waist circumference - cm	pe.pe210a, pe.pe210b, pe.pe210c
White blood cell - 1000/mm3	lr.lr311
White race	rg.rg114e
Waist-to-hip ratio	pe.pe210a, pe.pe210b, pe.pe210c, pe.pe211a, pe.pe211b, pe.pe211c

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

Characteristic	<=6 Months (n=491) [Manuscript]	<=6 Months (n=491) [DSIC]	<=6 Months (n=0) [Diff]	>6 Months (n=399) [Manuscript]	>6 Months (n=394) [DSIC]	>6 Months (n=5) [Diff]
Demographics						
Male (%)	38	38	0	34	34	0
Age, years (median ± SD)	49	50	1	52	52	0
White (%)	80	80	0	85	85	0
Hispanic (%)	13	13	0	7	6	1
Clinical						
Hypertension (%)	48	51	3	56	56	0
Type 2 diabetes (%)	32	34	2	43	41	2
Metabolic syndrome (%)	65	63	2	59	60	1
Acanthosis nigricans						
Positive (%)	14	14	0	9	8	1
Severity score (mean ± SD)	0.33	0.34	0.01	0.18	0.15	0.03
Anthropometric (median ± SD)						
Body mass index (kg/m²)	34	34	0	33	33	0
Waist circumference (cm)	109	109	0	106	105	1
Waist-to-hip ratio	0.94	0.94	0	0.93	0.93	0
Hepatology panel (median ± SD)						
AST (U/L)	44	44	0	37	37	0
Abnormal AST (%)	45	44	1	35	36	1
ALT (U/L)	60	60	0	44	44	0

Characteristic	<=6 Months (n=491) [Manuscript]	<=6 Months (n=491) [DSIC]	<=6 Months (n=0) [Diff]	>6 Months (n=399) [Manuscript]	>6 Months (n=394) [DSIC]	>6 Months (n=5) [Diff]
Abnormal ALT (%)	61	60	1	41	42	1
AST/ALT	0.74	0.75	0.1	0.84	0.84	0
Alkaline phosphatase (U/L)	83	83	0	85	86	1
Isolated abnormal alkaline phosphatase (%)	4	4	0	7	7	0
GGT (U/L)	50	51	1	46	47	1
Globulin (g/dL)	3	3	0	3	3	0
Albumin (g/dL)	4.2	4.2	0	4.3	4.3	0
Bilirubin, total (mg/dL)	0.7	0.7	0	0.7	0.7	0
Bilirubin, direct (mg/dL)	0.1	0.1	0	0.1	0.1	0
International normalized ratio (mean ± SD)	1.02	1	0.02	1.04	1	0.04
Hematology and other laboratory studies (median ± SD)						
Hematocrit (%)	42.3	42	0.3	41	41	0
White blood cells (1000/mm³)	6.8	6.7	0.1	6.3	6.3	0
Platelet count (1000/mm³)	246	244	2	225	225	0
Total cholesterol (mg/dL)	194	193	1	188.5	189	0.5
HDL cholesterol (mg/dL)	41	41	0	45	45	0
LDL cholesterol (mg/dL)	119	119	0	111	111	0
Triglycerides (mg/dL)	158	153	5	141	143	2
HbA1c (%)	5.7	5.8	0.1	5.8	5.8	0

Characteristic	<=6 Months (n=491) [Manuscript]	<=6 Months (n=491) [DSIC]	<=6 Months (n=0) [Diff]	>6 Months (n=399) [Manuscript]	>6 Months (n=394) [DSIC]	>6 Months (n=5) [Diff]
Fasting serum glucose (mg/dL)	97	97	9	99	100	1
Fasting serum insulin (µU/mL)	18.7	18.3	0.4	17	17	0
HOMA-IR (mg/dL µU/mL/405)	4.6	4.4	0.2	4.4	4.5	0.1
ANA (% positive)	25	24	1	20	21	1
ASMA (% positive)	14	15	1	16	14	2
ANA + ASMA (% both positive)	4	4	0	5	5	0
AMA (% positive)	2	1	1	1	1	0
Ferritin (ng/mL)	146	148	2	127	127	0
Histology						
Steatosis (% ≥ 34%)	56	52	4	46	48	2
Lobular inflammation (% ≥ grade 2)	43	41	2	43	46	3
Portal inflammation (% > mild)	20	22	2	29	27	2
Ballooning (% any)	62	59	3	64	65	1
NAFLD Activity Score (% ≥ 5)	43	42	1	43	44	1
Presence of NASH (% definite)	52	50	2	54	56	2
Fibrosis score (mean ± SD)	1.5	1.6	0.1	1.9	1.9	0
Mallory bodies (% present)	27	27	0	33	33	0
Biopsy length (% < 10 mm)	15	13	2	14	14	0

Characteristic	No Liver Biopsy (n=165) [Manuscript]	No Liver Biopsy (n=170) [DSIC]	No Liver Biopsy (n=5) [Diff]	Total (n=1055) [Manuscript]	Total (n=1055) [DSIC]	Total (n=0) [Diff]
Demographics						
Male (%)	33	35	2	36	36	0
Age, years (median ± SD)	52	52	0	51 ± 12	51 ± 12	0 ± 0
White (%)	79	79	0	82	82	0
Hispanic (%)	14	14	0	11	11	0
Clinical						
Hypertension (%)	52	47	5	52	52	0
Type 2 diabetes (%)	40	35	5	37	37	0
Metabolic syndrome (%)	57	53	4	61	60	1
Acanthosis nigricans						
Positive (%)	13	12	1	12	12	0
Severity score (mean ± SD)	0.28	0.27	0.01	0.26 ± 0.81	0.26 ± 0.80	0 ± 0.01
Anthropometric (median ± SD)						
Body mass index (kg/m²)	33	33	0	34 ± 6	34 ± 6	0 ± 0
Waist circumference (cm)	106	107	1	108 ± 15	108 ± 15	0 ± 0
Waist-to-hip ratio	0.92	0.92	0	0.93 ± 0.08	0.93 ± 0.08	0 ± 0
Hepatology panel (median ± SD)						
AST (U/L)	36	36	0	40 ± 34	40 ± 34	0 ± 0
Abnormal AST (%)	27	26	1	38	38	0

Characteristic	No Liver Biopsy (n=165) [Manuscript]	No Liver Biopsy (n=170) [DSIC]	No Liver Biopsy (n=5) [Diff]	Total (n=1055) [Manuscript]	Total (n=1055) [DSIC]	Total (n=0) [Diff]
ALT (U/L)	45	45	0	52 ± 48	52 ± 48	0 ± 0
Abnormal ALT (%)	42	40	2	50	50	0
AST/ALT	0.79	0.78	0.1	0.78 ± 0.39	0.78 ± 0.39	0 ± 0
Alkaline phosphatase (U/L)	82	83	1	83 ± 38	83 ± 39	0 ± 1
Isolated abnormal alkaline phosphatase (%)	2	2	0	5	5	0
GGT (U/L)	45	42	3	48 ± 90	48 ± 90	0 ± 0
Globulin (g/dL)	3	3	0	3 ± 0.6	3 ± 0.6	0 ± 0
Albumin (g/dL)	4.2	4.2	0	4.2 ± 0.5	4.2 ± 0.5	0 ± 0
Bilirubin, total (mg/dL)	0.7	0.7	0	0.7 ± 0.5	0.7 ± 0.5	0 ± 0
Bilirubin, direct (mg/dL)	0.1	0.1	0	0.1 ± 0.15	0.1 ± 0.15	0 ± 0
International normalized ratio (mean ± SD)	1.04	1	0.04	1.04 ± 0.20	1 ± 0.21	0.04 ± 0.01
Hematology and other laboratory studies (median ± SD)						
Hematocrit (%)	41	41	0	42 ± 4	42 ± 4	0 ± 0
White blood cells (1000/mm³)	6.5	6.6	0.1	6.6 ± 2.8	6.6 ± 2.9	0 ± 0.1
Platelet count (1000/mm³)	236	240.5	4.5	237 ± 80	237 ± 80	0 ± 0
Total cholesterol (mg/dL)	184	183	1	191 ± 44	190 ± 44	1 ± 0
HDL cholesterol (mg/dL)	43	43	0	43 ± 13	43 ± 13	0 ± 0
LDL cholesterol (mg/dL)	110	110	0	115 ± 37	115 ± 38	0 ± 1

Characteristic	No Liver Biopsy (n=165) [Manuscript]	No Liver Biopsy (n=170) [DSIC]	No Liver Biopsy (n=5) [Diff]	Total (n=1055) [Manuscript]	Total (n=1055) [DSIC]	Total (n=0) [Diff]
Triglycerides (mg/dL)	134	133	1	145 ± 124	144 ± 123	1 ± 1
HbA1c (%)	5.8	5.8	0	5.8 ± 1.3	5.8 ± 1.3	0 ± 0
Fasting serum glucose (mg/dL)	96	96	0	98 ± 41	98 ± 41	0 ± 0
Fasting serum insulin (μU/mL)	18.9	18.6	0.3	18 ± 23	18 ± 23	0 ± 0
HOMA-IR (mg/dL μU/mL/405)	4.7	4.6	0.1	4.5 ± 8.2	4.5 ± 8.2	0 ± 0
ANA (% positive)	30	29	1	24	24	0
ASMA (% positive)	27	25	2	17	16	1
ANA + ASMA (% both positive)	11	10	1	5	5	0
AMA (% positive)	2	2	0	1	1	0
Ferritin (ng/mL)	108	106.5	1.5	131 ± 242	131.5 ± 242	0.5 ± 0
Histology						
Steatosis (% ≥ 34%)	-	-	-	51	50	1
Lobular inflammation (% ≥ grade 2)	-	-	-	43	43	0
Portal inflammation (% > mild)	-	-	-	24	25	1
Ballooning (% any)	-	-	-	63	62	1
NAFLD Activity Score (% ≥ 5)	-	-	-	43	43	0
Presence of NASH (% definite)	-	-	-	53	53	0
Fibrosis score (mean ± SD)	-	-	-	1.7 ± 1.4	1.7 ± 1.4	0 ± 0
Mallory bodies (% present)	-	-	-	30	30	0
Biopsy length (% < 10 mm)	-	-	-	15	14	1

