Dataset Integrity Check for Careful Urinary Tract Evaluation (CUTIE) Data File

Prepared by Jane Wang
3901 Calverton Blvd, Suite 200 Calverton MD 20705
December 11, 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

The Careful Urinary Tract Infection Evaluation (CUTIE) study is an observational study designed to determine why some children develop kidney scars after UTIs. Ultimately, the study was aimed at understanding which children are at the greatest risk of developing renal scarring following a UTI, so physicians can provide targeted therapies and appropriate interventions. Children between 2 months and 6 years of age who had a first or second UTI in the past 4 months were eligible for enrollment in the study. Participants were followed for 2 years through the course of five study visits, during which they were required to undergo a physical exam and provide information about recent medical history and quality of life via a questionnaire completed by parents. Additionally, blood and urine biospecimens were collected to monitor kidney function, and renal scans were performed to assess the development of kidney damage.

3 Archived Datasets

The SAS data file, as provided by the Data Coordinating Center (DCC), are located in the data package. For this replication, variables were taken from the SAS file cutie_outcomes_manuscript, dm_derv_niddk1, and scan_basedv2.

4 Statistical Methods

Analyses were performed to duplicate results for the data published by Ron Keren, et al [1] in PEDIATRICS Volume 136, number 1, July 2015. To verify the integrity of the dataset, descriptive statistics were computed.

5 Results

For Table 1 in the publication [1], Baseline Demographic and Clinical Characteristics, Table A lists the variables that can be used in the replication. Table B compares the results calculated from the archived data file to the results published in Table 1. The results of the replication are similar.

For Table 2 in the publication [1], Clinical Outcomes According to VUR Status, Table C lists the variables that can be used in the replication. Table D compares the results calculated from the archived data file to the results published in Table 2. The results of the replication are similar when the data is available.

6 Conclusions

The NIDDK repository is confident that the CUTIE data files to be distributed are an exact match to the manuscript data when available.

7 References

Ron Keren, MD, MPH, Nader Shaikh, MD, MPH, Hans Pohl, MD, Lisa Gravens-Mueller, MS, Anastasia Ivanova, PhD, Lisa Zaoutis, MD, Melissa Patel, MD, Rachel deBerardinis, BA, Allison Parker, MA, Sonika Bhatnagar, MD, Mary Ann Haralam, MSN CRNP, Marcia Pope, RN, Diana Kearney, RN CCRC, Bruce Sprague, BS, Raquel Barrera, MS, Bernarda Viteri, MD, Martina Egigueron, MD, Neha Shah, MD, Alejandro Hoberman, MD. Risk Factors for Recurrent Urinary Tract Infection and Renal Scarring. PEDIATRICS Volume 136, number 1, July 2015.

Table A: Variables used to replicate Table 1**.** Baseline Demographic and Clinical Characteristics**.**

Table Variable	Variables Used in Replication from the Dataset cutie_outcomes_manuscript
Age	AGE0101
Age group	AGE06
Gender	PEF1
Race	RACE04
ethnicity	ETHNIC0101
Education	EDUCATION0101
Health insurance	INSURANCETYPE0101
No. of index UTIs	PRIORUTI0101
Type of index UTI	UTI_TYPE0103
Index UTI organism	PORG0101
Toilet-trained	TTUB0101
Bladder and bowel dysfunction at baseline	BBD0102
Constipation	CHR_CONST0102
Ultrasound abnormalities	ANYHYDRONEPHROSIS0101
Renal scarring (none)	ANYDUPLICATION0101
Renal scarring (category)	WORST_SCARRING0101

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

	No VUR (n = 195) [Manuscript]	No VUR (n = 195) [DSIC]	No VUR (n = 195) [Difference]	VUR (n = 305)[Manuscript]	VUR (n = 305) [DSIC]	VUR (n = 305) [Difference]
Age Median, mo	13	13	0	12	12	0
Age 25th and 75th quartile, mo	5–42	5–42	0	6–30	6–30	0
Age group 2–11 mo	91 (47)	91 (47)	0	147 (48)	147 (48)	0
Age group 12–23 mo	32 (16)	32 (16)	0	59 (19)	59 (19)	0
Age group 24–35 mo	12 (6)	12 (6)	0	36 (12)	36 (12)	0
Age group 36–71 mo	60 (31)	60 (31)	0	63 (21)	63 (21)	0
Male (circumcised)	6 (3)	6 (3)	0	7 (2)	7 (2)	0
Male (uncircumcised)	18 (9)	18 (9)	0	17 (6)	17 (6)	0
Female	171 (88)	171 (88)	0	281 (92)	281 (92)	0
White	131 (68)	131 (67)	0 (1)	237 (79)	237 (78)	0
African American	35 (18)	35 (18)	0	17 (6)	17 (6)	0
Multiracial	17 (9)	17 (9)	0	20 (7)	20 (7)	0
Other	10 (5)	10 (5)	0	25 (8)	25 (8)	0
Hispanic ethnicity group	41 (21)	41 (21)	0	46 (15)	46 (15)	0
High school graduate or lower	70 (36)	70 (36)	0	80 (26)	80 (26)	0
Some college or 2-y degree	42 (22)	42 (22)	0	78 (26)	78 (26)	0
College graduate or higher	82 (42)	82 (42)	0	145 (48)	145 (48)	0
Health insurance Commercial	103 (53)	103 (53)	0	211 (70)	211 (70)	0
Health insurance Public	92 (47)	92 (47)	0	91 (30)	91 (30)	0
No. of index UTIs First episode	174 (89)	174 (89)	0	279 (91)	279 (92)	0 (-1)
No. of index UTIs Second episode	21 (11)	21 (11)	0	26 (9)	26 (9)	0
Type of index UTI Febrile only	60 (31)	60 (31)	0	100 (33)	100 (33)	0

