Dataset Integrity Check for START Randomized Clinical Trial

Prepared by Dominick Parisi
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton, MD 20705
December 11, 2018

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

The aim of the START study was to determine whether corticosteroid treatment after portoenterostomy would improve bile drainage and reduce the need for liver transplantation, compared to surgery alone. The START study participants were recruited from the ChiLDREN prospective observational database study and randomized into either the corticosteroid or placebo group within 72 hours after the portoenterostomy procedure. Patients were given their assigned treatments daily over the course of 13 weeks. After the treatment period, patients underwent follow-up testing and assessments until age 24 months.

3 Archived Datasets

All the SAS data files, as provided by the Data Coordinating Center (DCC), are located in the START folder in the “Start Transfer Files” data package. For this replication, variables were taken from the “start_final_dd2.sas7bdat” dataset.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Jorge A Bezerra et al [1] in The Journal of American Medical Association 2014. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 2 in the publication [1], **Participant Characteristics at Enrollment in the Study,** 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 2. The results of the replication are an exact match to the published results, with the exception of the "White blood cell count" in which the differences are due to rounding error.

For Table 3 in the publication [1], **Primary and Secondary End Points,** 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 3. The results of the replication are an exact match to the published results.

For eTable 4 in the publication [1], **Comparison of demographic characteristics of START participants and those assessed for eligibility who did not consent to participate in START,** Table E lists the variables that were used in the replication and Table F compares the results calculated from the archived data files to the results published in eTable 4. The results of the replication are almost an exact match to the published results.

6 Conclusions

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

7 References

[1] Bezerra et. al. Use of Corticosteroids After Hepatoportoenterostomy for Bile Drainage in Infants With Biliary Atresia: The START Randomized Clinical Trial. JAMA. 2014;311(17):1750-1759

Table A: Variables used to replicate Table 2: Participant Characteristics at Enrollment in the Study

Table Variable	**dataset.variable**
Sex	start_final_dd2.gender
Race	start_final_dd2.race_collapsed
Ethnicity	start_final_dd2.ethnicity_n
BASM syndrome	start_final_dd2.BASM
Main types of Ohi classification system	start_final_dd2.ohitype
Age	start_final_dd2.age_at_kasai divided by 30.4
Weight, z-score	start_final_dd2.basewaz
Height, z-score	start_final_dd2.basehaz
Total bilirubin, mg/dL	start_final_dd2.base_bili
γ-Glutamyltransferase, U/L	start_final_dd2.ggtp
Alkaline phosphatase, U/L	start_final_dd2.alk_phos
Alanine aminotransferase, U/L	start_final_dd2.alt
Aspartate aminotransferase, U/L	start_final_dd2.ast
White blood cell count, /µL	start_final_dd2.wbc times 1000
Hemoglobin, g/dL	start_final_dd2.hemo
Platelet count, $x10^3$/ µL	start_final_dd2.platelets
International normalized ratio	start_final_dd2.inr
Albumin, g/dL	start_final_dd2.albumin

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

Variable	START Manuscript Steroids (n=70) N (%) or mean (SD)	START DISC Steroids (n=70) N (%) or mean (SD)	Diff. (n=0)	START Manuscript Placebo (n=70) N (%) or mean (SD)	START DSIC Placebo (n=70) N (%) or mean (SD)	Diff. (n=0)
- Male sex	38 (54)	38 (54)	0 (0)	30 (43)	30 (43)	0 (0)
Race						
- White	46 (66)	46 (66)	0 (0)	44 (63)	44 (63)	0 (0)
- Black	8 (11)	8 (11)	0 (0)	11 (16)	11 (16)	0 (0)
- Other	16 (23)	16 (23)	0 (0)	15 (21)	15 (21)	0 (0)
Ethnicity						
- Hispanic	14 (20)	14 (20)	0 (0)	22 (31)	22 (31)	0 (0)
- Non-Hispanic	55 (79)	55 (79)	0 (0)	48 (69)	48 (69)	0 (0)
- Refused to respond	1 (1)	1 (1)	0 (0)	0	0	0 (0)
BASM syndrome	2 (3)	2 (3)	0 (0)	3 (4)	3 (4)	0 (0)
Main types of Ohi classification system						
- I	5 (7)	5 (7)	0 (0)	8 (11)	8 (11)	0 (0)
- II	1 (1)	1 (1)	0 (0)	4 (6)	4 (6)	0 (0)
- III	64 (91)	64 (91)	0 (0)	57 (81)	57 (81)	0 (0)

