

Data Set Integrity Check for the African American Study of Kidney Disease and Hypertension (AASK) Cohort Files



Prepared by

RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709-2194

Revision History

Version	Author/Title	Date	Comments
1.0	Sylvia Tan	May, 2012	

Table of Contents

Table of Contents

1	INTRODUCTION	4
2	BACKGROUND	4
3	DATASETS	5
4	STATISTICAL METHODS	6
5	DSIC RESULTS	7
5.1	PARTICIPANT FLOW	7
5.2	AASK TRIAL BASELINE CHARACTERISTICS	8
5.3	AASK COHORT BASELINE CHARACTERISTICS	8
6	CONCLUSIONS	11
7	REFERENCES	11
8	TABLES	13
9	APPENDIX 1: FULL TEXT OF SIKI M, ET AL. FOR APPROVED DATA REQUESTORS	2
10	APPENDIX 2: SAS VERSION 9.2 LOG FOR PROGRAMMING CODE	3
11	APPENDIX 3: SAS VERSION 9.2 OUTPUT FOR PROGRAMMING CODE	25

TABLE 1: SOCIODEMOGRAPHIC CHARACTERISTICS	13
TABLE 2: HEALTH HABITS.....	15
TABLE 3: CLINICAL CHARACTERISTICS	16
TABLE 4: LABORATORY RESULTS.....	17
TABLE 7: ESTIMATED GFR AND URINARY PROTEIN-CREATININE RATIO AT BASELINE	18

1 Introduction

As a partial check of the integrity of the AASK Cohort datasets archived in the NIDDK data repository, a set of tabulations was performed to verify that published results from the AASK Cohort study can be reproduced using the archived datasets. Analyses were performed to duplicate published results for the data reported by the AASK Study Group [1] in the American Journal of Kidney Diseases (AJKD) in July, 2007. The results of this dataset integrity check (DSIC) are described below. The full text of the AJKD article can be found in Attachment 1, and the SAS code for our tabulations is included in Attachment 2.

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 on a first exercise 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. We do not attempt to resolve minor or inconsequential discrepancies with published results or discrepancies that involve complex analyses, unless staff of the NIDDK Repository 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 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 Background

The AASK Cohort Study was a prospective, observational study that was an extension of the AASK Clinical Trial. The AASK Trial itself was a multi-center, prospective, controlled, blinded, randomized study examining the impact of two different levels of blood pressure (BP) control and the impact of three antihypertensive regimens with different randomized agents (ramipril, metoprolol and amlodipine) on the rate of change in Glomerular Filtration Rate (GFR) in African American subjects with hypertension and established renal insufficiency.

Of the 1,094 randomized participants in the AASK Trial, it was anticipated that 650-750 individuals who had not reached end-stage renal disease (ESRD) would enroll in the observational AASK Cohort study.

During the Cohort Study:

- Exposure data on environmental, genetic, physiologic, and socio-economic factors were collected annually.
- The primary renal outcome was defined by the doubling of serum creatinine, ESRD, or death.
- Antihypertensive treatment recommended by current treatment guidelines was provided to all participants.
- The treatment goal was BP < 130/80, as this was recommended by current treatment guidelines for patients with chronic kidney disease (CKD).

3 Datasets

The AJKD article referenced in this integrity check reports on participant baseline characteristics at two distinct timepoints: Trial baseline (all randomized), and Cohort baseline (for the subset of Cohort enrollees.) Even though the purpose of this DSIC was to assess the integrity of AASK Cohort files, data from both the AASK Trial and Cohort periods were analyzed in tandem, as indicated by the AJKD article.

AASK Trial and Cohort data are housed separately in the NIDDK Repository. Users interested in both periods must explicitly request both archives. Here, we briefly describe the contents of each:

AASK Trial Archive: The AASK Trial archive has two sets of primary analysis datafiles, housed in separate subfolders

- **The AASK_Analysis_20020409** subfolder contains data and documentation files from the database as of April 9, 2002, which is used in the study primary paper (Wright, et al. JAMA, 2002). While these files were analyzed as part of the AASK Trial dataset integrity check [2], they are not utilized in the current DSIC.
- **The AASK_Analysis_Current** subfolder contains the data and documentation files from the database from after April 9, 2002 when the study was closed for the primary study analyses. The files include corrections on the previous set of analysis files. They were used for later papers, but not for the study primary paper. These files were utilized for analysis in the current DSIC.

AASK Cohort Archive: The AASK Cohort archive contains 51 forms (“raw”) datasets corresponding with 59 study forms. We refer the reader to the AASK Cohort archive Roadmap for an exact mapping of datasets to forms. There are more forms than datasets because: 1.) some individual datasets are associated with multiple forms, and 2.) some forms, such as the Suicide Question Response Form #163, did not have associated datasets submitted to the Repository. The archive also contains one analysis dataset (cohort_analysis.sas7bdat), which contains key outcome variables, such as time to ESRD, extending from Trial and Cohort periods. The analysis dataset does not contain covariate data.

Datasets in the AASK Trial and Cohort periods can be linked via a unique masked participant ID, <pid>.

4 Statistical Methods

For purposes of this DSIC, a portion of published results was replicated to assure the quality of the archived datasets.

Participant status flow from Trial to Cohort periods was assessed by analyzing indicator variables in the Cohort analysis dataset (cohort_analysis.sas7bdat); results were compared to the publication.

Baseline patient characteristics were summarized for all randomized participants (Trial baseline) and all Cohort enrollees (Cohort baseline). Of special note, Glomerular Filtration Rate (GFR), an AASK Trial outcome measure, was calculated from baseline serum creatinine via an equation developed using baseline AASK Trial data . As per the publication, calculated GFR was used for all analyses, even if other measures of GFR were available in archived data.

As in published results, baseline characteristics were summarized using mean(SD) for continuous variables and percents for categorical variables. An assessment of change over time (from Trial to Cohort baseline) was not reported in the publication; hence, it was not analyzed for this integrity check.

Further details as to how variables were assembled are described as part of the Results.

5 DSIC Results

As reported in the publication, 1094 participants were randomized in the AASK Trial phase. Relevant Trial analysis files (basecov.sas7bdat, evnt_all.sas7bdat) and the Cohort analysis file (cohort_analysis.sas7bdat) show data for 1094 participants.

The publication also states that 691 participants went on to enroll in the AASK Cohort. The Cohort forms dataset on participant enrollment, c_enroll_rep.sas7bdat, shows a commensurate 691 participants. This dataset was merged via <pid> with other datasets to create an indicator for Cohort enrollment in those datasets.

5.1 Participant Flow

Figure 1, “Participant Flow”, was replicated by first defining 3 study periods, using the variable <al_mos_xd> (Month to dialysis, death, or administrative censor date) in the Cohort analysis dataset, cohort_analysis.sas7bdat. Dialysis, according to the variable label, was taken to be the same as ESRD as described in Figure 1.

- a. Trial period: ESRD, death, or censor at or before the study end date of Sept. 30, 2001
- b. Transition period: ESRD, death, or censor beyond the study end date of Sept. 30, 2001, but without reaching the Cohort start date of April, 2002
- c. Cohort period: those reached the start date of April, 2002, and were either censored or enrolled into the Cohort.

Then, a frequency analysis was run on the indicators <al_evt_d> (ESRD alone) and <al_evt_xd> (ESRD or death), stratified by study period. In this way, it was found that of 1094 enrolled in the trial, 179 reached ESRD, 85 died (no ESRD) and 7 were lost to follow-up; of 823 who made it event-free into the transition period, 17 reached ESRD, 18 died and 24 were lost-to-followup; of the 764 who made it into the Cohort period, 691 enrolled and 73 did not. These numbers exactly replicated the numbers in Figure 1.

5.2 AASK Trial Baseline Characteristics

Baseline characteristics for AASK Trial participants were replicated using data in AASK_Analysis_Current subfolder. This was done by limiting demo.sas7bdat to the 1094 randomized participants and producing means/frequencies from the relevant variables. Serum hematocrit was obtained from the basecov.sas7bdat as it was not present in the demo dataset. Descriptive statistics replicated the publication for all measures analyzed, with allowances for rounding error. See Tables 1, 2, 3 and 4 for a comparison of archived versus published results, as well as a listing of variables used. The only exception was for evidence of left ventricular hypertrophy (LVH), which was 38.4% according to published counts versus 43.9% using archived data. This represents a 14% relative difference, which could possibly be due to differences in coding of missing data. In archived data, observations missing LVH (19.6%) were excluded from the analysis [Table 4].

Note: many of the baseline variables reported on in the Publication and present in demo.sas7bdat were also present in basecov.sas7bdat. However, when running the same analysis on the basecov dataset, many descriptive statistics did not replicate those in the publication as precisely as those in the demo dataset. Documentation in the AASK_Analysis_Current subfolder describes small differences in production of baseline variables in the two files, which may explain these small differences.

The numbers in Table 7, categories of estimated GFR (as calculated from baseline serum creatinine) by categories of urine protein:creatinine ratio for AASK Trial participants, exactly replicated numbers in the Publication.

5.3 AASK Cohort baseline characteristics

As previously mentioned, a basic analysis dataset was submitted for the AASK Cohort study period. This dataset did not include baseline covariate measures. Thus, a Cohort Baseline merged dataset was assembled from 8 (of 51 total) archived raw sas7bdat datasets:

- c_enroll_rep (enrollment)
- c_demo_rep (demographics)
- c_exposures_rep (exposures)
- c_bp_rep (blood pressure measurements)
- c_local_lab_rep (CBL data)
- c_cbl_result_serum_rep (serum/plasma testing results)
- c_cbl_mailing_urine_rep (CBL urine mailing form)
- c_cbl_result_urine_rep (urine testing results)

Unchanged measures from the AASK Trial baseline (such as: gender, years of hypertension) were also merged into the Cohort Baseline merged dataset.

Measures in Table 1, Sociodemographic Characteristics, were replicated using this Cohort Baseline merged dataset. Participants with a mix of public and private insurance had to be grouped in a ‘none of the above’ category in order to replicate published numbers on health insurance.