	No VUR (n = 195) [Manuscript]	No VUR (n = 195) [DSIC]	No VUR (n = 195) [Difference]	VUR (n = 305)[Manuscript]	VUR (n = 305) [DSIC]	VUR (n = 305) [Difference]
Type of index UTI Symptomatic only	46 (24)	46 (24)	0	37 (12)	37 (12)	0
Type of index UTI Febrile and symptomatic	89 (46)	89 (46)	0	168 (55)	168 (55)	0
Index UTI organism E coli	177 (91)	177 (91)	0	274 (90)	274 (90)	0
Index UTI organism Other	18 (9)	18 (9)	0	31 (10)	31 (10)	0
Toilet-trained	60 (31)	60 (31)	0	67 (22)	67 (22)	0
Bladder and bowel dysfunction at baseline	26 (46)	26 (46)	0	37 (59)	37 (59)	0
Constipation	8 (13)	8 (13)	0	8 (13)	8 (13)	0
Ultrasound abnormalities Hydronephrosisg	11 (6)	11 (6)	0	13 (4)	13 (4)	0
Ureter duplication	4 (2)	4 (2)	0	15 (5)	15 (5)	0
Renal scarring None	187 (98)	187 (98)	0	281 (97)	281 (97)	0
Renal scarring Mild	3 (2)	3 (2)	0	0	0	0
Renal scarring Moderate	0	0	0	2 (1)	2 (1)	0
Renal scarring Severe	0	0	0	2 (1)	2 (1)	0
Renal scarring Global atrophy	0	0	0	5 (2)	5 (2)	0

Table C: Variables used to replicate Table 2. Clinical Outcomes According to VUR Status**.**

Table Variable	Variables Used in Replication from the Dataset(s)
Recurrent F/SUTI	cutie_outcomes_manuscript.UTI05
Renal scarring	cutie_outcomes_manuscript.SCAR_SCAN0
Overall	cutie_outcomes_manuscript.SCAR01
Renal scarring Mild/ Moderate/ Severe/ Global atrophy for VUR	scan_basedv2. WORST_SCARRING
Renal scarring Mild/ Moderate/ Severe/ Global atrophy for NO VUR	dm_derv_niddk1 .WORST_SCARRING
New	cutie_outcomes_manuscript.SCAR_NEW04

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

	No VUR [Manuscript]	**No VUR [DSIC]**	**No VUR [Difference]**	**No VUR [Manuscript]**	**VUR [DSIC]**	**VUR [Difference]**
	(n = 195)	(n = 195)	0	(n = 305)	(n = 305)	0
Recurrent F/SUTI	33 (17.3)	33 (17.3)	0	72 (25.4)	72 (25.4)	0
Renal scarring	(n = 144)	(n = 144)	0	(n = 235)	(n = 235)	0
Overall	8 (5.6)	8 (5.6)	0	24 (10.2)	24 (10.2)	0
Mild	7 (4.9)	7 (4.9)	0	13 (5.5)	13 (5.5)	0
Moderate	1 (0.7)	1 (0.7)	0	5 (2.1)	5 (2.1)	0
Severe	0 (0)	0 (0)	0	4 (1.7)	4 (1.7)	0
Global atrophy	0 (0)	0 (0)	0	2 (0.9)	2 (0.9)	0
New	6 (4.3)	6 (4.3)	0	19 (8.4)	19 (8.4)	0

