Dataset Integrity Check for A Prospective Database of Infants with Cholestasis (PROBE) Hertel

Prepared by NIDDK-CR
September 1, 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

This 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.

3 Archived Datasets

All data files, as provided by the Data Coordinating Center (DCC), are located in PROBE folder in the data package. For this replication, variables were taken from the “inh_01jul19.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], Baseline characteristics of idiopathic cholestasis participants by clinical outcome, Table A lists the variables that were used in the replication, and Tables B1 and B2 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 data files to be distributed are a true copy of the study data.

7 References

[1] Hertel PM, Hawthorne K, Kim S, Finegold MJ, Shneider BL, Squires JE, Gupta NA, Bull LN, Murray KF, Kerkar N, Ng VL, Molleston JP, Bezerra JA, Loomes KM, Taylor SA, Schwarz KB, Turmelle YP, Rosenthal P, Magee JC, Sokol RJ. Presentation and Outcomes of Infants With Idiopathic Cholestasis: A Multicenter Prospective Study. Journal of Pediatric Gastroenterology and Nutrition, 73(4), 478-484, October 2021. doi: https://doi.org/10.1097/MPG.0000000000003248

Table A: Variables used to replicate Table 1 – Baseline characteristics of idiopathic cholestasis participants by clinical outcome

Table Variable	**dataset.variable**
Sex (male)	inh_01jul19.sex inh_01jul19.outcome
Race	inh_01jul19.race inh_01jul19.outcome
Ethnicity (Non-Hispanic)	inh_01jul19.ethnicity inh_01jul19.outcome
Age at enrollment (months)	inh_01jul19.age_consent_m inh_01jul19.outcome
Gestational age, weeks	inh_01jul19.gestagevalue inh_01jul19.outcome
Gestational age (< 37 weeks)	inh_01jul19.gestagevalue inh_01jul19.outcome
Birth weight (< 2500 grams)	inh_01jul19.birthweightoz inh_01jul19.outcome
White or pale stools by parent report	inh_01jul19.palestools inh_01jul19.outcome
Age at onset of white or pale stools (days)	inh_01jul19.palestoolagedays inh_01jul19.outcome
Stool color on physical examination	inh_01jul19.stoolcolor inh_01jul19.outcome
Total bilirubin (mg/dL)	inh_01jul19.totalbilirubinmgdl inh_01jul19.outcome
ALT, U/L	inh_01jul19.altunitsl inh_01jul19.outcome
GGT, U/L	inh_01jul19.ggtpunitsl inh_01jul19.outcome
GGT (< 100 U/L)	inh_01jul19.ggtpunitsl inh_01jul19.outcome
Platelets, x 10^3/µL	inh_01jul19.plateletscnt inh_01jul19.outcome
Labs at last follow-up	inh_01jul19.totalbilirubinmgdl inh_01jul19.altunitsl inh_01jul19.ggtpunitsl inh_01jul19.plateletscnt inh_01jul19.outcome

Table B1: Comparison of values computed in integrity check to reference article Table 1 (Total and Biochemical resolution)

Variable n (%) or median (IQR)	Total Publication (n=94)	Total DSIC (n=92)	Diff. (n=2)	Biochemical resolution Publication (n=51)	Biochemical resolution DSIC (n=51)	Diff. (n=0)
Sex (male)	66 (70)	65 (71)	1 (1)	35 (69)	35 (69)	0 (0)
Race						
Black	12 (13)	12 (13)	0 (0)	7 (14)	7 (14)	0 (0)
Other	20 (22)	18 (20)	2 (2)	8 (16)	8 (16)	0 (0)
White	59 (65)	59 (66)	0 (1)	34 (69)	34 (69)	0 (0)
Ethnicity (Non-Hispanic)	71 (76)	70 (77)	1 (1)	40 (78)	40 (78)	0 (0)
Age at enrollment (months)	1.5 (1.0-2.2)	1.5 (1.0-2.2)	0 (0-0)	1.5 (1.0-2.0)	1.5 (1.0-2.0)	0 (0-0)
Gestational age, weeks	38 (37-39)	38 (37-39)	0 (0-0)	38 (37-39)	38 (37-39)	0 (0-0)
Gestational age (< 37 weeks)	22 (24)	25 (27)	3 (3)	10 (20)	11 (22)	1 (2)
Birth weight (< 2500 grams)	25 (28)	29 (32)	4 (4)	14 (29)	17 (33)	3 (4)
White or pale stools by parent report	25 (28)	25 (29)	0 (1)	14 (30)	14 (30)	0 (0)
Age at onset of white or pale stools (days)	9 (4-28)	9 (4-28)	0 (0-0)	7 (4-21)	7 (4-21)	0 (0-0)
Stool color on physical examination						
Normal (yellow, brown, green)	66 (72)	65 (72)	1 (0)	38 (75)	38 (75)	0 (0)
Pale (less color than normal)	16 (17)	15 (17)	1 (0)	9 (18)	9 (18)	0 (0)
White or gray (acholic)	10 (11)	10 (11)	0 (0)	4 (8)	4 (8)	0 (0)
Total bilirubin (mg/dL)	6.8 (4.9-10.9)	6.8 (4.9-10.9)	0 (0-0)	6.5 (4.2-10.3)	6.5 (4.2-10.3)	0 (0-0)
ALT, U/L	104 (59-208)	104 (59-208)	0 (0-0)	95 (59-167)	95 (59-167)	0 (0-0)
GGT, U/L	124 (72-285)	124 (72-285)	0 (0-0)	127 (72-284)	127 (72-284)	0 (0-0)
GGT (< 100 U/L)	34 (41)	43 (47)	9 (6)	19 (40)	22 (43)	3 (3)
Platelets, x 10^3/μL	344 (245-493)	344 (245-493)	0 (0-0)	328 (220-458)	328 (220-458)	0 (0-0)
At last follow-up						
Total bilirubin, mg/dL	0.4 (0.2-0.7)	0.4 (0.2-0.7)	0 (0-0)	0.3 (0.2-0.5)	0.3 (0.2-0.5)	0 (0-0)
ALT, U/L	32 (26-46)	32 (26-46)	0 (0-0)	27 (21-31)	27 (21-31)	0 (0-0)
GGT, U/L	19 (13-60)	18 (13-60)	1 (0-0)	16 (11-22)	16 (11-22)	0 (0-0)
GGT (< 100 U/L)	62 (87)	85 (90)	23 (3)	42 (95)	49 (96)	7 (1)
Platelets, x 10^3/μL	345 (257-419)	345 (260-419)	0 (3-0)	337 (257-446)	337 (257-446)	0 (0-0)
Age at clinical outcome (months)	11 (7-13)	-	-	9 (6-13)	-	-