Measures in Table 2, Health Habits, were also replicated using this merged dataset. Measures on medication usage, available in non-coded format archived data (c_med_improved_rep.sas7bdat and c_med_list_improved_rep.sas7bdat), were not assembled for purposes of this DSIC.

Measures in Table 3, Clinical Characteristics, were replicated reasonably well using the merged data. Any differences were in the decimal points. For example, mean baseline Diastolic BP was 80.7(12.5) in published data, and 80.9(12.2) in archived data. Such small differences could be attributable to either rounding error, or to the method of calculating summary measures. In archived data, baseline Diastolic BP was derived by taking the earliest baseline measurement for each participant. Each participant typically had more than one blood pressure baseline measurement (i.e., more than one measurement for $\langle \text{vist} \rangle = 0$), however, and a summary measure could be derived in multiple ways, e.g., by taking the average.

Measures in Table 4, Laboratory Measures, posed more challenges for replication. First, it is not entirely clear how baseline laboratory measures were selected for each participant. Table 7, Estimated GFR and Urinary Protein-Creatinine Ratio at Baseline, suggests that at least 671 participants had a baseline serum creatinine measure (needed to obtain estimated GFR) and baseline urine protein-creatinine ratio.

However, the archived serum result dataset (c_cbl_result_serum_rep.sas7bdat) shows only 663 persons with measures at visit-month 0 ($\langle \text{visn} \rangle = 0$). Taking the earliest observation up to the twelfth month after baseline (up to $\langle \text{visn} \rangle = 12$) increased the sample size to $n=689$, but it is unknown if the published analysis was performed on the same sample as in the Publication.

Similarly, the archived urine result dataset (c_cbl_result_urine_rep.sas7bdat) shows 581 persons with measures at baseline ($\langle \text{visn} \rangle = 0$). Taking the earliest observation up to the twelfth month after baseline (up to $\langle \text{visn} \rangle = 12$) increased the sample size to $n=670$, which still does not match the reported minimum n of 671 in Table 7.

Additionally, archived data shows that many participants had two successive serums drawn at ≤ 0 , sometimes weeks apart. For the analysis using archived data, the first measure was used as the baseline measure. It is unknown how baseline measures were derived in published analyses.

Measures in the archived urine result dataset were expressed in absolute units, such as grams per dL. Urine measures in the publication were generally reported as rates, such as grams/24 hours. In order to convert the measures in the raw dataset, the volume of urine in 24 hours was obtained from the dataset `c_cbl_mailing_urine_rep.sas7bdat`.

In spite of these data issues, baseline serum and urine measures reported in the Publication were replicated reasonably well using archived data [Table 4]. For example, mean baseline HDL was 47.7(15.4) mg/dL in published data, and 46.3(14.9) mg/dL in archived data. Mean urinary protein was 0.53(1.05) g/day in published data, and 0.53(1.06) in archived data. We could not replicate data on urinary urea nitrogen (UUN), which was 7.96(4.85) g/day in published data, but 0.72(0.44) g/day in archived data. The discrepancy could be due to issues in data conversion, as the numbers are off by a power of (approximately) 10. The remaining minor discrepancies in laboratory measures could be attributable either to the selection of baseline laboratory measures, or in the way summary measures were calculated.

We were unable to replicate Cohort baseline evidence of LVH, which was 25.8% in published data and 20.8% in archived data. This represents a 19% relative difference. In archived data, LVH was derived from the demo dataset; observations with missing data (33%) were excluded from the analysis. Differences in coding of missings may partially explain the relatively large difference [Table 4].

The numbers in Table 7, categories of estimated GFR (as calculated from baseline serum creatinine) by categories of urine protein:creatinine ratio, approximately replicated numbers in the Publication for AASK Cohort participants. The total sample size for this analysis, using archived data, was 663, eight fewer subjects than the published sample size. The discrepancy in sample sizes are likely due to differences in selecting the baseline serum/urine samples, as described above. The slight differences in percentages (between archived and published data) may also be attributed to the differences in samples, too.

6 Conclusions

With the replication of selected results, the analysis of archived AASK Trial data closely matches published results, allowing for rounding error. Analyses of archived AASK Cohort data also reasonably matches published results, allowing for rounding error and variations expected from differences in sample selection. We were unable to replicate numbers for baseline evidence of left ventricular hypertrophy for either study period. We were also unable to replicate numbers for baseline urinary urea nitrogen using archived AASK Cohort data.

Replication of published statistics from the AASK Trial was facilitated by the presence of analysis datasets. There were no such analysis datasets for the Cohort period (except for simple indicators of outcome); still, with some effort, we were able to replicate most of our selected numbers. We are confident there were no errors in the transmission of AASK Cohort archived datasets from the DCC to the Repository. We are similarly confident there were no errors in transmission of the AASK Trial (post-April 2009) files to the Repository, as these files were also analyzed as part of this DSIC.

7 References

1. Sika M, Lewis J, Douglas J, Erlinger T, Dowie D, Lipkowitz , Lash J, Cornish-Zirker D, Peterson G, Toto R, Kusek J, Appel L, Kendrick C, Gassman J, on behalf of the AASK group. Baseline Characteristics of Participants in the African American Study of Kidney Disease and Hypertension (AASK) Clinical Trial and Cohort Study. *Am J Kidney Dis.* 50(1) [2007 July]: 78-89.
2. Integrity Check for the African American Study of Kidney Disease and Hypertension (AASK) Baseline and Follow-up Files, by Norma Pugh for the NIDDK Repository [July 2006]. https://www.niddkrepository.org/niddkdocs/AASK/integrity/AASK_IntegrityCheckMemo_NoFullText_V1.pdf.

8 Tables

Table 1: Sociodemographic Characteristics

	trial variables*	1. Trial Baseline			cohort variables**	3. Cohort Baseline		
		Published	Archived	Difference		Published	Archived	Difference
N		1,094	1,094	0		691	691	0
Age (y)	age_r	54.6 ± 10.7	54.6 ± 10.7	0.0 / 0.0	visdt, dob*	60.3 ± 10.2	60.2 ± 10.2	-0.1 / 0
Sex (% men)	gender	61.2	61.2	0.0	gender*	60.5	60.5	0.0
Marital Status	mar_stat				marital			
Never married		20.8	20.8	0.0		18.1	18.1	0.0
Married/married-like		36.0	36.0	0.0		37.1	37.1	0.0
Divorced or separated		30.9	30.9	0.0		29.7	29.7	0.0
Widow/widower		12.3	12.3	0.0		15.1	15.1	0.0
Patient lives alone (%)	alone	22.6	22.6	0.0	alone	30.1	30.1	0.0
Employment status	work				employ_stats			
Employed		37.0	37.0	0.0		36.2	36.2	0.0
Full-time homemaker		3.4	3.4	0.0		2.8	2.8	0.0
Retired		25.3	25.3	0.0		35.9	35.9	0.0
Unemployed		28.2	28.2	0.0		15.1	15.1	0.0
Student		0.8	0.8	0.0		0.3	0.3	0.0
Other		5.3	5.4	0.1		9.7	9.7	0.0
Annual income (\$)	incomg				tot_income			
<15,000 (%)		47.7	47.6	-0.1		41.0	41.0	0.0
15,000-39,999 (%)		25.6	25.6	0.0		26.2	26.2	0.0
40,000+ (%)		8.2	8.2	0.0		10.7	10.7	0.0
Declined		18.5	18.6	0.1		22.1	22.1	0.0

AASK Cohort

	trial variables*	1. Trial Baseline			cohort variables**	3. Cohort Baseline		
		Published	Archived	Difference		Published	Archived	Difference
Health insurance (%)	insure				***			
Private/HMO/Other		42.0	42.0	0.0		30.4	30.3	-0.1
Medicaid/Medicare only		24.5	24.5	0.0		40.4	40.3	-0.1
None of the above		33.5	33.5	0.0		29.1	29.4	0.3
Education (%)	edgrp					NA	NA	
Not a HS graduate		40.7	40.7	0.0				
HS graduate		29.9	29.9	0.0				
College or beyond		29.5	29.5	0.0				

* Source: demo.sas7bdat (AASK Trial), which is in the current analysis dataset directory (i.e., not in the previous 4/9/2002 directory).

** Source: c_demo_rep (AASK Cohort), which is the raw dataset associated with form 84, Cohort Demographic and Medical History Form.

***Variables private, hmo, oth_insure, medicaid, medicare in c_demo_rep. Participants with a mix of public and private insurance had to be coded as "none of the above" for numbers to match reasonably well.

Table 2: Health Habits

	trial variables*	1. Trial Baseline			cohort variables**	3. Cohort Baseline		
		Published	Archived	Difference		Published	Archived	Difference
N		1,094	1,094	0		691	691	0
Smoking status (%)	smoke				smoke			
Never smoked		42.1	42.1	0.0		40.7	40.7	0.0
Currently smoking		29.3	29.3	0.0		16.8	16.8	0.0
Past smoker		28.5	28.5	0.0		42.5	42.5	0.0
Patient drinks alcohol (%)	drink	27.8	27.8	0.0	alcohol=2	15.3	15.3	0.0
Patient exercise (%)	exercise	44.1	44.1	0.0	exer_freq***	60.9	60.9	0.0
Baseline ACE inhibitors (%)	ace	39.0	39.0	0.0		73.3	NG	
Baseline β -blockers (%)	beta	28.9	28.9	0.0		39.5	NG	
Baseline calcium channel blockers (%)	ccb	65.0	65.0	0.0		28.5	NG	
Baseline diuretics	diur	63.8	63.8	0.0		78.1	NG	

* Source: demo.sas7bdat (AASK Trial), which is in the current analysis dataset directory (i.e., not in the previous 4/9/2002 directory).

** Source: c_demo_rep (AASK Cohort), which is the raw dataset associated with form 84, Cohort Demographic and Medical History Form.

*** Source: c_exposures_rep (AASK Cohort), which is the raw dataset associated with form 85, Exposures Data Form.

NG = These summary measures were not generated for purposes of this DSIC.