	Grade I-II [Manuscript]	**Grade I-II [DSIC]**	**Grade I-II [Difference]**	**Grade III-IV [Manuscript]**	**Grade III-IV [DSIC]**	**Grade III-IV [Difference]**
	(n = 167)	(n = 167)	0	(n = 138)	(n = 138)	0
Recurrent F/SUTI	35 (22.4)	35 (22.4)	0	37 (28.9)	37 (28.9)	0
Renal scarring	(n = 126)	(n = 126)	0	(n = 109)	(n = 109)	0
Overall	9 (7.1)	9 (7.1)	0	15 (13.8)	15 (13.8)	0
Mild	5 (4.0)	5 (4.0)	0	8 (7.3)	8 (7.3)	0
Moderate	1 (0.8)	1 (0.8)	0	4 (3.7)	4 (3.7)	0
Severe	3 (2.4)	3 (2.4)	0	1 (0.9)	1 (0.9)	0
Global atrophy	0 (0)	0 (0)	0	2 (1.8)	2 (1.8)	0
New	8 (6.5)	8 (6.5)	0	11 (10.6)	11 (10.6)	0

Attachment A: SAS Code

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

options nofmterr source2 mprint symbolgen spool;

libname sasver '/prj/niddk/ims_analysis/CUTIE/private_orig_data/CUTIE_material_for_NIDDK/';
data cutie_outcomes_manuscript  ;  set sasver.cutie_outcomes_manuscript             ;


proc format;
  value yesnof
        1="Yes"
        2="No"
        ;
        ;


%macro freqdata1(order=, invar=, level=, popvar=, totalvl=);

%if &totalvl.=null %then %do;
  proc freq data=table1 noprint;
        tables &invar*TXGROUP/out=data1 outpct;
        format _all_;
        run;

  data data1(keep=LEVEL TXGROUP name CHARALL ORDERER);
    set data1(rename=(&invar=LEVEL));
    length name $100 CHARALL $100;
    name=upcase("&invar");
    PCT_DISP=round(PCT_COL,.1);
    CHARALL=compress(put(COUNT,8.))||" ("||compress(put(PCT_DISP,8.))||")";
    ORDERER=&order;
    if level in &level then output;
%end;
%else %do;
  proc freq data=table1 noprint;
        tables &invar*TXGROUP/out=data1 outpct;
        format _all_;
        where &popvar. in &totalvl.;
        run;

  data data1(keep=LEVEL TXGROUP name COUNT PCT_DISP ORDERER) total1(keep=accumer1 accumer2);
    set data1(rename=(&invar=LEVEL)) end=end1;
    length name $100;
    retain accumer1 accumer2;
    if _n_=1 then do;
      accumer1=0;
      accumer2=0;
    end;
    if TXGROUP=1 then accumer1=accumer1+count;
    else if TXGROUP=2 then accumer2=accumer2+count;
```

```
    name=upcase("&invar");
    PCT_DISP=round(PCT_COL,.1);
    ORDERER=&order;
    if level in &level then output data1;
    if end1 then output total1;

data total1(drop=accumer:);
  set total1;
  length TXGROUP total 8.;
  TXGROUP=1;
  total=accumer1;
  output;
  TXGROUP=2;
  total=accumer2;
  output;

data data1(drop=COUNT PCT_DISP total);
  merge data1 total1;
  by TXGROUP;
  length CHARALL $100;
  CHARALL=compress(put(COUNT,8.))||"/"||compress(put(total,8.))||" ("||compress(put(PCT_DISP,8.1))||")";


%end;
data accumfreq1;
  set accumfreq1 data1;

%mend freqdata1;

%macro meandata1(order=, invar=, roundvar=, digit=);
proc means data=table1 mean stddev noprint;
        var &invar;
        class TXGROUP;
        output out=data1 mean=mean stddev=stddev;
        run;

data data1(drop=_TYPE_ _FREQ_ mean stddev);
  set data1;
  length name CHARALL $100;
  name=upcase("&invar");
  mean=round(mean,&roundvar);
  stddev=round(stddev,&roundvar);
  CHARALL=compress(put(mean,8.&digit))||" ?"||compress(put(stddev,8.&digit));
  ORDERER=&order;

data accummean1;
  set accummean1 data1;

%mend meandata1;

%macro mediandata1(order=, invar=, roundvar=, digit=);
proc means data=table1 median p25 p75 min max noprint;
        var &invar;
        class TXGROUP;
        output out=data1 median=median p25=p25 p75=p75 min=min max=max;
        run;
```

```
data data1(drop=_TYPE_ _FREQ_ median p25 p75 min max);
  set data1;
  length name CHARALL $100;
  name=upcase("&invar");
  median=round(median,&roundvar);
  min=round(min,&roundvar);
  max=round(max,&roundvar);
  p25=round(p25,&roundvar);
  p75=round(p75,&roundvar);

  ORDERER=&order;
  CHARALL=compress(put(median,8.&digit));
  output;
  ORDERER=ORDERER+.01;
  CHARALL=compress(put(p25,8.&digit)||"-"||put(p75,8.&digit));
  output;

data accummedian1;
  set accummedian1 data1;