Table C: Variables used to replicate Table 2: Characteristics of Adult Patients with NAFLD with Contemporaneous Biopsies and Clinical Factors by Presence of Definite NASH

Table Variable	**dataset.variable**
Proximity of Liver Biopsy to Enrollment	bg.bg320a, cr.cr108
Acanthosis nigricans Positive (%)	pe.pe218
Acanthosis nigricans severity score	pe.pe218
Age - yrs	rg.rg110
Albumin - g/dL	lr.lr333
Alkaline phosphatase - U/L	lr.lr330
Isolated Alkaline phosphatase	lr.lr330, lr.lr330a, lr.lr328a, lr.lr329a
AST - U/L	lr.lr328
AMA	ls.ls125
ANA	ls.ls123
ANA and ASMA	ls.ls123, ls.ls124
ASMA	ls.ls124
ALT - U/L	lr.lr329
Ballooning	cr.cr118a
Bilirubin, direct - mg/dL	lr.lr327
Bilirubin, total - mg/dL	lr.lr326
Body Mass Index - kg/m²	pe.pe208a, pe.pe208b, pe.pe209a, pe.pe209b, pe208c, pe209c
Total cholesterol - mg/dL	lr.lr338b
Presence of NASH	cr.cr123
Type 2 diabetes	bg.bg349b
Ferritin ng/mL	ls.ls109c
GGT - U/L	lr.lr331
Globulin - g/dL	lr.lr332, lr.lr333
Fasting serum glucose mg/dL	lr.lr339a
HbA1c - %	lr.lr324
HDL cholesterol - mg/dL	lr.lr338c
Hematocrit - %	lr.lr310
Hispanic ethnicity	rg.rg112
HOMA-IR - mg/dL/uU/mL/405	lr.lr339a, lr.lr339b
Hypertension	bg.bg349af
Fibrosis score	cr.cr121
Steatosis	cr.cr116a

Table Variable	dataset.variable
Lobular inflammation	cr.cr117a
NAFLD Activity Score	cr.cr116a, cr.cr117a, cr.cr118a
International normalized ratio	lr.lr335
Fasting serum insulin uU/mL	lr.lr339b
Portal inflammation	cr.cr117d
LDL cholesterol - mg/dL	lr.lr338b
Biopsy length	cr.cr115
Male gender	rg.rg111
Mallory bodies	cr.cr119
Metabolic syndrome	rg.rg111, pe.pe210a, pe.pe210b, pe.pe210c, lr.lr338a, lr.lr338c, pe.pe214a, pe.pe214b, lr.lr339a
Platelets - 1000/mm3	lr.lr312
AST/ALT ratio	lr.lr328, lr.lr329
Triglycerides - mg/dL	lr.lr338a
Waist circumference - cm	pe.pe210a, pe.pe210b, pe.pe210c
White blood cell - 1000/mm3	lr.lr311
White race	rg.rg114e
Waist-to-hip ratio	pe.pe210a, pe.pe210b, pe.pe210c, pe.pe211a, pe.pe211b, pe.pe211c

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

Characteristic	No (n=236) [Manuscript]	No (n=238) [DSIC]	No (n=2) [Diff]	Yes (n=252) [Manuscript]	Yes (n=240) [DSIC]	Yes (n=12) [Diff]
Demographics						
Male (%)	42	43	1	35	33	2
Age, years (median ± SD)	48	48.5	0.5	50	50	0
White (%)	82	82	0	78	78	0
Hispanic (%)	13	13	0	13	13	0
Clinical						
Hypertension (%)	42	44	2	54	57	3
Type 2 diabetes (%)	21	23	2	41	45	4
Metabolic syndrome (%)	57	54	3	72	71	1
Acanthosis nigricans						
Positive (%)	14	13	1	15	17	2
Severity score (mean ± SD)	0.29	0.29	0	0.37	0.4	0.03
Anthropometric (median ± SD)						
Body mass index (kg/m²)	34	34	0	35	34	1
Waist circumference (cm)	109	108	1	110	110	0
Waist-to-hip ratio	0.93	0.93	0	0.94	0.94	0
Hepatology panel (median ± SD)						
AST (U/L)	36	37	1	54	53	1
ALT (U/L)	54.5	55	0.5	68	66	2
AST/ALT	0.72	0.71	0.01	0.76	0.78	0.02

Characteristic	No (n=236) [Manuscript]	No (n=238) [DSIC]	No (n=2) [Diff]	Yes (n=252) [Manuscript]	Yes (n=240) [DSIC]	Yes (n=12) [Diff]
Alkaline phosphatase (U/L)	79	79	0	86	87	1
Isolated abnormal alkaline phosphatase (%)	6	6	0	2	2	0
GGT (U/L)	39	40	1	60	61	1
Globulin (g/dL)	2.9	2.9	0	3	3.1	0.1
Albumin (g/dL)	4.2	4.2	0	4.3	4.2	0.1
Bilirubin, total (mg/dL)	0.7	0.7	0	0.6	0.6	0
Bilirubin, direct (mg/dL)	0.1	0.1	0	0.1	0.1	0
International normalized ratio (mean ± SD)	1.01	1	0.01	1.04	1	0.04
Hematology and other laboratory studies (median ± SD)						
Hematocrit (%)	42.05	42	0.05	42.55	42.15	0.4
White blood cells (1000/mm³)	6.65	6.6	0.05	7	7	0
Platelet count (1000/mm³)	249.5	250.5	1	239	237	2
Total cholesterol (mg/dL)	193	191.5	1.5	197	196	1
HDL cholesterol (mg/dL)	43	43	0	39	39	0
LDL cholesterol (mg/dL)	121	119	2	118	119.5	1.5

Characteristic	No (n=236) [Manuscript]	No (n=238) [DSIC]	No (n=2) [Diff]	Yes (n=252) [Manuscript]	Yes (n=240) [DSIC]	Yes (n=12) [Diff]
Triglycerides (mg/dL)	138.5	138	0.5	173	170	3
HbA1c (%)	5.6	5.7	0.1	6	6.05	0.05
Fasting serum glucose (mg/dL)	93.5	93	0.5	103	102.5	0.5
Fasting serum insulin (µU/mL)	16.45	15.75	0.7	21.3	20.6	0.7
HOMA-IR (mg/dL µU/mL/405)	3.83	3.71	0.12	5.42	5.43	0.01
ANA (% positive)	28	26	2	23	23	0
ASMA (% positive)	16	17	1	11	13	2
ANA + ASMA (% both positive)	6	6	0	2	3	1
AMA (% positive)	1	1	0	2	1	1
Ferritin (ng/mL)	122	128.5	6.5	179	179	0
Histology						
Steatosis (% ≥ 34%)	45	42	3	66	65	1
Lobular inflammation (% ≥ grade 2)	26	25	1	58	58	0
Portal inflammation (% > mild)	13	16	3	27	30	3
Ballooning (% any)	22	21	1	100	100	0
NAFLD Activity Score (% ≥ 5)	14	13	1	70	73	3
Fibrosis score (mean ± SD)	0.88	0.94	0.6	2.08	2.16	0.08
Mallory bodies (% present)	2	2	0	50	52	2
Biopsy length (% < 10 mm)	21	19	2	9	9	0

Table E: Variables used to replicate Table 3: Characteristics of Adult Patients with NAFLD with Contemporaneous Biopsies and Clinical Factors by Fibrosis Stage

Table Variable	**dataset.variable**
Proximity of Liver Biopsy to Enrollment	bg.bg320a, cr.cr108
Acanthosis nigricans Positive (%)	pe.pe218
Acanthosis nigricans severity score	pe.pe218
Age - yrs	rg.rg110
Albumin - g/dL	lr.lr333
Alkaline phosphatase - U/L	lr.lr330
Isolated Alkaline phosphatase	lr.lr330, lr.lr330a, lr.lr328a, lr.lr329a
AST - U/L	lr.lr328
AMA	ls.ls125
ANA	ls.ls123
ANA and ASMA	ls.ls123, ls.ls124
ASMA	ls.ls124
ALT - U/L	lr.lr329
Ballooning	cr.cr118a
Bilirubin, direct - mg/dL	lr.lr327
Bilirubin, total - mg/dL	lr.lr326
Body Mass Index - kg/m²	pe.pe208a, pe.pe208b, pe.pe209a, pe.pe209b, pe208c, pe209c
Total cholesterol - mg/dL	lr.lr338b
Presence of NASH	cr.cr123
Type 2 diabetes	bg.bg349b
Ferritin ng/mL	ls.ls109c
GGT - U/L	lr.lr331
Globulin - g/dL	lr.lr332, lr.lr333
Fasting serum glucose mg/dL	lr.lr339a
HbA1c - %	lr.lr324
HDL cholesterol - mg/dL	lr.lr338c
Hematocrit - %	lr.lr310
Hispanic ethnicity	rg.rg112
HOMA-IR - mg/dL/uU/mL/405	lr.lr339a, lr.lr339b
Hypertension	bg.bg349af
Fibrosis score	cr.cr121
Steatosis	cr.cr116a

