Dataset Integrity Check for Biliary Atresia Study in Infants and Children/A Prospective Database of Infants with Cholestasis (BASIC/PROBE) Venkat

Prepared by NIDDK-CR
September 2, 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 is a data analysis to develop prognostic models for survival with native liver and predictors of liver disease progression, despite liver transplant, among a subset of participants from the Biliary Atresia Study in Infants and Children (BASIC) and Prospective Database of Infants with Cholestasis (PROBE) study.

BASIC

The Biliary Atresia Study in Infants and Children (BASIC) is a prospective, observational study to collect pertinent clinical information and biospecimens to aid in the understanding of the disorder. Specific aims of the study include identifying the gene(s) implicated in the etiology of biliary atresia, identifying the polymorphisms that may influence disease progression, and characterizing the natural history of the older, non-transplanted patients with biliary atresia.

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.

3 Archived Datasets

All data files, as provided by the Data Coordinating Center (DCC), are located in the BASIC and PROBE folder in the data packages. For this replication, variables were taken from the "manuscript_data_25jun19.sas7bdat" dataset.

4 Statistical Methods

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

5 Results

For Table 1 in the publication [1], Demographics and baseline characteristics of participants with BA SNL at age 2 years, 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 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 BASIC/PROBE data files to be distributed are a true copy of the study data.

7 References

[1] Venkat V, Ng VL, Magee JC, Ye W, Hawthorne K, Harpavat S, Molleston JP, Murray KF, Wang KS, Soufi N, Bass LM, Alonso EM, Bezerra JA, Jensen MK, Kamath BM, Loomes KM, Mack CL, Rosenthal P, Shneider BL, Squires RH, Sokol RJ, Karpen SJ. Modeling Outcomes in Children With Biliary Atresia With Native Liver After 2 Years of Age. Hepatology Communications, 4(12), 1824-1834, December 2020. doi: https://doi.org/10.1002/hep4.1602

Table A: Variables used to replicate Table 1 – Demographics and baseline characteristics of participants with BA SNL at age 2 years

Table Variable	**dataset.variable**
Female	manuscript_data_25jun19.sex manuscript_data_25jun19.study
Race	manuscript_data_25jun19.race manuscript_data_25jun19.study
Hispanic ethnicity	manuscript_data_25jun19.hispanic manuscript_data_25jun19.study
Age at Kasai (days)	manuscript_data_25jun19.age_kasai_d manuscript_data_25jun19.study
Associated anomalies	manuscript_data_25jun19.asplenia_polysplenia manuscript_data_25jun19.HeartAnomaly manuscript_data_25jun19.GIAnomaly manuscript_data_25jun19.study
Clinical features	manuscript_data_25jun19.BleedHx manuscript_data_25jun19.CholangitisHx manuscript_data_25jun19.ascitehx manuscript_data_25jun19.spleenhx manuscript_data_25jun19.weightfailure manuscript_data_25jun19.heightfailure manuscript_data_25jun19.antibiotics manuscript_data_25jun19.urso manuscript_data_25jun19.study
Laboratory features	manuscript_data_25jun19.totalbilirubinmgdl manuscript_data_25jun19.ggtpunitsl manuscript_data_25jun19.plateletscnt manuscript_data_25jun19.astunitsl manuscript_data_25jun19.altunitsl manuscript_data_25jun19.albumingdl manuscript_data_25jun19.inr manuscript_data_25jun19.study

