

Dataset Integrity Check for
Outcomes of Childhood Cholestasis
in Alagille Syndrome: Results of a
Multicenter Observational Study –
Kamath et al

Prepared by Sabrina Chen
IMS Inc.

3901 Calverton Blvd, Suite 200 Calverton, MD 20705
September 8, 2020

Contents

1 Standard Disclaimer.....	2
2 Study Background.....	2
3 Archived Datasets.....	2
4 Statistical Methods.....	2
5 Results.....	2
6 Conclusions	3
7 References	3
Table A: Variables used to replicate Table 1: Characteristics at Baseline of Participants with ALGS	4
Table B: Comparison of values computed in integrity check to reference article Table 1 values	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

The Longitudinal Study of Genetic Causes of Intrahepatic Cholestasis (LOGIC; NCT00571272) enrolled children with one of four diagnoses: ALGS, PFIC, A1AT, or bile acid synthetic disorder (BASD). As part of the LOGIC study protocol, participants aged >5 years underwent DXA scanning, the results of which are reported here. The objectives of this study were to address an important knowledge gap by investigating the prevalence of bone mineral deficits in this cohort of children with chronic liver disease and to determine factors associated with lower bone density, including growth parameters, laboratory values, and clinical events such as fracture and biliary diversion.

3 Archived Datasets

All the SAS data files, as provided by the Data Coordinating Center (DCC), are located in the LOGIC folder in the data package. For this replication, variables were taken from the “algsbl.sas7bdat” dataset.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Kamath et al [1] in the journal *Hepatology Communications* in 2020. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], Characteristics at Baseline of Participants with ALGS, 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.

6 Conclusions

The NIDDK repository is confident that the algsbl data file to be distributed is a true copy of the study data.

7 References

[1] Kamath B, et al. Outcomes of Childhood Cholestasis in Alagille Syndrome: Results of a Multicenter Observational Study. *Hepatology Communications*. 2020 March; 4(3): 325-470.

Table A: Variables used to replicate Table 1: Characteristics at Baseline of Participants with ALGS

Table Variable	dataset.variable
Race	algsbl.race
JAG1 Mutation	algsbl.jag1
NOTCH2 Mutation	algsbl.notch2
Age	algsbl.ageatbaselinedisplay
Height z-score	algsbl.heightzscoredisplay
Weight z-score	algsbl.weightzscoredisplay
BMI z-score	algsbl.bmizscoredisplay
Gender	algsbl.sex
Ethnicity	algsbl.ethnicity
Cardiac features	algsbl.heartdefectcat
Renal features	algsbl.ihxanyrenalabnormal

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

		Manuscript	DSIC	Diff	Manuscript	DSIC	Diff	Manuscript	DSIC	Diff
		n/stats			Q1/SD/Percent			Q3		
Age (years)	n	293	293	0						
	Median (Q1, Q3)	3	3	0	1	1	0	8	8	0
	Range (min, max)	0	0	0	25	25	0			
Height z-score	n	266	266	0						
	Mean (SD)	-1.9	-1.9	0	1.4	1.4	0			
Weight z-score	n	269	269	0						
	Mean (SD)	-1.9	-1.9	0	1.5	1.5	0			
BMI z-score‡	n	171	171	0						
	Mean (SD)	-0.4	-0.4	0	1.2	1.2	0			
Gender	Male	175	175	0	60	60	0			
Mutational analysis	JAG1	187	187	0	64	64	0			
	NOTCH2	9	9	0	3	3	0			
	Neither JAG1 or NOTCH2	10	10	0	3	3	0			
	Not tested	87	87	0	30	30	0			
Race	White	181	181	0	62	62	0			
	Black or African American	33	33	0	11	11	0			
	Asian	13	13	0	4	4	0			
	Other	39	39	0	13	13	0			
	Missing or refused	27	27	0	9	9	0			
Ethnicity	Hispanic	65	65	0	22	22	0			
	Non-Hispanic	219	219	0	75	75	0			
	Missing or refused	9	9	0	3	3	0			
Cardiac features	Severe* heart defect	89	89	0	30	30	0			
	Mild† heart defect	170	170	0	58	58	0			
	No cardiac involvement	34	34	0	12	12	0			
Renal features	Any renal anomaly	63	63	0	22	22	0			

