

# Dataset Integrity Check for PROBE Anatomic Pattern of BA (Superina)

**Prepared by Dominick Parisi**

**IMS Inc.**

3901 Calverton Blvd, Suite 200 Calverton, MD 20705

**November 26, 2018**

## Contents

1 Standard Disclaimer .....	2
2 Study Background .....	2
3 Archived Datasets .....	2
4 Statistical Methods .....	2
5 Results .....	3
6 Conclusions .....	3
7 References .....	3
Table A: Variables used to replicate Table 1: Baseline Demographics of 244 BA Subjects Who Underwent Surgical Drainage .....	4
Table B: Comparison of values computed in integrity check to reference article Table 1 values .....	4
Table C: Variables used to replicate Table 2: Operative Details for 244 BA Subjects Who Underwent Surgical Drainage .....	5
Table D: Comparison of values computed in integrity check to reference article Table 2 values.....	6
Attachment A: SAS Code .....	7

## **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**

PROBE 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 in these liver problems. Children were screened and enrolled at presentation at the participating pediatric liver sites. Subjects diagnosed with biliary atresia were followed intensively for the first year, at 18 months of age, and then annually up to 15 years of age. Other subjects diagnosed with cholestasis were followed on the same schedule; if there was complete (clinical and biochemical) resolution of their underlying liver disease off all therapy, there was one follow up visit within one year (preferably scheduled at the time of the next planned follow up visit or at 12 months of age, whichever was later) for data collection and to obtain blood samples. The development of a serum and tissue bank of specimens from children with various neonatal cholestatic disorders are used for future investigations into the etiology and pathogenesis of hepatobiliary injury in the infant.

## **3 Archived Datasets**

All the SAS data files, as provided by the Data Coordinating Center (DCC), are located in the PROBE folder in the “PROBE START Dataset for Superina ” data package. For this replication, variables were taken from the “surgery\_dataset.sas7bdat” dataset.

## **4 Statistical Methods**

Analyses were performed to duplicate results for the data published by Riccardo Superina et al [1] in Annals of Surgery 2011. To verify the integrity of the dataset, descriptive statistics were computed.

## 5 Results

For Table 1 in the publication [1], **Baseline Demographics of 244 BA Subjects Who Underwent Surgical Drainage**, 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 almost an exact match to the published results.

For Table 2 in the publication [1], **Operative Details for 244 BA Subjects Who Underwent Surgical Drainage**, Table C lists the variables that were used in the replication and Table D compares the results calculated from the archived data files to the results published in Table 2. The results of the replication are an exact match to the published results.

## 6 Conclusions

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

## 7 References

[1] Riccardo Superina, MD, John C. Magee, MD, Mary L. Brandt, MD, Patrick J. Healey, MD, Greg Tiao, MD, Fred Ryckman, MD, Frederick M. Karrer, MD, Kishore Iyer, MD, Annie Fecteau, MD, Karen West, MD, R. Cartland Burns, MD, Alan Flake, MD, Hanmin Lee, MD, Jeff A. Lowell, MD, Pat Dillon, MD, Paul Colombani, MD, Richard Ricketts, MD, Yun Li, PhD, Jeffrey Moore, MS, Kasper S. Wang, MD, and the Childhood Liver Disease Research and Education Network. *Annals of Surgery* Volume 254, Number 4, October 2011; 577-585.

**Table A:** Variables used to replicate Table 1: Baseline Demographics of 244 BA Subjects Who Underwent Surgical Drainage

Table Variable	dataset.variable
Gender	surgery_dataset.gender_n
Race	surgery_dataset.race_n
Ethnicity	surgery_dataset.ethnicity_n
Age at first evaluation (days)	surgery_dataset.gest_age

