Dataset Integrity Check for the RIVUR Data Files

Prepared by Patty Griffin
IMS Inc.
3901 Calverton Blvd, Suite 200 Calverton MD 20705
October 29, 2015

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

This multicenter, randomized, double-blind, placebo-controlled trial was designed to determine whether daily antimicrobial prophylaxis is superior to placebo in preventing recurrence of urinary tract infection (UTI) in children with vesicoureteral reflux (VUR). Patients were randomly assigned to treatment for 2 years with daily antimicrobial prophylaxis (trimethoprim-sulfamethoxazole) or placebo. The study was designed to recruit 600 children (approximately 300 in each treatment group. The protocol encouraged prompt evaluation of children with UTI symptoms and early therapy of culture-proven UTIs. It was expected that approximately 10% of children would have to discontinue study medication due to allergic reactions. Assuming a 20% placebo event rate and 10% non-adherence rate, the study had 83% power to detect an absolute 10% event rate in the antimicrobial prophylaxis group. If the placebo event rate was instead 25%, power was 97% to detect an absolute 10% event rate in the treated group, even if non-adherence were as high as 15%. The primary analysis is intention-to-treat with missing outcome data analyzed as UTI.

3 Archived Datasets

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

4 Statistical Methods

Analyses were performed to duplicate results for the data published by The RIVUR Trial Investigators in The New England Journal of Medicine, June 19, 2014. To verify the integrity of the datasets, tables from the paper were checked.

5 Results

Table 1 in the publication [1], Baseline Demographic and Clinical Characteristics According to Study Group. 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 very similar to published results.

6 Conclusions

The NIDDK repository are confident that the RIVUR data files to be distributed are a copy of the manuscript data with only inconsequential discrepancies.

7 References

[1] Antimicrobial Prophylaxis for Children with Vesicoureteral Reflux, Hoberman Alejandro et al; The New England Journal of Medicine. 2014;370:2367-76. doi: /10.1056/NEJMoa1401811.

Table A: Variables used to replicate Table 1. Baseline Demographic and Clinical Characteristics According to Study Group.

Characteristic	File.Variable(s)
Treatment (columns)	Outcomes_manuscript.TXGROUP
Age	Outcomes_manuscript.AGE0101
Age Group	Outcomes_manuscript.AGE06
Sex	Outcomes_manuscript.PEFA1
Race	Outcomes_manuscript.RACE04
Hispanic ethnic group	Outcomes_manuscript.ETHNIC0101
Educational level of primary caregiver	Outcomes_manuscript.EDUCATION0101
Health insurance	Outcomes_manuscript.INSURANCETYPE0101
Index UTI – first/second	Outcomes_manuscript.PRIORUTI0101
Index UTI – febrile/symptomatic	Outcomes_manuscript.UTI_TYPE0103
Susceptibilty of index UTI to trimethoprim-sufamethoxazole	Outcomes_manuscript.URINE_RES_INDEX01
Bladder and bowel dysfunction	Outcomes_manuscript.BBD0102
Constipation	Outcomes_manuscript.CONST0102
Ultrasound abnormalities	Outcomes_manuscript. ANYHYDRONEPHROSIS and ANYDUPLICATION
Highest degree of vesicoureteral reflux	Outcomes_manuscript. HIGHEST_REFLUX0102
Bilateral vesicoureteral reflux	Outcomes_manuscript. VUR_LATERAL0101
Renal scarring	Outcomes_manuscript. WORST_SCARRING0101

