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

Prepared by NIDDK-CR
May 2, 2024

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

A subset of the study, A Prospective Database of Infants with Cholestasis (PROBE), was used in assessing whether histologic features in the liver have prognostic value in participants with biliary atresia.

PROBE

The PROBE study was a multi-center project to establish a prospective database of clinical information and a repository of blood and tissue specimens 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

A full listing of archived datasets included in the package can be found in the Roadmap document. All data files, as provided by the Data Coordinating Center (DCC), are located in the PROBE folder in the data package. For this replication, variables were taken from the “histopath_manuscript.sas7bdat” dataset.

4 Statistical Methods

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

5 Results

For Table 1 in the publication [1], Clinical and demographic features, 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 within expected variation to the published results.

6 Conclusions

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

7 References

[1] Russo P, Magee JC, Anders RA, Bove KE, Chung C, Cummings OW, Finegold MJ, Finn LS, Kim GE, Lovell MA, Magid MS, Melin-Aldana H, Ranganathan S, Shehata BM, Wang LL, White FV, Chen Z, Spino C. Key Histopathologic Features of Liver Biopsies That Distinguish Biliary Atresia From Other Causes of Infantile Cholestasis and Their Correlation With Outcome: A Multicenter Study. The American Journal of Surgical Pathology, 40(12), 1601-1615, December 2016. doi: https://doi.org/10.1097/PAS.0000000000000755

Table A: Variables used to replicate Table 1 – Clinical and demographic features

Table Variable	**dataset.variable**
Analysis group variable	histopath_manuscript.analysis_group
BA cases	histopath_manuscript.ba
Non-BA cases	histopath_manuscript.ba
Needle biopsies	histopath_manuscript.biopsy_type
Wedge biopsies	histopath_manuscript.biopsy_type
Sex	histopath_manuscript.sex
Race	histopath_manuscript.race_cat
Median age at biopsy	histopath_manuscript.age_d
Median age at HPE	histopath_manuscript.age_k
Extrahepatic anomalies	histopath_manuscript.extrahepatic_ano
Ohi subtype at HPE	histopath_manuscript.ohi
Observational study only	histopath_manuscript.trt
START study placebo	histopath_manuscript.trt
START study treatment	histopath_manuscript.trt
Transplant	histopath_manuscript.tx_flag
No transplant	histopath_manuscript.tx_flag
Death	histopath_manuscript.death_pretx
Median follow-up	histopath_manuscript.time_to_event

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

Characteristic	Pub: Correlation of Needle Biopsies With Clinical Diagnosis (n=227)	DSIC: Correlation of Needle Biopsies With Clinical Diagnosis (n=227)	Diff. (n=0)	Pub: Correlation of Liver Biopsy Features in BA With Clinical Outcome (n=316)	DSIC: Correlation of Liver Biopsy Features in BA With Clinical Outcome (n=316)	Diff. (n=0)
BA cases (n)	136	136	0	316	316	0
Non-BA cases (n)	91	91	0	--	--	--
Needle biopsies (n)	227	227	0	110	110	0
Wedge biopsies (n)	--	--	--	206	206	0
Sex (%)	Male (58)	Male (59)	(1)	Female (50.3)	Female (50.3)	(0)
Race (n[%])						
White	125 (55)	125 (55)	0 (0)	180 (57)	180 (57)	0 (0)
Black	27 (12)	27 (12)	0 (0)	38 (12)	38 (12)	0 (0)
Other or missing	75 (33)	75 (33)	0 (0)	98 (31)	98 (31)	0 (0)
Median age at biopsy (days, SD)	58 (29)	58 (29)	0 (0)	--	--	--
Median age at HPE (days, SD)	--	--	--	63 (25)	61 (29)	2 (4)
Extrahepatic anomalies (BA cases only) (n[%])						
None	113 (83.0)	123 (90.4)	10 (7.4)	280 (88.6)	280 (88.6)	0 (0)
Laterality defect (BASM)	13 (9.5)	9 (6.6)	4 (2.9)	27 (8.5)	27 (8.5)	0 (0)
Non-laterality defect	9 (6.6)	4 (2.9)	5 (3.7)	9 (2.8)	9 (2.8)	0 (0)
Ohi subtype at HPE (BA cases only) (n[%])						
Type I	12 (8.8)	12 (9.0)	0 (0.2)	30 (9.5)	30 (9.5)	0 (0)
Type II	11 (8.0)	11 (8.0)	0 (0)	23 (7.3)	23 (7.3)	0 (0)
Type III	111 (81.6)	111 (82.8)	0 (1.2)	260 (82.3)	260 (83.1)	0 (0.8)
Missing	2 (1.4)	0 (0)	2 (1.4)	3 (0.9)	0 (0)	3 (0.9)
Observational study only (n)	--	--	--	191	192	1
START study placebo (n)	--	--	--	62	62	0
START study treatment (n)	--	--	--	62	62	0
Transplant (n[%])	--	--	--	129 (40.8)	129 (40.8)	0 (0)
No transplant (n[%])	--	--	--	187 (59.2)	187 (59.2)	0 (0)
Death (n[%])	--	--	--	15 (4.7)	15 (4.7)	0 (0)
Median follow-up (years)	--	--	--	3.9	3.9	0

Attachment A: SAS Code

```
libname hist "X:\NIDDK\niddk-dr_studies6\PROBE\private_created_data\Russo\ChiLDReN PROBE Histopath Dataset 201710\Zip File\3. Analysis Data set for primary study publications";

/*************************/
/* PROBE Russo DSIC */
/*************************/

*creating a two flag variables for the analysis groups;
data one; set hist.histopath_manuscript;
if analysis_group = 0 OR analysis_group = 1 then analflag1 = 1;
if analysis_group = 0 OR analysis_group = 2 then analflag2 = 1;
run;

*BA cases and Non-BA cases;
proc freq data=one;
tables ba*(analflag1 analflag2)/norow nopercent;
run;

*Needle/Wedge biopsies;
proc freq data=one;
tables biopsy_type*analflag2/norow nopercent;
run;

*Sex;
proc freq data=one;
tables male*(analflag1 analflag2)/norow nopercent;
run;

*race;
proc freq data=one;
tables (race_cat)*(analflag1 analflag2)/norow nopercent;
run;

*median age at biopsy;
proc means data=one n median std q1 q3;
var age_d;
class analflag1;
run;

*median age at HPE;
```

```
proc means data=one n median q1 q3;
var age_k;
class analflag2;
run;

*Extrahepatic anomalies;
proc freq data=one;
tables extrahepatic_ano*(analflag1 analflag2)/norow nopercent;
where ba = 1;
run;

*ohi subtype;
proc freq data=one;
tables ohi*(analflag1 analflag2)/norow nopercent;
where ba = 1;
run;

*study type;
proc freq data=one;
tables trt*analflag2/norow nopercent;
run;

*transplant;
proc freq data=one;
tables tx_flag*analflag2/norow nopercent;
run;

*death;
proc freq data=one;
tables death_pretx*analflag2/norow nopercent;
run;

*median follow-up;
data two; set one;
fuyr = time_to_event/365.25;
run;

proc means data=two n median q1 q3 std;
var fuyr;
class analflag2;
where censor = 1;
run;
```