Table B2: Comparison of values computed in integrity check to reference article Table 1 (Partial resolution and Exited healthy)

Variable	Partial resolution Publication (n=7)	Partial resolution DSIC (n=7)	Diff. (n=0)	Exited Healthy Publication (n=34)	Exited Healthy DSIC (n=32)	Diff. (n=2)
Sex (male)	6 (86)	6 (86)	0 (0)	24 (71)	23 (72)	1 (1)
Race						
Black	0 (0)	0 (0)	0 (0)	5 (15)	5 (16)	0 (1)
Other	2 (29)	2 (29)	0 (0)	9 (27)	7 (23)	2 (4)
White	5 (71)	5 (71)	0 (0)	19 (58)	19 (61)	0 (3)
Ethnicity (Non-Hispanic)	4 (57)	4 (57)	0 (0)	25 (76)	24 (77)	1 (1)
Age at enrollment (months)	1.7 (0.5-2.4)	1.7 (0.5-2.4)	0 (0-0)	1.7 (0.9-2.2)	1.8 (1.0-2.2)	0.1 (0.1-0)
Gestational age, weeks	39 (37-40)	39 (37-40)	0 (0-0)	38 (35-39)	38 (34-39)	0 (1-0)
Gestational age (< 37 weeks)	1 (17)	2 (29)	1 (12)	10 (31)	11 (34)	1 (3)
Birth weight (< 2500 grams)	1 (14)	1 (14)	0 (0)	9 (28)	10 (31)	1 (3)
White or pale stools by parent report	0 (0)	0 (0)	0 (0)	10 (30)	10 (32)	0 (2)
Age at onset of white or pale stools (days)	–	-	-	9 (3-28)	9 (3-28)	0 (0-0)
Stool color on physical examination						
Normal (yellow, brown, green)	6 (100)	6 (100)	0 (0)	22 (67)	21 (68)	1 (1)
Pale (less color than normal)	0 (0)	0 (0)	0 (0)	7 (21)	6 (19)	1 (2)
White or gray (acholic)	0 (0)	0 (0)	0 (0)	4 (12)	4 (13)	0 (1)
Total bilirubin (mg/dL)	7.4 (5.3-10.2)	7.4 (5.3-10.2)	0 (0-0)	7.0 (5.2-13.2)	7.0 (5.2-13.2)	0 (0-0)
ALT, U/L	181 (26-246)	181 (26-246)	0 (0-0)	120 (59-235)	120 (59-235)	0 (0-0)
GGT, U/L	310 (73-518)	310 (73-518)	0 (0-0)	121 (68-211)	121 (68-211)	0 (0-0)
GGT (< 100 U/L)	3 (43)	3 (43)	0 (0)	12 (44)	17 (53)	5 (9)
Platelets, x 10^3/µL	396 (239-527)	396 (239-527)	0 (0-0)	425 (326-560)	425 (326-560)	0 (0-0)
At last follow-up						
Total bilirubin, mg/dL	0.4 (0.2-0.5)	0.4 (0.2-0.5)	0 (0-0)	0.5 (0.3-1.2)	0.5 (0.3-1.2)	0 (0-0)
ALT, U/L	54 (46-73)	49 (43-116)	5 (3-43)	46 (38-77)	46 (38-77)	0 (0-0)
GGT, U/L	58 (19-129)	45 (17-129)	13 (2-0)	61 (17-93)	61 (17-93)	0 (0-0)
GGT (< 100 U/L)	5 (71)	5 (71)	0 (0)	15 (79)	30 (88)	15 (9)
Platelets, x 10^3/µL	237 (179-311)	277 (179-311)	40 (0-0)	403 (326-566)	403 (326-566)	0 (0-0)
Age at clinical outcome (months)	24 (24-27)	-	-	12 (8-13)	-	-