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

Characteristic	Trimethoprim-Sulfamethoxazole		Placebo		Overall	
	Manuscript (N=302)	DSIC (N=302)	Manuscript (N=305)	DSIC (N=305)	Manuscript (N=607)	DSIC (N=607)
Age						
Median - mo	12	12	12	12	12	12
Interquartile range - mo	5-31	5-31	6-30	6-30	6-31	6-31
Group – no. (%)						
2-11 mo	151 (50.0)	151 (50.0)	147 (48.2)	147 (48.2)	298 (49.1)	298 (49.1)
12-23 mo	48 (15.9)	48 (15.9)	59 (19.3)	59 (19.3)	107 (17.6)	107 (17.6)
24-35 mo	43 (14.2)	43 (14.2)	36 (11.8)	36 (11.8)	79 (13.0)	79 (13.0)
36-72 mo	60 (19.9)	60 (19.9)	63 (20.7)	63 (20.7)	123 (20.3)	123 (20.3)
Sex – no (%)						
Male						
Circumcised	11 (3.6)	11 (3.6)	7 (2.3)	7 (2.3)	18 (3.0)	18 (3.0)
Uncircumcised	14 (4.6)	14 (4.6)	17 (5.6)	17 (5.6)	31 (5.1)	31 (5.1)
Female	277 (91.7)	277 (91.7)	281 (92.1)	281 (92.1)	558 (91.9)	558 (91.9)
Race – no/total no. (%)						
White	245/298 (82.2)	245/298 (82.2)	237/299 (79.3)	237/299 (79.3)	482/597 (80.7)	482/597 (80.7)
Black	10/298 (3.4)	10/298 (3.4)	17/299 (5.7)	17/299 (5.7)	27/597 (4.5)	27/597 (4.5)
Multiracial	20/298 (6.7)	20/298 (6.7)	20/299 (6.7)	20/299 (6.7)	40/597 (6.7)	40/597 (6.7)
Other	23/298 (7.7)	23/298 (7.7)	25/299 (8.4)	25/299 (8.4)	48/597 (8.0)	48/597 (8.0)
Hispanic ethnic group – no./total no. (%)	31/300 (10.3)	31/300 (10.3)	46/304 (15.1)	46/304 (15.1)	77/604 (12.7)	77/604 (12.7)
Educational level of primary caregiver – no./total no. (%)						
High-school graduate or lower	78/299 (26.1)	78/299 (26.1)	80/303 (26.4)	80/303 (26.4)	158/602 (26.2)	158/602 (26.2)
Some college or 2-yr degree	78/299 (26.1)	78/299 (26.1)	78/303 (25.7)	78/303 (25.7)	156/602 (27.0)	156/602 (27.0)
College graduate or higher	143/299 (47.8)	143/299 (47.8)	145/303 (47.9)	145/303 (47.9)	288/602 (47.8)	288/602 (47.8)
Health insurance – no./total no. (%)						
Commercial	228/299 (76.3)	228/299 (76.3)	211/302 (69.9)	211/302 (69.9)	439/601 (73.0)	439/601 (73.0)
Public	71/299 (23.7)	71/299 (23.7)	91/302 (30.1)	91/302 (30.1)	162/601 (27.0)	162/601 (27.0)
Index UTI – no. (%)						
First episode	275 (91.1)	275 (91.1)	279 (91.5	279 (91.5	554 (91.3)	554 (91.3)
Second episode	27 (8.9)	27 (8.9)	26 (8.5)	26 (8.5)	53 (8.7)	53 (8.7)
Febrile only	96 (31.8)	96 (31.8)	100 (32.8)	100 (32.8)	196 (32.3)	196 (32.3)
Symptomatic only	49 (16.2)	49 (16.2)	37 (12.1)	37 (12.1)	86 (14.2)	86 (14.2)
Febrile and symptomatic	157 (52.0)	157 (52.0)	168 (55.1)	168 (55.1)	325 (53.5)	325 (53.5)
Susceptibility of index UTI to trimethoprim-sulfmethocazole – no. /total no. (%)						
Resistant	55/278 (19.8)	55/278 (19.8)	65/293 (22.2)	65/293 (22.2)	120/571 (21.0)	120/571 (21.0)
Sensitive	223/278 980.2)	223/278 980.2)	228/293 (77.8)	228/293 (77.8)	451/571 (79.0)	451/571 (79.0)