Variable	START Manuscript Steroids (n=70) N (%) or mean (SD)	START DISC Steroids (n=70) N (%) or mean (SD)	Diff. (n=0)	START Manuscript Placebo (n=70) N (%) or mean (SD)	START DSIC Placebo (n=70) N (%) or mean (SD)	Diff. (n=0)
Age, mo	2.3 (0.93)	2.3 (0.93)	0 (0)	2.3 (0.84)	2.3 (0.85)	0 (0.01)
z Score						
- Weight	-0.8 (1.07)	-0.8 (1.07)	0 (0)	-0.8 (1.06)	-0.8 (1.06)	0 (0)
- Length	-0.7 (1.35)	-0.7 (1.35)	0 (0)	-0.6 (1.35)	-0.6 (1.35)	0 (0)
Total bilirubin, mg/dL	7.5 (2.6)	7.5 (2.6)	0 (0)	7.9 (2.8)	7.9 (2.8)	0 (0)
γ-Glutamyltransferase, U/L	929 (719)	929 (719)	0 (0)	731 (569)	731 (569)	0 (0)
Alkaline phosphatase, U/L	619 (341)	619 (341)	0 (0)	658 (290)	657 (290)	1 (0)
Alanine aminotransferase, U/L	154 (94)	154 (94)	0 (0)	178 (131)	178 (131)	0 (0)
Aspartate aminotransferase, U/L	236 (215)	236 (215)	0 (0)	235 (122)	235 (122)	0 (0)
White blood cell count, /μL	13200 (4300)	13230 (4305)	30 (5)	12900 (4300)	12947 (4267)	47 (33)
Hemoglobin, g/dL	10.8 (1.9)	10.8 (1.9)	0 (0)	10.4 (1.3)	10.4 (1.3)	0 (0)
Platelet count, $x10^3$/ μL	473 (179)	473 (179)	0 (0)	441 (164)	441 (164)	0 (0)
International normalized ratio	1.0 (0.2)	1.0 (0.2)	0 (0)	1.1 (0.4)	1.1 (0.4)	0 (0)
Albumin, g/dL	3.6 (0.5)	3.6 (0.5)	0 (0)	3.6 (0.5)	3.6 (0.5)	0 (0)

Table C: Variables used to replicate Table 3: Primary and Secondary End Points

Table Variable	**dataset.variable**
Total bilirubin <1.5 mg/dL and survival with native liver	start_final_dd2.success
Total bilirubin <1.5 mg/dL	start_final_dd2.success_bili
Survival with native liver	start_final_dd2.tx_6
Alive	start_final_dd2.death_6
Prevalence of ascites at age 12 mo	start_final_dd2.ascites12
Prevalence of ascites at age 24 mo	start_final_dd2.ascites24

Table D: Comparison of values computed in integrity check to reference article Table 3 values

Variable	START Manuscript Steroids (n=70)	START DSIC Steroids (n=70)	Diff (n=0)	START Manuscript Placebo (n=70)	START DSIC Placebo (n=70)	Diff (n=0)
At 6 mo posthepatoportoenterostomy						
- Total bilirubin <1.5 mg/dL and survival with native liver	41 (58.6)	41 (58.6)	0 (0, 0)	34 (48.6)	34 (48.6)	0 (0, 0)
- Total bilirubin <1.5 mg/dL	43 (61.4)	43 (61.4)	0 (0, 0)	38 (54.3)	38 (54.3)	0 (0, 0)
- survival with native liver	55 (78.6)	55 (78.6)	0 (0, 0)	52 (74.3)	52 (74.3)	0 (0, 0)
- Alive	68 (97.1)	68 (97.1)	0 (0, 0)	68 (97.1)	68 (97.1)	0 (0, 0)
At 24 mo posthepatoportoenterostomy						
- Survival with native liver and total bilirubin <1.5 mg/dL	49.4%	Not		39.8%	Not	
- Survival with native liver	58.7%	replicated	N/A	59.4%	replicated	N/A
Prevalence of ascites at 12 mo	5 (9.6)	5 (9.6)	0 (0, 0)	3 (6.4)	3 (6.4)	0 (0, 0)
Prevalence of ascites at 24 mo	1 (2.4)	1 (2.4)	0 (0, 0)	3 (7.0)	3 (7.0)	0 (0, 0)

