

# Dataset Integrity Check for the FibroScan in Pediatric Cholestatic Liver Disease (FORCE) study

Prepared by NIDDK CR  
May 14, 2021

## Contents

1 Standard Disclaimer .....	3
2 Study Background .....	3
3 Archived Datasets .....	3
4 Statistical Methods .....	3
5 Results .....	4
6 Conclusions .....	4
7 References .....	4
Table A: Variables used to replicate Table 1 – Biochemical Testing by Diagnosis for Participants with Valid FibroScans (n=458).....	5
Table B-1: Comparison of values computed in integrity check to reference article Table 1 values (n, mean, SD, median, quartiles) for BA and A1ATD groups.....	6
Table B-2: Comparison of values computed in integrity check to reference article Table 1 values (n, mean, SD, median, quartiles) for ALGS and All Diagnoses groups .....	7
Table B-3: Comparison of values computed in integrity check to reference article Table 1 values (n, mean, SD, median, quartiles, %) for BA and A1ATD groups .....	8
Table B-4: Comparison of values computed in integrity check to reference article Table 1 values (n, mean, SD, median, quartiles, %) for ALGS and All Diagnoses groups.....	9
Attachment A: SAS Code .....	10

# 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, a natural history study within ChiLDReN, 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 biosamples 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

All baseline data files, as provided by the Data Coordinating Center (DCC), are located in the FORCE baseline study data package. For this replication, variables were taken from the derived dataset: “forcebl.sas7bdat”.

## 4 Statistical Methods

Analyses were performed to duplicate results for the data reported by Dr. Shneider et al. [1] for Nonfasted Liver Stiffness Correlates with Liver Disease Parameters and Portal Hypertension in Pediatric Cholestatic Liver Disease. To verify the integrity of the datasets, descriptive statistics were computed.

## 5 Results

For Table 1 in the publication [1], Table 1 – Biochemical Testing by Diagnosis, Table A below lists the variables that were used in the replication, and Table B (B1-B4) compares the results calculated from the archived data file to the results published in Table 1.

## 6 Conclusions

The results of the replication are almost an exact match to the published results.

## 7 References

[1] Shneider BL, Goodrich NP, Ye W, et al. Nonfasted Liver Stiffness Correlates with Liver Disease Parameters and Portal Hypertension in Pediatric Cholestatic Liver Disease. *Hepatology Communications*, 4(11), 1694-1701, August 2020.

**Table A:** Variables used to replicate Table 1 – Biochemical Testing by Diagnosis for Participants with Valid FibroScans (n=458)

<b>Characteristic</b>	<b>dataset.variable</b>
Biliary Atresia (BA)	forcebl.forcedgnba
Alpha-1 Antitrypsin Deficiency (A1ATD)	forcebl.forcedgna1at
Alagille Syndrome (ALGS)	forcebl.forcedgnalgs
Total Bilirubin (TB)	forcebl.forcebili
Gamma-Glutamyl Transpeptidase (GGT)	forcebl.forceggtp
Aspartate Aminotransferase (AST)	forcebl.forceast
Alanine Aminotransferase (ALT)	forcebl.forcealt
Albumin	forcebl.forcealbumin
International Normalized Ratio (INR)	forcebl.forceinr
GGT to Platelet Ratio (GPR)	forcebl.gpr
Platelet Count	forcebl.forceplatelets
AST to Platelet Ratio Index (APRI)	forcebl.apri
Pediatric End-Stage Liver Disease (PELD) Score	forcebl.peld

**Table B-1:** Comparison of values computed in integrity check to reference article Table 1 values (n, mean, SD, median, quartiles) for BA and A1ATD groups