Characteristic	Trimethoprim-Sulfamethoxazole		Placebo		Overall	
Bladder and bowel dysfunction – no./total no. (%)	34/63 (54.0)	34/63 (54.0)	37/63 (58.7)	37/63 (58.7)	71/126 (56.3)	71/126 (56.3)
Constipation – no./total no. (%)	8/65 (12.3)	8/65 (12.3)	8/64 (12.5)	8/64 (12.5)	16/129 (12.4)	16/129 (12.4)
Ultrasound abnormalities – no./total no. (%)						
Hydronephrosis	19/300 (6.3)	19/300 (6.3)	13/302 (4.3)	13/302 (4.3)	32/602 (5.3)	32/602 (5.3)
Ureter duplication	18/300 (6.0)	18/300 (6.0)	15/302 (5.0)	15/302 (5.0)	33/602 (5.5)	33/602 (5.5)
Highest degree of vesicoureteral reflex – no./total no. (%)						
Grade I	35/301 (11.6)	35/301 (11.6)	33/301 (11.0)	33/301 (11.0)	68/602 (11.3)	68/602 (11.3)
Grade II	123/301 (40.9)	123/301 (40.9)	131/301 (43.5)	131/301 (43.5)	254/602 (42.2)	254/602 (42.2)
Grade III	118/301 (39.2)	118/301 (39.2)	112/301 (37.2)	112/301 (37.2)	230/602 (38.2)	230/602 (38.2)
Grade IV	25/301 (8.3)	25/301 (8.3)	25/301 (8.3)	25/301 (8.3)	50/602 (8.3)	50/602 (8.3)
Bilateral vesicoureteral reflux – no./total no. (%)	146/300 (48.7)	146/300 (48.7)	141/300 (47.0)	141/300 (47.0)	287/600 (47.8)	287/600 (47.8)
Renal scarring – no./total no. (%)						
None	280/292 (95.9)	280/292 (95.9)	281/290 (96.9)	281/290 (96.9)	561/582 (96.4)	561/582 (96.4)
Mild	1/292 (0.3)	1/292 (0.3)	0/290	0/290	1/582 (0.2)	1/582 (0.2)
Moderate	3/292 (1.0)	3/292 (1.0)	2/290 (0.7)	2/290 (0.7)	5/582 (0.9)	5/582 (0.9)
Severe	4/292 (1.40	4/292 (1.40	2/290 (0.7)	2/290 (0.7)	6/582 (1.0)	6/582 (1.0)
Global atrophy	4/292 (1.4)	4/292 (1.4)	5/290 (1.7)	5/290 (1.7)	9/582 (1.5)	9/582 (1.5)

Table C: Variables used to replicate Table 2. Clinical Outcomes According to Study Group.

Characteristic	**File.Variable(s)**
Treatment Group	Outcomes_manuscript.TXGROUP
Recurrent febrile or symptomatic UTI	
Children with missing 2-yr data classified as having had an event	Outcomes_manuscript.UTI02
Children with missing 2-yr data classified as not having had an event	Outcomes_manuscript.UTI05
Children with missing 2-yr data omitted	Outcomes_manuscript.UTI01
Treatment failure	Outcomes_manuscript.TF02
Renal scarring	
Overall	Outcomes_manuscript.SCAR01
Severe	Outcomes_manuscript.SCAR_SEV01
New	Outcomes_manuscript.SCAR_NEW04
Any cortical defect	Outcomes_manuscript.ANYABN01
Antimicrobial resistance	
Resistant E. coli in stool	Outcomes_manuscript. RSR_TMPSMZ04
First recurrent febrile or symptomatic UTI with resistant E. coli	Outcomes_manscript.URINE_RES07 and Outcomes_manuscript.USRORG13A01
First recurrent febrile or symptomatic UTI with any resistant pathogen	Outcomes_manuscript.URINE_RES07