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

Variable	PROBE Manuscript N (%)	PROBE DSIC N (%)	Diff. (n=0)
Gender			
- Male	113 (46.3)	113 (46.3)	0 (0)
- Female	131 (53.7)	131 (53.7)	0 (0)
Race			
- White	146 (59.8)	148 (60.7)	2 (0.9)
- Black/African American	38 (15.6)	38 (15.6)	0 (0)
- American Indian/Alaska Native	2 (0.8)	2 (0.8)	0 (0)
- Asian	18 (7.4)	18 (7.4)	0 (0)
- Native Hawaiian/Pacific Islander	3 (1.2)	3 (1.2)	0 (0)
- Other	29 (11.9)	27 (11.1)	2 (0.8)
- Unknown	3 (1.2)	3 (1.2)	0 (0)
- Refused to answer	5 (2.0)	5 (2.0)	0 (0)
Ethnicity			
- Hispanic	55 (22.5)	54 (22.1)	1 (0.4)
- Non-Hispanic	188 (77.0)	189 (77.5)	1 (0.5)
- Unknown	1 (0.4)	1 (0.4)	0 (0)
Gestational age, Weeks; mean $\pm$ SD (N)	38.1 $\pm$ 2.32 (233)	38.1 $\pm$ 2.31 (233)	0 $\pm$ 0.1 (0)
Gestational age categorical			
- $\leq$ 37 Weeks	65 (26.6)	65 (26.6)	0 (0)
- $>$ 37 Weeks	168 (68.9)	168 (68.9)	0 (0)

**Table C:** Variables used to replicate Table 2: Operative Details for 244 BA Subjects Who Underwent Surgical Drainage

<b>Table Variable</b>	<b>dataset.variable</b>
Age at Surgery	surgery_dataset.age_at_kasai
Liver appearance	surgery_dataset.liver_appearance
Ascites	surgery_dataset.ascites
Ascites volume	surgery_dataset.ascites_volume
Common bile duct inflamed	surgery_dataset.cbd_inflamed
No abnormality identified	surgery_dataset.abdominal_abnormality
Intestinal malrotation	surgery_dataset.intestinal_malrotation
Situs inversus	surgery_dataset.situs_inversus
Midline liver	surgery_dataset.midline_liver
Polysplenia	surgery_dataset.polysplenia
Asplenia	surgery_dataset.asplenia
Preduodenal portal vein	surgery_dataset.preduodenal_portal_vein
Other abdominal anatomy abnormalities	surgery_dataset.other_abnormality
Left-to-right (LR) dissection	surgery_dataset.lr_dissection
Anterior-to-posterior (AP) dissection	surgery_dataset.ap_dissection

Table D: Comparison of values computed in integrity check to reference article Table 2 values

Variable	PROBE Manuscript (n=244) N (%)	PROBE DISC (n=244) N (%)	Diff. (n=0)
Age at surgery, Days; Mean ± SD (N)	64.6 ± 25.1 (244)	64.9 ± 28.3 (243)	0.3 ± 3.2 (1)
Age at surgery adjusted for gestational age, Days; Mean ± SD (N)	50.8 ± 27.2 (233)	NOT CHECKED	
Liver Appearance			
- Firm	170 (69.7)	166 (68.0)	4 (1.7)
- Nodular	49 (20.1)	49 (20.1)	0 (0)
- Normal	24 (9.8)	23 (9.4)	1 (0.4)
Ascites			
- No ascites or ≤ 20 cc	205 (81.0)	197 (83.5)	8 (2.5)
- Ascites > 20 cc	39 (19.0)	39 (16.5)	0 (2.5)
Common bile duct inflamed			
- Yes	77 (31.6)	74 (30.3)	3 (1.3)
- No	143 (58.6)	142 (58.2)	1 (0.4)
Abdominal anatomy abnormalities			
- No abnormality identified	118 (48.4)	120 (49.2)	2 (0.8)
- Intestinal malrotation	23 (9.4)	22 (9.0)	1 (0.4)
- Situs inversus	7 (2.9)	7 (2.9)	0 (0)
- Midline liver	9 (3.7)	9 (3.7)	0 (0)
- Polysplenia	13 (5.3)	13 (5.3)	0 (0)
- Asplenia	1 (0.4)	1 (0.4)	0 (0)
- Preduodenal portal vein	10 (4.1)	10 (4.1)	0 (0)
- Other	23 (9.4)	22 (9.0)	1 (0.4)
Hilar dissection			
- Left-to-right (LR) dissection, mm; Mean ± SD (N)	18.0 ± 12.8 (182)	17.9 ± 12.8 (220)	0.1 ± 0 (38)
- Anterior-to-posterior (AP) dissection, mm; Mean ± SD (N)	9.9 ± 8.5 (182)	9.9 ± 8.4 (220)	0 ± 0.1 (38)
- Total dissection area (LR x AP) mm <sup>2</sup> , Mean ± SD (N)	285.9 ± 1026.6 (182)	272.4 ± 1019.7 (220)	13.5 ± 6.9 (38)