Variable	BA (n=254)			A1ATD (n=104)		
	Manuscript	DSIC	Diff	Manuscript	DSIC	Diff
TB, n	240	240	0	103	103	0
Mean (SD)	1.0 (1.3)	1.0 (1.3)	0	0.6 (0.7)	0.6 (0.7)	0
Median (Q1, Q3)	0.6 (0.4, 1.0)	0.6 (0.4, 1.0)	0	0.4 (0.3, 0.6)	0.4 (0.3, 0.6)	0
GGT, n	223	223	0	90	90	0
Mean (SD)	140.5 (190.5)	140.5 (190.5)	0	65.5 (102.0)	65.5 (102.0)	0
Median (Q1, Q3)	74.0 (26.0, 172.0)	74.0 (26.0, 172.0)	0	27.0 (17.0, 53.0)	27.0 (17.0, 53.0)	0
AST, n	241	241	0	103	103	0
Mean (SD)	78 (71)	78 (71)	0	60 (42)	60 (42)	0
Median (Q1, Q3)	51 (34, 92)	51 (34, 92)	0	48 (31, 76)	48 (31, 76)	0
ALT, n	241	241	0	103	103	0
Mean (SD)	81 (84)	81 (84)	0	73 (57)	73 (57)	0
Median (Q1, Q3)	51 (30, 94)	51 (30, 94)	0	59 (36, 85)	59 (36, 85)	0
INR, n	204	204	0	77	77	0
Mean (SD)	1.1 (0.1)	1.1 (0.1)	0	1.1 (0.1)	1.1 (0.1)	0
GPR, n	222	222	0	87	87	0
Mean (SD)	2.6 (4.4)	2.6 (4.4)	0	0.9 (1.9)	0.9 (1.9)	0
Median (Q1, Q3)	1.0 (0.3, 2.9)	1.0 (0.3, 2.9)	0	0.2 (0.1, 0.5)	0.2 (0.1, 0.5)	0

**Table B-2:** Comparison of values computed in integrity check to reference article Table 1 values (n, mean, SD, median, quartiles) for ALGS and All Diagnoses groups

Variable	ALGS (n=100)			All Diagnoses (n=458)		
	Manuscript	DSIC	Diff	Manuscript	DSIC	Diff
TB, n	97	97	0	440	440	0
Mean (SD)	3.0 (4.7)	3.0 (4.7)	0	1.3 (2.6)	1.3 (2.6)	0
Median (Q1, Q3)	1.1 (0.6, 3.0)	1.1 (0.6, 3.0)	0	0.6 (0.4, 1.1)	0.6 (0.4, 1.1)	0
GGT, n	84	84	0	397	397	0
Mean (SD)	451.0 (400.7)	451.0 (400.7)	0	189.6 (275.1)	189.2 (275.1)	0.4
Median (Q1, Q3)	326.5 (171.5, 609.0)	326.5 (171.5, 609.0)	0	77.0 (26.0, 234.0)	77.0 (26.0, 234.0)	0
AST, n	97	97	0	441	441	0
Mean (SD)	155 (120)	155 (121)	0 (1)	91 (87)	91 (87)	0
Median (Q1, Q3)	111 (74, 189)	111 (74, 189)	0	59 (36, 109)	59 (36, 109)	0
ALT, n	96	96	0	440	440	0
Mean (SD)	174 (125)	174 (125)	0	99 (98)	99 (98)	0
Median (Q1, Q3)	143 (89, 248)	143 (89, 248)	0	68 (36, 122)	68 (36, 122)	0
INR, n	84	84	0	365	365	0
Mean (SD)	1.1 (0.1)	1.1 (0.1)	0	1.1 (0.1)	1.1 (0.1)	0
GPR, n	82	82	0	391	391	0
Mean (SD)	4.6 (7.9)	4.6 (7.9)	0	2.6 (5.2)	2.6 (5.2)	0
Median (Q1, Q3)	2.7 (1.3, 5.6)	2.7 (1.3, 5.6)	0	0.9 (0.2, 3.0)	0.9 (0.2, 3.0)	0

**Table B-3:** Comparison of values computed in integrity check to reference article Table 1 values (n, mean, SD, median, quartiles, %) for BA and A1ATD groups