Table Variable	dataset.variable
Lobular inflammation	cr.cr117a
NAFLD Activity Score	cr.cr116a, cr.cr117a, cr.cr118a
International normalized ratio	lr.lr335
Fasting serum insulin uU/mL	lr.lr339b
Portal inflammation	cr.cr117d
LDL cholesterol - mg/dL	lr.lr338b
Biopsy length	cr.cr115
Male gender	rg.rg111
Mallory bodies	cr.cr119
Metabolic syndrome	rg.rg111, pe.pe210a, pe.pe210b, pe.pe210c, lr.lr338a, lr.lr338c, pe.pe214a, pe.pe214b, lr.lr339a
Platelets - 1000/mm3	lr.lr312
AST/ALT ratio	lr.lr328, lr.lr329
Triglycerides - mg/dL	lr.lr338a
Waist circumference - cm	pe.pe210a, pe.pe210b, pe.pe210c
White blood cell - 1000/mm3	lr.lr311
White race	rg.rg114e
Waist-to-hip ratio	pe.pe210a, pe.pe210b, pe.pe210c, pe.pe211a, pe.pe211b, pe.pe211c

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

Characteristic	None (n=152) [Manuscript]	None (n=144) [DSIC]	None (n=8) [Diff]	Mild/Moderate (n=206) [Manuscript]	Mild/ Moderate (n=197) [DSIC]	Mild/ Moderate (n=9) [Diff]
Demographics						
Male (%)	39	41	2	41	38	3
Age, years (median ± SD)	45	46	1	49	50	1
White (%)	82	82	0	79	79	0
Hispanic (%)	18	18	0	13	13	0
Clinical						
Hypertension (%)	36	37	1	51	55	4
Type 2 diabetes (%)	14	14	0	34	39	5
Metabolic syndrome (%)	57	54	3	71	69	2
Acanthosis nigricans						
Positive (%)	14	13	1	17	17	0
Severity score (mean ± SD)	0.3	0.3	0	0.4	0.5	0.1
Anthropometric (median ± SD)						
Body mass index (kg/m²)	34	33	1	34	34	0
Waist circumference (cm)	107.2	107.3	0.1	109.65	109.24	0.41
Waist-to-hip ratio	0.93	0.93	0	0.95	0.94	0.01
Hepatology panel (median ± SD)						
AST (U/L)	34.5	34	0.5	44	45	1
ALT (U/L)	55.5	55	0.5	67	67	0

Characteristic	None (n=152) [Manuscript]	None (n=144) [DSIC]	None (n=8) [Diff]	Mild/Moderate (n=206) [Manuscript]	Mild/ Moderate (n=197) [DSIC]	Mild/ Moderate (n=9) [Diff]
AST/ALT	0.68	0.67	0.01	0.7	0.7	0
Alkaline phosphatase (U/L)	79	79	0	82	82	0
Isolated abnormal alkaline phosphatase (%)	4	4	0	5	3	2
GGT (U/L)	36	36	0	49.5	49	0.5
Globulin (g/dL)	2.9	2.9	0	3	3	0
Albumin (g/dL)	4.2	4.2	0	4.3	4.3	0
Bilirubin, total (mg/dL)	0.7	0.7	0	0.7	0.7	0
Bilirubin, direct (mg/dL)	0.1	0.1	0	0.1	0.1	0
International normalized ratio (mean ± SD)	0.98	1	0.02	1.02	1	0.02
Hematology and other laboratory studies (median ± SD)						
Hematocrit (%)	42	42	0	43	43	0
White blood cells (1000/mm³)	6.7	6.7	0	7.2	7.1	0.1
Platelet count (1000/mm³)	257.5	258.5	1	257.5	257	0.5
Total cholesterol (mg/dL)	199.5	197.5	2	196	193.5	2.5
HDL cholesterol (mg/dL)	43	42.5	0.5	40	40	0

Characteristic	None (n=152) [Manuscript]	None (n=144) [DSIC]	None (n=8) [Diff]	Mild/Moderate (n=206) [Manuscript]	Mild/ Moderate (n=197) [DSIC]	Mild/ Moderate (n=9) [Diff]
LDL cholesterol (mg/dL)	123	122	1	121	120.5	0.5
Triglycerides (mg/dL)	151	150	1	165	164.5	0.5
HbA1c (%)	5.6	5.6	0	5.8	5.9	0.1
Fasting serum glucose (mg/dL)	92.5	92	0.5	101	101	0
Fasting serum insulin (µU/mL)	15.25	14.2	1.05	19.5	18.5	1
HOMA-IR (mg/dL µU/mL/405)	3.5	3.4	0.1	4.9	4.7	0.2
ANA (% positive)	23	22	1	25	25	0
ASMA (% positive)	14	15	1	11	12	1
ANA + ASMA (% both positive)	6	6	0	1	2	1
AMA (% positive)	2	1	1	1	1	1
Ferritin (ng/mL)	118	118.5	0.5	167	161	6
Histology						
Steatosis (% ≥ 34%)	44	42	2	69	68	1
Lobular inflammation (% ≥ grade 2)	22	22	0	59	57	2
Portal inflammation (% > mild)	7	7	0	15	16	1
Ballooning (% any)	30	28	2	71	70	1
NAFLD Activity Score (% ≥ 5)	14	14	0	58	59	1
Presence of NASH (% definite)	12	10	2	65	63	2
Mallory bodies (% present)	1	1	0	26	26	0

Characteristic	None (n=152) [Manuscript]	None (n=144) [DSIC]	None (n=8) [Diff]	Mild/Moderate (n=206) [Manuscript]	Mild/ Moderate (n=197) [DSIC]	Mild/ Moderate (n=9) [Diff]
Biopsy length (% < 10 mm)	21	20	1	15	14	1

Characteristic	Bridging (n=80) [Manuscript]	Bridging (n=81) [DSIC]	Bridging (n=1) [Diff]	Cirrhotic (n=49) [Manuscript]	Cirrhotic (n=54) [DSIC]	Cirrhotic (n=6) [Diff]
Demographics						
Male (%)	28	27	1	43	41	2
Age, years (median ± SD)	54	54	0	57	57	0
White (%)	75	75	0	88	91	3
Hispanic (%)	9	9	0	4	6	2
Clinical						
Hypertension (%)	58	56	2	61	63	2
Type 2 diabetes (%)	43	47	4	55	54	1
Metabolic syndrome (%)	69	69	0	55	54	1
Acanthosis nigricans						
Positive (%)	11	12	1	12	15	3
Severity score (mean ± SD)	0.3	0.2	0.1	0.2	0.3	0.1
Anthropometric (median ± SD)						
Body mass index (kg/m²)	34	34	0	36	36	0
Waist circumference (cm)	110.82	108.65	2.17	119.55	119.55	0
Waist-to-hip ratio	0.93	0.94	0.01	0.96	0.94	0.02

Characteristic	Bridging (n=80) [Manuscript]	Bridging (n=81) [DSIC]	Bridging (n=1) [Diff]	Cirrhotic (n=49) [Manuscript]	Cirrhotic (n=54) [DSIC]	Cirrhotic (n=6) [Diff]
Hepatology panel (median ± SD)						
AST (U/L)	55.5	51	4.5	50	50	0
ALT (U/L)	72.5	66	6.5	46	46.5	0.5
AST/ALT	0.81	0.84	0.03	1.1	1.1	0
Alkaline phosphatase (U/L)	89	87	2	101	99.5	1.5
Isolated abnormal alkaline phosphatase (%)	3	4	1	6	7	1
GGT (U/L)	67	64	3	79	74.5	4.5
Globulin (g/dL)	3.1	3.1	0	3.4	3.4	0
Albumin (g/dL)	4.2	4.3	0.1	4.1	4.1	0
Bilirubin, total (mg/dL)	0.6	0.6	0	0.8	0.9	0.1
Bilirubin, direct (mg/dL)	0.1	0.1	0	0.2	0.2	0
International normalized ratio (mean ± SD)	1.04	1	0.04	1.17	1.1	0.07
Hematology and other laboratory studies (median ± SD)						
Hematocrit (%)	41.75	41	0.75	41.8	41	0.8
White blood cells (1000/mm^3)	6.4	6.5	0.1	6	6	0

Characteristic	Bridging (n=80) [Manuscript]	Bridging (n=81) [DSIC]	Bridging (n=1) [Diff]	Cirrhotic (n=49) [Manuscript]	Cirrhotic (n=54) [DSIC]	Cirrhotic (n=6) [Diff]
Platelet count (1000/mm³)	229.5	237	7.5	146	147.5	1.5
Total cholesterol (mg/dL)	197.5	199	1.5	170	172	2
HDL cholesterol (mg/dL)	40	40	0	42	40	2
LDL cholesterol (mg/dL)	120	122.5	2.5	97	102.5	5.5
Triglycerides (mg/dL)	153.5	151	2.5	138	122.5	15.5
HbA1c (%)	6	6	0	5.9	5.8	0.1
Fasting serum glucose (mg/dL)	103	103	0	96	96	0
Fasting serum insulin (µU/mL)	21.8	20.3	1.5	25	25	0
HOMA-IR (mg/dL µU/mL/405)	5.7	5.4	0.3	6.1	6.6	0.5
ANA (% positive)	24	23	2	35	30	5
ASMA (% positive)	18	20	2	17	20	3
ANA + ASMA (% both positive)	6	6	0	4	6	2
AMA (% positive)	4	2	2	0	0	0
Ferritin (ng/mL)	211.5	216	4.5	163	164	1
Histology						
Steatosis (% ≥ 34%)	59	58	1	29	26	3
Lobular inflammation (% ≥ grade 2)	55	53	2	18	22	4
Portal inflammation (% > mild)	39	41	2	55	63	8