Attachment A: SAS Code

```
libname dsic "X:\NIDDK\niddk-dr_studies6\PROBE\private_orig_data\Hertel_PROBE_IHH";

/*****************************/
/* PROBE Analysis dataset */
/* DSIC (Hertel et al.) */
/*****************************/

*creating temp dataset;
data dsic; set dsic.inh_01jul19;
where visit = "Baseline";
run;

proc freq data=dsic;
tables outcome;
run;

*sex;
proc freq data=dsic;
tables sex*outcome/norow;
run;

*Race;
proc freq data=dsic;
tables race*outcome/norow;
run;

*ethnicity;
proc freq data=dsic;
tables Ethnicity*outcome/norow;
run;

*age at enrollment;
proc sort data=dsic;
by outcome;
run;

proc means data=dsic n median q1 q3;
var age_consent_m;
run;

proc means data=dsic n median q1 q3;
var age_consent_m;
by outcome;
run;
```

```
*gestational age;
proc means data=dsic n median q1 q3;
var GestAgeValue;
run;

proc means data=dsic n median q1 q3;
var GestAgeValue;
by outcome;
run;

*gestational age (<37);
data dsic1; set dsic;
if gestagevalue <= 36 then gestagecat = 1; else gestagecat = 0;
run;

proc freq data=dsic1 ;
tables gestagecat*outcome/norow;
run;

*birth weight;
data dsic2; set dsic;
bwg = birthweightoz * 28.3495;
run;

data dsic3; set dsic2;
if bwg <2500 then bwcat = 1; else bwcat = 0;
run;

proc freq data=dsic3;
tables bwcat*outcome/norow;
run;

*pale stools;
proc freq data=dsic;
tables PaleStools*outcome/norow;
run;

*pale stool age of onset;
proc means data=dsic n median q1 q3;
var PaleStoolAgeDays;
run;

proc means data=dsic n median q1 q3;
var PaleStoolAgeDays;
by outcome;
run;

*stool color on exam ;
```

```
proc freq data=dsic;
tables StoolColor*outcome/norow;
run;

*Total bilirubin;
proc means data=dsic n median q1 q3;
var TotalBilirubinMgdl;
run;

proc means data=dsic n median q1 q3;
var TotalBilirubinMgdl;
by outcome;
run;

*ALT;
proc means data=dsic n median q1 q3;
var ALTUnitsL;
run;

proc means data=dsic n median q1 q3;
var ALTUnitsL;
by outcome;
run;

*GGT);
proc means data=dsic n median q1 q3;
var GGTPUnitsL;
run;

proc means data=dsic n median q1 q3 ;
var GGTPUnitsL;
by outcome;
run;

*GGT <100;
data dsic4; set dsic;
if ggtpunitsl <=99 then ggtcat = 1; else ggtcat = 0;
run;

proc freq data=dsic4;
tables ggtcat*outcome/norow;
run;

*Platelets;
proc means data=dsic n median q1 q3;
var PlateletsCnt;
run;
```

```
proc means data=dsic n median q1 q3;
var PlateletsCnt;
by outcome;
run;

*at last follow-up;
data all; set dsic.inh_01jul19;
run;

proc sort data=all;
by ChiLDReN_StudyID;
run;

data last_visit; set all;
by ChiLDReN_StudyID;
if last.children_studyid;
run;

*bilirubin;
proc sort data=last_visit;
by outcome;
run;

proc means data=last_visit n median q1 q3;
var TotalBilirubinMgdl;
run;

proc means data=last_visit n median q1 q3;
var TotalBilirubinMgdl;
by outcome;
run;

*ALT;
proc means data=last_visit n median q1 q3;
var ALTUnitsL;
run ;

proc means data=last_visit n median q1 q3;
var ALTUnitsL ;
by outcome;
run;

*GGT;
proc means data=last_visit n median q1 q3;
var GGTPUnitsL;
run;

proc means data=last_visit n median q1 q3;
```

```
var GGTPUnitsL;
by outcome;
run;

*GGT <100;
data last_visit2; set last_visit;
if ggtpunitsl <=99 then ggtcat = 1; else ggtcat = 0;
run;

proc freq data=last_visit2;
tables ggtcat*outcome/norow;
run;

*Platelets;
proc means data=last_visit n median q1 q3;
var PlateletsCnt;
run;

proc means data=last_visit n median q1 q3;
var PlateletsCnt;
by outcome;
run;
```