Dataset Integrity Check for Prediction by Ultrasound of the Risk of Hepatic Cirrhosis in Cystic Fibrosis (PUSH) Pub100 Leung

Prepared by Sabrina Chen
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton, MD 20705
May 28, 2020

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

PUSH is a prospective longitudinal study that aims to determine the utility of abdominal ultrasound to predict the development of cirrhosis in patients with cystic fibrosis. Participants will undergo abdominal ultrasound at enrollment and, based on the outcome, will be placed in one of four groups. Within a five year period subjects will undergo other sample collection procedures in addition to abdominal ultrasound. This study will also monitor the effects of cystic fibrosis on associated pulmonary and nutritional issues.

3 Archived Datasets

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

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Leung et al [1] in Journal of Pediatrics in 2016. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], Demographics and diagnostic history by ultrasound pattern, 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 contain discrepancies to the published results.

For Table 2 in the publication [1], Clinical features at the time of US findings, 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 contain discrepancies to the published results.

6 Conclusions

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

7 References

[1] Daniel Leung, Wen Ye, Jean P Molleston, Alexander Weymann, Simon Ling, Shruti M Paranjape, Rene Romero, Sara Jane Schwarzenberg, Joseph Palermo, Estella M Alonso, Karen F Murray, Bruce C Marshall, Averell H Sherker, Marilyn J Siegel, Rajesh Krishnamurthy, Roger Harned, Boaz Karmazyn, John C Magee, Michael R Narkewicz for the CFLD Network Baseline ultrasound and clinical correlates in children with Cystic Fibrosis (CF) from the prospective cohort in the CF Liver Disease Network. Journal of Pediatrics. J Pediatr. 2015 Oct;167(4):862-868.e2. doi: 10.1016/j.jpeds.2015.06.062.

Table A: Variables used to replicate Table 1**:** Characteristics of the data analyzed

Table Variable	**dataset.variable**
Age at screening US (yrs)	analysis.age_us_screening
Ethnicity	analysis.ethnicity
Genotype F508Ddel	analysis.f508del1
Consensus grade	analysis.grade1
Sex	analysis.sex

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

Demographics and diagnostic history by ultrasound pattern

Ultrasound results at screening	Manuscript	DSIC	Diff	Manuscript	DSIC	Diff	Manuscript	DSIC	Diff	Manuscript	DSIC	Diff	Manuscript	DSIC	Diff
	NL	**NL**		**HTG**	**HTG**		**HMG**	**HMG**		**CIR**	**CIR**		**p-value£**	**p-value£**	
Number (%)	590 (82.1%)	590 (82.1%)	0(0)	64 (8.9%)	64 (8.9%)	0(0)	41 (5.7%)	41 (5.7%)	0(0)	24 (3.3%)	24 (3.3%)	0(0)			
Age at ultrasound,mean ±SD†	7.6±2.9	7.6±2.9	0±0	8.5±3.2§	8.5±3.2§	0±0	9.6±3.1§	9.6±3.1§	0±0	9.8±2.7§	9.8±2.7§	0±0	<0.0001	<0.0001	0
Female,count (%)†	328 (55.6%)	328 (55.6%)	0(0)	22 (34.4%)§	22 (34.4%)§	0(0)	23 (56.1%)	23 (56.1%)	0(0)	10 (41.7%)	10 (41.7%)	0(0)	0.0078	0.0078	0
Ethnicity, count (%)†															
Non-Hispanic White	509 (86.7%)	509 (86.7%)	0(0)	60 (93.8%)§	60 (93.8%)§	0(0)	33 (80.5%)	33 (80.5%)	0(0)	22 (91.7%)	22 (91.7%)	0(0)	0.025	0.029	(-0.004)
Non-Hispanic Black	12 (2%)	12 (2%)	0(0)	3 (4.6%)§	3 (4.7%)§	0(0.1)	0	0	0	0	0	0			
Hispanic	47 (8%)	47 (8%)	0(0)	0§	0§	0(0)	8 (19.5%)	8 (19.5%)	0(0)	1 (4.2%)	1 (4.2%)	0(0)			
Other	19 (3.2%)	19 (3.2%)	0(0)	1 (1.6%)§	1 (1.6%)§	0(0)	0	0	0(0)	1 (4.2%)	1 (4.2%)	0(0)			
Genotype, count (%)†															
ΔF508 homozygous	355 (60.2%)	355 (60.2%)	0(0)	41 (64.1%)	41 (64.1%)	0(0)	24 (58.5%)	24 (58.5%)	0(0)	14 (58.3%)	14 (58.3%)	0(0)	0.73	0.73	0
ΔF508 heterozygous	181 (30.7%)	181 (30.7%)	0(0)	21 (32.8%)	21 (32.8%)	0(0)	12 (29.3%)	12 (29.3%)	0(0)	7 (29.2%)	7 (29.2%)	0(0)			
Other	54 (9.2%)	54 (9.2%)	0(0)	2 (3.1%)	2 (3.1%)	0(0)	5 (12.2%)	5 (12.2%)	0(0)	3 (12.5%)	3 (12.5%)	0(0)			
Sweat Chloride Value (meq/l),mean ±SD† (n = count)	100.9±13.5 (n=456)	100.9±13.5 (n=456)	0±0 (0)	98.7±17.6 (n=44)	98.7±17.6 (n=44)	0±0 (0)	104.5±15.1 (n=28)	104.5±15.1 (n=28)	0±0 (0)	107.1±16.4 (n=21)	107.1±16.4 (n=21)	0±0 (0)	0.29	0.29	0