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

Characteristic	Trimethoprim-Sulfamethoxazole		
	Manuscript	DSIC	Diff
Recurrent febrile or symptomatic UTI			
Children with missing 2-yr data classified as	77/302 (25.5)	77/302 (25.5)	0
Children with missing 2-yr data classified as not	39/302 (12.8)	39/302 (12.8)	0
Children with missing 2-yr data omitted	39/264 (14.8)	39/264 (14.8)	0
Treatment failure	14/302 (5.0)	14/302 (5.0)	0
Renal scarring			
Overall	27/227 (11.9)	27/227 (11.9)	0
Severe	9/227 (4.0)	9/227 (4.0)	0
New	18/220 (8.2)	18/220 (8.2)	0
Any cortical defect	29/227 (12.8)	29/227 (12.8)	0
Antimicrobial resistance			
Resistant E. coli in stool	56/203 (27.6)	56/203 (27.6)	0
First recurrent febrile or symptomatic UTI with	19/30 (63.3)	19/30 (63.3)	0
First recurrent febrile or symptomatic UTI with	26/38 (68.4)	26/38 (68.4)	0

Characteristic	Placebo		
	Manuscript	DSIC	Diff
Recurrent febrile or symptomatic UTI			
Children with missing 2-yr data classified as having had an	114/305 (37.4)	114/305 (37.4)	0
Children with missing 2-yr data classified as not having had	72/305 (25.4)	72/305 (25.4)	0
Children with missing 2-yr data omitted	72/263 (27.4)	72/263 (27.4)	0
Treatment failure	27/305 (9.6)	27/305 (9.6)	0
Renal scarring			
Overall	24/235 (10.2)	24/235 (10.2)	0
Severe	6/235 (2.6)	6/235 (2.6)	0
New	19/227 (8.4)	19/227 (8.4)	0
Any cortical defect	25/235 (10.6)	25/235 (10.6)	0
Antimicrobial resistance			
Resistant E. coli in stool	41/210 (19.5)	41/210	0
First recurrent febrile or symptomatic UTI with resistant E.	11/57 (19.3)	11/57 (19.3)	0
First recurrent febrile or symptomatic UTI with any	17/69 (24.6)	17/69 (24.6)	0

Table E: Variables used to replicate counts in Figure 3. Effect of Antimicrobial Prophylaxis on the Risk of Febrile or Symptomatic UTI.

Characteristic	**File.Variable(s)**
No. of Patients with UTI	Outcomes_manuscript.UTI01
Sex	Outcomes_manuscript.SEX0101
Age	Outcomes_manuscript.AGE0106
VUR at baseline	Outcomes_manuscript.HIGHEST_REFLUX0103
VUR at 2 yr	Outcomes_manuscript.VUR_RESOLVED02
Index UTI	Outcomes_manuscript.PRIOR_UTI0101
Index UTI type	Outcomes_manuscript.UTI_FEBRILE0102
Index UTI resistance	Outcomes_manuscript.URINE_RES_INDEX01
BBD at baseline	Outcomes_manuscript.BBD0102
BBD during study	Outcomes_manuscript.BBD_EVER01
Constipation at baseline	Outcomes_manuscript.CHR_CONST0102
Constipation during study	Outcomes_manuscript.CHR_CONST_EVER02

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

Subgroup	No. of Patients with UTI				Total No. of Patients		
	Manuscript	DSIC	Diff		Manuscript	DSIC	Diff
Overall	111	111	0		607	607	0
Sex							
Female	107	107	0		558	558	0
Male	4	4	0		49	49	0
Age							
< 24 mo	65	65	0		405	405	0
>= 24 mo	46	46	0		202	202	0
VUR at baseline							
Grade I or II	46	46	0		322	322	0
Grade III or IV	64	64	0		280	280	0
VUR at 2 yr							
Resolved	34	34	0		218	218	0
Improved	17	17	0		100	100	0
Not improved	25	25	0		110	110	0
Index UTI							
First episode	94	94	0		554	554	0
Second episode	17	17	0		53	53	0
Index UTI type							
Febrile	94	94	0		521	521	0
Nonfrebrile	17	17	0		86	86	0
Index UTI resistance							
Sensitive to TMP-SMX	77	77	0		451	451	0
Resistant to TMP-SMX	24	24	0		120	120	0
BBD at baseline							
Present	25	25	0		71	71	0
Absent	12	12	0		55	55	0
BBD during study							
Present	41	41	0		154	154	0
Absent	24	24	0		164	164	0
Constipation at baseline							
Present	5	5	0		16	16	0
Absent	33	33	0		113	113	0
Constipation during study							
Present	11	11	0		51	51	0
Absent	58	58	0		280	280	0