%mend mediandata1;

%macro rangedata1(order=, invar=, roundvar=, digit=);
proc means data=table1 median p25 p75 min max noprint;
        var &invar;
        class TXGROUP;
        output out=data1 min=min max=max;
        run;

data data1(drop=_TYPE_ _FREQ_ min max);
  set data1;
  length name CHARALL $100;
  name=upcase("&invar");
  min=round(min,&roundvar);
  max=round(max,&roundvar);
  ORDERER=&order;
  CHARALL=compress(put(min,8.&digit)||"-"||put(max,8.&digit));
  output;

data accummedian1;
  set accummedian1 data1;

%mend rangedata1;

data table1;
  set cutie_outcomes_manuscript;
  if missing(TXGROUP) then TXGROUP= 'NP';
  AGE06_CHAR=strip(put(AGE06,8.));
  RACE04_CHAR=strip(put(RACE04,8.));
  EDUCATION0101_CHAR=strip(put(EDUCATION0101,8.));
  PRIORUTI0101_CHAR=strip(put(PRIORUTI0101,8.));
  if PORG0101 = 11 then PORG0101_char = 'E coli';
  else PORG0101_char = 'Other';
  if missing(BBD0102) then BBD0102_CHAR = '';
  else BBD0102_CHAR=strip(put(BBD0102,8.));
```

```
proc sort data = table1;
  by TXGROUP;

data accumfreq1 accummean1 accummedian1;
  set _null_;

%mediandata1(order=1,  invar=AGE0101             , roundvar=1, digit=0);
%freqdata1(order=2.1,  invar=AGE06_CHAR  , level=("1"),popvar=, totalvl=null);
%freqdata1(order=2.2,  invar=AGE06_CHAR  , level=("2"),popvar=, totalvl=null);
%freqdata1(order=2.3,  invar=AGE06_CHAR  , level=("3"),popvar=, totalvl=null);
%freqdata1(order=2.4,  invar=AGE06_CHAR  , level=("4"),popvar=, totalvl=null);
%freqdata1(order=3.1,  invar=PEF1  , level=("C"),popvar=, totalvl=null);
%freqdata1(order=3.2,  invar=PEF1  , level=("U"),popvar=, totalvl=null);
%freqdata1(order=3.3,  invar=PEF1  , level=("F"),popvar=, totalvl=null);
%freqdata1(order=4.1,  invar=RACE04_CHAR  , level=("1"),popvar=, totalvl=null);
%freqdata1(order=4.2,  invar=RACE04_CHAR  , level=("2"),popvar=, totalvl=null);
%freqdata1(order=4.3,  invar=RACE04_CHAR  , level=("3"),popvar=, totalvl=null);
%freqdata1(order=4.4,  invar=RACE04_CHAR  , level=("4"),popvar=, totalvl=null);
%freqdata1(order=5,  invar=ETHNIC0101  , level=("H"),popvar=, totalvl=null);
%freqdata1(order=6.1,  invar=EDUCATION0101_CHAR  , level=("1"),popvar=, totalvl=null);
%freqdata1(order=6.2,  invar=EDUCATION0101_CHAR  , level=("2"),popvar=, totalvl=null);
%freqdata1(order=6.3,  invar=EDUCATION0101_CHAR  , level=("3"),popvar=, totalvl=null);
%freqdata1(order=7.1,  invar=INSURANCETYPE0101  , level=("COMMERCIAL"),popvar=, totalvl=null);
%freqdata1(order=7.2,  invar=INSURANCETYPE0101  , level=("PUBLIC"),popvar=, totalvl=null);
%freqdata1(order=8.1,  invar=PRIORUTI0101_CHAR  , level=("0"),popvar=, totalvl=null);
%freqdata1(order=8.2,  invar=PRIORUTI0101_CHAR  , level=("1"),popvar=, totalvl=null);
%freqdata1(order=9.1,  invar=UTI_TYPE0103  , level=("Febrile"),popvar=, totalvl=null);
%freqdata1(order=9.2,  invar=UTI_TYPE0103  , level=("Symptomatic"),popvar=, totalvl=null);
%freqdata1(order=9.3,  invar=UTI_TYPE0103  , level=("Both"),popvar=, totalvl=null);
%freqdata1(order=10.1,  invar=PORG0101_char  , level=("E coli"),popvar=, totalvl=null);
%freqdata1(order=10.2,  invar=PORG0101_char  , level=("Other"),popvar=, totalvl=null);
%freqdata1(order=11,  invar=TTUB0101  , level=("Y"),popvar=, totalvl=null);
%freqdata1(order=12,  invar=BBD0102_CHAR  , level=("1"),popvar=, totalvl=null);
%freqdata1(order=13,  invar=CHR_CONST0102  , level=("Y"),popvar=, totalvl=null);
%freqdata1(order=14,  invar=ANYHYDRONEPHROSIS0101  , level=("Y"),popvar=, totalvl=null);
%freqdata1(order=15,  invar=ANYDUPLICATION0101  , level=("Y"),popvar=, totalvl=null);
%freqdata1(order=16.1,  invar=WORST_SCARRING0101  , level=("A"),popvar=, totalvl=null);
%freqdata1(order=16.2,  invar=WORST_SCARRING0101  , level=("B"),popvar=, totalvl=null);
%freqdata1(order=16.3,  invar=WORST_SCARRING0101  , level=("C"),popvar=, totalvl=null);
%freqdata1(order=16.4,  invar=WORST_SCARRING0101  , level=("D"),popvar=, totalvl=null);
%freqdata1(order=16.5,  invar=WORST_SCARRING0101  , level=("E"),popvar=, totalvl=null);



