Dataset Integrity Check for A Prospective Database of Infants with Cholestasis/Longitudinal Study of Genetic Causes of Intrahepatic Cholestasis (PROBE/LOGIC) Teckman

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

This is a data analysis of Alpha-1-Antitrypsin deficiency as a prediction biomarker for portal hypertension, liver transplantation, and death among a subset of participants from the Prospective Database of Infants with Cholestasis (PROBE) and the Longitudinal Study of Genetic Causes of Intrahepatic Cholestasis (LOGIC).

PROBE
The PROBE study is a multi-center project to establish a prospective database of clinical information and a repository of blood and tissue samples from children with diagnoses of neonatal liver diseases, such as biliary atresia and neonatal hepatitis, in order to perform research on these liver problems. Children were screened and enrolled at presentation at the participating pediatric liver sites.

LOGIC
The LOGIC study investigates the natural history and progression of four rare genetic disorders, Alagille syndrome (ALGS), alpha-1 antitrypsin (a-1AT) deficiency, bile acid synthesis defects, and progressive familial intrahepatic cholestasis (PFIC) , responsible for approximately 20% to 30% of all infant cases of cholestasis. Cholestasis, a rare condition involving a reduction or obstruction of bile flow from the liver to the small intestine, can cause significant growth problems, liver complications, the need for liver transplantation, and death.

3 Archived Datasets

All data files, as provided by the Data Coordinating Center (DCC), are located in PROBE and LOGIC folder in the data packages. For this replication, variables were taken from the "a1atlongv4.sas7bdat" dataset.

4 Statistical Methods

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

5 Results

For Table 1 in the publication [1], Patient characteristics at study entry and during follow-up, Table A lists the variables that were used in the replication, and Tables B1 through B5 compare the results calculated from the archived data files to the results 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 PROBE and LOGIC data files to be distributed are a true copy of the study data.

7 References

[1] Teckman J, Rosenthal P, Hawthorne K, Spino C, Bass LM, Murray KF, Kerkar N, Magee JC, Karpen S, Heubi JE, Molleston JP, Squires RH, Kamath BM, Guthery SL, Loomes KM, Sherker AH, Sokol RJ. Longitudinal Outcomes in Young Patients with Alpha-1-Antitrypsin Deficiency with Native Liver Reveal that Neonatal Cholestasis is a Poor Predictor of Future Portal Hypertension. The Journal of Pediatrics, 227, 81-86.e4, December 2020. doi: https://doi.org/10.1016/j.jpeds.2020.07.031

Table A: Variables used to replicate Table 1 – Patient characteristics at study entry and during follow-up

Table Variable	dataset.variable
Visits per participant	a1atlongv4.visit
Phenotype	a1atlongv4.PiSZPheno
	a1atlongv4.PiZZPheno
Age (years)	a1atlongv4.AgeAtExam
Age group (years)	a1atlongv4.AgeAtExam
Ethnicity	a1atlongv4.ethnicity
Sex	a1atlongv4.Gender
Race	a1atlongv4.race

Table B1: Comparison of values computed in integrity check to reference article Table 1 (Baseline and Year 1)

