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

Prepared by Allyson Mateja
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton, MD 20705
February 15, 2017

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

Optimizing outcome in biliary atresia (BA) requires timely diagnosis. Cholestasis is a presenting feature of BA, as well as other diagnoses (Non-BA). Identification of clinical features of neonatal cholestasis that would expedite decisions to pursue subsequent invasive testing to correctly diagnose or exclude BA would enhance outcomes. The analytical goal was to develop a predictive model for BA using data available at initial presentation. Infants at presentation with neonatal cholestasis were enrolled prior to surgical exploration in a prospective observational multi-centered study. Clinical features (physical findings, laboratory results, gallbladder sonography) at enrollment were analyzed. In PROBE April 2004-February 2014, 401 infants met criteria for BA and 259 for Non-BA. Univariate analysis identified 13 features that were significantly different between BA and Non-BA.   Accurate identification of BA in infants with neonatal cholestasis requires further evaluation, and BA should not be excluded based only on presenting clinical features.

3 Archived Datasets

All the SAS 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 "discriminate_ba.sas7bdat" dataset.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Shneider et al [1] in 2017. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], <u>Comparison of clinical information at presentation between infants with and without BA</u>, Table A lists the variables that were used in the replication and Table B compares the results calculated from the archived data file to the results published in Table 1. The results of the replication are almost 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] Shneider, B.L., Moore, J., Kerkar, N., Magee, J.C., Ye, W., Karpen, S.J., Kamath, B.M., Molleston, J.P., Bezerra, J.A., Murray, K.F., Loomes, K.M., Whitington, P.F., Rosenthal, P., Squires, R.H., Guthery, S.L., Arnon, R., Schwarz, K.B., Turmelle, Y.P., Sherker, A.H., Sokol, R.J., and the Childhood Liver Disease Research Network. "Initial Assessment of the Infant with Neonatal Cholestasis". (2017).

Table A: Variables used to replicate Table 1: Comparison of clinical information at presentation between infants with and without BA

Table Variable	Variable
Race	race_c
Sex	gender_c
Ethnicity	ethnicity_c
Age at First Evaluation (Days)	age_fev_d
Age at Disease Onset (Days)	age_onset_d
Weight (kg)	weight
Length (cm)	height
Head Circumference (cm)	headcir
Weight Z-Score	waz
Length Z-Score	haz
Head Circumference Z-Score	hcz
Acholic Stools	acholic_stools_3
Acholic Stools (3 Levels)	acholic_stools_3
Facial Features	facial_fea
Liver Edge Palpable	liver_pal
Liver Edge Below Costal Margin (cm)	liver_below
Spleen Palpable	spleen_pal
Direct Baseline Bilirubin (mg/dL)	d_bili
Conjugated Baseline Bilirubin (mg/dL)	c_bili
Total Baseline Bilirubin (mg/dL)	ct_bili
AST (u/L)	ast
ALT (u/L)	alt
Albumin (g/dL)	albumin
GGTP (u/L)	ggtp
Platelets (10³/mm³)	platelets
Alkaline Phosphatase (IU/L)	alk_phos
Total Cholesterol (mg/dL)	tchol
Gallbladder	gb
Gallbladder (Absent vs. Present)	gb

