Norma Pugh
December 7, 2006

Integrity Check for the Hemodialysis Study (HEMO) Files

As a partial check of the integrity of the HEMO datasets archived in the NIDDK data repository, a set of tabulations was performed to verify that published results from the HEMO study can be reproduced using the archived datasets. Analyses were performed to duplicate published results for the data reported by Eknoyan et al [1] in *The New England Journal of Medicine* in December 2002. The results of this integrity check are described below. The full text of the *New England Journal of Medicine* article can be found in Attachment 1, and the SAS code for our tabulations is included in Attachment 2.

Baseline Data. The HEMO Study evaluated the prevailing theory that a higher dose of dialysis or high-flux filters improve survival and reduce hospitalizations. Doctors at 15 medical centers recruited more than 1,800 hemodialysis patients and randomly assigned them to high or standard dialysis doses and high- or low-flux filters. Duration of survival was the major outcome studied, but hospitalizations, nutritional status and quality of life were also examined [2].

Study inclusion criteria included: (1) patient currently on hemodialysis three times a week; and (2) patient on hemodialysis constantly for the last three months. Study exclusion criteria included twenty items related to pregnancy, renal transplant status, concurrent illnesses, patient availability, use of investigational drugs or active interventions, likelihood of patient compliance and various biological measures [3].

Table 1 of the 2002 *New England Journal of Medicine* article reports on baseline characteristics of the study subjects. All variables summarized in the baseline table (Table 1. Base-Line Characteristics) can be found in a single baseline analysis dataset created for the HEMO study. Table A lists the variables used in our replication of the Table 1 variables.

Table A: Variables Used to Replicate Table 1 Variables

Table 1 Variable	Variables Used in Replication
Sample size	ktv_grp, flux_grp (all values = 1 or 2)
Age (yr)	age
Female sex	sex
Black race	black
Diabetes	dm_score (if value = 1, 2, 3)
ICED score	iced_mod
Cardiac disease	ihd_scor, chf_scor, arr_scor, card_scor (if any value = 1, 2, 3)
Yr of dialysis	duration
Residual urea clearance > 0	bkr (if value > 0)
Weight after dialysis (kg)	wt_post
Body water volume after dialysis (liters)	bv_sa
Systolic blood pressure (mm Hg)	bst_bp_s
Diastolic blood pressure (mm Hg)	bst_bp_d
Serum creatinine (mg/dl)	creatini
Serum total cholesterol (mg/dl)	tot_chol
Serum albumin (g/dl)	alb

Equilibrated normalized protein catabolic rate (g/kg/day)	benpcr_d
Equilibrated Kt/V	bektv_d
High-flux membrane	act_flux

In Table B, we compare the results for sample size and baseline characteristics calculated from the archived dataset to the results published in the 2002 *New England Journal of Medicine* article. As Table B shows, the results obtained from the archived data match the published tabulations.

Table B: Comparison of Baseline Table Values Computed in Integrity Check to Reference Article Values

Variable	All Patients		
	Eknoyan et al (2002)	**Integrity Check**	**Difference**
Sample size	1846	1846	0
Age (yr)	57.6 ± 14.0	57.6 ± 14.0	0
Female sex (%)	56.2	56.2	0
Black race (%)	62.6	62.7	0.1
Diabetes (%)	44.6	44.6	0
ICED score	2.0 ± 0.8	2.0 ± 0.8	0
Cardiac disease (%)	80.1	80.1	0
Yr of dialysis	3.7 ± 4.4	3.7 ± 4.4	0
Residual urea clearance > 0 (%)	32.9	32.9	0
Weight after dialysis (kg)	69.2 ± 14.7	69.1 ± 14.7	0.1 ± 0
Body water volume after dialysis (liters)	34.9 ± 6.1	34.9 ± 6.1	0
Measurements before dialysis			
Blood pressure (mm Hg)			
Systolic	151.8 ± 22.1	151.8 ± 22.1	0
Diastolic	81.4 ± 13.0	81.4 ± 13.0	0
Serum creatinine (mg/dl)	10.3 ± 2.9	10.3 ± 2.9	0
Serum total cholesterol (mg/dl)	172.7 ± 40.7	172.7 ± 40.7	0
Serum albumin (g/dl)	3.6 ± 0.4	3.6 ± 0.4	0
Equilibrated normalized protein catabolic rate (g/kg/day)	1.03 ± 0.24	1.03 ± 0.24	0
Equilibrated Kt/V	1.43 ± 0.21	1.43 ± 0.21	0
High-flux membrane (%)	60.2	60.3	0.1

Table B: Comparison of Baseline Table Values Computed in Integrity Check to Reference Article Values (cont.)