Table 3: Clinical Characteristics

	trial variables*	1. Trial Baseline			cohort variables**	3. Cohort Baseline		
		Published	Archived	Difference		Published	Archived	Difference
N		1,094	1,094	0		691	691	
Weight (kg)	wt	89.5 ± 20.7	89.5 ± 20.7	0 / 0	wt, wt_unit	91.7 ± 22.2	92.2 ± 22.3	0.5 / 0.1
Body mass index (kg/m ²)	bmi	30.6 ± 6.59	30.6 ± 6.59	0 / 0	wt, ht*	31.4 ± 7.10	31.5 ± 7.13	-0.1 / 0.3
Duration of hypertension (y)	htn_year	14.2 ± 10.1	14.2 ± 10.1	0 / 0	htn_year*, rand_dt*, visdt***	19.7 ± 9.86	19.7 ± 9.84	0.0 / 0.02
Systolic BP (mm Hg)	sys	150.3 ± 23.9	150.3 ± 23.9	0 / 0	sbf_avg	136.0 ± 22.0	135.8 ± 22.0	-0.2 / 0.0
Diastolic BP (mm Hg)	dia	95.5 ± 14.2	95.5 ± 14.2	0 / 0	dbp_avg	80.7 ± 12.5	80.9 ± 12.2	0.2 / -0.3
Heart rate (beats/min)	pulse	72.0 ± 12.6	72.0 ± 12.6	0 / 0	puls_sit	68.7 ± 12.0	68.7 ± 12.1	0.0 / 0.1

* Source: demo.sas7bdat (AASK Trial), which is in the current analysis dataset directory (i.e., not in the previous 4/9/2002 directory).

** Source: c_bp_rep (AASK Cohort), which is the raw dataset associated with form 110, Blood Pressure Form.

*** Variable visdt is from c_demo_rep (AASK Cohort).

Table 4: Laboratory Results

	trial variables*	1. Trial Baseline			cohort variables ^a	3. Cohort Baseline		
		Published	Archived	Difference		Published	Archived	Difference
N		1,094	1,094	0		691	691	0
Serum creatinine (mg/dL)	scr	2.0 ± 0.7	2.0 ± 0.7	0 / 0	scr	2.3 ± 1.5	2.3 ± 1.5	0.0 / 0.0
Serum urea nitrogen (mg/dL)	sun	25 ± 10	25 ± 10	0 / 0	sun	30 ± 18	30 ± 18	0 / 0
Serum albumin (g/dL)	alb	4.3 ± 0.4	4.3 ± 0.4	0 / 0	alb	4.0 ± 0.3	4.0 ± 0.3	0.0 / 0.0
Serum glucose (mg/dL)	gluc	95 ± 19	95 ± 18	0 / -1	gluc	101 ± 34	102 ± 35	1 / 1
Total cholesterol (mg/dL)	tchol	211.7 ± 45.5	211.7 ± 45.5	0 / 0	tot_s_c	201.4 ± 46.2	202.0 ± 47.6	0.6 / 0.4
HDL cholesterol (mg/dL)	hdl	48.3 ± 16.1	48.3 ± 16.1	0 / 0	hdl_s_c	47.7 ± 15.4	46.3 ± 14.9	-1.4 / -0.5
Triglycerides (mg/dL)	trig	140.5 ± 80.9	140.5 ± 80.9	0 / 0	trig	143.3 ± 91.3	143.8 ± 95.4	0.5 / 4.1
LDL cholesterol (mg/dL)	ldl	136.4 ± 41.0	136.4 ± 41.0	0 / 0	ldl_s_c	124.6 ± 41.5	126.4 ± 43.1	1.8 / 1.6
Uric acid (mg/dL)	uacid	8.3 ± 2.0	8.3 ± 2.0	0 / 0	uacid	8.8 ± 2.2	8.9 ± 2.2	0.1 / 0.0
Calcium (mg/dL)	cal	9.2 ± 0.5	9.2 ± 0.5	0 / 0	cal	9.6 ± 0.6	9.6 ± 0.6	0.0 / 0.0
Phosphorous (mg/dL)	phos	3.5 ± 0.7	3.5 ± 0.7	0 / 0	p	3.6 ± 0.8	3.6 ± 0.8	0.0 / 0.0
Calcium × phosphorous product	cal, phos	32 ± 7	32 ± 7	0 / 0	cal, p	34 ± 7	34 ± 7	0 / 0
Hematocrit (%)	hct**	39.7 ± 5.16	39.4 ± 4.95	-0.3 / -0.21	hematocrit ^b	38.3 ± 5.49	38.3 ± 5.49	0.0 / 0.00
Serum potassium (mEq/L)	pot	4.2 ± 0.6	4.2 ± 0.6	0 / 0	k	4.3 ± 0.6	4.3 ± 0.6	0.0 / 0.0
Urine total volume (L/day)	uvol	2.21 ± 0.94	2.21 ± 0.94	0 / 0	volume ^c	2.22 ± 1.08	2.22 ± 1.08	0.00 / 0.00
Urine creatinine (g/day)	ucre	1.61 ± 0.64	1.61 ± 0.64	0 / 0	ucr ^d	1.59 ± 0.85	1.59 ± 0.85	0.00 / 0.00
Urine protein (g/day)	upro	0.53 ± 0.94	0.53 ± 0.94	0 / 0	up ^d	0.53 ± 1.05	0.53 ± 1.06	0.00 / -0.01
Urine protein-creatinine ratio (mg/mg)	ratio	0.33 ± 0.52	0.33 ± 0.52	0 / 0	up, ucr	0.38 ± 0.82	0.38 ± 0.82	0.00 / 0.00
Urine urea nitrogen (g/day)	uun	8.29 ± 3.79	8.29 ± 3.79	0 / 0	uun ^d	7.96 ± 4.85	0.72 ± 0.44	7.24 / 4.41 -0.01 / -0.01
Urine sodium (g/day)	usod	3.68 ± 1.98	3.68 ± 1.98	0 / 0	una ^d	3.75 ± 2.89	3.76 ± 2.90	0.01
Urine potassium (g/day)	upot	1.82 ± 0.98	1.82 ± 0.98	0 / 0	uk ^d	1.87 ± 1.11	1.88 ± 1.11	-0.01 / 0.00
Estimated GFR (AASK equation) (mL/min/1.73 m ²)	***	46.8 ± 14.1	46.8 ± 14.1	0 / 0	***	43.8 ± 16.7	44.0 ± 16.7	0.2 / 0.0
Evidence of left ventricular hypertrophy (%)****	lvh_ekg	38.4	43.9	5.5	lvh**** ^e	25.8	20.8	-5.0

- * Source: demo.sas7bdat (AASK Trial), which is in the current analysis dataset directory (i.e., not in the previous 4/9/2002 directory).
- ** Source: basecov.sas7bdat (AASK Trial), which is in the current AASK analysis dataset directory (i.e., not in the previous 4/9/2002 directory).
- *** calculated from the corresponding baseline serum creatinine using formula specified in publication (see p. 81 and reference #9)
- **** excluding missings.

Table 4 Sources

- a Source: c_cbl_result_serum_rep (AASK Cohort), which is the raw dataset associated with form 124, CBL Serum and Plasma Data Form.
- b Source: c_local_lab_rep (AASK Cohort), which is the raw dataset associated with form 113, Local Lab Results (CBC) Form.
- c Source: c_cbl_mailing_urine_rep (AASK Cohort), which is the raw dataset associated with form 123, CBL Urine Mailing (24-hr urine) Form.
- d Source: c_cbl_result_urine_rep (AASK Cohort), which is the raw dataset associated with form 125, CBL Urine Data Form. Data on 24-hour urine volume was used to convert measures to the reported units. Note that the publication reports rates as L/d or g/d. These units were not defined in the publication, but were assumed to mean L/day or g/day.
- e Source: c_demo_rep (AASK Cohort), which is the raw dataset associated with form 84, Cohort Demographic and Medical History Form.

Table 7: Estimated GFR and Urinary Protein-Creatinine Ratio at Baseline

Urine protein-creatinine ratio (mg/mg)	Estimated GFR (mL/min/1.73 m ²)														
	<30			30-40			40-48			>48			All		
	Published n (%)	Archived n (%)	Difference n	Published n (%)	Archived n (%)	Difference n	Published n (%)	Archived n (%)	Difference n	Published n (%)	Archived n (%)	Difference n	Published n (%)	Archived n (%)	Difference n
AASK Main Trial															
<0.08	25 (2.29)	25 (2.29)	0	66 (6.06)	66 (6.06)	0	111 (10.18)	111 (10.18)	0	338 (31.01)	338 (31.01)	0	540 (49.54)	540 (49.54)	0
0.08-0.22	31 (2.84)	31 (2.84)	0	44 (4.04)	44 (4.04)	0	30 (2.75)	30 (2.75)	0	88 (8.07)	88 (8.07)	0	193 (17.71)	193 (17.71)	0
0.22-0.66	38 (3.49)	38 (3.49)	0	49 (4.50)	49 (4.50)	0	26 (2.39)	26 (2.39)	0	64 (5.87)	64 (5.87)	0	177 (16.24)	177 (16.24)	0
>0.66	64 (5.87)	64 (5.87)	0	52 (4.77)	52 (4.77)	0	28 (2.57)	28 (2.57)	0	36 (3.30)	36 (3.30)	0	180 (16.51)	180 (16.51)	0
All	158 (14.50)	158 (14.50)	0	211 (19.36)	211 (19.36)	0	195 (17.89)	195 (17.89)	0	526 (48.26)	526 (48.26)	0	1090 (100.00)	1090 (100.00)	0
AASK Cohort Study															
<0.08	36 (5.37)	37 (5.58)	-1	62 (9.24)	65 (9.80)	-3	66 (9.84)	66 (9.95)	0	206 (30.70)	199 (30.02)	7	370 (55.14)	367 (55.35)	3
0.08-0.22	19 (2.83)	19 (2.87)	0	21 (3.13)	21 (3.17)	0	18 (2.68)	18 (2.71)	0	38 (5.66)	36 (5.43)	2	96 (14.31)	94 (14.18)	2
0.22-0.66	32 (4.77)	32 (4.83)	0	21 (3.13)	22 (3.32)	-1	13 (1.94)	11 (1.66)	2	32 (4.77)	32 (4.83)	0	98 (14.61)	97 (14.63)	1
>0.66	59 (8.79)	56 (8.45)	3	19 (2.83)	20 (3.02)	-1	12 (1.79)	12 (1.81)	0	17 (2.53)	17 (2.56)	0	107 (15.95)	105 (15.84)	2
All	146 (21.76)	144 (21.72)	2	123 (18.33)	128 (19.31)	-5	109 (16.24)	107 (16.14)	2	293 (43.67)	284 (42.84)	9	671 (100.00)	663 (100.00)	8

Sources:

AASK Trial: demo.sas7bdat, which is in the current analysis dataset directory (i.e., not in the previous 4/9/2002 directory).

