Dataset Integrity Check for the Longitudinal Study of Genetic Causes of Intrahepatic Cholestasis (LOGIC) Shneider

Prepared by NIDDK-CR
October 17, 2022

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

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. The four rare genetic disorders, Alagille syndrome (ALGS), alpha-1 antitrypsin (AAT) deficiency, bile acid synthesis defects, and progressive familial intrahepatic cholestasis (PFIC), account for approximately 20% to 30% of all infant cases of cholestasis. Current knowledge concerning the etiology and outcomes of these diseases is limited. The Longitudinal Study of Genetic Causes of Intrahepatic Cholestasis (LOGIC) was established to investigate the natural history and progression of these four genetic disorders.

The LOGIC ALGS Natural History Controls study was constructed due to the difficulty in conducting long-term conventional randomized clinical trials in rare diseases. Criteria were pulled from a published trial of maralixibat in ALGS (The Evaluation of the Intestinal Bile Acid Transport Inhibitor LUM001 In the Reduction of Pruritus in Alagille Syndrome, a Cholestatic Liver Disease (ITCH) Study) and applied to a prospective longitudinal cohort of children from the LOGIC study in order to compile comparator data for clinical trials of intestinal bile acid transport inhibitors in ALGS.

3 Archived Datasets

All data files, as provided by the Data Coordinating Center (DCC), are located in the LOGIC folder in the data package. For this replication, variables were taken from the “algsnhcontrolsbl.sas7bdat” dataset.

4 Statistical Methods

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

5 Results

For Table 2 in the publication [1], Comparison of baseline parameters of ITCH participants and LOGIC ALGS natural history cohort, Table A lists the variables that were used in the replication, and Table B compares the results calculated from the archived data files to the results published in Table 2. The results of the replication are within expected variation to the published results.

6 Conclusions

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

7 References

[1] Shneider BL, Kamath BM, Magee JC, Goodrich NP, Loomes KM, Ye W, Spino C, Alonso EM, Molleston JP, Bezerra JA, Wang KS, Karpen SJ, Horslen SP, Guthery SL, Rosenthal P, Squires RH, Sokol RJ. Use of Funded Multicenter Prospective Longitudinal Databases to Inform Clinical Trials in Rare Diseases – Examination of Cholestatic Liver Disease in Alagille Syndrome. Hepatology Communications, 6(8), 1910-1921, August 2022. doi: https://doi.org/10.1002/hep4.1970

Table A: Variables used to replicate Table 2 – Comparison of baseline parameters of ITCH participants and LOGIC ALGS natural history cohort

Table Variable	**dataset.variable**
Age, years	algsnhcontrolsbl.ageatbaseline
Age, category groups	algsnhcontrolsbl.agegrp
Gender	algsnhcontrolsbl.gender
Race	algsnhcontrolsbl.race
Ethnicity	algsnhcontrolsbl.ethnicity
CSS, range 0-4	algsnhcontrolsbl.css
PesQL parent total score, range 0-100	algsnhcontrolsbl.pedsqltotalparent
Total bilirubin (mg/dL)	algsnhcontrolsbl.bilitotalmgdl
Cholesterol (mg/dL)	algsnhcontrolsbl.cholesterolmgdl
Serum bile acid (umol/L)	algsnhcontrolsbl.serumbileacidsumoll
ALT (IU/L)	algsnhcontrolsbl.altunitsl
GGT (IU/L)	algsnhcontrolsbl.ggtpunitsl
Albumin (g/dL)	algsnhcontrolsbl.albumingdl
Platelets ($10^3/mm^3$)	algsnhcontrolsbl.plateletscnt
Height z-score	algsnhcontrolsbl.heightzscore
Weight z-score	algsnhcontrolsbl.weightzscore

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

Characteristic **n (%) or mean (SD), median**	**Publication: LOGIC ALGS Natural History Cohort (unweighted) (n=59)**	**DSIC: LOGIC ALGS Natural History Cohort (unweighted) (n=59)**	**Diff. (n=0)**
Age, years	4.1 (3.2), 2.5	4.1 (3.2), 2.5	0 (0), 0
Age			
< 2 years	26 (44)	26 (44)	0 (0)
2-4 years	13 (22)	13 (22)	0 (0)
5-7 years	10 (17)	10 (17)	0 (0)
8-18 years	10 (17)	10 (17)	0 (0)
Gender			
Female	25 (42)	25 (42)	0 (0)
Male	33 (56)	33 (56)	0 (0)
Not reported	1 (2)	1 (2)	0 (0)
Race			
White	30 (51)	30 (51)	0 (0)
Black	13 (22)	13 (22)	0 (0)
Non-White, Non-Black	9 (15)	9 (15)	0 (0)
Not Reported	7 (12)	7 (12)	0 (0)
Ethnicity			
Hispanic	9 (15)	9 (15)	0 (0)
Non-Hispanic	48 (81)	48 (81)	0 (0)
Not reported	2 (3)	2 (3)	0 (0)
CSS, range 0-4	2.9 (0.7), 3.0	2.9 (0.7), 3.0	0 (0), 0
PesQL parent total score, range 0-100	78.8 (12.86), 84.7	78.8 (12.86), 84.7	0 (0), 0
Total bilirubin (mg/dL)	4.9 (5.42), 2.3	4.9 (5.42), 2.3	0 (0), 0
Cholesterol (mg/dL)	522.4 (342.6), 454	522.4 (342.6), 454	0 (0), 0
Serum bile acid (umol/L)	174.2 (9.0), 179.2	174.2 (79), 179.2	0 (70), 0
ALT (IU/L)	196.7 (99.5), 187	196.7 (99.5), 187	0 (0), 0
GGT (IU/L)	542.1 (460.4), 388	542.1 (460.4), 388	0 (0), 0
Albumin (g/dL)	4.1 (0.4), 4.1	4.1 (0.4), 4.1	0 (0), 0
Platelets (10^3/mm^3)	308.0 (110.4), 308	308.0 (110.4), 308	0 (0), 0
Height z-score	-2.3 (1.3), -2.3	-2.3 (1.3), -2.3	0 (0), 0
Weight z-score	-1.7 (0.9), -1.7	-1.7 (0.9), -1.7	0 (0), 0

Attachment A: SAS Code

```
libname algs "X:\NIDDK\niddk-
dr_studies2\LOGIC\private_orig_data\Shneider_ALGSNaturalHistoryControls\DataAndFiles";
libname library "X:\NIDDK\niddk-
dr_studies2\LOGIC\private_orig_data\Shneider_ALGSNaturalHistoryControls\DataAndFiles";

/*********************************/
/* LOGIC ALGS Nat Hist */
/* Controls DSIC Shneider et al. */
/*********************************/

proc contents data=algs.algsnhcontrolsbl;
run;

data dsic; set algs.algsnhcontrolsbl;
run;

*age;
proc means data=dsic n mean std median;
var AgeAtBaseline;
run;

*age cat., gender, race, and ethnicity;
proc freq data=dsic;
tables AgeGrp Gender Race Ethnicity;
run;

*CSS, PesQL, total bilirubin, cholesterol, serum bile acid, alt, ggt, albumin, platelets, height and weight z
scores;
proc means data=dsic n mean std median;
var css PedsQLTotalParent BiliTotalMgdl CholesterolMgdl SerumBileAcidsUmolL ALTUnitsL GGTPUnitsL
AlbuminGdl PlateletsCnt
HeightZScore WeightZScore;
run;
```