Attachment A: SAS Code

```
title1 "%sysfunc(getoption(sysin))";
title2 " ";

*** File containing macro for examining each dataset ***;
%include 'borrowed_macro.sas';

libname rivur "/prj/niddk/ims_analysis/RIVUR/private_orig_data/Final zip file";

%global caser;
%let caser=TGRP;

%macro datachunk();

%mediandata1(order=1, invar=AGE0101, roundvar=1, digit=0);
%freqdata1(order=2, invar=AGE06, level=("1"), levelname="Age group 2-11 mo" );
%freqdata1(order=3, invar=AGE06, level=("2"), levelname="Age group 12-23 mo");
%freqdata1(order=4, invar=AGE06, level=("3"), levelname="Age group 24-35 mo");
%freqdata1(order=5, invar=AGE06, level=("4"), levelname="Age group 36-72 mo");
%freqdata1(order=6, invar=PEFA1, level=("C"), levelname="Sex=Male-Circumsized");
%freqdata1(order=7, invar=PEFA1, level=("U"), levelname="Sex=Male-Uncircumsized");
%freqdata1(order=8, invar=PEFA1, level=("F"), levelname="Sex=Female");
*%freqdata1(order=9, invar=RACE04, level=("1"), levelname="Race=White");
*%freqdata1(order=10, invar=RACE04, level=("2"), levelname="Race=Black");
*%freqdata1(order=11, invar=RACE04, level=("3"), levelname="Race=Multi");
*%freqdata1(order=12, invar=RACE04, level=("4"), levelname="Race=Other");
*%freqdata1(order=13, invar=ETHNIC0101, level=("H"), levelname="Hispanic ethnic grp");
%freqdata1(order=14, invar=PRIORUTI0101, level=("0"), levelname="Index UTI - first");
%freqdata1(order=15, invar=PRIORUTI0101, level=("1"), levelname="Index UTI - second");
%freqdata1(order=16, invar=UTI_TYPE0103, level=("Febrile"), levelname="Index UTI - Febrile");
%freqdata1(order=17, invar=UTI_TYPE0103, level=("Symptomatic"), levelname="Index UTI - Symptomatic");
%freqdata1(order=18, invar=UTI_TYPE0103, level=("Both"), levelname="Index UTI - Both");


%mend datachunk;

************************************************;
**  Input                                      **;
************************************************;
data table1;
  set rivur.outcomes_manuscript;

  if TXGROUP = 'A' then TGRP = 1;   *trimethoprim;
  else TGRP = 0;                    *placebo;

  keep TXGROUP TGRP AGE0101 AGE06 PEFA1 ETHNIC0101 RACE04 EDUCATION0101 INSURANCETYPE0101
       UTI_TYPE0103 PRIORUTI0101 URINE_RES_INDEX01 BBD0102 CHR_CONST0102 ANYDUPLICATION
       ANYHYDRONEPHROSIS HIGHEST_REFLUX0102 VUR_LATERAL0101 WORST_SCARRING0101;
run;

proc freq data = table1 (where=(race04 ne .));
  table race04*&caser;
run;
```

```
proc freq data = table1 (where=(EDUCATION0101 ne .));
  table EDUCATION0101*&caser;
run;

proc freq data = table1 (where=(INSURANCETYPE0101 ne ''));
  table INSURANCETYPE0101*&caser;
run;

proc freq data = table1 (where=(URINE_RES_INDEX01 ne ''));
  table URINE_RES_INDEX01*&caser;
run;

proc freq data = table1 (where=(BBD0102 ne .));
  table BBD0102*&caser;
run;

proc freq data = table1 (where=(CHR_CONST0102 ne ''));
  table CHR_CONST0102*&caser;
run;

proc freq data = table1 (where=(ANYHYDRONEPHROSIS ne ''));
  table ANYHYDRONEPHROSIS*&caser;
run;

proc freq data = table1 (where=(ANYDUPLICATION ne ''));
  table ANYDUPLICATION*&caser;
run;