AASK Cohort: c_cbl_result_serum_rep.sas7bdat, which is the raw dataset associated with form 124, CBL Serum and Plasma Data Form.

Estimated GFR is calculated from the corresponding baseline serum creatinine using formula specified in publication (see publication, p. 81 and reference #9)

9 Appendix 1: Full Text of Sika M, et al. for Approved Data Requestors

Full Text of Sika M, et al. *Am J Kidney Dis.* 50(1), provided to approved data requestors

10 Appendix 2: SAS version 9.2 Log for Programming Code.

SAS version 9.2 Log for programming code submitted for the replication of results in Tables 1, 2, 3, 4, 7 of Sika M, et al. Am J Kidney Dis. 50(1).

AASK Cohort

NOTE: Copyright (c) 2002-2008 by SAS Institute Inc., Cary, NC, USA.

NOTE: SAS (r) Proprietary Software 9.2 (TS2M0)
Licensed to RTI INTL MAIN, Site 70006746.

NOTE: This session is executing on the XP_PRO platform.

NOTE: SAS initialization used:

real time	37.98 seconds
cpu time	2.93 seconds

```
1  options ps=55 ls=75 nonumber formchar='|----|+\---+=|~^<>*' mprint
orientation=portrait;
2
3  *****;
4  * to get AASK trial data *;
5  *****;
6
7  * according to the AASK Trial DSIC, the current Trial datasets, not
former versions,
8  should be used for analysis in the Cohort baseline paper.;
9
10 * these are current versions of the ANALYSIS datasets *;
11 libname asktri 'C:\Documents and Settings\stan\My
Documents\DATA\NIDDK\AASK_Trial';
NOTE: Libref AASKTRI was successfully assigned as follows:
      Engine:          V9
      Physical Name:  C:\Documents and Settings\stan\My
Documents\DATA\NIDDK\AASK_Trial
12
13
14 * n=1094 Trial participants, with rx date, time to
death/dialysis/administrative censoring,
14 ! etc. *;
15 data evnt_all; set asktri.evnt_all;
NOTE: Data file AASKTRI.EVNT_ALL.DATA is in a format that is native to
another host, or the file
      encoding does not match the session encoding. Cross Environment Data
Access will be used,
      which might require additional CPU resources and might reduce
performance.

NOTE: There were 1094 observations read from the data set AASKTRI.EVNT_ALL.
NOTE: The data set WORK.EVNT_ALL has 1094 observations and 25 variables.
NOTE: DATA statement used (Total process time):
      real time          1.23 seconds
      cpu time           0.17 seconds

16  proc sort; by pid;
17
18  * AASK trial demographic data *;
```

AASK Cohort

NOTE: There were 1094 observations read from the data set WORK.EVNT_ALL.

NOTE: The data set WORK.EVNT_ALL has 1094 observations and 25 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time	0.45 seconds
cpu time	0.01 seconds

```
19 data demo; set aasktri.demo;
```

NOTE: Data file AASKTRI.DEMO.DATA is in a format that is native to another host, or the file

encoding does not match the session encoding. Cross Environment Data Access will be used,

which might require additional CPU resources and might reduce performance.

NOTE: There were 2802 observations read from the data set AASKTRI.DEMO.

NOTE: The data set WORK.DEMO has 2802 observations and 102 variables.

NOTE: DATA statement used (Total process time):

real time	0.03 seconds
cpu time	0.03 seconds

```
20 proc sort; by pid;
```

```
21
```

```
22 * to limit demographic data to just the 1094 Trial participants *;
```

NOTE: There were 2802 observations read from the data set WORK.DEMO.

NOTE: The data set WORK.DEMO has 2802 observations and 102 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time	0.04 seconds
cpu time	0.01 seconds

```
23 data demo_trial; merge demo evnt_all(in=in1 keep=pid);
```

```
24 by pid;
```

```
25 if in1;
```

```
26
```

```
27 * calculating calcium*phosphorous product *;
```

```
28 calp=cal* phos;
```

```
29
```

```
30 * calculating eGFR (see
```

```
30 ! http://www.kidney.org/professionals/kdoqi/guidelines\_ckd/p5\_lab\_g4.htm
```

```
31 about the preference of calculating eGFR from estimating equations  
instead of obtaining
```

```
31 ! from creatinine clearance);
```

```
32 * per methods in current paper, this is calculated from AASK Baseline  
patients,
```

```
33 and reported in J Am Soc Nephrol 15: 3175-3183, 2004 *;
```

```
34
```

```
35 ratio_cat=1+(ratio>=.08)+(ratio>=0.22)+(ratio>=.66);
```

```
36 if ratio=. then ratio_cat=.;
```

```
37
```

```
38 if gender=1 then
```

```
39 eGFR_trialbase=329*(scr**-1.096)*(age_r**-0.294);
```

```
40 else if gender=2 then
```

```
41 eGFR_trialbase=329*(scr**-1.096)*(age_r**-0.294)*.736;
```

```
42
```

AASK Cohort

```
43
eGFR_trialbase_cat=1+(eGFR_trialbase>=30)+(eGFR_trialbase>=40)+(eGFR_trialbase>=48);
44     if eGFR_trialbase=. then eGFR_trialbase_cat=.;
45
46     * AASK trial lab data *;
47     * Contains 1094 records providing values for selected baseline variables
for the 1094
47 ! randomized patients *;
48     * however, the values in this dataset do not always match what is in the
publication *;
49     * values in the DEMO dataset do match the pub *;
```

```
NOTE: There were 2802 observations read from the data set WORK.DEMO.
NOTE: There were 1094 observations read from the data set WORK.EVNT_ALL.
NOTE: The data set WORK.DEMO_TRIAL has 1094 observations and 106 variables.
NOTE: DATA statement used (Total process time):
      real time          0.17 seconds
      cpu time           0.04 seconds
```

```
50     data basecov; set aasktri.basecov;
NOTE: Data file AASKTRI.BASECOV.DATA is in a format that is native to another
host, or the file
      encoding does not match the session encoding. Cross Environment Data
Access will be used,
      which might require additional CPU resources and might reduce
performance.
```

```
NOTE: There were 1094 observations read from the data set AASKTRI.BASECOV.
NOTE: The data set WORK.BASECOV has 1094 observations and 62 variables.
NOTE: DATA statement used (Total process time):
      real time          0.04 seconds
      cpu time           0.03 seconds
```

```
51     proc sort; by pid;
52
53     * add hematocrit to the dataset *;
```

```
NOTE: There were 1094 observations read from the data set WORK.BASECOV.
NOTE: The data set WORK.BASECOV has 1094 observations and 62 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time          0.04 seconds
      cpu time           0.03 seconds
```

```
54     data demo_trial; merge demo_trial(in=in1) basecov(keep=pid hct);
55     by pid;
56     if in1; run;
```

```
NOTE: There were 1094 observations read from the data set WORK.DEMO_TRIAL.
NOTE: There were 1094 observations read from the data set WORK.BASECOV.
NOTE: The data set WORK.DEMO_TRIAL has 1094 observations and 107 variables.
NOTE: DATA statement used (Total process time):
      real time          0.03 seconds
      cpu time           0.03 seconds
```


AASK Cohort

```
57
58
59 * table 1 col. 1 *;
60 title Table 1 col 1 (AASK Trial);
61 proc means data=demo_trial mean std; var age_r; run;
```

NOTE: There were 1094 observations read from the data set WORK.DEMO_TRIAL.
NOTE: PROCEDURE MEANS used (Total process time):
real time 1.09 seconds
cpu time 0.07 seconds

```
62 proc freq data=demo_trial; tables gender mar_stat alone work incomg
insure edgrp; run;
```

NOTE: There were 1094 observations read from the data set WORK.DEMO_TRIAL.
NOTE: PROCEDURE FREQ used (Total process time):
real time 0.04 seconds
cpu time 0.06 seconds

```
63 * table 2 col. 1 *;
64 title Table 2 col 1 (AASK Trial);
65 proc freq data=demo_trial; tables smoke drink exercise ace beta ccb
diur; run;
```

NOTE: There were 1094 observations read from the data set WORK.DEMO_TRIAL.
NOTE: PROCEDURE FREQ used (Total process time):
real time 0.06 seconds
cpu time 0.06 seconds

```
66 * table 3 col. 1 *;
67 title Table 3 col 1 (AASK Trial);
68 proc means data=demo_trial mean std; var wt bmi htn_year sys dia
pulse; run;
```

NOTE: There were 1094 observations read from the data set WORK.DEMO_TRIAL.
NOTE: PROCEDURE MEANS used (Total process time):
real time 0.04 seconds
cpu time 0.06 seconds

```
69 * table 4 col. 1... are all variables there? *;
70 title Table 4 col 1 (AASK Trial);
71 proc means data=demo_trial n mean std; var scr sun alb gluc tchol hdl
trig ldl UACID cal
71 ! phos calp hct
72 /* calcium phosphorus product not already in DEMO, calculated above */
73 /* hematocrit not in DEMO, grab from BASECOV dataset, but values don't
exactly match */
74 pot uvol ucre upro ratio UUN usod upot; run;
```

NOTE: There were 1094 observations read from the data set WORK.DEMO_TRIAL.
NOTE: PROCEDURE MEANS used (Total process time):

AASK Cohort

real time 0.04 seconds
cpu time 0.03 seconds

75 proc means data=demo_trial; var glgfr g2gfr MBGFR eGFR_trialbase; run;

NOTE: There were 1094 observations read from the data set WORK.DEMO_TRIAL.