## Attachment A: SAS Code

```

/*****/
/* Library statements */
/*****/
LIBNAME SASDATA '/prj/niddk/ims_analysis/PROBE/private_orig_data/PROBE_START Dataset for Superina/';

/*****/
/* Import datasets */
/*****/
DATA SURGERY_WDRAIN;
  SET SASDATA.SURGERY_DATASET (RENAME=(lr_dissection=lr_dissection_char ap_dissection=ap_dissection_char));
  WHERE SURGERY='Surgery W/Drain';

  IF .<GEST_AGE<=37 THEN GEST_AGE_CAT=1;
  ELSE IF GEST_AGE>37 THEN GEST_AGE_CAT=2;
  ELSE GEST_AGE_CAT=.;

  IF lr_dissection_char IN('ND','DK','UNK') THEN lr_dissection=.;
  ELSE lr_dissection=input(lr_dissection_char,8.);

  IF ap_dissection_char IN('ND','DK','UNK') THEN ap_dissection=.;
  ELSE ap_dissection=input(ap_dissection_char,8.);

  IF ascites_volume='UNK' THEN ascites_vol=.;
  ELSE ascites_vol=INPUT(ascites_volume,8.);

  total_dissection=lr_dissection*ap_dissection;

  if ohi_subtype_a^='' or ohi_subtype_b^='' or ohi_subtype_c^='' or ohi_subtype_d^='' then do;
    if ohi_subtype_a='' then ohi_subtype_a='0';
    if ohi_subtype_b='' then ohi_subtype_b='0';
    if ohi_subtype_c='' then ohi_subtype_c='0';
    if ohi_subtype_d='' then ohi_subtype_d='0';
  end;
RUN;

/*****/
/* Table 1 */
/*****/
TITLE2 'Table 1';
PROC FREQ DATA=SURGERY_WDRAIN;
  TABLE GENDER_N RACE_N ETHNICITY_N;
```



```

RUN;

PROC MEANS DATA=SURGERY_WDRAIN N NMISS MEAN STD MIN MAX;
  VAR GEST_AGE;
RUN;

PROC FREQ DATA=SURGERY_WDRAIN;
  TABLE GEST_AGE_CAT /LIST MISSING;
RUN;

/*****/
/* Table 2 */
/*****/
TITLE2 'Table 2';
PROC MEANS DATA=SURGERY_WDRAIN N NMISS MEAN STD MIN MAX;
  VAR AGE_AT_KASAI;
RUN;

PROC FREQ DATA=SURGERY_WDRAIN;
  TABLE liver_appearance /list missing;
RUN;

PROC FREQ DATA=SURGERY_WDRAIN;
  TABLE ascites * ascites_vol /LIST MISSING;
  WHERE (ASCITES='1') OR (ASCITES='2' AND ASCITES_VOL^=.);
RUN;

PROC FREQ DATA=SURGERY_WDRAIN;
  TABLE cbd_inflamed
    abdominal_abnormality
    intestinal_malrotation
    situs_inversus
    midline_liver
    polysplenia
    asplenia
    preduodenal_portal_vein
    other_abnormality /LIST MISSING;
RUN;

PROC MEANS DATA=SURGERY_WDRAIN N NMISS MEAN STD MIN MAX;
  VAR lr_dissection
    ap_dissection
    total_dissection;
RUN;

```