Attachment A: SAS Code

```
options nocenter validvarname=upcase;
title '/prj/niddk/ims_analysis/PUSH_Children/prog_initial_analysis/push_integrity_check_20200519.sas';
run;


*** PUSH M100 DSIC;
*** Programmer: Sabrina Chen;
*** Date: 8/25/17;

** 5/19/20, SEC - an updated analysis file was sent that should correct the counts for F508 heterozygous. Table 2 var, Pseudomonas, was dropped.;


**********;
* INPUT   ;
**********;

libname sas_data '/prj/niddk/ims_analysis/PUSH_Children/private_orig_data/PUSH.Baseline.Feb.2020.Delivery/';



***********;
* MACRO    ;
***********;

* This macro creates a contents and print of first 10 recs;
* ds = dataset;
%macro preview(ds);
  data &ds;
    set sas_data.&ds;
  run;

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

  proc print data=&ds (obs=10);
  run;
%mend;


************;
* FORMATS   ;
************;
proc format;
  value grade1gp2f
  0 = 'normal'
  1 = 'abnormal'
  ;
```

```
  value nomiss
  .     = 'no value'
  other = '   value'
  ;
run;



** Take a look at the files used for the paper.;
%preview(analysis);
%preview(dmga   );
%preview(dop    );
%preview(elg    );
%preview(mch    );
%preview(usg    );


proc freq data=analysis;
  tables age_us_screening
         ethnicity
         f508del1
         grade1
        /* id                  */
         mutation
        /*       pseudomonas   */              /* This var was dropped. */
         sex
         sweat_chloride_value/missing;
title3 'analysis file';
run;

**********;
* Table 1 ;
**********;

* Counts, SD and percentages;
proc freq data=analysis;
  tables grade1/missing;
  tables sex*grade1/missing norow;
  tables f508del1 * grade1/missing norow;
  title3 'Table 1';
  title4 'Counts and Percentages';
run;

proc freq data=analysis;
  where ethnicity ne '';
  tables ethnicity *grade1/missing norow;
  title3 'Table 1';
  title4 'Counts and Percentages (NOTE: subset to subjects not missing ethnicity value)';
run;

* Diagnosis via newborn or prenatal screening var not in analysis file. ;

proc sort data=analysis;
  by grade1;
run;

proc univariate data=analysis;
```

```
  var age_us_screening;
  by grade1;
  title3 'Table 1';
  title4 'Mean/SD';
run;

proc univariate data=analysis;
  var sweat_chloride_value;
  by grade1;
  title3 'Table 1';
  title4 'Mean/SD';
run;

* p-values;
proc freq data=analysis;
  tables sex*grade1/missing chisq;
  tables f508del1*grade1/missing chisq;
  title3 'Table 1';
  title4 'p-values';
run;

proc freq data=analysis;
  where ethnicity ne '';
  tables ethnicity*grade1/missing chisq;
  title3 'Table 1';
  title4 'p-values (NOTE: subset to subjects not missing ethnicity value)';
run;

proc npar1way wilcoxon data=analysis;
  class grade1;
  var age_us_screening;
  title3 'Table 1';
  title4 'p-values';
run;

proc npar1way wilcoxon data=analysis;
  class grade1;
  var sweat_chloride_value;
  title3 'Table 1';
  title4 'p-values';
run;
```