Characteristics	Baseline Publication (n=350)	Baseline DSIC (n=350)	Diff. (n=0)	Year 1 Publication (n=229)	Year 1 DSIC (n=229)	Diff. (n=0)
Visits per participant						
Mean (SD)	3.55 (2.48)	3.55 (2.48)	0 (0)	-	-	-
Median (IQR)	3 (4)	3 (4)	0 (0)	-	-	-
Min, Max	1, 11	1, 11	0, 0	-	-	-
Phenotype						
SZ	32 (10.1)	33 (11.5)	1 (1.4)	-	-	-
ZZ	286 (89.9)	286 (84.9)	0 (5)	-	-	-
Age (years)						
Mean (SD)	6.05 (5.49)	6.05 (5.49)	0 (0)	6.82 (5.09)	6.82 (5.09)	0 (0)
Median (IQR)	4.23 (7.98)	4.23 (7.98)	0 (0)	5.59 (7.38)	5.59 (7.38)	0 (0)
Min, Max	0.12, 24.87	0.12, 24.87	0, 0	0.96, 22.48	0.96, 22.48	0, 0
Age group (years)						
0 - < 5	191 (54.6)	191 (54.6)	0 (0)	107 (46.7)	107 (46.7)	0 (0)
5 - < 10	83 (23.7)	83 (23.7)	0 (0)	70 (30.6)	70 (30.6)	0 (0)
10 - < 15	43 (12.3)	43 (12.3)	0 (0)	31 (13.5)	31 (13.5)	0 (0)
15 - < 20	28 (8)	28 (8)	0 (0)	17 (7.4)	17 (7.4)	0 (0)
20 - < 25	5 (1.4)	5 (1.4)	0 (0)	4 (1.7)	4 (1.7)	0 (0)
≥ 25	-	-	-	-	-	-
Ethnicity						
Hispanic	21 (6.1)	21 (6.1)	0 (0)	15 (6.6)	15 (6.6)	0 (0)
Non-Hispanic	324 (93.9)	324 (93.9)	0 (0)	212 (93.4)	212 (93.4)	0 (0)
Sex						
Male	211 (60.5)	211 (60.5)	0 (0)	144 (62.9)	144 (62.9)	0 (0)
Female	138 (39.5)	138 (39.5)	0 (0)	85 (37.1)	85 (37.1)	0 (0)
Race						
White	322 (93.6)	322 (93.6)	0 (0)	210 (92.5)	210 (92.5)	0 (0)
Black	3 (0.9)	3 (0.9)	0 (0)	3 (1.3)	3 (1.3)	0 (0)
Other	19 (5.5)	19 (5.5)	0 (0)	14 (6.2)	14 (6.2)	0 (0)

Table B2: Comparison of values computed in integrity check to reference article Table 1 (Year 2 and Year 3)

Characteristics	Year 2 Publication (n=179)	Year 2 DSIC (n=179)	Diff. (n=0)	Year 3 Publication (n=151)	Year 3 DSIC (n=151)	Diff. (n=0)
Age (years)						
Mean (SD)	7.31 (4.87)	7.31 (4.87)	0 (0)	8.91 (4.99)	8.91 (4.99)	0 (0)
Median (IQR)	5.93 (6.38)	5.93 (6.38)	0 (0)	7.7 (7.07)	7.7 (7.07)	0 (0)
Min, Max	1.77, 22.85	1.77, 22.85	0, 0	2.78, 23.9	2.78, 23.9	0, 0
Age group (years)						
0 - < 5	77 (43)	77 (43)	0 (0)	42 (27.8)	42 (27.8)	0 (0)
5 - < 10	60 (33.5)	60 (33.5)	0 (0)	56 (37.1)	56 (37.1)	0 (0)
10 - < 15	25 (14)	25 (14)	0 (0)	31 (20.5)	31 (20.5)	0 (0)
15 - < 20	13 (7.3)	13 (7.3)	0 (0)	17 (11.3)	17 (11.3)	0 (0)
20 - < 25	4 (2.2)	4 (2.2)	0 (0)	5 (3.3)	5 (3.3)	0 (0)
≥ 25	-	-	-	-	-	-
Ethnicity						
Hispanic	13 (7.3)	13 (7.3)	0 (0)	9 (6)	9 (6)	0 (0)
Non-Hispanic	165 (92)	165 (92.7)	0 (0.7)	141 (94)	141 (94)	0 (0)
Sex						
Male	112 (62.6)	112 (62.6)	0 (0)	86 (57)	86 (57)	0 (0)
Female	67 (37.4)	67 (37.4)	0 (0)	65 (43)	65 (43)	0 (0)
Race						
White	167 (93.3)	167 (93.3)	0 (0)	140 (92.7)	140 (92.7)	0 (0)
Black	1 (0.6)	1 (0.6)	0 (0)	1 (0.7)	1 (0.7)	0 (0)
Other	11 (6.1)	11 (6.1)	0 (0)	10 (6.6)	10 (6.6)	0 (0)

Table B3: Comparison of values computed in integrity check to reference article Table 1 (Year 4 and Year 5)

