

Dataset Integrity Check for FibroScan in Pediatric Cholestatic Liver Disease (FORCE) Leung

Prepared by NIDDK-CR
May 6, 2024

Contents

1 Standard Disclaimer	2
2 Study Background	2
3 Archived Datasets	2
4 Statistical Methods	2
5 Results	3
6 Conclusions	3
7 References	3
Table A1: Variables used to replicate Supplemental Table 1 - Comparison of Demographics between Biomarker study participants vs excluded FORCE participants	4
Table A2: Variables used to replicate Table 1 - Clinical characteristics of study participants by diagnosis .	5
Table B1: Comparison of values computed in integrity check to Supplemental Table 1	6
Table B2: Comparison of values computed in integrity check to Table 1.....	7
Attachment A: SAS Code	9

1 Standard Disclaimer

The intent of this DSIC is to provide confidence that the data distributed by the NIDDK repository is a true copy of the study data. Our intent is not to assess the integrity of the statistical analyses reported by study investigators. As with all statistical analyses of complex datasets, complete replication of a set of statistical results should not be expected in secondary analysis. This occurs for a number of reasons including differences in the handling of missing data, restrictions on cases included in samples for a particular analysis, software coding used to define complex variables, etc. Experience suggests that most discrepancies can ordinarily be resolved by consultation with the study data coordinating center (DCC), however this process is labor-intensive for both DCC and Repository staff. It is thus not our policy to resolve every discrepancy that is observed in an integrity check. Specifically, we do not attempt to resolve minor or inconsequential discrepancies with published results or discrepancies that involve complex analyses, unless NIDDK Repository staff suspect that the observed discrepancy suggests that the dataset may have been corrupted in storage, transmission, or processing by repository staff. We do, however, document in footnotes to the integrity check those instances in which our secondary analyses produced results that were not fully consistent with those reported in the target publication.

2 Study Background

The FibroScan in Pediatric Cholestatic Liver Disease (FORCE) study was a cross-sectional and longitudinal assessment of the utility of liver stiffness measurements (LSM) in children with chronic cholestatic liver disease. Study participants were from 13 ChiLDReN sites in the U.S. and Canada, and were also enrolled in the PROBE, BASIC, or LOGIC studies. FORCE participants were evaluated for a period of up to 24 months to assess a non-invasive ultrasound tool (FibroScan™) to detect and quantify global liver fibrosis in children with biliary atresia (BA), alpha-1 antitrypsin deficiency (A1ATD), and Alagille syndrome (ALGS). The participants were non-fasted and non-sedated during data collection. There were three visits in the study: baseline, 12-month follow-up, and 24-month follow-up. Clinical data and biospecimens were collected at each visit along with repeated FibroScan measurements such as LSM to quantify liver fibrosis, and controlled attenuation parameter (CAP) to quantify liver steatosis.

3 Archived Datasets

A full listing of archived datasets included in the package can be found in the Roadmap document. All data files, as provided by the Data Coordinating Center (DCC), are located in the FORCE folder in the data package. For this replication, variables were taken from the “forcebiomarker.sas7bdat” dataset.

4 Statistical Methods

Analyses were performed to replicate results for the data in the publication by Leung et al. [1]. To verify the integrity of the data, only descriptive statistics were computed.

5 Results

For Supplemental Table 1 in the publication [1], Comparison of Demographics between Biomarker study participants vs excluded FORCE participants, Table A1 lists the variables that were used in the replication, and Table B1 compares the results calculated from the archived data files to the results published in Supplemental Table 1. For Table 1 in the publication [1], Clinical characteristics of study participants by diagnosis, Table A2 lists the variables that were used in the replication, and Table B2 compares the results calculated from the archived data files to the results published in Table 1. The results of the replication are within expected variation to the published results.

6 Conclusions

The NIDDK Central Repository is confident that the FORCE Leung data files to be distributed are a true copy of the study data.

7 References

[1] Leung DH, Devaraj S, Goodrich NP, Chen X, Rajapakshe D, Ye W, Andreev V, Minard CG, Guffey D, Molleston JP, Bass LM, Karpen SJ, Kamath BM, Wang KS, Sundaram SS, Rosenthal P, McKiernan P, Loomes KM, Jensen MK, Horslen S, Bezerra JA, Magee JC, Merion RM, Sokol RJ, Shneider BL. Serum Biomarkers Correlated with Liver Stiffness Assessed in a Multi-center Study of Pediatric Cholestatic Liver Disease. *Hepatology*, 77(2), 530-545, February 2023. doi: <https://doi.org/10.1002/hep.32777>

Table A1: Variables used to replicate Supplemental Table 1 – Comparison of Demographics between Biomarker study participants vs excluded FORCE participants