NOTE: PROCEDURE MEANS used (Total process time):

real time 0.04 seconds
cpu time 0.04 seconds

76 * MBGFR is the same as in asktri.evnt_all... but it is not what is reported in paper *;

77 * calculated GFR from the equation appears to be reported in the paper *;

78

79 * CANNOT REPLICATE EVIDENCE OF LVH, EVEN WHEN ASSUMING MISSING LVH = NO EVIDENCE OF LVH *

79 ! ;

80 proc freq data=demo_trial; tables lvh_ekg; run;

NOTE: There were 1094 observations read from the data set WORK.DEMO_TRIAL.

NOTE: PROCEDURE FREQ used (Total process time):

real time 0.04 seconds
cpu time 0.01 seconds

81

82 * table 7, part 1 (exact match) *;

83 title Table 1 part 1 (AASK Trial);

84 proc freq; tables ratio_cat*eGFR_trialbase_cat/nocol norow; run;

NOTE: There were 1094 observations read from the data set WORK.DEMO_TRIAL.

NOTE: PROCEDURE FREQ used (Total process time):

real time 0.12 seconds
cpu time 0.06 seconds

85

86

87

88 *****;

89 * to get AASK cohort data *;

90 *****;

91

92 libname askform 'C:\Documents and Settings\stan\My Documents\DATA\NIDDK\AASK_Cohort\SAS

92 ! forms data';

NOTE: Libref AASKFORM was successfully assigned as follows:

Engine: V9
Physical Name: C:\Documents and Settings\stan\My Documents\DATA\NIDDK\AASK_Cohort\SAS forms data

93 libname askkana 'C:\Documents and Settings\stan\My

93 ! Documents\DATA\NIDDK\AASK_Cohort\AnalyData';

NOTE: Libref AASKANA was successfully assigned as follows:

AASK Cohort

```
Engine:          V9
Physical Name:   C:\Documents and Settings\stan\My
Documents\DATA\NIDDK\AASK_Cohort\AnalyData
94
95     title AASK Cohort analyses;
96
97     * get AASK rx date *;
98     data randomdate; set evnt_all;
99     keep pid rand_dt;
```

NOTE: There were 1094 observations read from the data set WORK.EVNT_ALL.
NOTE: The data set WORK.RANDOMDATE has 1094 observations and 2 variables.
NOTE: DATA statement used (Total process time):

real time	0.04 seconds
cpu time	0.01 seconds

```
100    proc sort; by pid;
101
102    * analysis dataset for all 1094 AASK trial subjects: has outcomes and
survival times
102! through the COHORT period *;
103    * does not have GFR data *;
```

NOTE: There were 1094 observations read from the data set WORK.RANDOMDATE.
NOTE: The data set WORK.RANDOMDATE has 1094 observations and 2 variables.
NOTE: PROCEDURE SORT used (Total process time):

real time	0.03 seconds
cpu time	0.03 seconds

```
104    data aaskcohort; set aaskana.cohort_analysis;
NOTE: Data file AASKANA.COHORT_ANALYSIS.DATA is in a format that is native to
another host, or
the file encoding does not match the session encoding. Cross
Environment Data Access will
be used, which might require additional CPU resources and might reduce
performance.
```

NOTE: There were 1094 observations read from the data set
AASKANA.COHORT_ANALYSIS.
NOTE: The data set WORK.AASKCOHORT has 1094 observations and 9 variables.
NOTE: DATA statement used (Total process time):

real time	0.04 seconds
cpu time	0.04 seconds

```
105    proc sort; by pid;
106
107    * has consent date for 691 AASK Cohort subjects, use this to determine
who enrolled in
107! Cohort study *;
```

NOTE: There were 1094 observations read from the data set WORK.AASKCOHORT.
NOTE: The data set WORK.AASKCOHORT has 1094 observations and 9 variables.
NOTE: PROCEDURE SORT used (Total process time):

real time	0.03 seconds
-----------	--------------

AASK Cohort

cpu time 0.03 seconds

```
108 data cohortenroll; set aaskform.c_enroll_rep;
```

NOTE: There were 691 observations read from the data set
AASKFORM.C_ENROLL_REP.

NOTE: The data set WORK.COHORTENROLL has 691 observations and 10 variables.

NOTE: DATA statement used (Total process time):

real time 0.07 seconds

cpu time 0.03 seconds

```
109 proc sort; by pid;
```

```
110
```

```
111 * to get reasons not enrolled in AASK study (form 83) *;
```

NOTE: There were 691 observations read from the data set WORK.COHORTENROLL.

NOTE: The data set WORK.COHORTENROLL has 691 observations and 10 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.01 seconds

cpu time 0.00 seconds

```
112 data cohort_notenroll; set aaskform.c_reason_not_enroll_rep;
```

NOTE: There were 95 observations read from the data set
AASKFORM.C_REASON_NOT_ENROLL_REP.

NOTE: The data set WORK.COHORT_NOTENROLL has 95 observations and 6 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.00 seconds

```
113 proc sort; by pid;
```

```
114
```

NOTE: There were 95 observations read from the data set
WORK.COHORT_NOTENROLL.

NOTE: The data set WORK.COHORT_NOTENROLL has 95 observations and 6 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time 0.01 seconds

cpu time 0.00 seconds

```
115 data aaskcohort; merge aaskcohort(in=in1) randomdate cohortenroll(in=in2  
keep=pid)
```

```
116 cohort_notenroll(in=in3 keep=pid--reason_no);
```

```
117 by pid;
```

```
118 if in1;
```

```
119 if in2 then cohortenroll=1; else cohortenroll=0;
```

```
120 if in3 then reasonnotenroll=1; else reasonnotenroll=0;run;
```

NOTE: There were 1094 observations read from the data set WORK.AASKCOHORT.

NOTE: There were 1094 observations read from the data set WORK.RANDOMDATE.

NOTE: There were 691 observations read from the data set WORK.COHORTENROLL.

AASK Cohort

NOTE: There were 95 observations read from the data set

WORK.COHORT_NOTENROLL.

NOTE: The data set WORK.AASKCOHORT has 1094 observations and 14 variables.

NOTE: DATA statement used (Total process time):

real time	0.04 seconds
cpu time	0.04 seconds

```
121
122 data aaskcohort; set aaskcohort;
123   daystoend=al_mos_xd*30.4; *Month to Dialysis, Death or ADMDT (admin
124   censor date) from AASK
123! cohort_analysis dataset *;
124   dateend=rand_dt+daystoend;
125   format dateend MMDDYY10.;
126
127   * for Figure 1: participant flow *;
128   period=.;
129
130   if '01AUG1995'D<dateend<='30SEP2001'D then period=1; * censored or
131   reached outcome during
130! Trial *;
131   else if cohortenroll=1 then period=3; * censored or reached outcome
132   during Cohort *;
132   else period=2; * censored or reached outcome during Transition period
133   *;
133
134   if period=2 and al_evt_d=0 and al_evt_xd=0 then do;
135     * ltfu during transition period, versus not LTFU (but did not enroll
136     in Cohort study) *;
136     * the 24 LTFU during transition period are those with enddate in
137     January 2002 *;
137     * the 73 who did not enroll in cohort study have an enddate after
138     Cohort start date of
137! April 2002 *;
138     if .<dateend<'01APR2002'D then ltfu_trans=1;
139     else if dateend>='01APR2002'D then ltfu_trans=0;
140   end;
141
142
```

NOTE: There were 1094 observations read from the data set WORK.AASKCOHORT.

NOTE: The data set WORK.AASKCOHORT has 1094 observations and 18 variables.

NOTE: DATA statement used (Total process time):

real time	0.15 seconds
cpu time	0.00 seconds

```
143   proc sort; by period;
144   * to replicate Figure 1, Participant flow *;
145   * al_evt_d=ESRD alone (0/1)
146     al_evt_xd=ESRD or death (0/1)
147     to get #s of persons who reached ESRD, Deaths (no ESRD), ltfu;
148   title AASK Cohort analyses;
149   title2 to replicate Figure 1, Participant flow [al_evt_d=ESRD alone
150   (0/1), al_evt_xd=ESRD
151   or death (0/1)];
```

AASK Cohort

NOTE: There were 1094 observations read from the data set WORK.AASKCOHORT.
NOTE: The data set WORK.AASKCOHORT has 1094 observations and 18 variables.
NOTE: PROCEDURE SORT used (Total process time):
 real time 0.03 seconds
 cpu time 0.03 seconds

```
150  proc freq; by period; tables al_evt_d al_evt_xd ltfu_trans  
150! al_evt_d*al_evt_xd*ltfu_trans/list missing; run;
```

NOTE: There were 1094 observations read from the data set WORK.AASKCOHORT.
NOTE: PROCEDURE FREQ used (Total process time):
 real time 0.04 seconds
 cpu time 0.01 seconds

```
151  
152  
153  * Cohort demographics *;  
154  data demo_cohort; set askform.c_demo_rep;  
155  cohort_demo_date=DATEPART(visdt);  
156  format cohort_demo_date MMDDYY10.;
```

NOTE: There were 680 observations read from the data set AASKFORM.C_DEMO_REP.
NOTE: The data set WORK.DEMO_COHORT has 680 observations and 110 variables.
NOTE: DATA statement used (Total process time):
 real time 0.26 seconds
 cpu time 0.01 seconds

```
157  proc sort; by pid;  
158  
159  * Exposures *;
```

NOTE: There were 680 observations read from the data set WORK.DEMO_COHORT.
NOTE: The data set WORK.DEMO_COHORT has 680 observations and 110 variables.
NOTE: PROCEDURE SORT used (Total process time):
 real time 0.01 seconds
 cpu time 0.00 seconds

```
160  data exp_cohort; set askform.c_exposures_rep;
```

NOTE: There were 3292 observations read from the data set
AASKFORM.C_EXPOSURES_REP.
NOTE: The data set WORK.EXP_COHORT has 3292 observations and 23 variables.
NOTE: DATA statement used (Total process time):
 real time 0.03 seconds
 cpu time 0.00 seconds

```
161  proc sort; by pid visn;
```

NOTE: There were 3292 observations read from the data set WORK.EXP_COHORT.
NOTE: The data set WORK.EXP_COHORT has 3292 observations and 23 variables.
NOTE: PROCEDURE SORT used (Total process time):

AASK Cohort

```
real time          0.03 seconds
cpu time           0.01 seconds
```

```
162 data exp_cohort_base; set exp_cohort; by pid visn;
163   if first.pid;
```

NOTE: There were 3292 observations read from the data set WORK.EXP_COHORT.
NOTE: The data set WORK.EXP_COHORT_BASE has 685 observations and 23 variables.