Characteristics	Year 4 Publication (n=102)	Year 4 DSIC (n=102)	Diff. (n=0)	Year 5 Publication (n=73)	Year 5 DSIC (n=73)	Diff. (n=0)
Age (years)						
Mean (SD)	9.83 (4.85)	9.83 (4.85)	0 (0)	9.99 (4.21)	9.99 (4.21)	0 (0)
Median (IQR)	8.64 (7.27)	8.64 (7.27)	0 (0)	9.04 (5.3)	9.04 (5.3)	0 (0)
Min, Max	4.05, 24.96	4.05, 24.96	0, 0	4.95, 25.97	4.95 25.97	0, 0
Age group (years)						
0 - <5	13 (12.7)	13 (12.7)	0 (0)	1 (1.4)	1 (1.4)	0 (0)
5 - <10	47 (46.1)	47 (46.1)	0 (0)	43 (58.9)	43 (58.9)	0 (0)
10 - <15	26 (25.5)	26 (25.5)	0 (0)	23 (31.5)	23 (31.5)	0 (0)
15 - <20	12 (11.8)	12 (11.8)	0 (0)	4 (5.5)	4 (5.5)	0 (0)
20 - <25	4 (3.9)	4 (3.9)	0 (0)	1 (1.4)	1 (1.4)	0 (0)
≥ 25	-	-	-	1 (1.4)	1 (1.4)	0 (0)
Ethnicity						
Hispanic	8 (7.8)	8 (7.8)	0 (0)	5 (6.8)	5 (6.8)	0 (0)
Non-Hispanic	94 (92.2)	94 (92.2)	0 (0)	68 (93.2)	68 (93.2)	0 (0)
Sex						
Male	61 (59.8)	61 (59.8)	0 (0)	39 (53.4)	39 (53.4)	0 (0)
Female	41 (40.2)	41 (40.2)	0 (0)	34 (46.6)	34 (46.6)	0 (0)
Race						
White	96 (94.1)	96 (94.1)	0 (0)	70 (95.9)	70 (95.9)	0 (0)
Black	-	-	-	-	-	-
Other	6 (5.9)	6 (5.9)	0 (0)	3 (4.1)	3 (4.1)	0 (0)

Table B4: Comparison of values computed in integrity check to reference article Table 1 (Year 6 and Year 7)

Characteristics	Year 6 Publication (n=56)	Year 6 DSIC (n=56)	Diff. (n=0)	Year 7 Publication (n=47)	Year 7 DSIC (n=47)	Diff. (n=0)
Age (years)						
Mean (SD)	10.86 (4.04)	10.86 (4.04)	0 (0)	11.52 (3.75)	11.52 (3.75)	0 (0)
Median (IQR)	9.94 (5.58)	9.94 (5.58)	0 (0)	10.72 (4.49)	10.72 (4.49)	0 (0)
Min, Max	6.06, 24.54	6.06, 24.54	0, 0	7.35, 25.79	7.35, 25.79	0, 0
Age group (years)						
0 - <5	-	-	-	-	-	-
5 - <10	28 (50)	28 (50)	0 (0)	20 (42.6)	20 (42.6)	0 (0)
10 - <15	21 (37.5)	21 (37.5)	0 (0)	20 (42.6)	20 (42.6)	0 (0)
15 - <20	5 (8.9)	5 (8.9)	0 (0)	5 (10.6)	5 (10.6)	0 (0)
20 - <25	2 (3.6)	2 (3.6)	0 (0)	1 (2.1)	1 (2.1)	0 (0)
≥ 25	-	-	-	1 (2.1)	1 (2.1)	0 (0)
Ethnicity						
Hispanic	3 (5.4)	3 (5.4)	0 (0)	4 (8.5)	4 (8.5)	0 (0)
Non-Hispanic	53 (94.6)	53 (94.6)	0 (0)	43 (91.5)	43 (91.5)	0 (0)
Sex						
Male	32 (57.1)	32 (57.1)	0 (0)	22 (46.8)	22 (46.8)	0 (0)
Female	24 (42.9)	24 (42.9)	0 (0)	25 (53.2)	25 (53.2)	0 (0)
Race						
White	52 (92.9)	52 (92.9)	0 (0)	43 (91.5)	43 (91.5)	0 (0)
Black	-	-	-	-	-	-
Other	4 (7.1)	4 (7.1)	0 (0)	4 (8.5)	4 (8.5)	0 (0)

Table B5: Comparison of values computed in integrity check to reference article Table 1 (Year 8 and Year 9)