Characteristic	Bridging (n=80) [Manuscript]	Bridging (n=81) [DSIC]	Bridging (n=1) [Diff]	Cirrhotic (n=49) [Manuscript]	Cirrhotic (n=54) [DSIC]	Cirrhotic (n=6) [Diff]
Ballooning (% any)	88	85	3	80	80	0
NAFLD Activity Score (% ≥ 5)	66	64	2	29	33	4
Presence of NASH (% definite)	88	84	4	59	59	0
Mallory bodies (% present)	61	58	3	53	54	1
Biopsy length (% < 10 mm)	3	1	2	15	15	0

Appendix A: SAS Code

```
**** NASH NAFLD Adult DSIC;
**** Programmer: Allyson Mateja;
**** Date: March 17, 2016;

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

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

libname inlib1 xport '/prj/niddk/ims_analysis/NASH/private_orig_data/NASHCRN_Data_Sharing_AdultNAFLDDatabase_Hepatology_2010/Datasets/link_id.xpt';
proc copy in=inlib1 out = work;
libname inlib2 xport '/prj/niddk/ims_analysis/NASH/private_orig_data/NASHCRN_Data_Sharing_AdultNAFLDDatabase_Hepatology_2010/Datasets/table1.xpt';
proc copy in=inlib2 out = work;

data adult_ids;
       set ids;

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_adult;
       set table1;

proc sort data = table1_adult;
       by id;

proc sort data = adult_ids;
       by id;

data table1_adult;
       merge table1_adult (in=val1)
             adult_ids    (in=val2);
       by id;
       if val1 and val2 then output table1_adult;

options nofmterr;

data ad;
       set nafldata.ad;

data cr;
       set nafldata.cr;

data dr;
       set nafldata.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 nafldata.qf;

data rg;
       set nafldata.rg;

data ie;
       set nafldata.ie;

data bg;
       set nafldata.bg;

data nash_id;
       set nafldata.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_adult;
proc contents data = peds;

proc freq data = rg;
       tables rg110 rg123;

proc sort data = table1_adult;
       by nash;

proc print data = table1_adult (obs=20);
       var nash;

proc sort data = rg;
```

```
       by nash;

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

data in_rg_and_table1_adult in_rg_only in_table1_adult_only;
       merge rg           (in=val1 keep=nash rg110 rg123 rg130a)
             table1_adult (in=val2);
       by nash;
       if val1 and val2 then output in_rg_and_table1_adult;
       else if val1 and not val2 then output in_rg_only;
       else if val2 and not val1 then output in_table1_adult_only;

proc freq data = in_rg_and_table1_adult;
       tables rg110 rg123 proximit proximit*rg130a /list missing;

proc sort data = in_rg_and_table1_adult;
       by proximit;

proc freq data = in_rg_and_table1_adult;
       tables male;
       by proximit;
       title3 'Manuscript Male (%)';

proc freq data = in_rg_and_table1_adult;
       tables male;

proc means data = in_rg_and_table1_adult n median std;
       var age;
       class proximit;
       types () proximit;
       title3 'Manuscript Age, years';

proc freq data = in_rg_and_table1_adult;
       tables white;
       by proximit;
       title3 'Manuscript White (%)';

proc freq data = in_rg_and_table1_adult;
       tables white;

proc freq data = in_rg_and_table1_adult;
       tables hispanic;
       by proximit;
       title3 'Manuscript Hispanic (%)';

proc freq data = in_rg_and_table1_adult;
       tables hispanic;

proc freq data = in_rg_and_table1_adult;
       tables htn;
       by proximit;
       title3 'Manuscript Hypertension (%)';

proc freq data = in_rg_and_table1_adult;
       tables htn;
```

```
proc freq data = in_rg_and_table1_adult;
        tables diab2;
        by proximit;
        title3 'Manuscript Type 2 diabetes (%)';

proc freq data = in_rg_and_table1_adult;
        tables diab2;

proc freq data = in_rg_and_table1_adult;
        tables meta;
        by proximit;
        title3 'Manuscript Metabolic syndrome (%)';

proc freq data = in_rg_and_table1_adult;
        tables meta;

proc freq data = in_rg_and_table1_adult;
        tables acanth;
        by proximit;
        title3 'Manuscript Acanthosis nigricans positive (%)';

proc freq data = in_rg_and_table1_adult;
        tables acanth;

proc means data = in_rg_and_table1_adult n mean std;
        var acanthn;
        class proximit;
        types () proximit;
        title3 'Manuscript Acanthosis nigricans Severity score';

proc means data = in_rg_and_table1_adult n median std;
        var bmi;
        class proximit;
        types () proximit;
        title3 'Manuscript Body mass index';

proc means data = in_rg_and_table1_adult n median std;
        var waist;
        class proximit;
        types () proximit;
        title3 'Manuscript Waist circumference';

proc means data = in_rg_and_table1_adult n median std;
        var wthip;
        class proximit;
        types () proximit;
        title3 'Manuscript Waist-to-hip ratio';

proc means data = in_rg_and_table1_adult n median std;
        var ast;
        class proximit;
        types () proximit;
        title3 'Manuscript AST';

proc freq data = in_rg_and_table1_adult;
```

```
        tables astuln1;
        by proximit;
        title3 'Manuscript Abnormal AST (%)';

proc freq data = in_rg_and_table1_adult;
        tables astuln1;

proc means data = in_rg_and_table1_adult n median std;
        var alt;
        class proximit;
        types () proximit;
        title3 'Manuscript ALT';

proc freq data = in_rg_and_table1_adult;
        tables altuln1;
        by proximit;
        title3 'Manuscript Abnormal ALT (%)';

proc freq data = in_rg_and_table1_adult;
        tables altuln1;

proc means data = in_rg_and_table1_adult n median std;
        var ratio;
        class proximit;
        types () proximit;
        title3 'Manuscript AST/ALT';

proc means data = in_rg_and_table1_adult n median std;
        var alka;
        class proximit;
        types () proximit;
        title3 'Manuscript Alkaline phosphatase';

proc freq data = in_rg_and_table1_adult;
        tables alkaiso;
        by proximit;
        title3 'Manuscript Isolated abnormal alkaline phosphatase (%)';

proc freq data = in_rg_and_table1_adult;
        tables alkaiso;

proc means data = in_rg_and_table1_adult n median std;
        var ggt;
        class proximit;
        types () proximit;
        title3 'Manuscript GGT';

proc means data = in_rg_and_table1_adult n median std;
        var glob;
        class proximit;
        types () proximit;
        title3 'Manuscript Globulin';

proc means data = in_rg_and_table1_adult n median std;
        var alb;
        class proximit;
```

```
        types () proximit;
        title3 'Manuscript Albumin';

proc means data = in_rg_and_table1_adult n median std;
        var bilit;
        class proximit;
        types () proximit;
        title3 'Manuscript Bilirubin, total';

proc means data = in_rg_and_table1_adult n median std;
        var bilid;
        class proximit;
        types () proximit;
        title3 'Manuscript Bilirubin, direct';

proc means data = in_rg_and_table1_adult n mean std;
        var inr;
        class proximit;
        types () proximit;
        title3 'Manuscript Inernational normalized ratio';

proc means data = in_rg_and_table1_adult n median std;
        var hema;
        class proximit;
        types () proximit;
        title3 'Manuscript Hematocrit';

proc means data = in_rg_and_table1_adult n median std;
        var wbc;
        class proximit;
        types () proximit;
        title3 'Manuscript White blood cells';

proc means data = in_rg_and_table1_adult n median std;
        var plat;
        class proximit;
        types () proximit;
        title3 'Manuscript Platelet count';

proc means data = in_rg_and_table1_adult n median std;
        var chol;
        class proximit;
        types () proximit;
        title3 'Manuscript Total cholesterol';

proc means data = in_rg_and_table1_adult n median std;
        var hdl;
        class proximit;
        types () proximit;
        title3 'Manuscript HDL cholesterol';

proc means data = in_rg_and_table1_adult n median std;
        var ldl;
        class proximit;
        types () proximit;
        title3 'Manuscript LDL Cholesterol';
```

```
proc means data = in_rg_and_table1_adult n median std;
        var tri;
        class proximit;
        types () proximit;
        title3 'Manuscript Triglycerides';

proc means data = in_rg_and_table1_adult n median std;
        var hba1c;
        class proximit;
        types () proximit;
        title3 'Manuscript HbA1c';

proc means data = in_rg_and_table1_adult n median std;
        var gluc;
        class proximit;
        types () proximit;
        title3 'Manuscript Fasting serum glucose';

proc means data = in_rg_and_table1_adult n median std;
        var insu;
        class proximit;
        types () proximit;
        title3 'Manuscript Fasting serum insulin';

proc means data = in_rg_and_table1_adult n median std;
        var homa;
        class proximit;
        types () proximit;
        title3 'Manuscript HOMA-IR';

proc freq data = in_rg_and_table1_adult;
        tables ana;
        by proximit;
        title3 'Manuscript ANA (%)';

proc freq data = in_rg_and_table1_adult;
        tables ana;

proc freq data = in_rg_and_table1_adult;
        tables asma;
        by proximit;
        title3 'Manuscript ASMA (%)';

proc freq data = in_rg_and_table1_adult;
        tables asma;

proc freq data = in_rg_and_table1_adult;
        tables anaasma;
        by proximit;
        title3 'Manuscript ANA and ASMA (%)';

proc freq data = in_rg_and_table1_adult;
        tables anaasma;

proc freq data = in_rg_and_table1_adult;
```

```
        tables ama;
        by proximit;
        title3 'Manuscript AMA (%)';

proc freq data = in_rg_and_table1_adult;
        tables ama;

proc means data = in_rg_and_table1_adult n median std;
        var ferr;
        class proximit;
        types () proximit;
        title3 'Manuscript Ferritin';


proc freq data = in_rg_and_table1_adult;
        tables igrade;
        by proximit;
        title3 'Manuscript Steatosis (%)';

proc freq data = in_rg_and_table1_adult;
        tables igrade;

proc freq data = in_rg_and_table1_adult;
        tables iinflam;
        by proximit;
        title3 'Manuscript Lobular inflammation (%)';

proc freq data = in_rg_and_table1_adult;
        tables iinflam;

proc freq data = in_rg_and_table1_adult;
        tables iportal;
        by proximit;
        title3 'Manuscript Portal inflammation (%)';

proc freq data = in_rg_and_table1_adult;
        tables iportal;

proc freq data = in_rg_and_table1_adult;
        tables iball;
        by proximit;
        title3 'Manuscript Ballooning (%)';

proc freq data = in_rg_and_table1_adult;
        tables iball;

proc freq data = in_rg_and_table1_adult;
        tables inas;
        by proximit;
        title3 'Manuscript NAFLD Activity Score (%)';

proc freq data = in_rg_and_table1_adult;
        tables inas;

proc freq data = in_rg_and_table1_adult;
        tables defnash;
```

```
        by proximit;
        title3 'Manuscript Presence of NASH (%)';

proc freq data = in_rg_and_table1_adult;
        tables defnash;

proc means data = in_rg_and_table1_adult n mean std;
        var nfibro;
        class proximit;
        types () proximit;
        title3 'Manuscript Fibrosis score (%)';

proc freq data = in_rg_and_table1_adult;
        tables mallory;
        by proximit;
        title3 'Manuscript Mallory bodies (%)';

proc freq data = in_rg_and_table1_adult;
        tables mallory;

proc freq data = in_rg_and_table1_adult;
        tables leng10;
        by proximit;
        title3 'Manuscript Biopsy length (%)';

proc freq data = in_rg_and_table1_adult;
        tables leng10;

proc sort data = in_rg_and_table1_adult;
        by nash;

proc freq data = pe;
        tables visit;
        title3 'pe';

proc freq data = lr;
        tables visit;
        title3 'lr';

proc freq data = ls;
        tables visit;
        title3 'ls';

proc freq data = pf;
        tables visit;
        title3 'pf';

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

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

data cr;
```