data accumtab1;
  set accumfreq1 accummean1 accummedian1;
  if TXGROUP=" " then delete;

proc sort data=accumtab1;
  by TXGROUP orderer;

proc print data=accumtab1 noobs;
  by TXGROUP;
  title3 'Table 1 stats (list)';
```

```
*************************************************************************************************************************
***Program:
***Programmer: Corey Del Vecchio
***Date Created: 7/8/2013
***Purpose:
***
***
***
***Source of Request:
***Input Files:
***
***Output Files:
***
***History
***Updated by: Jane Wang
***Date Modified: 11/11/2015
*************************************************************************************************************************;

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

options nofmterr source2 mprint symbolgen spool;

libname sasver '/prj/niddk/ims_analysis/CUTIE/private_orig_data/CUTIE_material_for_NIDDK/';
data cutie_outcomes_manuscript  ;  set sasver.cutie_outcomes_manuscript             ;


proc format;
  value yesnof
        1="Yes"
        2="No"
        ;
        ;


%macro freqdata2(order=, invar=, level=, popvar=, totalvl=);

%if &totalvl.=null %then %do;
  proc freq data=table2 noprint;
        tables &invar*TXGROUP/out=data2 outpct;
        format _all_;
        run;

  data data2(keep=LEVEL TXGROUP name CHARALL ORDERER);
    set data2(rename=(&invar=LEVEL));
    length name $100 CHARALL $100;
    name=upcase("&invar");
    PCT_DISP=round(PCT_COL,.1);
    CHARALL=compress(put(COUNT,8.)) ||" ("||compress(put(PCT_DISP,8.1))||")" ;
    ORDERER=&order;
    if level in &level then output;
%end;
%else %do;
  proc freq data=table2 noprint;
        tables &invar*TXGROUP/out=data2 outpct;
```

```
        format _all_;
        where &popvar. in &totalvl.;
        run;

  data data2(keep=LEVEL TXGROUP name COUNT PCT_DISP ORDERER) total1(keep=accumer1 accumer2);
    set data2(rename=(&invar=LEVEL)) end=end1;
    length name $100;
    retain accumer1 accumer2;
    if _n_=1 then do;
      accumer1=0;
      accumer2=0;
    end;
    if TXGROUP=1 then accumer1=accumer1+count;
    else if TXGROUP=2 then accumer2=accumer2+count;
    name=upcase("&invar");
    PCT_DISP=round(PCT_COL,.1);
    ORDERER=&order;
    if level in &level then output data2;
    if end1 then output total1;

data total1(drop=accumer:);
  set total1;
  length TXGROUP total 8.;
  TXGROUP=1;
  total=accumer1;
  output;
  TXGROUP=2;
  total=accumer2;
  output;

data data2(drop=COUNT PCT_DISP total);
  merge data2 total1;
  by TXGROUP;
  length CHARALL $100;
  CHARALL=compress(put(COUNT,8.))||"/"||compress(put(total,8.))||" ("||compress(put(PCT_DISP,8.1))||")";


%end;
data accumfreq2;
  set accumfreq2 data2;

%mend freqdata2;

data table2;
  set cutie_outcomes_manuscript;
  if missing(TXGROUP) then TXGROUP= 'NP';

  UTI05_CHAR=strip(put(UTI05,8.));
  if missing(UTI05) then UTI05_char = '';
  SCAR_SCAN01_CHAR=strip(put(SCAR_SCAN01,8.));
  if missing(SCAR_SCAN01) then SCAR_SCAN01_char = '';

  SCAR01_CHAR=strip(put(SCAR01,8.));
  if missing(SCAR01) then SCAR01_char = '';

  SCAR_SEV01_CHAR=strip(put(SCAR_SEV01,8.));
```

```
  if missing(SCAR_SEV01) then SCAR_SEV01_char = '';