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

Variable	Group 1 (BA Included) % or Mean (SD) N=401 Manuscript	Group 1 (BA Included) % or Mean (SD) N=401 DSIC	Diff. N=0
Race			
White	244 (63%)	244 (63.1%)	0 (0.1%)
Black	61 (15.8%)	61 (15.8%)	0 (0%)
Asian	36 (9.3%)	36 (9.3%)	0 (0%)
Other	46 (11.9%)	46 (11.9%)	0 (0%)
Sex			
Male	191 (47.6%)	191 (47.6%)	0 (0%)
Female	210 (52.4%)	210 (52.4%)	0 (0%)
Ethnicity			
Hispanic	92 (23%)	92 (23%)	0 (0%)
Non-Hispanic	308 (77%)	308 (77%)	0 (0%)
Age at First Evaluation (Days)	N=401	N=401	N=0
	63.5 (30.9)	63.5 (30.9)	0 (0)
Age at Disease Onset (Days)	N=401	N=401	N=0
	12.8 (18.5)	12.8 (18.5)	0 (0)
Weight (kg)	N=398	N=398	N=0
	4.5 (0.9)	4.5 (0.9)	0 (0)
Length (cm)	N=381	N=381	N=0
	55.5 (4)	55.5 (4)	0 (0)
Head Circumference (cm)	N=336	N=336	N=0
	37.6 (2.2)	37.6 (2.2)	0 (0)
Weight Z-Score	N=398	N=398	N=0
	-1 (1)	-1 (1)	0 (0)
Length Z-Score	N=381	N=381	N=0
	-0.8 (1.5)	-0.8 (1.5)	0 (0)
Head Circumference Z-Score	N=336	N=336	N=0
	-1.1 (1.6)	-1.1 (1.6)	0 (0)
Acholic Stools			
Absent	69 (17.6%)	69 (17.7%)	0 (0.1%)
Present	322 (82.4%)	322 (82.4%)	0 (0%)
Acholic Stools (3 Levels)			
Normal	69 (17.6%)	69 (17.7%)	0 (0.1%)
White or Gray	184 (47.1%)	184 (47.1%)	0 (0%)
Pale	138 (35.3%)	138 (35.3%)	0 (0%)
Facial Features			

Variable	Group 1 (BA Included) % or Mean (SD) N=401 Manuscript	Group 1 (BA Included) % or Mean (SD) N=401 DSIC	Diff. N=0
Normal	380 (95.2%)	380 (95.2%)	0 (0%)
Abnormal	19 (4.8%)	19 (4.8%)	0 (0%)
Liver Edge Palpable			
Not Palpable	26 (7.3%)	26 (7.3%)	0 (0%)
Palpable	332 (92.7%)	332 (92.7%)	0 (0%)
Liver Edge Below Costal Margin (cm)	N=334	N=334	N=0
	3.3 (1.6)	3.3 (1.6)	0 (0)
Spleen Palpable			
Not Palpable	188 (50%)	188 (50%)	0 (0%)
Palpable	188 (50%)	188 (50%)	0 (0%)
Direct Baseline Bilirubin (mg/dL)	N=239	N=239	N=0
	5.7 (2.2)	5.7 (2.2)	0 (0)
Conjugated Baseline Bilirubin (mg/dL)	N=215	N=215	N=0
	4.3 (1.6)	4.3 (1.6)	0 (0)
Total Baseline Bilirubin (mg/dL)	N=401	N=401	N=0
	8.3 (3.1)	8.3 (3.1)	0 (0)
AST (u/L)	N=397	N=397	N=0
	232.1 (206.4)	232.1 (206.4)	0 (0)
ALT (u/L)	N=400	N=400	N=0
	154.7 (124.3)	154.7 (124.3)	0 (0)
Albumin (g/dL)	N=391	N=391	N=0
	3.6 (0.5)	3.6 (0.5)	0 (0)
GGTP (u/L)	N=379	N=379	N=0
	711.9 (537.5)	711.9 (537.5)	0 (0)
Platelets (10^3/mm^3)	N=380	N=380	N=0
	445.2 (180.2)	445.2 (180.2)	0 (0)
Alkaline Phosphatase (IU/L)	N=395	N=395	N=0
	568.6 (320.7)	568.6 (320.7)	0 (0)
Total Cholesterol (mg/dL)	N=33	N=33	N=0
	184.2 (61.3)	184.2 (61.3)	0 (0)
Gallbladder			
Absent	125 (39.9%)	125 (39.9%)	0 (0%)
Present	5 (1.6%)	5 (1.6%)	0 (0%)
Present (Small)	142 (45.4%)	142 (45.4%)	0 (0%)
Normal	41 (13.1%)	41 (13.1%)	0 (0%)
Gallbladder (Absent vs. Present)			
Absent	125 (39.9%)	125 (39.9%)	0 (0%)
Present	188 (60.1%)	188 (60.1%)	0 (0%)