Variable	Standard-Dose Group			High-Dose Group		
	Eknoyan et al (2002)	**Integrity check**	**Difference**	**Eknoyan et al (2002)**	**Integrity check**	**Difference**
Sample size	926	926	0	920	920	0
Age (yr)	57.8 ± 13.9	57.8 ± 13.9	0	57.4 ± 14.1	57.4 ± 14.1	0
Female sex (%)	56.3	56.3	0	56.2	56.2	0
Black race (%)	64.1	64.2	0.1	61.1	61.2	0.1
Diabetes (%)	44.7	44.7	0	44.5	44.5	0
ICED score	2.0 ± 0.8	2.0 ± 0.8	0	2.0 ± 0.8	2.0 ± 0.8	0
Cardiac disease (%)	81.5	81.5	0	78.7	78.7	0
Yr of dialysis	3.9 ± 4.5	3.9 ± 4.5	0	3.6 ± 4.3	3.6 ± 4.3	0
Residual urea clearance > 0 (%)	31.9	31.9	0	33.9	33.9	0
Weight after dialysis (kg)	69.6 ± 14.8	69.5 ± 14.8	0.1 ± 0	68.7 ± 14.6	68.7 ± 14.7	0 ± 0.1
Body water volume after dialysis (liters)	35.1 ± 6.2	35.1 ± 6.2	0	34.8 ± 6.0	34.8 ± 6.0	0
Measurements before dialysis						
Blood pressure (mm Hg)						
Systolic	151.7 ± 22.0	151.7 ± 22.0	0	151.9 ± 22.3	151.9 ± 22.3	0
Diastolic	81.4 ± 12.8	81.4 ± 12.8	0	81.5 ± 13.1	81.5 ± 13.1	0
Serum creatinine (mg/dl)	10.3 ± 2.8	10.3 ± 2.8	0	10.3 ± 3.0	10.3 ± 3.0	0
Serum total cholesterol (mg/dl)	170.9 ± 40.4	170.9 ± 40.4	0	174.4 ± 41.0	174.4 ± 41.0	0
Serum albumin (g/dl)	3.6 ± 0.4	3.6 ± 0.4	0	3.6 ± 0.4	3.6 ± 0.4	0
Equilibrated normalized protein catabolic rate (g/kg/day)	1.04 ± 0.24	1.04 ± 0.24	0	1.03 ± 0.24	1.03 ± 0.24	0
Equilibrated Kt/V	1.43 ± 0.21	1.43 ± 0.21	0	1.43 ± 0.21	1.43 ± 0.21	0
High-flux membrane (%)	59.5	59.6	0.1	60.9	60.9	0

Table B: Comparison of Baseline Table Values Computed in Integrity Check to Reference Article Values (cont.)