  SCAR_NEW04_CHAR=strip(put(SCAR_NEW04,8.));
  if missing(SCAR_NEW04) then SCAR_NEW04_char = '';

proc sort data = table2;
  by TXGROUP;

data accumfreq2  accummean2 accummedian2;
  set _null_;

%freqdata2(order=1,  invar=UTI05_CHAR  , level=("1"),popvar=, totalvl=null);
%freqdata2(order=2,  invar=SCAR_SCAN01_CHAR  , level=("1"),popvar=, totalvl=null);
%freqdata2(order=3,  invar=SCAR01_CHAR  , level=("1"),popvar=, totalvl=null);
%freqdata2(order=8,  invar=SCAR_NEW04_CHAR  , level=("1"),popvar=, totalvl=null);



data accumtab2;
  set accumfreq2 accummean2 accummedian2;
  if TXGROUP=" " then delete;

proc sort data=accumtab2;
  by TXGROUP orderer;

proc print data=accumtab2 noobs;
  by TXGROUP;
  title3 'Table 2 stats (list)';
%macro freqdata3(order=, invar=, level=, popvar=, totalvl=);

%if &totalvl.=null %then %do;
  proc freq data=table2 noprint;
        tables &invar*HIGHEST_REFLUXC01/out=data3 outpct;
        format _all_;
        run;

  data data3(keep=LEVEL HIGHEST_REFLUXC01 name CHARALL ORDERER);
    set data3(rename=(&invar=LEVEL));
    length name $100 CHARALL $100;
    name=upcase("&invar");
    PCT_DISP=round(PCT_COL,.1);
    CHARALL=compress(put(COUNT,8.)) ||" ("||compress(put(PCT_DISP,8.1))||")" ;
    ORDERER=&order;
    if level in &level then output;
%end;
%else %do;
  proc freq data=table2 noprint;
        tables &invar*HIGHEST_REFLUXC01/out=data3 outpct;
        format _all_;
        where &popvar. in &totalvl.;
        run;

  data data3(keep=LEVEL HIGHEST_REFLUXC01 name COUNT PCT_DISP ORDERER) total1(keep=accumer1 accumer2);
    set data3(rename=(&invar=LEVEL)) end=end1;
    length name $100;
    retain accumer1 accumer2;
```

```
    if _n_=1 then do;
      accumer1=0;
      accumer2=0;
    end;
    if HIGHEST_REFLUXC01=1 then accumer1=accumer1+count;
    else if HIGHEST_REFLUXC01=2 then accumer2=accumer2+count;
    name=upcase("&invar");
    PCT_DISP=round(PCT_COL,.1);
    ORDERER=&order;
    if level in &level then output data3;
    if end1 then output total1;

data total1(drop=accumer:);
  set total1;
  length TXGHIGHEST_REFLUXC01ROUP total 8.;
  HIGHEST_REFLUXC01=1;
  total=accumer1;
  output;
  HIGHEST_REFLUXC01=2;
  total=accumer2;
  output;

data data3(drop=COUNT PCT_DISP total);
  merge data3 total1;
  by HIGHEST_REFLUXC01;
  length CHARALL $100;
  CHARALL=compress(put(COUNT,8.))||"/"||compress(put(total,8.))||" ("||compress(put(PCT_DISP,8.1))||")";


%end;
data accumfreq3;
  set accumfreq3 data3;

%mend freqdata3;

data accumfreq3  accummean3 accummedian3;
  set _null_;

%freqdata3(order=1,  invar=UTI05_CHAR  , level=("1"),popvar=, totalvl=null);
%freqdata3(order=2,  invar=SCAR_SCAN01_CHAR  , level=("1"),popvar=, totalvl=null);
%freqdata3(order=3,  invar=SCAR01_CHAR  , level=("1"),popvar=, totalvl=null);
%freqdata3(order=8,  invar=SCAR_NEW04_CHAR  , level=("1"),popvar=, totalvl=null);



data accumtab3;
  set accumfreq3 accummean3 accummedian3;
  if HIGHEST_REFLUXC01=" " then delete;

proc sort data=accumtab3;
  by HIGHEST_REFLUXC01 orderer;

proc print data=accumtab3 noobs;
  by HIGHEST_REFLUXC01;
  title3 'Table 2-2 stats (list)';
```

```
*************************************************************************************************************************
***Program:
***Programmer: Patty Griffin
***Date Created: 11/17/2015
***Purpose: Modify Jane Wang's program dsic.t2.sas to calculate the KM estimates in Table 2 of the
***         CUTIE outcomes manuscript.
***
***
*************************************************************************************************************************;