Variable	Group 2 (Non-BA Included) % or Mean (SD) N=259 Manuscript	Group 2 (Non-BA Included) % or Mean (SD) N=259 DSIC	Diff. N=0
Race			
White	156 (61.2%)	156 (61.2%)	0 (0%)
Black	51 (20%)	51 (20%)	0 (0%)
Asian	13 (5.1%)	13 (5.1%)	0 (0%)
Other	35 (13.7%)	35 (13.7%)	0 (0%)
Sex			
Male	164 (63.3%)	164 (63.3%)	0 (0%)
Female	95 (36.7%)	95 (36.7%)	0 (0%)
Ethnicity			
Hispanic	60 (23.3%)	60 (23.4%)	0 (0.1%)
Non-Hispanic	197 (76.7%)	197 (76.7%)	0 (0%)
Age at First Evaluation (Days)	N=259	N=259	N=0
	60 (33.3)	60 (33.3)	0 (0)
Age at Disease Onset (Days)	N=259	N=259	N=0
	18.7 (22.1)	18.7 (22.1)	0 (0)
Weight (kg)	N=257	N=257	N=0
	4.1 (1.1)	4.1 (1.1)	0 (0)
Length (cm)	N=252	N=252	N=0
	54.3 (4.3)	54.3 (4.3)	0 (0)
Head Circumference (cm)	N=215	N=215	N=0
	37.1 (2.6)	37.1 (2.6)	0 (0)
Weight Z-Score	N=257	N=257	N=0
	-1.5 (1.2)	-1.5 (1.2)	0 (0)
Length Z-Score	N=252	N=252	N=0
	-1.4 (1.5)	-1.4 (1.5)	0 (0)
Head Circumference Z-Score	N=215	N=215	N=0
	-1.4 (1.2)	-1.4 (1.2)	0 (0)
Acholic Stools			
Absent	165 (66%)	165 (66%)	0 (0%)
Present	85 (34%)	85 (34%)	0 (0%)
Acholic Stools (3 Levels)			
Normal	165 (66%)	165 (66%)	0 (0%)
White or Gray	30 (12%)	30 (12%)	0 (0%)
Pale	55 (22%)	55 (22%)	0 (0%)
Facial Features			
Normal	207 (81.2%)	207 (81.2%)	0 (0%)