Variable	Low-Flux Group			High-Flux Group		
	Eknoyan et al (2002)	**Integrity check**	**Difference**	**Eknoyan et al (2002)**	**Integrity check**	**Difference**
Sample size	925	925	0	921	921	0
Age (yr)	57.6 ± 14.2	57.6 ± 14.2	0	57.7 ± 13.9	57.7 ± 13.9	0
Female sex (%)	55.8	55.8	0	56.7	56.7	0
Black race (%)	62.6	62.6	0	62.6	62.7	0.1
Diabetes (%)	44.4	44.4	0	44.7	44.7	0
ICED score	2.0 ± 0.8	2.0 ± 0.8	0	2.0 ± 0.8	2.0 ± 0.8	0
Cardiac disease (%)	80.5	80.5	0	79.7	79.7	0
Yr of dialysis	3.7 ± 4.2	3.7 ± 4.2	0	3.8 ± 4.5	3.8 ± 4.5	0
Residual urea clearance > 0 (%)	31.2	31.2	0	34.5	34.5	0
Weight after dialysis (kg)	69.0 ± 14.7	69.0 ± 14.7	0	69.3 ± 14.7	69.2 ± 14.7	0.1 ± 0
Body water volume after dialysis (liters)	34.9 ± 6.1	34.9 ± 6.1	0	34.9 ± 6.0	34.9 ± 6.0	0
Measurements before dialysis						
Blood pressure (mm Hg)						
Systolic	151.7 ± 22.7	151.7 ± 22.7	0	151.9 ± 21.6	151.9 ± 21.6	0
Diastolic	81.5 ± 13.3	81.5 ± 13.3	0	81.4 ± 12.6	81.4 ± 12.6	0
Serum creatinine (mg/dl)	10.3 ± 2.9	10.3 ± 2.9	0	10.2 ± 2.9	10.2 ± 2.9	0
Serum total cholesterol (mg/dl)	172.7 ± 40.4	172.7 ± 40.4	0	172.6 ± 41.0	172.6 ± 41.0	0
Serum albumin (g/dl)	3.6 ± 0.4	3.6 ± 0.4	0	3.6 ± 0.4	3.6 ± 0.4	0
Equilibrated normalized protein catabolic rate (g/kg/day)	1.03 ± 0.24	1.03 ± 0.24	0	1.03 ± 0.24	1.03 ± 0.24	0
Equilibrated Kt/V	1.43 ± 0.21	1.43 ± 0.21	0	1.44 ± 0.21	1.44 ± 0.21	0
High-flux membrane (%)	59.0	59.0	0	61.3	61.5	0.2

Follow-up Data. Table 3 of the *New England Journal of Medicine* article (Table 3. Primary and Secondary Outcomes) reports on the primary outcome, main secondary outcomes and additional secondary outcomes for randomized patients - overall, by dose group and by flux group. Variables summarized in this follow-up table are found in two analysis datasets created for the HEMO study. The NIDDK repository is currently awaiting one of these analysis datasets so that the table may be replicated in its entirety. Table C lists the variables used in our replication of the Table 3 variables.

Table C: Variables Used to Replicate Table 3 Variables

Table 3 Variable	Variables Used in Replication
Sample size	ktv_grp, flux_grp (all values = 1 or 2)
Death from any cause	ev_d
First hospitalization for cardiac causes or death from any cause	Awaiting analysis file of hospitalization data.
First hospitalization due to infection or death from any cause	ev_inf
First >15% decrease in albumin level or death from any cause	ev_decl
All hospitalizations not related to vascular access	Awaiting analysis file of hospitalization data.
Death due to cardiac causes	ev_cvd
First hospitalization for cardiac causes or death due to cardiac causes	Awaiting analysis file of hospitalization data.
Death due to infection	ev_infd
First hospitalization due to infection or death due to infection	Awaiting analysis file of hospitalization data.

In Table D, we compare the results for sample size and follow-up outcomes calculated from the archived dataset to the results published in the 2002 *New England Journal of Medicine* article. As Table D shows, the results obtained from the archived data match the published tabulations.

Table D: Comparison of Follow-up Table Values Computed in Integrity Check to Reference Article Values

Variable	All Patients		
	Eknoyan et al (2002)	**Integrity Check**	**Difference**
Sample size	1846	1846	0
Primary outcome			
Death from any cause	871	871	0
Main secondary outcomes			
First hospitalization for cardiac causes or death from any cause	1079	*	*
First hospitalization due to infection or death from any cause	1104	1104	0
First >15% decrease in albumin level or death from any cause	1011	1011	0
All hospitalizations not related to vascular access	6155	*	*
Additional secondary outcomes			
Death due to cardiac causes	343	343	0
First hospitalization for cardiac causes or death due to cardiac causes	835	*	*
Death due to infection	201	201	0
First hospitalization due to infection or death due to infection	802	*	*

* Awaiting analysis file of hospitalization data.