Table Variable	dataset.variable
Age	forcebiomarker.inelisa forcebiomarker.age
Sex	forcebiomarker.inelisa forcebiomarker.sex
Ethnicity	forcebiomarker.inelisa forcebiomarker.ethnicity
Race	forcebiomarker.inelisa forcebiomarker.race

Table A2: Variables used to replicate Table 1 – Clinical characteristics of study participants by diagnosis

Table Variable	dataset.variable
FibroScan LSM	forcebiomarker.emedian
Height z-score	forcebiomarker.heightz
Weight z-score	forcebiomarker.weightz
BMI z-score	forcebiomarker.bmiz
Total bilirubin	forcebiomarker.forcebili
GGTP	forcebiomarker.forceggtp
AST	forcebiomarker.forceast
ALT	forcebiomarker.forcealt
Albumin	forcebiomarker.forcealbumin
INR	forcebiomarker.forceinr
Spleen size	forcebiomarker.spleensize
Platelet count	forcebiomarker.forceplatelets
APRI	forcebiomarker.apri
PELD	forcebiomarker.peld

Table B1: Comparison of values computed in integrity check to Supplemental Table 1

Characteristic	Pub: Biomarker Participants (n=330)	DSIC: Biomarker Participants (n=330)	Diff. (n=0)	Pub: FORCE Participants (n=128)	DSIC: FORCE Participants (n=128)	Diff. (n=0)
Age (years)						
Median (Q1, Q3)	9.2 (5.5, 13.0)	9.2 (5.5, 13.0)	0 (0)	6.1 (3.1, 11.4)	6.1 (3.1, 11.4)	0 (0)
0 to 4 years	76 (23%)	76 (23%)	0 (0)	52 (41%)	52 (41%)	0 (0)
5 to 9 years	113 (34%)	113 (34%)	0 (0)	38 (30%)	38 (30%)	0 (0)
10 to 15 years	92 (28%)	92 (28%)	0 (0)	29 (23%)	29 (23%)	0 (0)
15 years and older	49 (15%)	49 (15%)	0 (0)	9 (7%)	9 (7%)	0 (0)
Sex						
Male	174 (53%)	174 (53%)	0 (0)	64 (50%)	64 (50%)	0 (0)
Female	156 (47%)	156 (47%)	0 (0)	64 (50%)	64 (50%)	0 (0)
Ethnicity						
Refused/Unknown	2 (1%)	2 (1%)	0 (0)	1 (1%)	1 (1%)	0 (0)
Hispanic/Latino	65 (20%)	65 (20%)	0 (0)	18 (14%)	18 (14%)	0 (0)
Non-Hispanic/Latino	263 (80%)	263 (80%)	0 (0)	109 (85%)	109 (85%)	0 (0)
Race						
White	223 (68%)	223 (68%)	0 (0)	83 (65%)	83 (65%)	0 (0)
Black or African-American	34 (10%)	34 (10%)	0 (0)	13 (10%)	13 (10%)	0 (0)
Asian	22 (7%)	22 (7%)	0 (0)	8 (6%)	8 (6%)	0 (0)
Multiracial	25 (8%)	25 (8%)	0 (0)	15 (12%)	15 (12%)	0 (0)
Other	10 (3%)	10 (3%)	0 (0)	5 (4%)	5 (4%)	0 (0)
Refused/Unknown	16 (5%)	16 (5%)	0 (0)	4 (3%)	4 (3%)	0 (0)