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

Variable	BASIC Publication (n=44)	BASIC DSIC (n=44)	Diff. (n=0)	PROBE Publication (n=196)	PROBE DSIC (n=196)	Diff. (n=0)	Total Publication (n=240)	Total DSIC (n=240)	Diff. (n=0)
Female	28 (63.6)	28 (63.6)	0 (0)	104 (53.1)	104 (53.1)	0 (0)	132 (55.0)	132 (55.0)	0 (0)
Race									
White	21 (48.8)	21 (48.8)	0 (0)	108 (55.4)	108 (55.4)	0 (0)	129 (54.2)	129 (54.2)	0 (0)
Black	7 (16.3)	7 (16.3)	0 (0)	29 (14.9)	29 (14.9)	0 (0)	36 (15.1)	36 (15.1)	0 (0)
Other	15 (34.9)	15 (34.9)	0 (0)	58 (29.7)	58 (29.7)	0 (0)	73 (30.7)	73 (30.7)	0 (0)
Hispanic ethnicity	5 (11.9)	5 (11.9)	0 (0)	48 (24.5)	48 (24.5)	0 (0)	53 (22.3)	53 (22.3)	0 (0)
Age at Kasai (days)	61 (40, 76)	61 (40, 76)	0 (0,0)	57 (43, 74)	57 (43, 74)	0 (0,0)	59 (42, 74)	59 (42, 74)	0 (0,0)
Associated anomalies									
Asplenia/polysplenia	1 (2.7)	1 (2.7)	0 (0)	8 (4.1)	8 (4.1)	0 (0)	9 (3.9)	9 (3.9)	0 (0)
Cardiovascular anomaly	5 (13.5)	5 (13.5)	0 (0)	30 (15.4)	30 (15.4)	0 (0)	35 (15.1)	35 (15.1)	0 (0)
Gastrointestinal anomaly	8 (21.6)	8 (21.6)	0 (0)	18 (9.2)	18 (9.2)	0 (0)	26 (11.2)	26 (11.2)	0 (0)
Clinical features									
History of GI bleed	1 (2.3)	1 (2.3)	0 (0)	10 (5.1)	10 (5.1)	0 (0)	11 (4.6)	11 (4.6)	0 (0)
History of cholangitis	23 (52.3)	23 (52.3)	0 (0)	81 (41.3)	81 (41.3)	0 (0)	104 (43.3)	104 (43.3)	0 (0)
History of ascites	8 (18.2)	8 (18.2)	0 (0)	34 (17.3)	34 (17.3)	0 (0)	42 (17.5)	42 (17.5)	0 (0)
History of splenomegaly	21 (47.7)	21 (47.7)	0 (0)	117 (59.7)	117 (59.7)	0 (0)	138 (57.5)	138 (57.5)	0 (0)
Weight growth failure	4 (11.4)	4 (11.4)	0 (0)	6 (3.4)	6 (3.4)	0 (0)	10 (4.7)	10 (4.7)	0 (0)
Height growth failure	0 (0)	0 (0)	0 (0)	13 (7.3)	13 (7.3)	0 (0)	13 (6.1)	13 (6.1)	0 (0)
Antibiotics	11 (40.7)	11 (40.7)	0 (0)	58 (29.9)	58 (29.9)	0 (0)	69 (31.2)	69 (31.2)	0 (0)
Ursodeoxycholic acid	17 (63.0)	17 (63.0)	0 (0)	123 (65.1)	123 (65.1)	0 (0)	140 (64.8)	140 (64.8)	0 (0)
Laboratory features									
TB (mg/dL)	0.5 (0.3, 1.3)	0.5 (0.3, 1.3)	0 (0,0)	0.5 (0.3, 0.9)	0.5 (0.3, 0.9)	0 (0,0)	0.5 (0.3, 0.9)	0.5 (0.3, 0.9)	0 (0,0)
GGT (U/L)	127 (50, 239)	127 (50, 239)	0 (0,0)	134 (43, 332)	134 (43, 332)	0 (0,0)	130 (43, 327)	130 (43, 327)	0 (0,0)
Platelet count (x 10^3/µL)	189 (132, 246)	189 (132, 246)	0 (0,0)	206 (134, 280)	206 (134, 280)	0 (0,0)	205 (134, 271)	205 (134, 271)	0 (0,0)
APRI	1.1 (0.7, 2.7)	1.1 (0.7, 2.7)	0 (0,0)	1.2 (0.6, 2.1)	1.2 (0.6, 2.1)	0 (0,0)	1.2 (0.7, 2.2)	1.2 (0.7, 2.2)	0 (0,0)
AST (U/L)	80 (56, 138)	80 (56, 138)	0 (0,0)	88 (56, 159)	88 (56, 159)	0 (0,0)	84 (56, 154)	84 (56, 154)	0 (0,0)
ALT (U/L)	65 (39 , 117)	65 (39, 117)	0 (0,0)	85 (45, 163)	85 (45, 163)	0 (0,0)	83 (42, 155)	83 (42, 155)	0 (0,0)
Albumin (g/dL)	4.1 (3.8, 4.4)	4.1 (3.8, 4.4)	0 (0,0)	4.1 (3.8, 4.4)	4.1 (3.8, 4.4)	0 (0,0)	4.1 (3.8, 4.4)	4.1 (3.8, 4.4)	0 (0,0)
INR	1.0 (1.0, 1.2)	1.0 (1.0, 1.2)	0 (0,0)	1.0 (1.0, 1.1)	1.0 (1.0, 1.1)	0 (0,0)	1.0 (1.0, 1.1)	1.0 (1.0, 1.1)	0 (0,0)

Attachment A: SAS Code

```
libname dsic
"X:\NIDDK\new_data_packages_received\ChiLDREN\private_orig_data\Venkat_BA_Natural History";

/***********************************/
/* BASIC/PROBE Pub. Venkat et al. */
/* Modeling Outcomes in Children... */
/***********************************/

*create temp dataset;
data dsic; set dsic.manuscript_data_25jun19;
run;

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

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

*Hispanci ethnicity;
proc freq data=dsic;
tables hispanic*study/norow;
run;

*age at kasai;
proc sort data=dsic;
by study;
run;

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

proc means data=dsic n median q1 q3;
var age_kasai_d;
by study;
run;

*Associated anomolies;
proc freq data=dsic;
tables (asplenia_polysplenia HeartAnomaly GIAnomaly)*study/norow;
run;
```

```
*Clinical features;
proc freq data=dsic;
tables (BleedHx CholangitisHx asciteshx SpleenHx WeightFailure HeightFailure Antibiotics
Urso)*study/norow;
run;

*Laboratory features;
*create APRI variable;
data dsic1; set dsic;
apri = ((Astunitsl/40)*100)/(plateletscnt);
run;

proc means data=dsic1 n median q1 q3;
var TotalBilirubinMgdl GGTPUnitsL PlateletsCnt apri ASTUnitsL ALTUnitsL AlbuminGdl inr;
by study;
run;

proc means data=dsic1 n median q1 q3;
var TotalBilirubinMgdl GGTPUnitsL PlateletsCnt apri ASTUnitsL ALTUnitsL AlbuminGdl inr;
run;
```