Table D: Comparison of Follow-up Table Values Computed in Integrity Check to Reference Article Values (cont.)

Variable	Standard-Dose Group			High-Dose Group		
	Eknoyan et al (2002)	**Integrity check**	**Difference**	**Eknoyan et al (2002)**	**Integrity check**	**Difference**
Sample size	926	926	0	920	920	0
Primary outcome						
Death from any cause	440	440	0	431	431	0
Main secondary outcomes						
First hospitalization for cardiac causes or death from any cause	545	*	*	534	*	*
First hospitalization due to infection or death from any cause	557	557	0	547	547	0
First >15% decrease in albumin level or death from any cause	502	502	0	509	509	0
All hospitalizations not related to vascular access	3107	*	*	3048	*	*
Additional secondary outcomes						
Death due to cardiac causes	169	169	0	174	174	0
First hospitalization for cardiac causes or death due to cardiac causes	417	*	*	418	*	*
Death due to infection	99	99	0	102	102	0
First hospitalization due to infection or death due to infection	410	*	*	392	*	*

***** Awaiting analysis file of hospitalization data.

Table D: Comparison of Follow-up Table Values Computed in Integrity Check to Reference Article Values (cont.)

Variable	Low-Flux Group			High-Flux Group		
	Eknoyan et al (2002)	**Integrity check**	**Difference**	**Eknoyan et al (2002)**	**Integrity check**	**Difference**
Sample size	925	925	0	921	921	0
Primary outcome						
Death from any cause	442	442	0	429	429	0
Main secondary outcomes						
First hospitalization for cardiac causes or death from any cause	550	*	*	529	*	*
First hospitalization due to infection or death from any cause	562	562	0	542	542	0
First >15% decrease in albumin level or death from any cause	521	521	0	490	490	0
All hospitalizations not related to vascular access	3018	*	*	3137	*	*
Additional secondary outcomes						
Death due to cardiac causes	187	187	0	156	156	0
First hospitalization for cardiac causes or death due to cardiac causes	435	*	*	400	*	*
Death due to infection	104	104	0	97	97	0
First hospitalization due to infection or death due to infection	407	*	*	395	*	*

* Awaiting analysis file of hospitalization data.

Notes

1. Two of the four analysis datasets provided are examined in this replication analysis (BASE and DEMSUM). The BASE dataset contains baseline measurements on all enrolled patients. The DEMSUM dataset contains follow-up times and event indicators required to perform time-to-event analyses of key study outcomes. At least one additional analysis file for the hospitalization data will be provided at a later date. This hospitalization data is needed to complete the replication of the follow-up table, Table D above.

References

1. Eknoyan G, Beck GJ, Cheung AK, Daugirdas JT, Greene T, Kusek JW, Allon M, Bailey J, Delmez JA, Depner TA, Dwyer JT, Levey AS, Levin NW, Milford E, Ornt DB, Rocco MV, Schulman G, Schwab SJ, Teehan BP, Toto R; Hemodialysis (HEMO) Study Group. Effect of dialysis dose and membrane flux in maintenance hemodialysis. **The New England Journal of Medicine** 2002 Dec 19;347(25):2010-19.

2. National Institute of Diabetes & Digestive & Kidney Diseases (NIDDK) of the National Institutes of Health website: News Briefs. NIDDK : News Briefs : Hemodialysis Study Results Published Confirms Current Recommended Practice

3. HEMO Study Form1, Screening Form. Archived in the NIDDK data repository.

ATTACHMENT 1

"The full text of the article referenced will be provided to approved requestors along with the data archive."

Eknoyan G, Beck GJ, Cheung AK, Daugirdas JT, Greene T, Kusek JW, Allon M, Bailey J, Delmez JA, Depner TA, Dwyer JT, Levey AS, Levin NW, Milford E, Ornt DB, Rocco MV, Schulman G, Schwab SJ, Teehan BP, Toto R; Hemodialysis (HEMO) Study Group. Effect of dialysis dose and membrane flux in maintenance hemodialysis. The New England Journal of Medicine 2002 Dec 19;347(25):2010-19.