Table B2: Comparison of values computed in integrity check to Table 1

Characteristic	Pub: BA (n=187)	DSIC: BA (n=187)	Diff. (n=0)	Pub: A1AT (n=78)	DSIC: A1AT (n=78)	Diff. (n=0)	Pub: ALGS (n=65)	DSIC: ALGS (n=65)	Diff. (n=0)
FibroScan LSM (kPa)									
n	187	187	0	78	78	0	65	65	0
Median (Q1, Q3)	12.8 (7.3, 20.0)	12.8 (7.3, 20.0)	0 (0)	6.3 (5.1, 8.5)	6.3 (5.1, 9.5)	0 (0, 1)	9.0 (6.9, 12.5)	9.0 (6.9, 12.5)	0 (0)
Height z-score									
n	186	186	0	76	76	0	65	65	0
Median (Q1, Q3)	0.3 (-0.6, 0.7)	0.3 (-0.6, 0.7)	0 (0)	0.3 (-0.3, 1.0)	0.3 (-0.3, 1.0)	0 (0)	-1.2 (-2.0, -0.6)	-1.2 (-2.0, -0.6)	0 (0)
Weight z-score									
n	186	186	0	76	76	0	65	65	0
Median (Q1, Q3)	0.4 (-0.3, 1.0)	0.4 (-0.3, 1.0)	0 (0)	0.5 (-0.1, 0.9)	0.5 (-0.1, 0.9)	0 (0)	-1.1 (-2.1, -0.5)	-1.1 (-2.1, -0.5)	0 (0)
BMI z-score									
n	186	186	0	76	76	0	65	65	0
Median (Q1, Q3)	0.5 (-0.1, 1.1)	0.5 (-0.1, 1.1)	0 (0)	0.3 (-0.3, 0.9)	0.3 (-0.3, 0.9)	0 (0)	-0.5 (-1.3, 0.2)	-0.5 (-1.3, 0.2)	0 (0)
Total bilirubin (mg/dL)									
n	177	177	0	78	78	0	64	64	0
Median (Q1, Q3)	0.6 (0.4, 1.1)	0.6 (0.4, 1.1)	0 (0)	0.4 (0.2, 0.6)	0.4 (0.2, 0.6)	0 (0)	1.1 (0.6, 2.9)	1.1 (0.6, 2.9)	0 (0)
GGTP									
n	164	164	0	72	72	0	56	56	0
Median (Q1, Q3)	79.5 (29.5, 175.5)	79.5 (29.5, 175.5)	0 (0)	26.0 (17.0, 50.0)	26.0 (17.0, 50.0)	0 (0)	352.5 (185.0, 797.0)	352.5 (185.0, 797.0)	0 (0)
AST									
n	177	177	0	78	78	0	64	64	0
Median (Q1, Q3)	54 (34, 94)	54 (34, 94)	0 (0)	46 (31, 60)	46 (31, 60)	0 (0)	110 (74, 178)	110 (74, 178)	0 (0)
ALT									
n	178	178	0	78	78	0	64	64	0
Median (Q1, Q3)	54 (35, 94)	54 (35, 94)	0 (0)	53 (35, 77)	53 (35, 77)	0 (0)	140 (88, 243)	140 (88, 243)	0 (0)
Albumin (g/dL)									
n	176	176	0	76	76	0	63	63	0
Median (Q1, Q3)	4.3 (3.9, 4.5)	4.3 (3.9, 4.5)	0 (0)	4.4 (4.2, 4.6)	4.4 (4.2, 4.6)	0 (0)	4.4 (3.9, 4.5)	4.4 (3.9, 4.5)	0 (0)
n (%) < 3.0	3 (2%)	3 (2%)	0 (0)	1 (1%)	1 (1%)	0 (0)	2 (3%)	2 (3%)	0 (0)
INR									
n	148	148	0	60	60	0	56	56	0
Median (Q1, Q3)	1.1 (1.0, 1.2)	1.1 (1.0, 1.2)	0 (0)	1.1 (1.0, 1.1)	1.1 (1.0, 1.1)	0 (0)	1.0 (1.0, 1.1)	1.0 (1.0, 1.1)	0 (0)

Characteristic	Pub: BA (n=187)	DSIC: BA (n=187)	Diff. (n=0)	Pub: A1AT (n=78)	DSIC: A1AT (n=78)	Diff. (n=0)	Pub: ALGS (n=65)	DSIC: ALGS (n=65)	Diff. (n=0)
Spleen size (cm below costal margin)									
n	187	187	0	78	78	0	65	65	0
Median (Q1, Q3)	1 (0, 5)	1 (0, 5)	0 (0)	0 (0, 0)	0 (0, 0)	0 (0)	0 (0, 0)	0 (0, 0)	0 (0)
n (%) > 2 cm	75 (40%)	75 (40%)	0 (0)	3 (4%)	3 (4%)	0 (0)	11 (17%)	11 (17%)	0 (0)
Platelet count									
n	183	183	0	75	75	0	62	62	0
Median (Q1, Q3)	145 (84, 242)	145 (85, 242)	0 (1, 0)	269 (220, 332)	269 (220, 332)	0 (0)	251 (198, 325)	251 (198, 325)	0 (0)
n (%) < 150	92 (50%)	92 (50%)	0 (0)	7 (9%)	7 (9%)	0 (0)	9 (15%)	9 (15%)	0 (0)
APRI									
n	176	176	0	75	74	1	62	62	0
Median (Q1, Q3)	0.9 (0.4, 2.4)	0.9 (0.4, 2.4)	0 (0)	0.4 (0.3, 0.6)	0.4 (0.3, 0.6)	0 (0)	1.2 (0.6, 2.1)	1.2 (0.6, 2.1)	0 (0)
n (%) < 1.5	106 (60%)	106 (60%)	0 (0)	66 (88%)	66 (88%)	0 (0)	39 (63%)	39 (63%)	0 (0)
PELD									
n	141	141	0	58	58	0	55	55	0
Median (Q1, Q3)	-10.3 (-12.7, -5.6)	-10.3 (-12.7, -5.6)	0 (0, 0)	-13.2 (-16.0, -10.7)	-13.2 (-16.0, -10.7)	0 (0)	-6.3 (-10.1, 0.8)	-6.3 (-10.1, 0.8)	0 (0)

