Dataset Integrity Check for the Gut Microbiome and p-Inulin in CKD
TarGut CKD study

Prepared by Sabrina Chen
3901 Calverton Blvd, Suite 200 Calverton MD 20705
November 16, 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

The Gut Microbiome and p-Inulin in CKD (TarGut) Study was formed to determine if the gut microbiome of patients with CKD could be improved with treatment of p-inulin, a type of prebiotic. After recruitment and enrollment, TarGut participants are evaluated for a period of 28 weeks. This 28-week study period consists of three phases: an 8-week pre-treatment phase, 12-week active treatment phase, and 8-week post-treatment phase. During the active treatment phase participants take the p-inulin supplement daily, but participants don't receive any intervention during the pre- and post-treatment phases. Biological samples are collected from study participants every week during the 28-week study period and assessed to determine microbial composition and function.

3 Archived Datasets

All data files, as provided by the Data Coordinating Center (DCC), are located in the TarGUT study data package. For this replication, variables were taken from the derived datasets: "FORM105_TARGUT.sas7bdat".

4 Statistical Methods

Analyses were performed to duplicate results for the data reported by Dr. Gassman for the Pilot Clinical Trials in CKD Consortium. To verify the integrity of the datasets, descriptive statistics were computed.

5 Results

For Table L17A in the publication [1], Table L17A – Demographic and Other Patient Characteristics of n=18 Screened Participants, Table A lists the variables that were used in the replication and Table B compares the results calculated from the archived data file to the results published in Table L17A.

6 Conclusions

The results of the replication are almost an exact match to the published results.

7 References

[1] Gassman J. for the Pilot Clinical Trials in CKD Consortium. DSMB Report, Sept 2017.

Table A: Variables used to replicate Table L17A – Demographic and Other Patient Characteristics of n=18 Screened Participants

Characteristic	**dataset.variable**
Gender	FORM105_TARGUT.gender
Race	FORM105_TARGUT.race
Hispanic/Latino Ethnicity	FORM105_TARGUT.ethnicity
How Participant heard about study	FORM105_TARGUT.hear_study
Age (years)	FORM105_TARGUT.age
Baseline Serum Creatinine (mg/dL)	FORM105_TARGUT.sc
Estimated eGFR (mUmin/1,73 m)	FORM105_TARGUT.egfr
Baseline Urine Creatinine (mg/dL)	FORM105_TARGUT.urine_cr
Baseline Urine Albumin (mg/L)	FORM105_TARGUT.urine_alb2
Urine Albumin/Creatinine Ratio (mg/g)	FORM105_TARGUT.urine_alb_cr_ratio

Table B-1: Comparison of values computed in integrity check to reference article Table L17A values. (n, %)

Variable	**n**			**%**		
	Manuscript	**DSIC**	**Diff**	**Manuscript**	**DSIC**	**Diff**
Male	12	12	0	66.7	66.7	0
Female	6	6	0	33.3	33.3	0
Black or African American	15	15	0	83.3	83.3	0
White or Caucasian	3	3	0	16.7	16.7	0
Hispanic/Latino Ethnicity	0	0	0	0	0	0
How Participant heard about study						
Personal physician	3	3	0	16.7	16.7	0
CKD Pilot Study physician	3	3	0	16.7	16.7	0
Other CKD Pilot Study staff member	8	8	0	44.4	44.4	0
Other physician or health professional	2	2	0	11.1	11.1	0
From the VA	2	2	0	11.1	11.1	0

Table B-2: Comparison of values computed in integrity check to reference article Table L17A values. (mean, sd)

Variable	mean			sd		
	Manuscript	**DSIC**	**Diff**	**Manuscript**	**DSIC**	**Diff**
Age (years)	63.6	62.5	1.1	12	12	0
Baseline Serum Creatinine (mg/dL)	2.6	2.6	0	0.93	0.93	0
Estimated eGFR (mUmin/1,73 m)	29.1	29.14	-0.04	9.27	9.27	0
Baseline Urine Creatinine (mg/dL)	92.2	92.22	-0.02	39.7	39.71	-0.01
Baseline Urine Albumin (mg/L)	1041	1041.44	-0.44	1329	1328.83	0.17
Urine Albumin/Creatinine Ratio (mg/g)	1106	1105.72	0.28	1209	1208.94	0.06

Attachment A: SAS Code

```
title '/prj/niddk/ims_analysis/TarGUT/TarGUT_dsic.sas';
run;

* 2TarGut Results Tables L16 to L 25_09282017.pdf;


************;
* INPUT     ;
************;
libname redact '/prj/niddk/ims_analysis/TarGUT/private_created_data/TarGUT_sas_data/redacted/';

********;
* FORMAT;
********;
proc format;

  value missn
  .     = 'No Value'
  other = '   Value'
  ;

  value $missc
  ' '   = 'No Value'
  other = '   Value'
  ;

  value genderf
  1 = 'male'
  2 = 'female'
  ;

  value ethnicf
  0='no'
  1='yes'
  9='unknown or not reported'
  ;

  value racef
  1='American Indian/Alaska Native'
  2='Asian'
  3='Native Hawaiian or Other Pacific Islander'
  4='Black or African American'
  5='White'
  6='More than one race'
  9='Unknown or not reported'
  ;

  value hearf
   1="Personal physician or personal physician's office"
   2="CKD Pilot Study physician"
   3="Other CKD Pilot Study staff member"
```

```
   4="Other physician or health professional"
   5="Relative/Friend"
   6="Saw a poster or brochure"
   7="Received information in mail"
   8="Health program or health fair"
   9="Saw a newspaper article"
  10="Saw a newspaper advertisement"
  11="This participant is from the Washington DC VA"
  98="Other"
  99="Unknown"
  ;

run;


************;
* MACROS    ;
************;

* read in sas;
%macro readsas(ds, dsname);

  data &ds;
    set redact.&dsname;
  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;




%readsas(FORM105_TARGUTmod        , FORM105_TARGUTmod       );

data analy;
  set FORM105_TARGUTmod;
  analy=1;
run;

proc freq data=analy;
  tables gender
       age
       race
        ethnicity
        sc
        egfr
        urine_cr
        urine_alb_units*urine_alb*URINE_ALB2
         urine_alb_cr_ratio
       hear_study
  /list missing;
  title3 "preliminary freqs";
run;



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

%univ(2   , age       , analy , all);
%univ(5   , sc        , analy , all);
```

```
%univ(6   , egfr      , analy , all);
%univ(7   , urine_cr  , analy , all);
%univ(8   , urine_alb2 , analy , all);
%univ(9   , urine_alb_cr_ratio , analy , all);

data prntunivall;
  set prntunivall;
   _median_  = round(_median_ , 0.1);
   _q1_      = round(_q1_     , 0.1);
   _q3_      = round(_q3_     , 0.1);
   _mean_    = round(_mean_   , 0.01);
   _std_     = round(_std_    , 0.01);
run;

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


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

%npercent(1,  gender    , genderf  , analy, all);
%npercent(3,  race      , racef    , analy, all);
%npercent(4,  ethnicity , ethnicf  , analy, all);
%npercent(10, hear_study, hearf    , analy, all);

data prntall;
  set prntall;
  percent = round(percent,0.1);
run;

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