Variable	BA (n=254)			A1ATD (n=104)		
	Manuscript	DSIC	Diff	Manuscript	DSIC	Diff
Albumin, n	236	236	0	99	99	0
Mean (SD)	4.2 (0.5)	4.2 (0.5)	0	4.4 (0.4)	4.4 (0.4)	0
n (%) < 3.0	4 (2%)	4 (2%)	0	1 (1%)	1 (1%)	0
n (%) < 2.5	0 (0%)	0 (0%)	0	0 (0%)	0 (0%)	0
n (%) < 2.0	0 (0%)	0 (0%)	0	0 (0%)	0 (0%)	0
Platelet Count, n	247	247	0	97	97	0
Mean (SD)	173 (104)	173 (104)	0	273 (102)	273 (102)	0
n (%) < 150	118 (48%)	118 (48%)	0	10 (10%)	10 (10%)	0
n (%) < 100	84 (34%)	84 (34%)	0	6 (6%)	6 (6%)	0
n (%) < 50	18 (7%)	18 (7%)	0	3 (3%)	3 (3%)	0
APRI, n	239	239	0	97	97	0
Mean (SD)	1.9 (2.5)	1.9 (2.5)	0	0.8 (1.1)	0.8 (1.1)	0
Median (Q1, Q3)	0.9 (0.4, 2.3)	0.9 (0.4, 2.3)	0	0.4 (0.3, 0.8)	0.4 (0.3, 0.8)	0
n (%) > 1.0	109 (46%)	109 (46%)	0	18 (19%)	18 (19%)	0
n (%) > 1.5	86 (36%)	86 (36%)	0	13 (13%)	13 (13%)	0
n (%) > 2.0	68 (29%)	68 (29%)	0	8 (8%)	8 (8%)	0
PELD Score, n	193	193	0	75	75	0
Mean (SD)	-9.3 (5.7)	-9.3 (5.7)	0	-12.1 (5.4)	-12.1 (5.4)	0
Median (Q1, Q3)	-10.5 (-13.0, -5.7)	-10.5 (-13.0, -5.7)	0	-13.1 (-15.8, -10.3)	-13.1 (-15.8, -10.3)	0
n (%) > 0	13 (7%)	13 (7%)	0	3 (4%)	3 (4%)	0
n (%) > 10	1 (1%)	1 (1%)	0	0 (0%)	0 (0%)	0
n (%) > 20	0 (0%)	0 (0%)	0	0 (0%)	0 (0%)	0



**Table B-4:** Comparison of values computed in integrity check to reference article Table 1 values (n, mean, SD, median, quartiles, %) for ALGS and All Diagnoses groups

Variable	ALGS (n=100)			All Diagnoses (n=458)		
	Manuscript	DSIC	Diff	Manuscript	DSIC	Diff
Albumin, n	96	96	0	431	431	0
Mean (SD)	4.2 (0.5)	4.2 (0.5)	0	4.2 (0.5)	4.2 (0.5)	0
n (%) < 3.0	3 (3%)	3 (3%)	0	8 (2%)	8 (2%)	0
n (%) < 2.5	1 (1%)	1 (1%)	0	1 (0%)	1 (0%)	0
n (%) < 2.0	0 (0%)	0 (0%)	0	0 (0%)	0 (0%)	0
Platelet Count, n	95	95	0	439	439	0
Mean (SD)	262 (100)	262 (100)	0	214 (113)	214 (113)	0
n (%) < 150	12 (13%)	12 (13%)	0	140 (32%)	140 (32%)	0
n (%) < 100	5 (5%)	5 (5%)	0	95 (22%)	95 (22%)	0
n (%) < 50	1 (1%)	1 (1%)	0	22 (5%)	22 (5%)	0
APRI, n	95	95	0	431	431	0
Mean (SD)	2.2 (4.4)	2.2 (4.4)	0	1.7 (2.9)	1.7 (2.9)	0
Median (Q1, Q3)	1.2 (0.7, 2.6)	1.2 (0.7, 2.6)	0	0.8 (0.4, 1.9)	0.8 (0.4, 1.9)	0
n (%) > 1.0	54 (57%)	54 (57%)	0	181 (42%)	181 (42%)	0
n (%) > 1.5	35 (37%)	35 (37%)	0	134 (31%)	134 (31%)	0
n (%) > 2.0	27 (28%)	27 (28%)	0	103 (24%)	103 (24%)	0
PELD Score, n	83	83	0	351	351	0
Mean (SD)	-4.3 (8.4)	-4.3 (8.4)	0	-8.7 (6.9)	-8.7 (6.9)	0
Median (Q1, Q3)	-6.6 (-11.1, 0.8)	-6.6 (-11.1, 0.8)	0	-10.4 (-13.4, -5.2)	-10.4 (-13.4, -5.2)	0
n (%) > 0	24 (29%)	24 (29%)	0	40 (11%)	40 (11%)	0
n (%) > 10	6 (7%)	6 (7%)	0	7 (2%)	7 (2%)	0
n (%) > 20	0 (0%)	0 (0%)	0	0 (0%)	0 (0%)	0