Attachment A: SAS Code

```
/******  
/* FORCE Biomarker (Leung) */  
/* DSIC */  
/******  
  
ods escapechar='^';  
*-----;  
*-----Libraries-----;  
*-----;  
libname Repo "X:\NIDDK\niddk-dr_studies1\FORCE\private_created_data\Leung FORCE Biomarker  
Submission\DataAndFiles" access=readonly;  
ods graphics off;  
  
*set options and point to format library;  
options fmtsearch=(Repo.Formats) nodate;  
ods noproctitle;  
  
*Read in dataset for two purposes: sample selection and main analysis;  
Data SampleSelect;  
set Repo.ForceBiomarker;  
run;  
  
Data Biomarker;  
set Repo.ForceBiomarker(where=(InElisa));  
run;  
  
/*======  
Begin Code: Supp Table 1 - Biomarker vs not demographic comparison  
=====*/  
  
*Note: table made manually based on following output;  
title "age comparison for biomarker vs non-biomarker";  
proc means data=SampleSelect maxdec=1 median p25 p75;  
where 0 lt Dgn; *exclude controls;  
class InElisa;  
Var Age;  
run;  
  
proc npar1way data=SampleSelect Wilcoxon;  
where 0 lt Dgn;  
class InElisa;  
var Age;  
run;  
  
title "Age, sex, ethnicity, race by biomarker inclusion chi-square tests";
```

```

proc freq data=SampleSelect;
where 0 lt Dgn;
table (Age Sex Ethnicity Race)*InElisa/nopercent norow chisq missing;
format Age AgeCatFmt.
Ethnicity EthnicityFmt.
Race RaceFmt.
Sex SexFmt.;
run;

/*=====
End Code: Supp Table 1 - Biomarker vs not demographic comparison
=====*/

/*=====
Begin Code: Table 1 - Clinical characteristics
=====*/

*Note: table made based on following output;
title "Table 1 Clinical characteristics: Numbers for continuous variables";
proc means data=Biomarker n median q1 q3 maxdec=1;
where ForceDgn ne .;
*where Dgn ne 0;
class Dgn;
var eMedian HeightZ WeightZ BMIZ ForceBili ForceGGTP ForceAST ForceALT ForceAlbumin ForceINR
SpleenSize
ForcePlatelets APRI PELD;
run;

title "Table 1 Clinical characteristics: Tests for continuous variables";
proc npar1way data=Biomarker Wilcoxon;
where 0 lt Dgn;
class Dgn;
var eMedian HeightZ WeightZ BMIZ ForceBili ForceGGTP ForceAST ForceALT ForceAlbumin ForceINR
SpleenSize
ForcePlatelets APRI PELD;
run;

title "Table 1 Clinical characteristics: dichotomous variables";
proc format;
value AlbuminLt3Fmt
low - <3 = "<3"
3 - high = "3 or higher";
value SpleenGt2Fmt
low - 2 = "2 or lower"
2 - high = ">2";
value PlateLt150Fmt
low - <150 = "<150"
150 - high = "150 and higher";

```

```

value APRILt1Pt5Fmt
low - <1.5 = "<1.5"
1.5 - high = "1.5 and higher";
run; quit;

```

```

title2 "Albumin <3";
proc freq data=Biomarker;
where 0 lt Dgn;
table ForceAlbumin*Dgn/chisq nopercnt norow;
format ForceAlbumin AlbuminLt3Fmt.;
run;

```

```

title2 "Spleen size > 2";
proc freq data=Biomarker;
where 0 lt Dgn and .z lt SpleenSize;
table SpleenSize*Dgn/chisq;
format SpleenSize SpleenGt2Fmt.;
run;

```

```

title2 "Platelet count <150";
proc freq data=Biomarker;
where 0 lt Dgn;
table ForcePlatelets*Dgn/chisq;
format ForcePlatelets PlateLt150Fmt.;
run;

```

```

title2 "APRI < 1.5";
proc freq data=Biomarker;
where 0 lt Dgn;
table APRI*Dgn/chisq;
format APRI APRILt1Pt5Fmt.;
run;

```

```

/*=====
End Code: Table 1 - Clinical characteristics
=====*/

```