Variable	Group 2 (Non-BA Included) % or Mean (SD) N=259 Manuscript	Group 2 (Non-BA Included) % or Mean (SD) N=259 DSIC	Diff. N=0
Abnormal	48 (18.8%)	48 (18.8%)	0 (0%)
Liver Edge Palpable			
Not Palpable	34 (14.5%)	34 (14.5%)	0 (0%)
Palpable	201 (85.5%)	201 (85.5%)	0 (0%)
Liver Edge Below Costal Margin (cm)	N=202	N=202	N=0
	2.5 (1.4)	2.5 (1.4)	0 (0)
Spleen Palpable			
Not Palpable	149 (59.6%)	149 (59.6%)	0 (0%)
Palpable	101 (40.4%)	101 (40.4%)	0 (0%)
Direct Baseline Bilirubin (mg/dL)	N=162	N=162	N=0
	5.8 (3.2)	5.8 (3.2)	0 (0)
Conjugated Baseline Bilirubin (mg/dL)	N=121	N=121	N=0
	4.6 (2.6)	4.6 (2.6)	0 (0)
Total Baseline Bilirubin (mg/dL)	N=259	N=259	N=0
	8.6 (4.3)	8.6 (4.3)	0 (0)
AST (u/L)	N=254	N=254	N=0
	284.2 (347.7)	284.2 (347.7)	0 (0)
ALT (u/L)	N=255	N=255	N=0
	190.7 (232.5)	190.7 (232.5)	0 (0)
Albumin (g/dL)	N=246	N=246	N=0
	3.5 (0.6)	3.5 (0.6)	0 (0)
GGTP (u/L)	N=238	N=238	N=0
	299 (380.5)	299 (380.5)	0 (0)
Platelets (10^3/mm^3)	N=243	N=243	N=0
	419.7 (197.3)	419.7 (197.3)	0 (0)
Alkaline Phosphatase (IU/L)	N=254	N=254	N=0
	572.1 (252.1)	572.1 (252.1)	0 (0)
Total Cholesterol (mg/dL)	N=54	N=54	N=0
	190.6 (82.3)	190.6 (82.3)	0 (0)
Gallbladder			
Absent	13 (6.5%)	13 (6.5%)	0 (0%)
Present	1 (1.4%)	0 (0%)	1 (1.4%)
Present (Small)	81 (40.7%)	81 (40.7%)	0 (0%)
Normal	105 (52.8%)	105 (52.8%)	0 (0%)
Gallbladder (Absent vs. Present)			
Absent	13 (6.5%)	13 (6.5%)	0 (0%)
Present	186 (93.5%)	186 (93.5%)	0 (0%)