NOTE: DATA statement used (Total process time):

```
real time          0.01 seconds
cpu time           0.01 seconds
```

```
164   proc freq; tables visn; run;
```

NOTE: There were 685 observations read from the data set WORK.EXP_COHORT_BASE.

NOTE: PROCEDURE FREQ used (Total process time):

```
real time          0.03 seconds
cpu time           0.00 seconds
```

```
165 data exp_cohort_base; set exp_cohort_base;
166   if .<visn<=12; run;
```

NOTE: There were 685 observations read from the data set WORK.EXP_COHORT_BASE.

NOTE: The data set WORK.EXP_COHORT_BASE has 682 observations and 23 variables.

NOTE: DATA statement used (Total process time):

```
real time          0.01 seconds
cpu time           0.00 seconds
```

```
167
168   * data on BP (longitudinal) *;
169 data bslnbp; set aaskform.c_bp_rep; if visn=0;
170   bpdt=DATEPART(bp_dt);
171   format bpdt mmddy10.;
172   if wt_unit=1 then wtkg=wt;
173   else if wt_unit=2 then wtkg=wt/2.2046; run;
```

NOTE: There were 13349 observations read from the data set AASKFORM.C_BP_REP.
NOTE: The data set WORK.BSLNBP has 1288 observations and 30 variables.

NOTE: DATA statement used (Total process time):

```
real time          0.01 seconds
cpu time           0.01 seconds
```

```
174   proc sort data=bslnbp; by pid bpdt;
```

NOTE: There were 1288 observations read from the data set WORK.BSLNBP.
NOTE: The data set WORK.BSLNBP has 1288 observations and 30 variables.

NOTE: PROCEDURE SORT used (Total process time):

```
real time          0.06 seconds
```

AASK Cohort

cpu time 0.03 seconds

```
175 data bslnbp; set bslnbp; by pid bpdt; IF FIRST.PID;
176 *directions on form: if several are entered on one day, enter the
first and fax the rest
176! to the DCC *;
177
178 * get hematocrit from the CBL which is a separate dataset from serum
results *;
```

NOTE: There were 1288 observations read from the data set WORK.BSLNBP.

NOTE: The data set WORK.BSLNBP has 628 observations and 30 variables.

NOTE: DATA statement used (Total process time):

real time	0.01 seconds
cpu time	0.00 seconds

```
179 data cbl; set aaskform.c_local_lab_rep;
180 cbl_drawdt=DATEPART(DRAW_DT); format cbl_drawdt mmdyy10.;
181 visitnum=.;
182 visitnum=visn;
```

NOTE: There were 3130 observations read from the data set

AASKFORM.C_LOCAL_LAB_REP.

NOTE: The data set WORK.CBL has 3130 observations and 12 variables.

NOTE: DATA statement used (Total process time):

real time	0.03 seconds
cpu time	0.00 seconds

```
183 proc sort data=cbl; by pid cbl_drawdt;
```

NOTE: There were 3130 observations read from the data set WORK.CBL.

NOTE: The data set WORK.CBL has 3130 observations and 12 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time	0.01 seconds
cpu time	0.01 seconds

```
184 data cblbase; set cbl; by pid cbl_drawdt;
185 if first.pid;
```

NOTE: There were 3130 observations read from the data set WORK.CBL.

NOTE: The data set WORK.CBLBASE has 682 observations and 12 variables.

NOTE: DATA statement used (Total process time):

real time	0.06 seconds
cpu time	0.00 seconds

```
186 data cblbase1; set cblbase; if .<visitnum<=12; run;
```

NOTE: There were 682 observations read from the data set WORK.CBLBASE.

NOTE: The data set WORK.CBLBASE1 has 676 observations and 12 variables.

NOTE: DATA statement used (Total process time):

real time	0.03 seconds
cpu time	0.00 seconds

AASK Cohort

```
187
188   * CBL serum/plasma results *;
189 data cbl_blood_base; set aaskform.c_cbl_result_serum_rep;
190   * if visn=0; *if select visn=0, then only get 663 obs... table 7 shows
n=671 min *;
191   serum_drawdt=DATEPART(DRAW_DT); format serum_drawdt mmddy10.;
192   visitnum=.;
193   visitnum=visn;
```

NOTE: There were 6259 observations read from the data set
AASKFORM.C_CBL_RESULT_SERUM_REP.

NOTE: The data set WORK.CBL_BLOOD_BASE has 6259 observations and 44
variables.

NOTE: DATA statement used (Total process time):
real time 0.06 seconds
cpu time 0.01 seconds

```
194 proc sort; by pid visitnum serum_drawdt;
```

NOTE: There were 6259 observations read from the data set
WORK.CBL_BLOOD_BASE.

NOTE: The data set WORK.CBL_BLOOD_BASE has 6259 observations and 44
variables.

NOTE: PROCEDURE SORT used (Total process time):
real time 0.04 seconds
cpu time 0.01 seconds

```
195 data cbl_blood_base; set cbl_blood_base; by pid visitnum serum_drawdt;
196   if first.pid; run;
```

NOTE: There were 6259 observations read from the data set
WORK.CBL_BLOOD_BASE.

NOTE: The data set WORK.CBL_BLOOD_BASE has 690 observations and 44 variables.

NOTE: DATA statement used (Total process time):
real time 0.04 seconds
cpu time 0.00 seconds

```
197 data cbl_blood_base1; set cbl_blood_base; if .<visitnum<=12; run;
```

NOTE: There were 690 observations read from the data set WORK.CBL_BLOOD_BASE.

NOTE: The data set WORK.CBL_BLOOD_BASE1 has 689 observations and 44
variables.

NOTE: DATA statement used (Total process time):
real time 0.03 seconds
cpu time 0.01 seconds

```
198
199   * CBL urine volume *;
200 data cbl_urine_vol; set aaskform.c_cbl_mailing_urine_rep;
201   urine_drawdt=DATEPART(START_DT); format urine_drawdt mmddy10.;
202   visitnum=.;
```

AASK Cohort

```
203 visitnum=visn;
```

NOTE: There were 3063 observations read from the data set
AASKFORM.C_CBL_MAILING_URINE_REP.

NOTE: The data set WORK.CBL_URINE_VOL has 3063 observations and 17 variables.

NOTE: DATA statement used (Total process time):

```
real time          0.00 seconds
cpu time           0.00 seconds
```

```
204 proc sort; by pid urine_drawdt;
```

```
205
```

```
206 * CBL urine results *;
```

NOTE: There were 3063 observations read from the data set WORK.CBL_URINE_VOL.

NOTE: The data set WORK.CBL_URINE_VOL has 3063 observations and 17 variables.

NOTE: PROCEDURE SORT used (Total process time):

```
real time          0.04 seconds
cpu time           0.01 seconds
```

```
207 data cbl_urine_base; set aaskform.c_cbl_result_urine_rep;
```

```
208 urine_drawdt=DATEPART(START_DT); format urine_drawdt mmddyy10.;
```

```
209 visitnum=.;
```

```
210 visitnum=visn;
```

```
211 * if visitnum=0;
```

NOTE: Character values have been converted to numeric values at the places
given by:

```
(Line):(Column).
210:12
```

NOTE: There were 3063 observations read from the data set

AASKFORM.C_CBL_RESULT_URINE_REP.

NOTE: The data set WORK.CBL_URINE_BASE has 3063 observations and 21
variables.

NOTE: DATA statement used (Total process time):

```
real time          0.01 seconds
cpu time           0.00 seconds
```

```
212 proc sort; by pid urine_drawdt;
```

```
213
```

```
214 * perfect match when match on drawdt, visitnum will not always match!
*;
```

NOTE: There were 3063 observations read from the data set
WORK.CBL_URINE_BASE.

NOTE: The data set WORK.CBL_URINE_BASE has 3063 observations and 21
variables.

NOTE: PROCEDURE SORT used (Total process time):

```
real time          0.04 seconds
cpu time           0.00 seconds
```

```
215 data cbl_urine_base1; merge cbl_urine_base(rename=(visitnum=visitnum1))
```

```
215! cbl_urine_vol(keep=pid visitnum urine_drawdt volume);
```

```
216 by pid urine_drawdt; run;
```

AASK Cohort

NOTE: There were 3063 observations read from the data set
WORK.CBL_URINE_BASE.

NOTE: There were 3063 observations read from the data set WORK.CBL_URINE_VOL.

NOTE: The data set WORK.CBL_URINE_BASE1 has 3063 observations and 23
variables.

NOTE: DATA statement used (Total process time):

real time	0.03 seconds
cpu time	0.01 seconds

```
217
218 title AASK Cohort analyses;
219 title2 to compare visit number in urine mailing form versus visit
number in data form;
220 proc freq ; tables visitnum1*visitnum/list missing; run;
```

NOTE: There were 3063 observations read from the data set
WORK.CBL_URINE_BASE1.

NOTE: PROCEDURE FREQ used (Total process time):

real time	0.07 seconds
cpu time	0.00 seconds

```
221 * visitnum in mailing form is sometimes lower (closer to bsln) than
visitnum in data form *
221! ;
222 * choose visitnum in mailing form.... so can ID bsln visit in more
subjects *;
223
224 data cbl_urine_base1; set cbl_urine_base1;
225 * if visitnum=0;*if select visitnum=0, then only get 581 obs... table 7
shows n=671 min *;
226 by pid urine_drawdt;
227 if first.pid;
228 if .<visitnum<=12;
229 /* 8 had baseline visit more than a year after baseline, assume that
these
230 will be deleted **/; run;
```

NOTE: There were 3063 observations read from the data set
WORK.CBL_URINE_BASE1.