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

proc sort data = cr;
        by nash visit;

data cr;
        set cr;
        by nash;
        if last.nash then output;

data dsic_values;
        merge in_rg_and_table1_adult (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.);
        ratio = lr328/lr329;
        if rg114e = '1' and rg114a = '' and rg114b = '' and rg114c = '' and rg114d = '' and rg114f = '' then white = 1;
        else white = 0;
        if pe218 in ('', ' ') then pe218 = 0;
        if pe218 >= 1 then acanth = 1;
        else acanth = 0;
        acanthn = input(pe218, 8.);
        globulin = lr332-lr333;
        homa_ir = (lr339a*lr339b)/405;
        platelet = lr312/1000;
        if ls123 = '1' then ana = 1;
        else ana = 0;
        if ls124 = '1' then asma = 1;
        else asma = 0;
        if ls125 = '1' then ama = 1;
        else ama = 0;
        if ls123 = 1 and ls124 = 1 then anaasma = 1;
        else anaasma = 0;
        if pe210c = 1 then do;
                pe210a = 2.54*pe210a;
                pe210b = 2.54*pe210b;
        end;
        if pe210a ne . and pe210b ne . then waist_circum = (pe210a+pe210b)/2;
        else if pe210a ne . and pe210b = . then waist_circum = pe210a;
        else if pe210a = . and pe210b ne . then waist_circum = pe210b;
        if pe211c = 1 then do;
                pe211a = 2.54*pe211a;
                pe211b = 2.54*pe211b;
        end;
        if pe211a ne . and pe211b ne . then hip_circum = (pe211a+pe211b)/2;
        else if pe211a ne . and pe211b = . then hip_circum = pe211a;
        else if pe211a = . and pe211b ne . then hip_circum = pe211b;
        wthip = waist_circum/hip_circum;
        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 = 1 and waist_circum > 102 then rf1 = 1;
else if rg111 = 2 and waist_circum > 88 then rf1 = 1;
else rf1 = 0;
if lr338a >= 150 then rf2 = 1;
else rf2 = 0;
if rg111=1 and lr338c < 40 then rf3 = 1;
else if rg111 = 2 and lr338c < 50 then rf3 = 1;
else rf3 = 0;
if (pe214a >=130 or pe214b >= 85) then rf4 = 1;
else rf4 = 0;
if lr339a >= 110 then rf5 = 1;
else rf5 = 0;
total_rf = sum(rf1, rf2, rf3, rf4, rf5);
if total_rf >= 3 then meta = 1;
else meta = 0;
if lr328 > lr328a then abnormal_ast = 1;
else abnormal_ast = 0;
if lr329 > lr329a then abnormal_alt = 1;
else abnormal_alt = 0;
if (lr330 >= lr330a) and (lr328 < lr328a) and (lr329 < lr329a) then isolated_ab_alk_phos = 1;
else isolated_ab_alk_phos = 0;
bg_biopsy_months = ceil((bg320a-registdt)/30.416666667);
cr_biopsy_months = ceil((cr108-registdt)/30.416666667);
if cr_biopsy_months = . and bg_biopsy_months ne . then biopsy_months = bg_biopsy_months;
else if cr_biopsy_months ne . then do;
        if cr_biopsy_months >= bg_biopsy_months then biopsy_months = cr_biopsy_months;
        else if cr_biopsy_months < bg_biopsy_months then biopsy_months = bg_biopsy_months;
end;
else biopsy_months = .;
if biopsy_months = . then biopsy_group = 0;
else if biopsy_months < -6 then biopsy_group = 2;
else if biopsy_months >= -6 then biopsy_group = 1;
if biopsy_group in (1,2) then do;
        if cr116a >= 2 then igrade = 1;
        else igrade = 0;
        if cr117a >= 2 then iinflam = 1;
        else iinflam = 0;
        if cr117d = 2 then iportal = 1;
        else iportal = 0;
        if cr118a >= 1 then iball=1;
        else iball=0;
        nas = cr116a+cr117a+cr118a;
        if nas >= 5 then inas = 1;
```

```
                else inas = 0;
                if cr123 = 2 then defnash = 1;
                else if cr123 ne '' then defnash = 0;
                nfibro = input(substr(cr121, 1, 1), 8.);
                if 0 <= cr115 < 10 then leng10 = 1;
                else leng10 = 0;
        end;
        if val1 then output dsic_values;

proc freq data = dsic_values;
        tables bg_biopsy_months*cr_biopsy_months /list missing;

data dsic_values;
        length proximity $12.;
        set dsic_values;
        if biopsy_group = 0 then proximity = "no biopsy";
        if biopsy_group = 1 then proximity = "<= 6 months";
        if biopsy_group = 2 then proximity = "> 6 months";

proc freq data = dsic_values;
        tables proximity*cr123 /list missing;
        title3 'DSIC proximity';

proc sort data = dsic_values;
        by proximity;

proc freq data = dsic_values;
        tables rg111 /list missing;
        by proximity;
        title3 'DSIC Male (%)';

proc freq data = dsic_values;
        tables rg111 /list missing;

proc means data = dsic_values n median std;
        var age;
        class proximity;
        types () proximity;
        title3 'DSIC Age';

proc freq data = dsic_values;
        tables white;
        by proximity;
        title3 'DSIC White (%)';

proc freq data = dsic_values;
        tables white;

proc freq data = dsic_values;
        tables rg112;
        by proximity;
        title3 'DSIC Hispanic (%)';

proc freq data = dsic_values;
        tables rg112;
```

```
proc freq data = dsic_values;
       tables bg349af /missing;
       by proximity;
       title3 'DSIC Hypertension (%)';

proc freq data = dsic_values;
       tables bg349af /missing;

proc freq data = dsic_values;
       tables bg349b /missing;
       by proximity;
       title3 'DSIC Type 2 Diabetes (%)';

proc freq data = dsic_values;
       tables bg349b /missing;

proc freq data = dsic_values;
       tables meta /missing;
       by proximity;
       title3 'DSIC Metabolic syndrome';

proc freq data = dsic_values;
       tables meta /missing;

proc freq data = dsic_values;
       tables acanth;
       by proximity;
       title3 'DSIC Acanthosis nigricans, Positive (%)';

proc freq data = dsic_values;
       tables acanth;

proc means data = dsic_values n mean std;
       var acanthn;
       class proximity;
       types () proximity;
       title3 'DSIC Acanthosis nigricans, Severity score';

proc means data = dsic_values n median std;
       var bmi;
       class proximity;
       types () proximity;
       title3 'DSIC BMI';

proc means data = dsic_values n median std;
       var waist_circum;
       class proximity;
       types () proximity;
       title3 'DSIC Waist circumference';

proc means data = dsic_values n median std;
       var wthip;
       class proximity;
       types () proximity;
       title3 'DSIC Waist to hip ratio';
```