ATTACHMENT 2

SAS Code for Baseline and Follow-Up Tabulations from HEMO Datasets in the NIDDK Repository

```
/*************************************************************************************/
/*
/* Program: R:\05_Users\Norma\HEMO\table1.sas
/* Author:  Norma Pugh
/* Date:    07 December 06
/* Purpose: Table 1 results for NEJM replication.
/*
/*************************************************************************************/
/* Location of SAS anlaysis files */
libname data 'R:\05_Users\Norma\HEMO\MigratedData';

/*************************/
/* Get table 1 variables */
/*************************/
data baseline;
 set data.base;

 /* Keep randomized subjects only */
 if ktv_grp in(1,2) or flux_grp in(1,2);

 /* Create y/n variable: diabetes */
 if dm_score in(1,2,3) then have_diab=1; else if dm_score=0 then have_diab=0;
 label have_diab = 'Have diabetes, 1=y/0=n';

 /* Create y/n variable: residual urea clearance > 0 */
 if bkr>0 then ruc_gt0=1; else if bkr<=0 then ruc_gt0=0;
 label ruc_gt0 = 'Residual urea clearance > 0, 1=y/0=n';

 /* Create y/n variable: Cardiac disease */
 if ihd_scor in(1,2,3) or chf_scor in(1,2,3) or arr_scor in(1,2,3) or card_sco in(1,2,3)
  then card_dx=1; else card_dx=0;

 keep id ktv_grp flux_grp age sex black have_diab iced_mod card_dx duration ruc_gt0
wt_post bv_sa bst_bp_s bst_bp_d
      creatini tot_chol alb benpcr_d bektv_d act_flux;
run;

/************************/
/* Generate statistics */
/************************/
/* All Patients */
proc means data=baseline n mean std;
 var age iced_mod duration wt_post bv_sa bst_bp_s bst_bp_d creatini tot_chol alb benpcr_d
bektv_d;
run;

proc freq data=baseline;
 tables sex black have_diab card_dx ruc_gt0 act_flux;
run;

/* By Standard-Dose/High-Dose */
proc sort data=baseline; by ktv_grp; run;
proc means data=baseline n mean std;
 class ktv_grp;
 var age iced_mod duration wt_post bv_sa bst_bp_s bst_bp_d creatini tot_chol alb benpcr_d
bektv_d;
run;

proc freq data=baseline;
 tables ktv_grp*(sex black have_diab card_dx ruc_gt0 act_flux);
run;

/* By Low-Flux/High-Flux */
proc sort data=baseline; by flux_grp; run;
proc means data=baseline n mean std;
 class flux_grp;
 var age iced_mod duration wt_post bv_sa bst_bp_s bst_bp_d creatini tot_chol alb benpcr_d
bektv_d;
run;

proc freq data=baseline;
 tables flux_grp*(sex black have_diab card_dx ruc_gt0 act_flux);
run;
```

```
/*************************************************************************************/
/*
/* Program: R:\05_Users\Norma\HEMO\table3.sas
/* Author:  Norma Pugh
/* Date:    30 November 06
/* Purpose: Table 3 results for NEJM replication.
/*
/*************************************************************************************/
/* Location of SAS files */
libname data 'R:\05_Users\Norma\HEMO\MigratedData';

/*************************/
/* Get table 3 variables */
/*************************/
proc contents data=data.demsum; run;

data outcomes;
 set data.demsum;
 keep id ktv_grp flux_grp ev_d ev_inf ev_decl ev_cvd ev_infd;
run;

/************************/
/* Generate statistics */
/************************/
proc freq;
tables ev_d ev_inf ev_decl ev_cvd ev_infd;
title'All Patients';
run;

proc freq;
tables ktv_grp*(ev_d ev_inf ev_decl ev_cvd ev_infd);
title'Dose Groups';
run;

proc freq;
tables flux_grp*(ev_d ev_inf ev_decl ev_cvd ev_infd);
title'Flux Groups';
run;
```