NOTE: The data set WORK.CBL_URINE_BASE1 has 670 observations and 23
variables.

NOTE: DATA statement used (Total process time):

real time	0.03 seconds
cpu time	0.01 seconds

```
231
232 data cbl_urine_base1; set cbl_urine_base1;
233 up24=up*volume/100;
234 ucr24=ucr*volume/100;
235 ual24=ual*volume/100;
236 uun24=uun*volume/100;
237
```

AASK Cohort

```

238   una24mol=una*volume*22.99/1000000; * converting from mmol/L to g/24hr
*;
239           * molar mass of sodium is 22.99g/mol
*;
240           * 1000 mmol in 1 mole *;
241   uk24mol=uk*volume*39.1/1000000; * converting from mmol/L to g/24hr *;
242           * molar mass of potassium is
39.10g/mol *;
243           * 1000 mmol in 1 mole *;
244
245   label up24=urinary protein, mg/24hr
246         ucr24=urinary creatinine, mg/24hr
247         ual24=urinary albumine, mg/24hr
248         uun24=urinary urea nitrogen, mg/24hr
249         una24mol=urinary sodium, g/24hr
250         uk24mol=urinary potassium, g/24hr;
251   run;

```

NOTE: Missing values were generated as a result of performing an operation on missing values.

Each place is given by: (Number of times) at (Line):(Column).

1 at 235:12

NOTE: There were 670 observations read from the data set

WORK.CBL_URINE_BASE1.

NOTE: The data set WORK.CBL_URINE_BASE1 has 670 observations and 29 variables.

NOTE: DATA statement used (Total process time):

real time 0.01 seconds

cpu time 0.03 seconds

```

252
253   data demo_cohort; merge demo_cohort cohortenroll(in=in2 keep=pid)
bslnbp(drop=wt wt_unit)
254   cblbasel(keep=pid hematocrit)
255   cbl_blood_basel(drop=visn rename=(cc_n=blood_cc_n))
cbl_urine_basel(drop=visn visitnum1)
256   demo_trial(keep=pid dob gender ht lvh_ekg htn_year RAND_DT) /* fixed
variables from aask
256! trial */
257   exp_cohort_base(keep=pid exer_freq);
258   by pid;
259   if in2;
260
261   age_cohort_base=round(((cohort_demo_date-dob)/365.25),0.1);
262   yearsdiff=round(((cohort_demo_date-rand_dt)/365.25),0.1);
263
264   if exer_freq>0 then exercise=1; else if exer_freq=0 then exercise=0;
265
266   htn_year_cohort=htn_year+yearsdiff;
267
268   bmi_cohort_base=wtkg*10000/(ht*ht);
269   rename wtkg=wt_cohort_base;
270
271   lvh_cohort=lvh; * lvh is from cohort demo dataset *;
272
273   ANYLVHsum=sum(lvh_ekg,lvh_cohort);

```

AASK Cohort

```

274   if anylvhsum>0 then ANYLVH=1;
275   else if anylvhsum=0 then ANYLVH=0;
276
277   run;

```

WARNING: Multiple lengths were specified for the variable VIST by input data set(s). This may cause truncation of data.

NOTE: Missing values were generated as a result of performing an operation on missing values.

Each place is given by: (Number of times) at (Line):(Column).
 11 at 261:19 11 at 261:43 11 at 261:48 11 at 262:13 11 at
 262:37 11 at 262:46
 17 at 266:27 83 at 268:23 56 at 273:13

NOTE: There were 680 observations read from the data set WORK.DEMO_COHORT.
 NOTE: There were 691 observations read from the data set WORK.COHORTENROLL.
 NOTE: There were 628 observations read from the data set WORK.BSLNBP.
 NOTE: There were 676 observations read from the data set WORK.CBLBASE1.
 NOTE: There were 689 observations read from the data set
 WORK.CBL_BLOOD_BASE1.
 NOTE: There were 670 observations read from the data set
 WORK.CBL_URINE_BASE1.
 NOTE: There were 1094 observations read from the data set WORK.DEMO_TRIAL.
 NOTE: There were 682 observations read from the data set
 WORK.EXP_COHORT_BASE.
 NOTE: The data set WORK.DEMO_COHORT has 691 observations and 208 variables.
 NOTE: DATA statement used (Total process time):
 real time 0.35 seconds
 cpu time 0.06 seconds

```

278 * comparing LVH at trial bsln (lvh_ekg) to LVH at cohort bsln (lvh) *;
279   title AASK Cohort analyses;
280   title2 to compare LVH at trial bsln (lvh_ekg) to LVH at cohort bsln
(lvh);
281 proc freq data=demo_cohort; tables lvh_ekg lvh anylvh
lvh_ekg*lvh*anylvh ANYLVH/list missing
281! ; run;

```

NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.
 NOTE: PROCEDURE FREQ used (Total process time):
 real time 0.04 seconds
 cpu time 0.01 seconds

```

282
283 data demo_cohort; set demo_cohort;
284
285 * attempt to replicate insurance variable from variables in demo
dataset *;
286   if private=1 or hmo=1 or oth_insure=1 then privhmooth=1;
287   else if private=0 and hmo=0 and oth_insure=0 then privhmooth=0;
288
289   if private=1 or hmo=1 then privhmo=1;
290   else if private=0 and hmo=0 then privhmo=0;
291
292 * switch to medicaid/medicare only *;

```

AASK Cohort

```

293   insurancetype=.;
294   if (private=1 or hmo=1 or oth_insure=1) and (medicaid=0 and
medicare=0) then
294! insurancetype=1;
295   else if (medicaid=1 or medicare=1) and (private=0 and hmo=0 and
oth_insure=0) then
295! insurancetype=2;
296   else if private>. and hmo>. and oth_insure>. and medicaid>. and
medicare>. then
296! insurancetype=3;
297   label insurancetype='1=private hmo other 2=medicaid medicare ONLY
3=none of above/missing';
298
299   * employment *;
300   if employ_stats in (1,2,9) then employ_6=1;
301   else if employ_stats in (3) then employ_6=2;
302   else if employ_stats in (4) then employ_6=3;
303   else if employ_stats in (5,6) then employ_6=4;
304   else if employ_stats in (7,8) then employ_6=5;
305   else if employ_stats in (10) then employ_6=6;
306   label employ_6 = '1=working 2=homemaker 3=retired 4=unemployed
5=student 6=other';
307
308   * income *;
309   if 1<=tot_income<=3 then tot_inc_3=1;
310   else if 4<=tot_income<=6 then tot_inc_3=2;
311   else if 7<=tot_income<=10 then tot_inc_3=3;
312   else if tot_income=11 then tot_inc_3=4;
313   label tot_inc_3='1=lt 15K 2=15 to <40K 3=40k+ 4=declined';
314
315   * marital status *;
316   if marital=1 then mari_4=1;
317   else if marital in (2,3) then mari_4=2;
318   else if marital=4 then mari_4=3;
319   else if marital=5 then mari_4=4;
320   label mari_4='1=never 2=married/married like 3=div/sep 4=wid';
321
322   calp=cal*p;
323   * to get calcium*phosphorous product *;
324
325   uprocreat=up/ucr;
326   * to get urine protein:creatinine, mg/mg *;
327
328   uprocreat_cat=1+(uprocreat>=.08)+(uprocreat>=0.22)+(uprocreat>=.66);
329   if uprocreat=. then uprocreat_cat=.;
330
331   * calculating eGFR (see
331! http://www.kidney.org/professionals/kdoqi/guidelines\_ckd/p5\_lab\_g4.htm
332   about the preference of calculating eGFR from estimating equations
instead of obtaining
332! from creatinine clearance);
333   * per methods in current paper, this is calculated from AASK Baseline
patients,
334   and reported in J Am Soc Nephrol 15: 3175-3183, 2004 *;
335   if gender=1 then
336     eGFR=329*(scr**-1.096)*(age_cohort_base**-0.294);
337   else if gender=2 then

```

AASK Cohort

```
338     eGFR=329*(scr**-1.096)*(age_cohort_base**-0.294)*.736;
339
340     eGFR_cat=1+(eGFR>=30)+(eGFR>=40)+(eGFR>=48);
341     if eGFR=. then eGFR_cat=.;
342
343
344     * for table 1, col. 3 (all are rep'd except for insurance) *;
345     title AASK Cohort analyses;
346     title2 for table 1 col 3;
```

NOTE: Missing values were generated as a result of performing an operation on missing values.