```
proc means data = dsic_values n median std;
       var lr328;
       class proximity;
       types () proximity;
       title3 'DSIC AST';

proc freq data = dsic_values;
       tables abnormal_ast /missing;
       by proximity;
       title3 'DSIC Abnormal AST';

proc freq data = dsic_values;
       tables abnormal_ast /missing;

proc means data = dsic_values n median std;
       var lr329;
       class proximity;
       types () proximity;
       title3 'DSIC ALT';

proc freq data = dsic_values;
       tables abnormal_alt /missing;
       by proximity;
       title3 'DSIC Abnormal ALT';

proc freq data = dsic_values;
       tables abnormal_alt /missing;

proc means data = dsic_values n median std;
       var ratio;
       class proximity;
       types () proximity;
       title3 'DSIC AST/ALT';

proc means data = dsic_values n median std;
       var lr330;
       class proximity;
       types () proximity;
       title3 'DSIC Alkaline phosphatase';

proc freq data = dsic_values;
       tables isolated_ab_alk_phos /missing;
       by proximity;
       title3 'DSIC Isolated abnormal alkaline phosphatase';

proc freq data = dsic_values;
       tables isolated_ab_alk_phos;

proc means data = dsic_values n median std;
       var lr331;
       class proximity;
       types () proximity;
       title3 'DSIC GGT';

proc means data = dsic_values n median std;
       var globulin;
```

```
        class proximity;
        types () proximity;
        title3 'DSIC Globulin';

proc means data = dsic_values n median std;
        var lr333;
        class proximity;
        types () proximity;
        title3 'DSIC Albumin';

proc means data = dsic_values n median std;
        var lr326;
        class proximity;
        types () proximity;
        title3 'DSIC Bilirubin, total';

proc means data = dsic_values n median std;
        var lr327;
        class proximity;
        types () proximity;
        title3 'DSIC Bilirubin, direct';

proc means data = dsic_values n median std;
        var lr335;
        class proximity;
        types () proximity;
        title3 'DSIC INR';

proc means data = dsic_values n median std;
        var lr310;
        class proximity;
        types () proximity;
        title3 'DSIC Hematocrit';

proc means data = dsic_values n median std;
        var lr311;
        class proximity;
        types () proximity;
        title3 'DSIC White blood cells';

proc means data = dsic_values n median std;
        var platelet;
        class proximity;
        types () proximity;
        title3 'DSIC Platelet count';

proc means data = dsic_values n median std;
        var lr338b;
        class proximity;
        types () proximity;
        title3 'DSIC Total cholesterol';

proc means data = dsic_values n median std;
        var lr338c;
        class proximity;
        types () proximity;
```

```
        title3 'DSIC HDL cholesterol';

proc means data = dsic_values n median std;
        var lr338d;
        class proximity;
        types () proximity;
        title3 'DSIC LDL Cholesterol';

proc means data = dsic_values n median std;
        var lr338a;
        class proximity;
        types () proximity;
        title3 'DSIC Triglycerides';

proc means data = dsic_values n median std;
        var lr324;
        class proximity;
        types () proximity;
        title3 'DSIC HbA1c';

proc means data = dsic_values n median std;
        var lr339a;
        class proximity;
        types () proximity;
        title3 'DSIC Fasting serum glucose';

proc means data = dsic_values n median std;
        var lr339b;
        class proximity;
        types () proximity;
        title3 'DSIC Fasting serum insulin';



proc means data = dsic_values n median std;
        var homa_ir;
        class proximity;
        types () proximity;
        title3 'DSIC HOMA-IR';

proc freq data = dsic_values;
        tables ana;
        by proximity;
        title3 'DSIC ANA (% positive)';

proc freq data = dsic_values;
        tables ana;

proc freq data = dsic_values;
        tables asma;
        by proximity;
        title3 'DSIC ASMA (% positive)';

proc freq data = dsic_values;
        tables asma;
```

```
proc freq data = dsic_values;
       tables anaasma;
       by proximity;
       title3 'DSIC ANA+ASMA (% both positive)';

proc freq data = dsic_values;
       tables anaasma;

proc freq data = dsic_values;
       tables ama;
       by proximity;
       title3 'DSIC AMA (% positive)';

proc freq data = dsic_values;
       tables ama;

proc means data = dsic_values n median std;
       var ls109c;
       class proximity;
       types () proximity;
       title3 'DSIC Ferritin';

proc freq data = dsic_values;
       tables igrade;
       by proximity;
       title3 'DSIC Steatosis';

proc freq data = dsic_values;
       tables igrade;

proc freq data = dsic_values;
       tables iinflam;
       by proximity;
       title3 'DSIC Lobular inflammation';

proc freq data = dsic_values;
       tables iinflam;

proc freq data = dsic_values;
       tables iportal;
       by proximity;
       title3 'DSIC Portal inflammation';

proc freq data = dsic_values;
       tables iportal;

proc freq data = dsic_values;
       tables iball;
       by proximity;
       title3 'DSIC Ballooning';

proc freq data = dsic_values;
       tables iball;

proc freq data = dsic_values;
       tables inas;
```

```
        by proximity;
        title3 'DSIC NAFLD Activity Score';

proc freq data = dsic_values;
        tables inas;

proc freq data = dsic_values;
        tables defnash ;
        by proximity;
        title3 'DSIC Definite NASH';

proc freq data = dsic_values;
        tables defnash ;

proc means data = dsic_values;
        var nfibro;
        class proximity;
        types () proximity;
        title3 'DSIC Fibrosis score';

proc freq data = dsic_values;
        tables cr119;
        by proximity;
        title3 'DSIC Mallory bodies';

proc freq data = dsic_values;
        tables cr119;

proc freq data = dsic_values;
        tables leng10;
        by proximity;
        title3 'DSIC Biopsy length';

proc freq data = dsic_values;
        tables leng10;

data table2_manuscript;
        set in_rg_and_table1_adult;
        if proximit = '<= 6 mos';

proc freq data = table2_manuscript;
        tables defnash;
        title3 'Table 2';

data table2_manuscript;
        set table2_manuscript;
        if defnash ne .;

proc sort data = table2_manuscript;
        by defnash;

proc freq data = table2_manuscript;
        tables male;
        by defnash;
        title3 'Manuscript Male (%)';
```

```
proc means data = table2_manuscript n median;
       var age;
       class defnash;
       title3 'Manuscript Age, years';

proc freq data = table2_manuscript;
       tables white;
       by defnash;
       title3 'Manuscript White (%)';

proc freq data = table2_manuscript;
       tables hispanic;
       by defnash;
       title3 'Manuscript Hispanic (%)';

proc freq data = table2_manuscript;
       tables htn;
       by defnash;
       title3 'Manuscript Hypertension (%)';

proc freq data = table2_manuscript;
       tables diab2;
       by defnash;
       title3 'Manuscript Type 2 diabetes (%)';

proc freq data = table2_manuscript;
       tables meta;
       by defnash;
       title3 'Manuscript Metabolic syndrome (%)';

proc freq data = table2_manuscript;
       tables acanth;
       by defnash;
       title3 'Manuscript Acanthosis nigricans positive (%)';

proc means data = table2_manuscript n mean;
       var acanthn;
       class defnash;
       title3 'Manuscript Acanthosis nigricans Severity score';

proc means data = table2_manuscript n median;
       var bmi;
       class defnash;
       title3 'Manuscript Body mass index';

proc means data = table2_manuscript n median;
       var waist;
       class defnash;
       title3 'Manuscript Waist circumference';

proc means data = table2_manuscript n median;
       var wthip;
       class defnash;
       title3 'Manuscript Waist-to-hip ratio';

proc means data = table2_manuscript n median;
```

```
        var ast;
        class defnash;
        title3 'Manuscript AST';

proc means data = table2_manuscript n median;
        var alt;
        class defnash;
        title3 'Manuscript ALT';

proc means data = table2_manuscript n median;
        var ratio;
        class defnash;
        title3 'Manuscript AST/ALT';

proc means data = table2_manuscript n median;
        var alka;
        class defnash;
        title3 'Manuscript Alkaline phosphatase';

proc freq data = table2_manuscript;
        tables alkaiso;
        by defnash;
        title3 'Manuscript Isolated abnormal alkaline phosphatase (%)';

proc means data = table2_manuscript n median;
        var ggt;
        class defnash;
        title3 'Manuscript GGT';

proc means data = table2_manuscript n median;
        var glob;
        class defnash;
        title3 'Manuscript Globulin';

proc means data = table2_manuscript n median;
        var alb;
        class defnash;
        title3 'Manuscript Albumin';

proc means data = table2_manuscript n median;
        var bilit;
        class defnash;
        title3 'Manuscript Bilirubin, total';

proc means data = table2_manuscript n median;
        var bilid;
        class defnash;
        title3 'Manuscript Bilirubin, direct';

proc means data = table2_manuscript n mean;
        var inr;
        class defnash;
        title3 'Manuscript Inernational normalized ratio';

proc means data = table2_manuscript n median;
        var hema;
```

```
        class defnash;
        title3 'Manuscript Hematocrit';

proc means data = table2_manuscript n median;
        var wbc;
        class defnash;
        title3 'Manuscript White blood cells';

proc means data = table2_manuscript n median;
        var plat;
        class defnash;
        title3 'Manuscript Platelet count';

proc means data = table2_manuscript n median;
        var chol;
        class defnash;
        title3 'Manuscript Total cholesterol';

proc means data = table2_manuscript n median;
        var hdl;
        class defnash;
        title3 'Manuscript HDL cholesterol';

proc means data = table2_manuscript n median;
        var ldl;
        class defnash;
        title3 'Manuscript LDL Cholesterol';

proc means data = table2_manuscript n median;
        var tri;
        class defnash;
        title3 'Manuscript Triglycerides';

proc means data = table2_manuscript n median;
        var hba1c;
        class defnash;
        title3 'Manuscript HbA1c';

proc means data = table2_manuscript n median;
        var gluc;
        class defnash;
        title3 'Manuscript Fasting serum glucose';

proc means data = table2_manuscript n median;
        var insu;
        class defnash;
        title3 'Manuscript Fasting serum insulin';