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

options nofmterr source2 mprint symbolgen spool;

libname sasver '/prj/niddk/ims_analysis/CUTIE/private_orig_data/CUTIE_material_for_NIDDK/';

proc format;
  value yesnof
        1="Yes"
        2="No"
        ;

data table2  ;
  set sasver.cutie_outcomes_manuscript  ;
run;
/*
STUDY01
HIGHEST_REFLUXC01
*/

ods graphics on;

proc lifetest data = table2 method=km plots=(survival) timelim=730 timelist=730;
  time UTI01_TTFC01*UTI05(0);
  strata STUDY01 ;
  title4 'Recurrent UTI - KM 2yr estimate';

proc lifetest data = table2 method=km plots=(survival) timelim=730 timelist=730;
  time UTI01_TTFC01*UTI05(0);
  strata HIGHEST_REFLUXC01 ;
  title4 'Recurrent UTI - KM 2yr estimate';


ods graphics off;

*************************************************************************************************************************
***Program:
***Programmer: Corey Del Vecchio
***Date Created: 7/8/2013
***Purpose:
***
***
***Source of Request:
***Input Files:
```

```
***
***Output Files:
***
***History
***Updated by: Jane Wang
***Date Modified: 11/11/2015
*****************************************************************************************************************;

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

options nofmterr source2 mprint symbolgen spool;

libname sasver '/prj/niddk/ims_analysis/CUTIE/private_orig_data/CUTIE_material_for_NIDDK/';
libname sasadd '/prj/niddk/ims_analysis/CUTIE/private_orig_data/scan_basedv2/';
data cutie_outcomes_manuscript  ;  set sasver.cutie_outcomes_manuscript             ;
data dm_derv_niddk1             ;  set sasver.dm_derv_niddk1             ;
data scan_basedv2             ;  set sasadd.scan_basedv2             ;
proc sort data = scan_basedv2;
  where SCAR_SCAN01 = 1;
  by BlindID;

proc sort data = cutie_outcomes_manuscript (keep =BlindID  TXGROUP HIGHEST_REFLUXC01);
  by BlindID;


data   scan_basedv2 prob;
  merge scan_basedv2 (in = in1)  cutie_outcomes_manuscript (in = in2);
  by BlindID;
  if in1 then output  scan_basedv2;
  if in1 and not in2 then output prob;

proc freq data = scan_basedv2;
  where SCAR_SCAN01 = 1;
  tables TXGROUP TXGROUP * WORST_SCARRING TXGROUP * HIGHEST_REFLUXC01 * WORST_SCARRING/list missing;



proc format;
  value yesnof
        1="Yes"
        2="No"
        ;
        ;


%macro freqdata2(order=, invar=, level=, popvar=, totalvl=);

%if &totalvl.=null %then %do;
  proc freq data=table2 noprint;
        tables &invar*TXGROUP/out=data2 outpct;
        format _all_;
        run;

  data data2(keep=LEVEL TXGROUP name CHARALL ORDERER);
    set data2(rename=(&invar=LEVEL));
```

```
    length name $100 CHARALL $100;
    name=upcase("&invar");
    PCT_DISP=round(PCT_COL,.1);
    CHARALL=compress(put(COUNT,8.)) ||" ("||compress(put(PCT_DISP,8.1))||")" ;
    ORDERER=&order;
    if level in &level then output;
%end;
%else %do;
  proc freq data=table2 noprint;
       tables &invar*TXGROUP/out=data2 outpct;
       format _all_;
       where &popvar. in &totalvl.;
       run;

  data data2(keep=LEVEL TXGROUP name COUNT PCT_DISP ORDERER) total1(keep=accumer1 accumer2);
    set data2(rename=(&invar=LEVEL)) end=end1;
    length name $100;
    retain accumer1 accumer2;
    if _n_=1 then do;
      accumer1=0;
      accumer2=0;
    end;
    if TXGROUP=1 then accumer1=accumer1+count;
    else if TXGROUP=2 then accumer2=accumer2+count;
    name=upcase("&invar");
    PCT_DISP=round(PCT_COL,.1);
    ORDERER=&order;
    if level in &level then output data2;
    if end1 then output total1;

data total1(drop=accumer:);
  set total1;
  length TXGROUP total 8.;
  TXGROUP=1;
  total=accumer1;
  output;
  TXGROUP=2;
  total=accumer2;
  output;

data data2(drop=COUNT PCT_DISP total);
  merge data2 total1;
  by TXGROUP;
  length CHARALL $100;
  CHARALL=compress(put(COUNT,8.))||"/"||compress(put(total,8.))||" ("||compress(put(PCT_DISP,8.1))||")";


%end;
data accumfreq2;
  set accumfreq2 data2;

%mend freqdata2;

data table2;
  set dm_derv_niddk1;
  if visit = 13;
```

```
  TXGROUP= 'NP';

proc sort data = table2;
  by TXGROUP;

data accumfreq2  accummean2 accummedian2;
  set _null_;


