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

Prepared by NIDDK-CR
September 16, 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 (a-1AT) 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 main objective of the LOGIC PFIC study was to advance the understanding of monogenic forms of intrahepatic cholestasis.

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 “pfic.sas7bdat” dataset.

4 Statistical Methods

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

5 Results

For Table 1 in the publication [1], Native liver participants with history of sEHC at least 1 year before enrollment, 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 1. The results of the replication are an exact match to the published results.

6 Conclusions

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

7 References

[1] Hertel PM, Bull LN, Thompson RJ, Goodrich NP, Ye W, Magee JC, Squires RH, Bass LM, Heubi JE, Kim GE, Ranganathan S, Schwarz KB, Bozic MA, Horslen SP, Clifton MS, Turmelle YP, Suchy FJ, Superina RA, Wang KS, Loomes KM, Kamath BM, Sokol RJ, Shneider BL. Mutation Analysis and Disease Features at Presentation in a Multi-Center Cohort of Children With Monogenic Cholestasis. Journal of Pediatric Gastroenterology and Nutrition, 73(2), 169-177, August 2021. doi: https://doi.org/10.1097/MPG.0000000000003153

Table A: Variables used to replicate Table 1 – Native liver participants with history of sEHC at least 1 year before enrollment

Table Variable	**dataset.variable**
Participants enrolled pre-transplant	pfic.logicgroup pfic.geneticdgn
Participants with drainage procedure performed at least 1 year before baseline	pfic.preblgt1yrselectdrain pfic.geneticdgn
Time (years) between drainage procedure and baseline	pfic.daystoselectdrain pfic.geneticdgn pfic.preblgt1yrselectdrain
Total bilirubin (mg/dL)	pfic.bilitotalmgdl pfic.geneticdgn pfic.preblgt1yrselectdrain
Pruritus	pfic.pruritus pfic.geneticdgn pfic.preblgt1yrselectdrain
Platelet count ($10^3/mm^3$)	pfic.plateletscnt pfic.geneticdgn pfic.preblgt1yrselectdrain
Successful surgical interruption of enterohepatic circulation (sEHC)	pfic.successfuldrain pfic.geneticdgn pfic.preblgt1yrselectdrain

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

n (%) or mean (SD)	Publication: FIC1	DSIC: FIC1	Diff.	Publication: BSEP	DSIC: BSEP	Diff.	Publication: Total	DSIC: Total	Diff.
Participants enrolled pre-transplant	24	24	0	34	34	0	58	58	0
Participants with drainage procedure performed at least 1 year before baseline	10 (42%)	10 (42%)	0 (0)	9 (26%)	9 (26%)	0 (0)	19 (33%)	19 (33%)	0 (0)
Time (years) between drainage procedure and baseline	7.2 (4.2)	7.2 (4.2)	0 (0)	7.2 (4.6)	7.2 (4.6)	0 (0)	7.2 (4.3)	7.2 (4.3)	0 (0)
Total bilirubin (mg/dL)									
≤ 1	7 (70%)	7 (70%)	0 (0)	4 (44%)	4 (44%)	0 (0)	11 (58%)	11 (58%)	0 (0)
> 1	3 (30%)	3 (30%)	0 (0)	5 (56%)	5 (56%)	0 (0)	8 (42%)	8 (42%)	0 (0)
Pruritus									
None or mild	5 (50%)	5 (50%)	0 (0)	5 (56%)	5 (56%)	0 (0)	10 (53%)	10 (53%)	0 (0)
Active scratching or cutaneous mutilation	5 (50%)	5 (50%)	0 (0)	4 (44%)	4 (44%)	0 (0)	9 (47%)	9 (47%)	0 (0)
Platelet count ($10^3/mm^3$)									
< 150	2 (20%)	2 (20%)	0 (0)	3 (33%)	3 (33%)	0 (0)	5 (26%)	5 (26%)	0 (0)
≥ 150	6 (60%)	6 (60%)	0 (0)	4 (44%)	4 (44%)	0 (0)	10 (53%)	10 (53%)	0 (0)
Missing	2 (20%)	2 (20%)	0 (0)	2 (22%)	2 (22%)	0 (0)	4 (21%)	4 (21%)	0 (0)
Successful sEHC									
No	6 (60%)	6 (60%)	0 (0)	6 (67%)	6 (67%)	0 (0)	12 (63%)	12 (63%)	0 (0)
Yes	4 (40%)	4 (40%)	0 (0)	3 (33%)	3 (33%)	0 (0)	7 (37%)	7 (37%)	0 (0)

Attachment A: SAS Code

```
libname dsic "X:\NIDDK\niddk-
dr_studies2\LOGIC\private_orig_data\Shneider_Hertel_PFIC\DataAndFiles";

/*****************************/
/* DSIC for LOGIC PFIC */
/* Hertel et al. pub */
/*****************************/

*temp dataset;
data pfic; set dsic.pfic;
run;

*limiting data to the specific participants for publication;
data pfic_1; set pfic;
where GeneticDgn ^= 3;
if logicgroup ^= 3;
run;

data pfic_2; set pfic_1;
if logicgroup = 1 OR logicgroup = 2 then loggrp = 1;
run;

*Participants enrolled pre-transplant;
proc freq data=pfic_2;
tables Loggrp*GeneticDgn;
run;

*participants with drainage procedure performed prior to baseline;
proc freq data=pfic_2;
tables PreBLGt1YrSelectDrain*GeneticDgn/norow;
run;

*Time (years) between drainage procedure and baseline;
data pfic_3; set pfic_2;
/*if preblandtxpselectdrain = 1 then daystoselectdrain = daystoselectdrain*-1; */
yearsToSelectDrain = (daysToSelectDrain/365.25);
run;

proc means data=pfic_3 n mean std;
var yearsToSelectDrain;
/*class GeneticDgn; */
where PreBLGt1YrSelectDrain = 1;
run;

*total bilirubin;
```

```
data pfic_4; set pfic_3;
if bilitotalmgdl <= 1 then bilicat = 0;
if bilitotalmgdl > 1 then bilicat = 1;
run;

proc freq data=pfic_4;
tables bilicat*GeneticDgn/norow;
where PreBLGt1YrSelectDrain = 1;
run;

*pruritis;
data pfic_5; set pfic_4;
prurcat = 0;
if pruritus = 1 OR pruritus = 2 then prurcat = 1;
if pruritus >= 3 then prurcat = 2;
run;

proc freq data=pfic_5;
tables prurcat*GeneticDgn/norow;
where PreBLGt1YrSelectDrain = 1;
run;

*platelet count;
data pfic_6; set pfic_5;
platecat = .;
if plateletscnt > 0 and plateletscnt < 150 then platecat = 0;
if plateletscnt >= 150 then platecat = 1;
run;

proc freq data=pfic_6;
tables platecat*GeneticDgn/norow missing;
where PreBLGt1YrSelectDrain = 1;
run;

*Successful sEHC;
proc freq data=pfic_6;
tables SuccessfulDrain*GeneticDgn/norow missing;
where PreBLGt1YrSelectDrain = 1;
run;
```