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

proc freq data = table2_manuscript;
        tables ana;
        by defnash;
```

```
        title3 'Manuscript ANA (%)';

proc freq data = table2_manuscript;
        tables asma;
        by defnash;
        title3 'Manuscript ASMA (%)';

proc freq data = table2_manuscript;
        tables anaasma;
        by defnash;
        title3 'Manuscript ANA and ASMA (%)';

proc freq data = table2_manuscript;
        tables ama;
        by defnash;
        title3 'Manuscript AMA (%)';

proc means data = table2_manuscript n median;
        var ferr;
        class defnash;
        title3 'Manuscript Ferritin';


proc freq data = table2_manuscript;
        tables igrade;
        by defnash;
        title3 'Manuscript Steatosis (%)';

proc freq data = table2_manuscript;
        tables iinflam;
        by defnash;
        title3 'Manuscript Lobular inflammation (%)';

proc freq data = table2_manuscript;
        tables iportal;
        by defnash;
        title3 'Manuscript Portal inflammation (%)';

proc freq data = table2_manuscript;
        tables iball;
        by defnash;
        title3 'Manuscript Ballooning (%)';

proc freq data = table2_manuscript;
        tables inas;
        by defnash;
        title3 'Manuscript NAFLD Activity Score (%)';

proc means data = table2_manuscript n mean ;
        var nfibro;
        class defnash;
        title3 'Manuscript Fibrosis score (%)';

proc freq data = table2_manuscript;
        tables mallory;
        by defnash;
```

```
        title3 'Manuscript Mallory bodies (%)';

proc freq data = table2_manuscript;
        tables leng10;
        by defnash;
        title3 'Manuscript Biopsy length (%)';

data table2_dsic;
        set dsic_values;
        if proximity = '<= 6 months';

proc freq data = table2_dsic;
        tables defnash;
        title3 'Table 2 DSIC';

data table2_dsic;
        set table2_dsic;
        if defnash ne .;

proc sort data = table2_dsic;
        by defnash;

proc freq data = table2_dsic;
        tables rg111 /list missing;
        by defnash;
        title3 'DSIC Male (%)';

proc means data = table2_dsic n median;
        var age;
        class defnash;
        title3 'DSIC Age';

proc freq data = table2_dsic;
        tables white;
        by defnash;
        title3 'DSIC White (%)';

proc freq data = table2_dsic;
        tables rg112;
        by defnash;
        title3 'DSIC Hispanic (%)';

proc freq data = table2_dsic;
        tables bg349af /missing;
        by defnash;
        title3 'DSIC Hypertension (%)';

proc freq data = table2_dsic;
        tables bg349b /missing;
        by defnash;
        title3 'DSIC Type 2 Diabetes (%)';

proc freq data = table2_dsic;
        tables meta /missing;
        by defnash;
        title3 'DSIC Metabolic syndrome';
```

```
proc freq data = table2_dsic;
        tables acanth;
        by defnash;
        title3 'DSIC Acanthosis nigricans, Positive (%)';

proc means data = table2_dsic n mean;
        var acanthn;
        class defnash;
        title3 'DSIC Acanthosis nigricans, Severity score';

proc means data = table2_dsic n median;
        var bmi;
        class defnash;
        title3 'DSIC BMI';

proc means data = table2_dsic n median;
        var waist_circum;
        class defnash;
        title3 'DSIC Waist circumference';

proc means data = table2_dsic n median;
        var wthip;
        class defnash;
        title3 'DSIC Waist to hip ratio';

proc means data = table2_dsic n median;
        var lr328;
        class defnash;
        title3 'DSIC AST';

proc means data = table2_dsic n median;
        var lr329;
        class defnash;
        title3 'DSIC ALT';

proc means data = table2_dsic n median;
        var ratio;
        class defnash;
        title3 'DSIC AST/ALT';

proc means data = table2_dsic n median;
        var lr330;
        class defnash;
        title3 'DSIC Alkaline phosphatase';

proc freq data = table2_dsic;
        tables isolated_ab_alk_phos /missing;
        by defnash;
        title3 'DSIC Isolated abnormal alkaline phosphatase';

proc means data = table2_dsic n median;
        var lr331;
        class defnash;
        title3 'DSIC GGT';
```

```
proc means data = table2_dsic n median;
        var globulin;
        class defnash;
        title3 'DSIC Globulin';

proc means data = table2_dsic n median;
        var lr333;
        class defnash;
        title3 'DSIC Albumin';

proc means data = table2_dsic n median;
        var lr326;
        class defnash;
        title3 'DSIC Bilirubin, total';

proc means data = table2_dsic n median;
        var lr327;
        class defnash;
        title3 'DSIC Bilirubin, direct';

proc means data = table2_dsic n median;
        var lr335;
        class defnash;
        title3 'DSIC INR';

proc means data = table2_dsic n median;
        var lr310;
        class defnash;
        title3 'DSIC Hematocrit';

proc means data = table2_dsic n median;
        var lr311;
        class defnash;
        title3 'DSIC White blood cells';

proc means data = table2_dsic n median;
        var platelet;
        class defnash;
        title3 'DSIC Platelet count';

proc means data = table2_dsic n median;
        var lr338b;
        class defnash;
        title3 'DSIC Total cholesterol';

proc means data = table2_dsic n median;
        var lr338c;
        class defnash;
        title3 'DSIC HDL cholesterol';

proc means data = table2_dsic n median;
        var lr338d;
        class defnash;
        title3 'DSIC LDL Cholesterol';

proc means data = table2_dsic n median;
```

```
        var lr338a;
        class defnash;
        title3 'DSIC Triglycerides';

proc means data = table2_dsic n median;
        var lr324;
        class defnash;
        title3 'DSIC HbA1c';

proc means data = table2_dsic n median;
        var lr339a;
        class defnash;
        title3 'DSIC Fasting serum glucose';

proc means data = table2_dsic n median;
        var lr339b;
        class defnash;
        title3 'DSIC Fasting serum insulin';



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

proc freq data = table2_dsic;
        tables ana;
        by defnash;
        title3 'DSIC ANA (% positive)';

proc freq data = table2_dsic;
        tables asma;
        by defnash;
        title3 'DSIC ASMA (% positive)';

proc freq data = table2_dsic;
        tables anaasma;
        by defnash;
        title3 'DSIC ANA+ASMA (% both positive)';

proc freq data = table2_dsic;
        tables ama;
        by defnash;
        title3 'DSIC AMA (% positive)';

proc means data = table2_dsic n median;
        var ls109c;
        class defnash;
        title3 'DSIC Ferritin';

proc freq data = table2_dsic;
        tables igrade;
        by defnash;
        title3 'DSIC Steatosis';
```

```
proc freq data = table2_dsic;
       tables iinflam;
       by defnash;
       title3 'DSIC Lobular inflammation';

proc freq data = table2_dsic;
       tables iportal;
       by defnash;
       title3 'DSIC Portal inflammation';

proc freq data = table2_dsic;
       tables iball;
       by defnash;
       title3 'DSIC Ballooning';

proc freq data = table2_dsic;
       tables inas;
       by defnash;
       title3 'DSIC NAFLD Activity Score';

proc means data = table2_dsic;
       var nfibro;
       class defnash;
       title3 'DSIC Fibrosis score';

proc freq data = table2_dsic;
       tables cr119;
       by defnash;
       title3 'DSIC Mallory bodies';

proc freq data = table2_dsic;
       tables leng10;
       by defnash;
       title3 'DSIC Biopsy length';

proc freq data = table2_manuscript;
       tables ifibro;
       title3 'Table 3';

data table3_manuscript;
       set table2_manuscript;
       if ifibro ne .;

proc sort data = table3_manuscript;
       by ifibro;

proc freq data = table3_manuscript;
       tables male;
       by ifibro;
       title3 'Manuscript Male (%)';

proc means data = table3_manuscript n median;
       var age;
       class ifibro;
       title3 'Manuscript Age, years';
```

```
proc freq data = table3_manuscript;
        tables white;
        by ifibro;
        title3 'Manuscript White (%)';

proc freq data = table3_manuscript;
        tables hispanic;
        by ifibro;
        title3 'Manuscript Hispanic (%)';

proc freq data = table3_manuscript;
        tables htn;
        by ifibro;
        title3 'Manuscript Hypertension (%)';

proc freq data = table3_manuscript;
        tables diab2;
        by ifibro;
        title3 'Manuscript Type 2 diabetes (%)';

proc freq data = table3_manuscript;
        tables meta;
        by ifibro;
        title3 'Manuscript Metabolic syndrome (%)';

proc freq data = table3_manuscript;
        tables acanth;
        by ifibro;
        title3 'Manuscript Acanthosis nigricans positive (%)';

proc means data = table3_manuscript n mean;
        var acanthn;
        class ifibro;
        title3 'Manuscript Acanthosis nigricans Severity score';

proc means data = table3_manuscript n median;
        var bmi;
        class ifibro;
        title3 'Manuscript Body mass index';

proc means data = table3_manuscript n median;
        var waist;
        class ifibro;
        title3 'Manuscript Waist circumference';

proc means data = table3_manuscript n median;
        var wthip;
        class ifibro;
        title3 'Manuscript Waist-to-hip ratio';

proc means data = table3_manuscript n median;
        var ast;
        class ifibro;
        title3 'Manuscript AST';

proc means data = table3_manuscript n median;
```

```
        var alt;
        class ifibro;
        title3 'Manuscript ALT';

proc means data = table3_manuscript n median;
        var ratio;
        class ifibro;
        title3 'Manuscript AST/ALT';

proc means data = table3_manuscript n median;
        var alka;
        class ifibro;
        title3 'Manuscript Alkaline phosphatase';

proc freq data = table3_manuscript;
        tables alkaiso;
        by ifibro;
        title3 'Manuscript Isolated abnormal alkaline phosphatase (%)';

proc means data = table3_manuscript n median;
        var ggt;
        class ifibro;
        title3 'Manuscript GGT';

proc means data = table3_manuscript n median;
        var glob;
        class ifibro;
        title3 'Manuscript Globulin';

proc means data = table3_manuscript n median;
        var alb;
        class ifibro;
        title3 'Manuscript Albumin';

proc means data = table3_manuscript n median;
        var bilit;
        class ifibro;
        title3 'Manuscript Bilirubin, total';

proc means data = table3_manuscript n median;
        var bilid;
        class ifibro;
        title3 'Manuscript Bilirubin, direct';

proc means data = table3_manuscript n mean;
        var inr;
        class ifibro;
        title3 'Manuscript Inernational normalized ratio';

proc means data = table3_manuscript n median;
        var hema;
        class ifibro;
        title3 'Manuscript Hematocrit';