Attachment A: SAS Code

```
*** PROBE DSIC;
*** Programmer: Allyson Mateja;
*** Date: 2/13/2017;

title 'PROBE DSIC';
title2 ' ';

proc format;
      value yesnof 0 = 'Absent'
                   1 = 'Present';
      value stoolsf 0 = 'Normal'
                    1 = 'White or Gray'
                    2 = 'Pale';
      value liverf 0 = 'Not Palpable'
                   1 = 'Palpable';
      value gbf 1 = 'Absent'
                2 = 'Present'
                3 = 'Present (Small)'
                4 = 'Normal';

libname probedat '/prj/niddk/ims_analysis/PROBE/private_orig_data/BA discriminating features/';

data discriminate_ba;
      set probedat.discriminate_ba;

proc contents data=discriminate_ba;

proc freq data = discriminate_ba;
      tables group;

data groups_1_2;
      length gallbladder $10.
             facial_fea $8.;
      set discriminate_ba;
      if ethnicity_c = "Don't Know" then ethnicity_c = ' ';
      if race_c in ('Other', 'Native Hawaiian or Other Pacific Islander', 'American Indian or Alaska Native') then race_c = 'Other';
      if race_c in ("Don't Know", 'Refused') then race_c = ' ';
      if acholic_stools_3 = 0 then acholic_stools = 0;
      else if acholic_stools_3 in (1,2) then acholic_stools = 1;
      if gb = 1 then do;
             gallbladder = 'Absent';
             present_gb=1;
      end;
      else if gb in (2,3,4) then do;
             gallbladder = 'Present';
        present_gb = 0;
      end;
```

```
        else if gb = . then present_gb=0;
        if facial_fea = 'Dysmor' then do;
                facial_fea = 'Abnormal';
                abnormal_fea = 0;
        end;
        else if facial_fea = 'Normal' then abnormal_fea = 1;
        else if facial_fea = ' ' then abnormal_fea = 1;
        if gender_c = 'Female' then male = 0;
        else male = 1;
        if acholic_stools_3 = 2 then pale_stools = 1;
        else if acholic_stools_3 in (0,1, .) then pale_stools = 0;
        if acholic_stools_3 = 1 then white_gray_stools = 1;
        else if acholic_stools_3 in (0,2, .) then white_gray_stools = 0;
        where group in (1,2);

proc sort data = groups_1_2;
        by group;

proc freq data = groups_1_2;
        tables race_c;
        by group;
        title3 'Table 1 - Race';

proc freq data = groups_1_2;
        tables gender_c;
        by group;
        title3 'Table 1 - Sex';

proc freq data = groups_1_2;
        tables ethnicity_c;
        by group;
        title3 'Table 1 - Ethnicity';

proc means data = groups_1_2 n mean std;
        var age_fev_d;
        class group;
        title3 'Table 1 - Age at First Evaluation (Days)';

proc means data = groups_1_2 n mean std;
        var age_onset_d;
        class group;
        title3 'Table 1 - Age at Disease Onset (Days)';

proc means data = groups_1_2 n mean std;
        var weight;
        class group;
        title3 'Table 1 - Weight';

proc means data = groups_1_2 n mean std;
        var height;
        class group;
```

```
        title3 'Table 1 - Length';

proc means data = groups_1_2 n mean std;
        var headcir;
        class group;
        title3 'Table 1 - Head Circumference';

proc means data = groups_1_2 n mean std;
        var waz;
        class group;
        title3 'Table 1 - Weight Z-Score';

proc means data = groups_1_2 n mean std;
        var haz;
        class group;
        title3 'Table 1 - Length Z-Score';

proc means data = groups_1_2 n mean std;
        var hcz;
        class group;
        title3 'Table 1 - Head Circumference Z-Score';

proc freq data = groups_1_2;
        tables acholic_stools;
        by group;
        format acholic_stools yesnof.;
        title3 'Table 1 - Acholic Stools';

proc freq data = groups_1_2;
        tables acholic_stools_3;
        by group;
        format acholic_stools_3 stoolsf.;
        title3 'Table 1 - Acholic Stools (3 Levels)';

proc freq data = groups_1_2;
        tables facial_fea;
        by group;
        title3 'Table 1 - Facial Features';

proc freq data = groups_1_2;
        tables liver_pal;
        by group;
        format liver_pal liverf.;
        title3 'Table 1 - Liver Edge Palpable';

proc means data = groups_1_2 n mean std;
        var liver_below;
        class group;
        title3 'Table 1 - Liver Edge Below Costal Margin';

proc freq data = groups_1_2;
```

```
        tables spleen_pal;
        by group;
        format spleen_pal liverf.;
        title3 'Table 1 - Spleen Palpable';

proc means data = groups_1_2 n mean std;
        var d_bili;
        class group;
        title3 'Table 1 - Direct Baseline Bilirubin';

proc means data = groups_1_2 n mean std;
        var c_bili;
        class group;
        title3 'Table 1 - Conjugated Baseline Bilirubin';

proc means data = groups_1_2 n mean std;
        var ct_bili;
        class group;
        title3 'Table 1 - Total Baseline Bilirubin';

proc means data = groups_1_2 n mean std;
        var ast;
        class group;
        title3 'Table 1 - AST';

proc means data = groups_1_2 n mean std;
        var alt;
        class group;
        title3 'Table 1 - ALT';

proc means data = groups_1_2 n mean std;
        var albumin;
        class group;
        title3 'Table 1 - Albumin';

proc means data = groups_1_2 n mean std;
        var ggtp;
        class group;
        title3 'Table 1 - GGTP';

proc means data = groups_1_2 n mean std;
        var platelets;
        class group;
        title3 'Table 1 - Platelets';

proc means data = groups_1_2 n mean std;
        var alk_phos;
        class group;
        title3 'Table 1 - Alkaline Phosphatase';

proc means data = groups_1_2 n mean std;
```

```
            var tchol;
            class group;
            title3 'Table 1 - Total Cholesterol';

proc freq data = groups_1_2;
            tables gb;
            by group;
            format gb gbf.;
            title3 'Table 1 - Gallbladder';

proc freq data = groups_1_2;
            tables gallbladder;
            by group;
            title3 'Table 1 - Gallbladder (Absent vs. Present)';
```