proc freq data = table1 (where=(HIGHEST_REFLUX0102 ne .));
  table HIGHEST_REFLUX0102*&caser;
run;

proc freq data = table1 (where=(VUR_LATERAL0101 ne ''));
  table VUR_LATERAL0101*&caser;
run;

proc freq data = table1 (where=(WORST_SCARRING0101 ne ''));
  table WORST_SCARRING0101*&caser;
run;


************************************************;
*** Column processing;                          **;
************************************************;

data accumfreq1 accummean1 accummedian1 accuminert1;
  set _null_;

%datachunk();

data accumtab1;
  set accumfreq1 accummean1 accummedian1 accuminert1;
  if &caser=. then delete;
*** Total processing ***;

proc freq data=table1;
```

```
  tables &caser/missing list;
  title3 'Case Counts';

data table1;
  set table1;
  &caser=-1;

data accumfreq1 accummean1 accummedian1 accuminert1;
  set _null_;

%datachunk();

data accumtab2;
  set accumfreq1 accummean1 accummedian1 accuminert1;
  if &caser=. then delete;
*** Display processing ***;

proc sort data=accumtab2;
  by descending &caser orderer;

proc print data=accumtab2 noobs;
  by descending &caser;
  pageby &caser;
  title3 'Table 1 Total stats (list)';
  where &caser=-1;

proc sort data=accumtab1;
  by descending &caser orderer;

proc print data=accumtab1 noobs;
  by descending &caser;
  pageby &caser;
  title3 'Table 1 stats (list)';
  where &caser in (0 1);

************************************************;
**  Table 2 of paper                           **;
************************************************;
title3 'Table 2 Stats';

data table2;
  set rivur.outcomes_manuscript;

  if TXGROUP = 'A' then TGRP = 1;   *trimethoprim;
  else TGRP = 0;                    *placebo;

  TX_FAILURE = 'N';
  if TFFA4A = '' or TFFA4C = '' or TFFA4D = '' then TX_FAILURE = '';
  if TFFA4A = 'Y' or TFFA4C = 'Y' or TFFA4D = 'Y' then TX_FAILURE = 'Y';
run;

proc freq data = table2 ;
  table (TX_FAILURE TF02)*&caser / missing;
run;

proc freq data = table2 ;
```

```
  table (UTI01 UTI02 UTI05 SCAR01 SCAR_SEV01 SCAR_NEW04  ANYABN01 RSR_TMPSMZ04 URINE_RES07)*&caser ;
run;

proc freq data = table2(where=(USRORG13A01='11')) ;
  table URINE_RES07*&caser ;
  title4 "USRORG13A01";
run;

proc sort data = table2;
  by &caser;
run;

ods graphics on;

proc lifetest data = table2 method=km plots=(survival) timelim=730  timelist=730;
  time  TF01_TTFC01*TF02(0);
  strata &caser ;
  title4 'Treatment Failure - KM 2yr estimate';
run;

proc lifetest data = table2 method=km plots=(survival) timelim=730 timelist=730;
  time UTI01_TTFC01*UTI05(0);
  strata &caser ;
  title4 'Recurrent UTI - KM 2yr estimate';

ods graphics off;
title4 "";
************************************************;
**  Figure 3 of paper                          **;
************************************************;
title3 'Figure 3 Stats';

data figure3;
  set rivur.outcomes_manuscript;


run;

proc freq data = figure3 ;
  table  UTI01 (SEX0101 AGE0106 HIGHEST_REFLUX0103 VUR_RESOLVED02 PRIORUTI0101 UTI_FEBRILE0102
                URINE_RES_INDEX01 BBD0102 BBD_EVER01 CHR_CONST0102 CHR_CONST_EVER02)*UTI01 / missing;
run;
```