

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

## Contents

1 Standard Disclaimer .....	2
2 Study Background .....	2
3 Archived Datasets .....	2
4 Statistical Methods .....	3
5 Results .....	3
6 Conclusions .....	3
7 References .....	3
Table A: Variables used to replicate Table 1 – Baseline characteristics .....	4
Table B: Comparison of values computed in integrity check to reference article Table 1 .....	5
Attachment A: SAS Code .....	6

## 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 for risk factors, incidence, and outcomes of variceal hemorrhage 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 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 “ba\_vh\_18aug22.sas7bdat” dataset.

## 4 Statistical Methods

Analyses were performed to replicate results for the data in the publication by Bass 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, 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] Bass LM, Ye W, Hawthorne K, Leung DH, Murray KF, Molleston JP, Romero R, Karpen S, Rosenthal P, Loomes KM, Wang KS, Squires RH, Miethke A, Ng VL, Horslen S, Jensen MK, Sokol RJ, Magee JC, Shneider BL. Risk of Variceal Hemorrhage and Pretransplant Mortality in Children with Biliary Atresia. *Hepatology*, 76(3), 712-726, September 2022. doi: <https://doi.org/10.1002/hep.32451>

**Table A:** Variables used to replicate Table 1 – Baseline characteristics

<b>Table Variable</b>	<b>dataset.variable</b>
Age at baseline visit (years)	ba_vh_18aug22.basic ba_vh_18aug22.age_bsl_y
Age at baseline visit (months)	ba_vh_18aug22.basic ba_vh_18aug22.age_bsl_m
Age at HPE (days)	ba_vh_18aug22.basic ba_vh_18aug22.kasaiage
Sex	ba_vh_18aug22.basic ba_vh_18aug22.sex
Race	ba_vh_18aug22.basic ba_vh_18aug22.race
PHT features	ba_vh_18aug22.basic ba_vh_18aug22.spleenplate ba_vh_18aug22.spleenplate2
History of VH before baseline	ba_vh_18aug22.basic ba_vh_18aug22.bleedhx
Height z-score	ba_vh_18aug22.basic ba_vh_18aug22.haz
Weight z-score	ba_vh_18aug22.basic ba_vh_18aug22.waz
Spleen size (cm below costal margin)	ba_vh_18aug22.basic ba_vh_18aug22.spleensize
Platelet count (10 <sup>3</sup> /mm <sup>3</sup> )	ba_vh_18aug22.basic ba_vh_18aug22.plateletsct
AST (U/L)	ba_vh_18aug22.basic ba_vh_18aug22.astunitsl
ALT (U/L)	ba_vh_18aug22.basic ba_vh_18aug22.altunitsl
AST/ALT	ba_vh_18aug22.basic ba_vh_18aug22.astunitsl ba_vh_18aug22.altunitsl
GGT (U/L)	ba_vh_18aug22.basic ba_vh_18aug22.ggtpunitsl
Alkaline phosphatase (U/L)	ba_vh_18aug22.basic ba_vh_18aug22.alkphosphunitsl
Total bilirubin (mg/dL)	ba_vh_18aug22.basic ba_vh_18aug22.totalbilirubinmdgl
Functioning HPE	ba_vh_18aug22.basic ba_vh_18aug22.shpe
INR	ba_vh_18aug22.basic ba_vh_18aug22.inr
Albumin (g/dL)	ba_vh_18aug22.basic ba_vh_18aug22.albumingdl
APRI	ba_vh_18aug22.basic ba_vh_18aug22.astunitsl ba_vh_18aug22.plateletsct