Table E: Variables used to replicate eTable 4: Comparison of demographic characteristics of START participants and those assessed for eligibility who did not consent to participate in START

Table Variable	**dataset.variable**
Age at Kasai	start_final_dd2.age_at_kasai
Sex	start_final_dd2.gender
Race	start_final_dd2.race_collapsed
Ethnicity	start_final_dd2.ethnicity_n

Table F: Comparison of values computed in integrity check to reference article eTable 4 values

Variable	START Manuscript No Consent (n=116) N (%) or mean (SD)	START DISC No Consent (n=???) N (%) or mean (SD)	START Manuscript (n=141) N (%) or mean (SD)	START DISC (n=141) N (%) or mean (SD)	Diff. (n=0)
Age, days	63.4 (30.18)	Not replicated	68.7 (29.93)	68.7 (26.93%)	0 (3)
Male sex	47 (41%)	Not replicated	69 (49%)	69 (49%)	0 (0)
Race		Not replicated			
- White	68 (59%)		91 (65%)	91 (65%)	0 (0)
- Black	20 (17%)		19 (13%)	19 (13%)	0 (0)
- Other	28 (24%)		31 (22%)	31 (22%)	0 (0)
Ethnicity		Not replicated			
- Hispanic	32 (28%)		36 (26%)	36 (26%)	0 (0)
- Non-Hispanic	84 (72%)		104 (74%)	104 (74%)	0 (0)
- Refused to respond	0 (0%)		1 (<1%)	1 (<1%)	0 (0)

Attachment A: SAS Code

```
/******************/
/* Import datasets */
/******************/
LIBNAME SASDATA '/prj/niddk/ims_analysis/START/private_orig_data/START Transfer Files/START Transfer Files';

DATA START_FINAL_DD2;
  SET SASDATA.start_final_dd2;

  *** Convert age in days to age in months;
  AGE_AT_KASAI_MO=AGE_AT_KASAI/30.4;
  *** Multiply WBC by 1000;
  wbc_x1000 = WBC * 1000;
RUN;

/***********/
/* Table 2 */
/***********/
TITLE2 'Table 2';
PROC FREQ DATA=START_FINAL_DD2;
  TABLE TREATMENT * (gender
                     race_collapsed
                     ethnicity_n
                     BASM
                     ohitype) /MISSING NOCOL NOPERCENT;
  WHERE TREATMENT^='';
RUN;

PROC MEANS DATA=START_FINAL_DD2;
  VAR age_at_kasai
      age_at_kasai_mo
      basewaz
      basehaz
      base_bili
      ggtp
      alk_phos
      alt
      ast
      wbc
      wbc_x1000
      hemo
      platelets
      inr
      albumin;
  CLASS TREATMENT;
  WHERE TREATMENT^='';
RUN;
```

```
/***********/
/* Table 3 */
/***********/
TITLE2 'Table 3';
PROC FREQ DATA=START_FINAL_DD2;
  TABLE TREATMENT * (success
                     success_bili
                    tx_6
                    death_6) /NOCOL NOPERCENT;
  WHERE TREATMENT^='';
RUN;

PROC FREQ DATA=START_FINAL_DD2;
  TABLE TREATMENT * ascites12 /NOCOL NOPERCENT;
  WHERE TREATMENT^='' AND ASCITES12^=9999;
RUN;

PROC FREQ DATA=START_FINAL_DD2;
  TABLE TREATMENT * ascites24 /NOCOL NOPERCENT;
  WHERE TREATMENT^='' AND ASCITES24^=9999;
RUN;

/************/
/* eTable 4 */
/************/
TITLE2 'eTable 4';
PROC MEANS DATA=START_FINAL_DD2;
  VAR AGE_AT_KASAI;
RUN;

PROC FREQ DATA=START_FINAL_DD2;
  TABLE gender
        race_collapsed
       ethnicity_n;
RUN;
```