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

```
        class ifibro;
        title3 'Manuscript White blood cells';

proc means data = table3_manuscript n median;
        var plat;
        class ifibro;
        title3 'Manuscript Platelet count';

proc means data = table3_manuscript n median;
        var chol;
        class ifibro;
        title3 'Manuscript Total cholesterol';

proc means data = table3_manuscript n median;
        var hdl;
        class ifibro;
        title3 'Manuscript HDL cholesterol';

proc means data = table3_manuscript n median;
        var ldl;
        class ifibro;
        title3 'Manuscript LDL Cholesterol';

proc means data = table3_manuscript n median;
        var tri;
        class ifibro;
        title3 'Manuscript Triglycerides';

proc means data = table3_manuscript n median;
        var hba1c;
        class ifibro;
        title3 'Manuscript HbA1c';

proc means data = table3_manuscript n median;
        var gluc;
        class ifibro;
        title3 'Manuscript Fasting serum glucose';

proc means data = table3_manuscript n median;
        var insu;
        class ifibro;
        title3 'Manuscript Fasting serum insulin';

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

proc freq data = table3_manuscript;
        tables ana;
        by ifibro;
        title3 'Manuscript ANA (%)';

proc freq data = table3_manuscript;
        tables asma;
        by ifibro;
```

```
        title3 'Manuscript ASMA (%)';

proc freq data = table3_manuscript;
        tables anaasma;
        by ifibro;
        title3 'Manuscript ANA and ASMA (%)';

proc freq data = table3_manuscript;
        tables ama;
        by ifibro;
        title3 'Manuscript AMA (%)';

proc means data = table3_manuscript n median;
        var ferr;
        class ifibro;
        title3 'Manuscript Ferritin';


proc freq data = table3_manuscript;
        tables igrade;
        by ifibro;
        title3 'Manuscript Steatosis (%)';

proc freq data = table3_manuscript;
        tables iinflam;
        by ifibro;
        title3 'Manuscript Lobular inflammation (%)';

proc freq data = table3_manuscript;
        tables iportal;
        by ifibro;
        title3 'Manuscript Portal inflammation (%)';

proc freq data = table3_manuscript;
        tables iball;
        by ifibro;
        title3 'Manuscript Ballooning (%)';

proc freq data = table3_manuscript;
        tables inas;
        by ifibro;
        title3 'Manuscript NAFLD Activity Score (%)';

proc freq data = table3_manuscript ;
        tables defnash;
        by ifibro;
        title3 'Manuscript Definite NASH (%)';

proc freq data = table3_manuscript;
        tables mallory;
        by ifibro;
        title3 'Manuscript Mallory bodies (%)';

proc freq data = table3_manuscript;
        tables leng10;
        by ifibro;
```

```
        title3 'Manuscript Biopsy length (%)';

proc freq data = table2_dsic;
        tables cr121;
        title3 'Table 3 DSIC';

data table3_dsic;
        set table2_dsic;
        if cr121 = '0' then fibro_stage = 0;
        if cr121 in ('1a', '1b', '1c', '2') then fibro_stage = 1;
        if cr121 = '3' then fibro_stage = 2;
        if cr121 = '4' then fibro_stage = 3;
        if cr121 = 'm' then fibro_stage = .;

data table3_dsic;
        set table3_dsic;
        if fibro_stage ne .;

proc sort data = table3_dsic;
        by fibro_stage;

proc freq data = table3_dsic;
        tables rg111 /list missing;
        by fibro_stage;
        title3 'DSIC Male (%)';

proc means data = table3_dsic n median;
        var age;
        class fibro_stage;
        title3 'DSIC Age';

proc freq data = table3_dsic;
        tables white;
        by fibro_stage;
        title3 'DSIC White (%)';

proc freq data = table3_dsic;
        tables rg112;
        by fibro_stage;
        title3 'DSIC Hispanic (%)';

proc freq data = table3_dsic;
        tables bg349af /missing;
        by fibro_stage;
        title3 'DSIC Hypertension (%)';

proc freq data = table3_dsic;
        tables bg349b /missing;
        by fibro_stage;
        title3 'DSIC Type 2 Diabetes (%)';

proc freq data = table3_dsic;
        tables meta /missing;
        by fibro_stage;
        title3 'DSIC Metabolic syndrome';
```

```
proc freq data = table3_dsic;
        tables acanth;
        by fibro_stage;
        title3 'DSIC Acanthosis nigricans, Positive (%)';

proc means data = table3_dsic n mean;
        var acanthn;
        class fibro_stage;
        title3 'DSIC Acanthosis nigricans, Severity score';

proc means data = table3_dsic n median;
        var bmi;
        class fibro_stage;
        title3 'DSIC BMI';

proc means data = table3_dsic n median;
        var waist_circum;
        class fibro_stage;
        title3 'DSIC Waist circumference';

proc means data = table3_dsic n median;
        var wthip;
        class fibro_stage;
        title3 'DSIC Waist to hip ratio';

proc means data = table3_dsic n median;
        var lr328;
        class fibro_stage;
        title3 'DSIC AST';

proc means data = table3_dsic n median;
        var lr329;
        class fibro_stage;
        title3 'DSIC ALT';

proc means data = table3_dsic n median;
        var ratio;
        class fibro_stage;
        title3 'DSIC AST/ALT';

proc means data = table3_dsic n median;
        var lr330;
        class fibro_stage;
        title3 'DSIC Alkaline phosphatase';

proc freq data = table3_dsic;
        tables isolated_ab_alk_phos /missing;
        by fibro_stage;
        title3 'DSIC Isolated abnormal alkaline phosphatase';

proc means data = table3_dsic n median;
        var lr331;
        class fibro_stage;
        title3 'DSIC GGT';

proc means data = table3_dsic n median;
```

```
        var globulin;
        class fibro_stage;
        title3 'DSIC Globulin';

proc means data = table3_dsic n median;
        var lr333;
        class fibro_stage;
        title3 'DSIC Albumin';

proc means data = table3_dsic n median;
        var lr326;
        class fibro_stage;
        title3 'DSIC Bilirubin, total';

proc means data = table3_dsic n median;
        var lr327;
        class fibro_stage;
        title3 'DSIC Bilirubin, direct';

proc means data = table3_dsic n median;
        var lr335;
        class fibro_stage;
        title3 'DSIC INR';

proc means data = table3_dsic n median;
        var lr310;
        class fibro_stage;
        title3 'DSIC Hematocrit';

proc means data = table3_dsic n median;
        var lr311;
        class fibro_stage;
        title3 'DSIC White blood cells';

proc means data = table3_dsic n median;
        var platelet;
        class fibro_stage;
        title3 'DSIC Platelet count';

proc means data = table3_dsic n median;
        var lr338b;
        class fibro_stage;
        title3 'DSIC Total cholesterol';

proc means data = table3_dsic n median;
        var lr338c;
        class fibro_stage;
        title3 'DSIC HDL cholesterol';

proc means data = table3_dsic n median;
        var lr338d;
        class fibro_stage;
        title3 'DSIC LDL Cholesterol';

proc means data = table3_dsic n median;
        var lr338a;
```

```
        class fibro_stage;
        title3 'DSIC Triglycerides';

proc means data = table3_dsic n median;
        var lr324;
        class fibro_stage;
        title3 'DSIC HbA1c';

proc means data = table3_dsic n median;
        var lr339a;
        class fibro_stage;
        title3 'DSIC Fasting serum glucose';

proc means data = table3_dsic n median;
        var lr339b;
        class fibro_stage;
        title3 'DSIC Fasting serum insulin';



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

proc freq data = table3_dsic;
        tables ana;
        by fibro_stage;
        title3 'DSIC ANA (% positive)';

proc freq data = table3_dsic;
        tables asma;
        by fibro_stage;
        title3 'DSIC ASMA (% positive)';

proc freq data = table3_dsic;
        tables anaasma;
        by fibro_stage;
        title3 'DSIC ANA+ASMA (% both positive)';

proc freq data = table3_dsic;
        tables ama;
        by fibro_stage;
        title3 'DSIC AMA (% positive)';

proc means data = table3_dsic n median;
        var ls109c;
        class fibro_stage;
        title3 'DSIC Ferritin';

proc freq data = table3_dsic;
        tables igrade;
        by fibro_stage;
        title3 'DSIC Steatosis';

proc freq data = table3_dsic;
```

```
        tables iinflam;
        by fibro_stage;
        title3 'DSIC Lobular inflammation';

proc freq data = table3_dsic;
        tables iportal;
        by fibro_stage;
        title3 'DSIC Portal inflammation';

proc freq data = table3_dsic;
        tables iball;
        by fibro_stage;
        title3 'DSIC Ballooning';

proc freq data = table3_dsic;
        tables inas;
        by fibro_stage;
        title3 'DSIC NAFLD Activity Score';

proc freq data = table3_dsic;
        tables defnash;
        by fibro_stage;
        title3 'DSIC Fibrosis score';

proc freq data = table3_dsic;
        tables cr119;
        by fibro_stage;
        title3 'DSIC Mallory bodies';

proc freq data = table3_dsic;
        tables leng10;
        by fibro_stage;
        title3 'DSIC Biopsy length';
```