%freqdata2(order=16.2,  invar=WORST_SCARRING  , level=("B"),popvar=, totalvl=null);
%freqdata2(order=16.3,  invar=WORST_SCARRING  , level=("C"),popvar=, totalvl=null);


data accumtab2;
  set accumfreq2 accummean2 accummedian2;
  if TXGROUP=" " then delete;

proc sort data=accumtab2;
  by TXGROUP orderer;

proc print data=accumtab2 noobs;
  by TXGROUP;
  title3 'Table 2 stats (list) for No VUR and Renal scarring mild and moderate';

* reset table 2 data set;
data table2;
  set scan_basedv2;
  if SCAR_SCAN01 = 1 and txgroup="P";

proc freq data= table2;
  tables   WORST_SCARRING HIGHEST_REFLUXC01 * WORST_SCARRING/list missing;

%freqdata2(order=16.2,  invar=WORST_SCARRING  , level=("B"),popvar=, totalvl=null);
%freqdata2(order=16.3,  invar=WORST_SCARRING  , level=("C"),popvar=, totalvl=null);
%freqdata2(order=16.4,  invar=WORST_SCARRING  , level=("D"),popvar=, totalvl=null);
%freqdata2(order=16.5,  invar=WORST_SCARRING  , level=("E"),popvar=, totalvl=null);

data accumtab2;
  set accumfreq2 accummean2 accummedian2;
  if TXGROUP=" " then delete;

proc sort data=accumtab2;
  by TXGROUP orderer;

proc print data=accumtab2 noobs;
  by TXGROUP;
  title3 'Table 2 stats (list) for VUR and Renal scarring mild and moderate';



%macro freqdata3(order=, invar=, level=, popvar=, totalvl=);

%if &totalvl.=null %then %do;
  proc freq data=table2 noprint;
        tables &invar*HIGHEST_REFLUXC01/out=data3 outpct;
        format _all_;
```

```
        run;

  data data3(keep=LEVEL HIGHEST_REFLUXC01 name CHARALL ORDERER);
    set data3(rename=(&invar=LEVEL));
    length name $100 CHARALL $100;
    name=upcase("&invar");
    PCT_DISP=round(PCT_COL,.1);
    CHARALL=compress(put(COUNT,8.)) ||" ("||compress(put(PCT_DISP,8.1))||")" ;
    ORDERER=&order;
    if level in &level then output;
%end;
%else %do;
  proc freq data=table2 noprint;
        tables &invar*HIGHEST_REFLUXC01/out=data3 outpct;
        format _all_;
        where &popvar. in &totalvl.;
        run;

  data data3(keep=LEVEL HIGHEST_REFLUXC01 name COUNT PCT_DISP ORDERER) total1(keep=accumer1 accumer2);
    set data3(rename=(&invar=LEVEL)) end=end1;
    length name $100;
    retain accumer1 accumer2;
    if _n_=1 then do;
      accumer1=0;
      accumer2=0;
    end;
    if HIGHEST_REFLUXC01=1 then accumer1=accumer1+count;
    else if HIGHEST_REFLUXC01=2 then accumer2=accumer2+count;
    name=upcase("&invar");
    PCT_DISP=round(PCT_COL,.1);
    ORDERER=&order;
    if level in &level then output data3;
    if end1 then output total1;

data total1(drop=accumer:);
  set total1;
  length TXGHIGHEST_REFLUXC01ROUP total 8.;
  HIGHEST_REFLUXC01=1;
  total=accumer1;
  output;
  HIGHEST_REFLUXC01=2;
  total=accumer2;
  output;

data data3(drop=COUNT PCT_DISP total);
  merge data3 total1;
  by HIGHEST_REFLUXC01;
  length CHARALL $100;
  CHARALL=compress(put(COUNT,8.))||"/"||compress(put(total,8.))||" ("||compress(put(PCT_DISP,8.1))||")";


%end;
data accumfreq3;
  set accumfreq3 data3;

%mend freqdata3;
```

```
data accumfreq3  accummean3 accummedian3;
  set _null_;


%freqdata3(order=16.2,  invar=WORST_SCARRING  , level=("B"),popvar=, totalvl=null);
%freqdata3(order=16.3,  invar=WORST_SCARRING  , level=("C"),popvar=, totalvl=null);
%freqdata3(order=16.4,  invar=WORST_SCARRING  , level=("D"),popvar=, totalvl=null);
%freqdata3(order=16.5,  invar=WORST_SCARRING  , level=("E"),popvar=, totalvl=null);


data accumtab3;
  set accumfreq3 accummean3 accummedian3;
  if HIGHEST_REFLUXC01=" " then delete;

proc sort data=accumtab3;
  by HIGHEST_REFLUXC01 orderer;

proc print data=accumtab3 noobs;
  by HIGHEST_REFLUXC01;
  title3 'Table 2-2 stats (list)for VUR and Renal scarring mild and moderate  by grade';
```