## Attachment A: SAS Code

```
*****;  
*INPUT ;  
*****;  
libname data "Z:\NIDDK\niddk-dr_studies1\FORCE\private_created_data";  
  
*****;  
*FORMAT ;  
*****;  
proc format;  
  
    value group  
        1 = 'BA'  
        2 = 'A1ATD'  
        3 = 'ALGS'  
    ;  
  
    value groupall  
        1 = 'All diagnoses'  
    ;  
  
run;  
  
*****;  
*TABLE CHARACTERISTICS ;  
*****;  
  
*Define and flag diagnosis groups;  
data dsic_force;  
    set data.forcebl;  
        if forcedgnba=1 and validexambl=1 then diagnosis=1;  
        else if forcedgnalat=1 and validexambl=1 then diagnosis=2;  
        else if forcedgnalgs=1 and validexambl=1 then diagnosis=3;  
        if diagnosis=1 or diagnosis=2 or diagnosis=3 then diagnosis_all=1;  
        format diagnosis group. diagnosis_all groupall.;  
  
run;  
  
*Output table results by diagnosis groups;  
proc means data=dsic_force n mean std median q1 q3 maxdec=1;  
    var forcebili forcegtp forceast forcealt forcealbumin forceinr gpr forceplatelets apri peld;  
    class diagnosis;  
  
run;  
  
*Output table results by total;  
proc means data=dsic_force n mean std median q1 q3 maxdec=1;  
    var forcebili forcegtp forceast forcealt forcealbumin forceinr gpr forceplatelets apri peld;  
    class diagnosis_all;  
  
run;  
  
*****;  
*ALBUMIN BREAKDOWNS ;  
*****;  
  
*Define and flag albumin parameters;  
data albumin;  
    set dsic_force;  
        if forcealbumin<3.0 and forcealbumin ne . then albumin1=1;  
        if forcealbumin<2.5 and forcealbumin ne . then albumin2=1;  
        if forcealbumin<2.0 and forcealbumin ne . then albumin3=1;  
        label albumin1 = 'Albumin < 3.0'  
              albumin2 = 'Albumin < 2.5'  
              albumin3 = 'Albumin < 2.0';
```

```

run;

*Output table results by diagnosis groups;
proc freq data=albumin;
    tables albumin1*diagnosis;
    tables albumin2*diagnosis;
    tables albumin3*diagnosis;
run;

*****;
*PLATELET BREAKDOWNS ;
*****;

*Define and flag platelet parameters;
data platelet;
    set dsic_force;
    if forceplatelets>0 and forceplatelets<150 then platelet1=1;
    if forceplatelets>0 and forceplatelets<100 then platelet2=1;
    if forceplatelets>0 and forceplatelets<50 then platelet3=1;
    label platelet1 = 'Platelet < 150'
           platelet2 = 'Platelet < 100'
           platelet3 = 'Platelet < 50';
run;

*Output table results by diagnosis groups;
proc freq data=platelet;
    tables platelet1*diagnosis;
    tables platelet2*diagnosis;
    tables platelet3*diagnosis;
run;

*****;
*APRI BREAKDOWNS ;
*****;

*Define and flag apri parameters;
data apri;
    set dsic_force;
    if apri>1.0 and apri ne . then apri1=1;
    if apri>1.5 and apri ne . then apri2=1;
    if apri>2.0 and apri ne . then apri3=1;
    label apri1 = 'APRI > 1.0'
           apri2 = 'APRI > 1.5'
           apri3 = 'APRI > 2.0';
run;

*Output table results by diagnosis groups;
proc freq data=apri;
    tables apri1*diagnosis;
    tables apri2*diagnosis;
    tables apri3*diagnosis;
run;

*****;
*PELD BREAKDOWNS ;
*****;

*Define and flag peld parameters;
data peld;
    set dsic_force;
    if peld>0 and peld ne . then peld1=1;
    if peld>10 and peld ne . then peld2=1;
    if peld>20 and peld ne . then peld3=1;
    label peld1 = 'PELD > 0'
           peld2 = 'PELD > 10'

```

```
        peld3 = 'PELD > 20';  
run;  
  
*Output table results by diagnosis groups;  
proc freq data=peld;  
    tables peld1*diagnosis;  
    tables peld2*diagnosis;  
    tables peld3*diagnosis;  
run;
```