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

Variable	PROBE Publication (n=521)	PROBE DSIC (n=521)	Diff. (n=0)	BASIC Publication (n=348)	BASIC DSIC (n=348)	Diff. (n=0)
Age at baseline visit (years)	0.4 (0.4, 0.5)	0.4 (0.4, 0.5)	0 (0, 0)	8.5 (5.5, 12.9)	8.5 (5.5, 12.9)	0 (0, 0)
Age at baseline visit (months)	5.0 (4.4, 5.5)	5.0 (4.4, 5.5)	0 (0, 0)	-	-	-
Age at HPE (days)	62 (45, 74)	62 (45, 74)	0 (0, 0)	56 (42, 73)	56 (42, 73)	0 (0, 0)
Sex (Female)	284 (54.5)	284 (54.5)	0 (0)	189 (54.3)	189 (54.3)	0 (0)
Race						
Black	71 (14.0)	71 (14.0)	0 (0)	40 (11.6)	40 (11.6)	0 (0)
Non-Black, Non-White	140 (27.6)	140 (27.6)	0 (0)	74 (21.4)	74 (21.4)	0 (0)
White	296 (58.4)	296 (58.4)	0 (0)	231 (67.0)	231 (67.0)	0 (0)
PHT features						
0 features	312 (59.9)	312 (59.9)	0 (0)	133 (38.2)	133 (38.2)	0 (0)
1 feature	176 (33.8)	176 (33.8)	0 (0)	94 (27.0)	94 (27.0)	0 (0)
Splénomegaly only	153 (29.4)	153 (29.4)	0 (0)	48 (13.8)	48 (13.8)	0 (0)
Thrombocytopenia only	23 (4.4)	23 (4.4)	0 (0)	46 (13.2)	46 (13.2)	0 (0)
2 features	33 (6.3)	33 (6.3)	0 (0)	121 (34.8)	121 (34.8)	0 (0)
History of VH before baseline	5 (1.0)	5 (1.0)	0 (0)	11 (3.2)	11 (3.2)	0 (0)
Height z-score	-1.01 (-1.72, -0.23)	-1.01 (-1.72, -0.23)	0 (0, 0)	-0.04 (-0.77, 0.70)	-0.04 (-0.77, 0.70)	0 (0, 0)
Weight z-score	-1.28 (-1.95, -0.53)	-1.28 (-1.95, -0.53)	0 (0, 0)	0.37 (-0.36, 0.98)	0.37 (-0.36, 0.98)	0 (0, 0)
Spleen size (cm below costal margin)	2.0 (0.0, 3.0)	2.0 (0.0, 3.0)	0 (0, 0)	3.0 (0.0, 6.0)	3.0 (0.0, 6.0)	0 (0, 0)
Platelet count (10 <sup>3</sup> /mm <sup>3</sup> )	262 (188, 347)	262 (188, 347)	0 (0, 0)	129 (79, 217)	129 (79, 217)	0 (0, 0)
AST (U/L)	148 (102, 211)	148 (102, 211)	0 (0, 0)	65 (43, 117)	65 (43, 117)	0 (0, 0)
ALT (U/L)	112 (76, 173)	112 (76, 173)	0 (0, 0)	63 (35, 108)	63 (35, 108)	0 (0, 0)
AST/ALT	1.32 (1.03, 1.66)	1.32 (1.03, 1.66)	0 (0, 0)	1.12 (0.84, 1.42)	1.12 (0.84, 1.42)	0 (0, 0)
GGT (U/L)	776 (337, 1344)	776 (337, 1344)	0 (0, 0)	76 (33, 188)	76 (33, 188)	0 (0, 0)
Alkaline phosphatase (U/L)	533 (390, 729)	533 (390, 729)	0 (0, 0)	303 (208, 449)	303 (208, 449)	0 (0, 0)
Total bilirubin (mg/dL)	2.5 (0.7, 9.0)	2.5 (0.7, 9.0)	0 (0, 0)	0.7 (0.5, 1.3)	0.7 (0.5, 1.3)	0 (0, 0)
Functioning HPE	262 (51.0)	262 (51.0)	0 (0, 0)	-	-	-
INR	1.1 (1.0, 1.3)	1.1 (1.0, 1.3)	0 (0, 0)	1.1 (1.0, 1.2)	1.1 (1.0, 1.2)	0 (0, 0)
Albumin (g/dL)	3.7 (3.2, 4.0)	3.7 (3.2, 4.0)	0 (0, 0)	4.2 (3.7, 4.5)	4.2 (3.7, 4.5)	0 (0, 0)
APRI	1.4 (0.9, 2.4)	1.4 (0.9, 2.4)	0 (0, 0)	1.4 (0.7, 2.7)	1.4 (0.7, 2.7)	0 (0, 0)

## Attachment A: SAS Code

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

```
/******  
/* PROBE/BASIC DSIC for */  
/* Bass et al. Pub */  
/******
```

```
*creating temp dataset;  
data dsic; set dsic.ba_vh_18aug22;  
run;
```

```
*total from each study;  
proc freq data=dsic;  
tables basic;  
run;
```

```
/*age at baseline by study*/  
*years;  
proc sort data=dsic;  
by basic;  
run;
```

```
proc means data=dsic n median q1 q3;  
var age_bsl_y;  
by basic;  
run;
```

```
*months;  
proc means data=dsic n median q1 q3;  
var age_bsl_m;  
by basic;  
run;
```

```
*at HPE;  
proc means data=dsic n median q1 q3;  
var KasaiAge;  
by basic;  
run;
```

```
/* Sex: Female */  
proc freq data=dsic;  
tables sex*basic/norow nopercnt;  
run;
```

```
/* Race */
```

```

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

/* PHT Features */
proc freq data=dsic;
tables (spleenplate spleenplate2)*basic/norow nopercnt;
run;

/* History of VH */
proc freq data=dsic;
tables bleedhx*basic/norow nopercnt;
run;

/* height z score */
proc means data=dsic n median q1 q3;
var haz;
by basic;
run;

/* weight z score */
proc means data=dsic n median q1 q3;
var waz;
by basic;
run;

/*spleen size */
proc means data=dsic n median q1 q3;
var SpleenSize;
by basic;
run;

/* Platelet count */
proc means data=dsic n median q1 q3;
var PlateletsCnt;
by basic;
run.;

/* AST */;
proc means data=dsic n median q1 q3;
var ASTUnitsL;
by basic;
run;

/* ALT */
proc means data=dsic n median q1 q3;
var ALTUnitsL;
by basic;

```

```

run;

/*AST/ALT*/
data dsic1; set dsic;
ast_alt = ASTUnitsL/ALTUnitsL;
run;

proc means data=dsic1 n median q1 q3;
var ast_alt;
by basic;
run;

/* GGT */
proc means data=dsic n median q1 q3;
var GGTPUnitsL;
by basic;
run;

/* alkaline phosphatase */
proc means data=dsic n median q1 q3;
var AlkPhosphUnitsL;
by basic;
run;

/* total bilirubin */
proc means data=dsic n median q1 q3;
var TotalBilirubinMgdl;
by basic;
run;

/* functioning hpe */
proc freq data=dsic;
tables sHPE*basic/ norow nopercnt;
run;

/* INR */
proc means data=dsic n median q1 q3;
var inr;
by basic;
run;

/* Albumin */
proc means data=dsic n median q1 q3;
var AlbuminGdl;
by basic;
run;

/* APRI */

```

```
data dsic3; set dsic;  
apri = ((ASTUnitsL/40)*100)/PlateletsCnt;  
run;
```

```
proc means data=dsic3 n median q1 q3;  
var apri;  
by basic;  
run;
```