Characteristics	Year 8 Publication (n=37)	Year 8 DSIC (n=37)	Diff. (n=0)	Year 9 Publication (n=18)	Year 9 DSIC (n=18)	Diff. (n=0)
Age (years)						
Mean (SD)	12.96 (3.97)	12.96 (3.97)	0 (0)	13.24 (4.24)	13.24 (4.24)	0 (0)
Median (IQR)	12.79 (4.76)	12.79 (4.76)	0 (0)	12.13 (4.25)	12.13 (4.25)	0 (0)
Min, Max	8.25, 26.88	8.25, 26.88	0, 0	9.77, 28.07	9.77, 28.07	0, 0
Age group (years)						
0 - <5	-	-	-	-	-	-
5 - <10	11 (29.7)	11 (29.7)	0 (0)	2 (11.1)	2 (11.1)	0 (0)
10 - <15	18 (48.6)	18 (48.6)	0 (0)	13 (72.2)	13 (72.2)	0 (0)
15 - <20	6 (16.2)	6 (16.2)	0 (0)	2 (11.1)	2 (11.1)	0 (0)
20 - <25	1 (2.7)	1 (2.7)	0 (0)	-	-	-
≥ 25	1 (2.7)	1 (2.7)	0 (0)	1 (5.6)	1 (5.6)	0 (0)
Ethnicity						
Hispanic	3 (8.1)	3 (8.1)	0 (0)	3 (16.7)	3 (16.7)	0 (0)
Non-Hispanic	34 (91.9)	34 (91.9)	0 (0)	15 (83.3)	15 (83.3)	0 (0)
Sex						
Male	20 (54.1)	20 (54.1)	0 (0)	11 (61.1)	11 (61.1)	0 (0)
Female	17 (45.9)	17 (45.9)	0 (0)	7 (38.9)	7 (38.9)	0 (0)
Race						
White	34 (91.9)	34 (91.9)	0 (0)	17 (94.4)	17 (94.4)	0 (0)
Black	-	-	-	-	-	-
Other	3 (8.1)	3 (8.1)	0 (0)	1 (5.6)	1 (5.6)	0 (0)

Attachment A: SAS Code

libname dsic "X:\NIDDK\niddk-dr_studies6\PROBE\private_orig_data\Teckman_A1AT Longitudinal";

```
/***************************/
/* DSIC for PROBE/LOGIC */
/* Publication Teckman et al */
/***************************/

*create temp dataset;
data dsic; set dsic.a1atlongv4;
run;

*count for each visit;
proc freq data=dsic;
tables visit;
run;

*visits per participant;
*excluding non-regular visits and transplant visit;
data dsic1; set dsic;
 if visit ^= 100 AND visit ^=-1;
run;

proc sort data=dsic1 out=visit_count;
by SubjectID;
run;

*counting the number of visits by SubjectID;
data count; set visit_count;
by subjectID;
retain N;
if first.SubjectID then N=1;
else N=N+1;
if Last.subjectID then output;
keep subjectID N;
run;

proc means data=count mean std median q1 q3 min max;
var N;
run;

*phenotype;
proc freq data=dsic;
tables PiSZPheno PiZZPheno;
where visit = 0;
run;
```

```sas
*Age in year;
*sorting by visit;
proc sort data=dsic1;
by visit;
run;

*descriptive stats for continuous age;
proc means data=dsic1 n mean std median q1 q3 min max;
var AgeAtExam;
by visit;
run;

*creating categorical age variable;
data dsic2; set dsic1;
if ageatexam < 5 then agecat = 1;
if ageatexam >= 5 AND ageatexam < 10 then agecat = 2;
if ageatexam >= 10 AND ageatexam < 15 then agecat = 3;
if ageatexam >= 15 AND ageatexam < 20 then agecat = 4;
if ageatexam >= 20 AND ageatexam < 25 then agecat = 5;
if ageatexam >= 25 then agecat = 6;
run;

*categorical age values;
proc freq data=dsic2;
tables agecat*visit/norow nopercent;
run;

*Ethnicity;
proc freq data=dsic2;
tables ethnicity*visit/norow nopercent;
where visit ^=10;
run;

*sex;
proc freq data=dsic2;
tables Gender*visit/norow nopercent;
where visit^=10;
run;

*Race;
proc freq data=dsic2;
tables race*visit/norow nopercent;
where visit^=10;
run;
```