Attachment A: SAS Code

```
options nocenter validvarname=uppercase fmtsearch=(formats) nofmterr;

title '/prj/niddk/ims_analysis/LOGIC/prog_initial_analysis/LOGIC.dsic.20200812.sas';
run;

*****;
* INPUT ;
*****;
libname orig '/prj/niddk/ims_analysis/LOGIC/private_orig_data/2020_08_11_delivery/';

libname fmts '/prj/niddk/ims_analysis/LOGIC/private_created_data/';

PROC FORMAT CNTLIN = fmts.algsformats;

*****;
* MACROS ;
*****;
%macro readin(ds);
  data &ds;
    set sasfile.&ds;
  run;

  proc contents data=&ds;
    title3 "&ds";
  run;
%mend;

* produce n and %;
%macro npercent(rownum, var, varf, subset, subsetname);
  proc freq data=analy noprint;
    where &subset = 1;
    tables &var/list missing out=tbl1&subsetname;
  run;

  data tbl1&subsetname;
    length covar covarf $100;
    set tbl1&subsetname;
    covar = "&var";
    covarf = put(&var,&varf.);
    rownum = &rownum;
  run;

  data prnt&subsetname;
    set prnt&subsetname tbl1&subsetname;
```

```

run;

%mend;

%macro univ(rownum, var, subset, subsetname);

proc univariate data=analy outtable= univ&subsetname noprint;
  where &subset=1;
  var &var
  ;
run;

data univ&subsetname;
  length covarf $100 _var_ $25;
  set univ&subsetname;
  covarf = "&subset";
  rownum = &rownum;
run;

data prntuniv&subsetname;
  set prntuniv&subsetname univ&subsetname;
run;

%mend;

*****;
* FORMATS ;
*****;
proc format;
  value novalue
    . = "No Value"
  other = " Value"
  ;

  value genof
    1 = " SZ"
    2 = " ZZ"
    99 = "Missing"
  ;

  value mutatef
    1 = 'JAG1'
    2 = 'NOTCH2'
    3 = 'Neither JAG1 or NOTCH2'
    4 = 'Not tested'
  ;

  value racegf

```



```

1='White'
2='Black or African-American'
3='Asian'
4.5='other'
8.5='missing/refused'
;

run;

data analy;
  set orig.algsbl;
  analy=1;

  * mutational analysis;
  if jag1=1 then JAG1_NOTCH2 = 1;
  else if notch2=1 then JAG1_NOTCH2 = 2;
  else if jag1=0 and notch2=0 then JAG1_NOTCH2 = 3;
  else JAG1_NOTCH2 = 4;

  * group race;
  if race in(4,5,6,7) then racegp = 4.5;
  else if race in(.,8) then racegp = 8.5;
  else racegp = race;

run;

proc freq data=analy;
  tables JAG1_NOTCH2*JAG1*NOTCH2/list missing;
  tables racegp*race/list missing;
run;

proc contents data=analy;
run;

* med, q1, q3;
data prntunivbl;
  * length _VAR_ $100;
  set _null_;
run;

%univ(1   , AgeAtBaselineDisplay , analy , bl);
%univ(8   , HeightZScoreDisplay  , analy , bl);
%univ(9   , WeightZScoreDisplay  , analy , bl);
%univ(10  , BMIZScoreDisplay     , analy , bl);

data prntunivbl;

```

```

set prntunivbl;
  _median_ = round(_median_ , 0.1);
  _q1_     = round(_q1_     , 0.1);
  _q3_     = round(_q3_     , 0.1);
  _mean_   = round(_mean_   , 0.1);
  _std_    = round(_std_    , 0.1);
run;

proc print data=prntunivbl;
  var rownum _var_ covarf _nobs_ _median_ _q1_ _q3_ _min_ _max_ _mean_ _std_;
run;

data prntbl;
  * length _VAR_ $100;
  set _null_;
run;

%npercent(2, Sex, SEXFMT, analy, bl);
%npercent(3, JAG1_NOTCH2, mutateg, analy, bl);
%npercent(4, Racegp, RACEgF, analy, bl);
%npercent(5, Ethnicity, ETHNICITYFMT, analy, bl);
%npercent(6, HeartDefectCat, HEARTDEFECTCATFMT, analy, bl);
%npercent(7, IHxAnyRenalAbnormal, YNMFMT, analy, bl);

data prntbl;
  set prntbl;
  percent = round(percent);
run;

proc print data=prntbl;
  var rownum covar covarf count percent;
run;

```