Each place is given by: (Number of times) at (Line):(Column).
95 at 322:11 21 at 325:15 1 at 336:13 1 at 336:18 5 at 336:27
5 at 336:44
1 at 338:13 1 at 338:18 6 at 338:27 6 at 338:44

NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.
NOTE: The data set WORK.DEMO_COHORT has 691 observations and 219 variables.
NOTE: DATA statement used (Total process time):
real time 0.07 seconds
cpu time 0.03 seconds

```
347 proc means; var age_cohort_base; run;
```

NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.
NOTE: PROCEDURE MEANS used (Total process time):
real time 0.01 seconds
cpu time 0.01 seconds

```
348 proc freq; tables gender marital mari_4 alone employ_stats employ_6
tot_income tot_inc_3
348! insurancetype; run;
```

NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.
NOTE: PROCEDURE FREQ used (Total process time):
real time 0.01 seconds
cpu time 0.01 seconds

```
349 proc freq; tables tot_income tot_inc_3; *where tot_income^=11; *decline
to say *; run;
```

NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.
NOTE: PROCEDURE FREQ used (Total process time):
real time 0.03 seconds
cpu time 0.01 seconds

```
350 proc freq; tables insurancetype insurancetype*private insurancetype*hmo
350! insurancetype*medicaid insurancetype*medicare
351 insurancetype*oth_insure
352 insurancetype*private*hmo*medicaid*medicare*oth_insure/list; run;
```

NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.
NOTE: PROCEDURE FREQ used (Total process time):

AASK Cohort

```
real time          0.09 seconds
cpu time           0.03 seconds
```

```
353
354   * for table 2, col. 3 *;
355   title2 for table 2 col 3;
356 proc freq; tables smoke alcohol; run;
```

```
NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.
NOTE: PROCEDURE FREQ used (Total process time):
      real time          0.01 seconds
      cpu time           0.01 seconds
```

```
357 proc freq; tables exercise; run;
```

```
NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.
NOTE: PROCEDURE FREQ used (Total process time):
      real time          0.01 seconds
      cpu time           0.00 seconds
```

```
358 proc means; var htn_year_cohort; run;
```

```
NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.
NOTE: PROCEDURE MEANS used (Total process time):
      real time          0.01 seconds
      cpu time           0.01 seconds
```

```
359
360   * exercise data can be found in form 85, exposures *;
361
362   * for table 3, col. 3 *;
363   title2 for table 3 col 3;
364 proc means mean std; var wt_cohort_base bmi_cohort_base sbp_avg dbp_avg
puls_sit; run;
```

```
NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.
NOTE: PROCEDURE MEANS used (Total process time):
      real time          0.01 seconds
      cpu time           0.00 seconds
```

```
365
366   * for table 4, col. 3 *;
367 data demo_cohort; set demo_cohort;
368   title2 for table 4 col 3;
```

```
NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.
NOTE: The data set WORK.DEMO_COHORT has 691 observations and 219 variables.
NOTE: DATA statement used (Total process time):
      real time          0.03 seconds
      cpu time           0.00 seconds
```


AASK Cohort

```
369 proc means n mean std; var scr sun alb gluc tot_s_c hdl_s_c trig ldl_s_c
uacid cal p calp
369! hematocrit k; run;
```

NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.

NOTE: PROCEDURE MEANS used (Total process time):

```
real time      0.01 seconds
cpu time       0.03 seconds
```

```
370 proc means n mean std; var uprocreat; run;
```

NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.

NOTE: PROCEDURE MEANS used (Total process time):

```
real time      0.01 seconds
cpu time       0.01 seconds
```

```
371 proc means n mean std; var volume up24 ucr24 uun24 una24mol uk24mol;
run;
```

NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.

NOTE: PROCEDURE MEANS used (Total process time):

```
real time      0.01 seconds
cpu time       0.00 seconds
```

```
372 proc means n mean std; var eGFR; run;
```

NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.

NOTE: PROCEDURE MEANS used (Total process time):

```
real time      0.01 seconds
cpu time       0.03 seconds
```

```
373 * this is hand-calculated eGFR from serum creat and aforementioned
reference *;
```

```
374
```

```
375 proc freq data=demo_cohort; tables lvh_cohort; run;
```

NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.

NOTE: PROCEDURE FREQ used (Total process time):

```
real time      0.01 seconds
cpu time       0.00 seconds
```

```
376 * cannot replicate LVH in cohort *;
```

```
377
```

```
378 proc means n min max; class uprocreat_cat; var uprocreat; run;
```

NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.

NOTE: PROCEDURE MEANS used (Total process time):

```
real time      0.01 seconds
cpu time       0.01 seconds
```

```
379 proc means n min max; class eGFR_cat; var eGFR; run;
```

AASK Cohort

NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.

NOTE: PROCEDURE MEANS used (Total process time):

real time	0.01 seconds
cpu time	0.03 seconds

380 * for table 7 *;

381 * these n's do not match *;

382 * cant tell which serum&urine records were from aask cohort baseline
*;

383 title2 for table 7;

384 proc freq data=demo_cohort; tables uprocreat_cat*egfr_cat/norow nocol;
run;

NOTE: There were 691 observations read from the data set WORK.DEMO_COHORT.

NOTE: PROCEDURE FREQ used (Total process time):

real time	0.03 seconds
cpu time	0.01 seconds

11 Appendix 3: SAS version 9.2 Output for Programming Code.

SAS version 9.2 Output for programming code submitted for the replication of results in Tables 1, 2, 3, 4, 7 of Sika M, et al. Am J Kidney Dis. 50(1)

Table 1 col 1 (AASK Trial)

The MEANS Procedure

03:45 Wednesday, June 06, 2012

Analysis Variable : age_r Age at Randomization (yrs)	
Mean	Std Dev
54.5956602	10.6690069

Table 1 col 1 (AASK Trial)

The FREQ Procedure

Gender (1=Male, 2=Female)				
GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	669	61.15	669	61.15
2	425	38.85	1094	100.00

Marital Status at SV2, code defined from form 4				
MAR_STAT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	227	20.77	227	20.77
2	352	32.20	579	52.97
3	42	3.84	621	56.82
4	338	30.92	959	87.74
5	134	12.26	1093	100.00

Frequency Missing = 1

Living Alone (0:no,1:yes) at SV2, from form 4				
alone	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	837	77.43	837	77.43
1	244	22.57	1081	100.00

Frequency Missing = 13

Work Status as of SV2, from form 4 (1:Wking,2:Unemployed,3:Retired,4:Homemkr,5:Student/Wking,6:Othr/No Info)				
work	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	404	36.93	404	36.93
2	308	28.15	712	65.08
3	277	25.32	989	90.40
4	37	3.38	1026	93.78
5	9	0.82	1035	94.61
6	59	5.39	1094	100.00

Income Grp (as of SV2, from form 4; 1:<15000,2:15000-39999,3:>=40000,4:decline to provide income)				
incomg	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	521	47.62	521	47.62
2	280	25.59	801	73.22
3	90	8.23	891	81.44
4	203	18.56	1094	100.00

Table 1 col 1 (AASK Trial)

The FREQ Procedure

Insurance (at SV2, from form 4; 1:Private/HMO/Other,2:Medicaid/Medicare,3:None)				
insure	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	459	41.96	459	41.96
2	268	24.50	727	66.45
3	367	33.55	1094	100.00

Education Grp (as of SV2, from form 4; 1:Not HS Grdt,2:HS Grdt,3:College or Beyond)				
edgrp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	444	40.66	444	40.66
2	326	29.85	770	70.51
3	322	29.49	1092	100.00

Frequency Missing = 2

Table 2 col 1 (AASK Trial)

The FREQ Procedure

Smoking Status (0:Never,1:Current,2:Past) as of SV2 (from form 4)				
smoke	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	461	42.14	461	42.14
1	321	29.34	782	71.48
2	312	28.52	1094	100.00

Drinking Status (0:no,1:beer/wine/liquor) as of SV2 (from form 4)				
drink	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	786	72.18	786	72.18
1	303	27.82	1089	100.00

Frequency Missing = 5

Exercising (at SV2; 0:no,1:yes, Fm12 Q11-12)				
exercise	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	607	55.94	607	55.94
1	478	44.06	1085	100.00

Frequency Missing = 9

Bslne ACE Use (at SV2; 0:no,1=yes)				
ACE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	650	61.03	650	61.03
1	415	38.97	1065	100.00

Frequency Missing = 29

Bslne Beta Bkr Use (at SV2; 0:no,1=yes)				
BETA	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	757	71.08	757	71.08
1	308	28.92	1065	100.00

Frequency Missing = 29

Table 2 col 1 (AASK Trial)

The FREQ Procedure

Bsline CCB Use (at SV2; 0:no,1:yes)				
CCB	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	373	35.02	373	35.02
1	692	64.98	1065	100.00

Frequency Missing = 29

Bsline Diuretic Use (at SV2; 0:no,1:yes)				
DIUR	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	386	36.24	386	36.24
1	679	63.76	1065	100.00

Frequency Missing = 29

Table 3 col 1 (AASK Trial)

The MEANS Procedure

Variable	Label	Mean	Std Dev
WT	Body Wt (kg) at SV2, G1 if missing	89.4622645	20.6770600
bmi	Body Mass Index (kg/m ² , at SV2, G1 if missing)	30.5681406	6.5850452
HTN_YEAR	Yrs with hypertension (at SV2), from form 4	14.1568266	10.1417169
sys	Systolic BP (mm Hg) at SV2, G1 if SV2 MAP is missing	150.3025594	23.8524523
dia	Diastolic BP (mm Hg) at SV2, G1 if SV2 MAP is missing	95.5210238	14.2119950
pulse	Seated Pulse (beats/min) at SV2, G1 if SV2 MAP is missing	72.0301921	12.5767972

Table 4 col 1 (AASK Trial)

The MEANS Procedure

Variable	Label	N	Mean	Std Dev
SCR	Serum Creatinine (mg/dL), at SV2, G1 if missing	1094	2.0237660	0.7193458
SUN	Serum Urea Nitrogen (mg/dL) at SV2, G1 if missing	1094	24.6005484	10.1074031
ALB	Serum Albumin (g/dL), at SV2, G1 if missing	1094	4.2527422	0.3527266
GLUC	Serum Glucose (mg/dL) at SV2, G1 if missing	1094	94.9936015	18.4644500
tchol	Serum Total Cholesterol (mg/dL), at SV2, G1 if missing	1081	211.6975023	45.5040707
hdl	Serum HDL Cholesterol (mg/dL), at SV2, G1 if missing	1081	48.3006475	16.1487670
TRIG	Serum Triglycerides (mg/dL), at SV2, G1 if missing	842	140.5130641	80.9207004
ldl	Serum LDL Cholesterol (mg/dL), at SV2, G1 if missing	842	136.4133017	41.0283126
UACID	Serum Uric Acid (mg/dL), at SV2, G1 if missing	1094	8.3360146	1.9687915
CAL	Serum Calcium (mg/dL), at SV2, G1 if missing	1094	9.1551188	0.4918981
phos	Serum Phosphorus (mg/dL), at SV2, G1 if missing	1094	3.5221207	0.6987276
calp		1094	32.2451828	6.6321861
HCT	Mean Baseline Hematocrit (%)	1092	39.4290751	4.9487334
pot	Serum Potassium (mmol/L) at SV2, G1 if missing	1094	4.2437843	0.6286487
uvol	U. Total Volume (L) from SV2 or G1	1093	2.2132205	0.9431965
ucre	U. Creatinine (g/d) from SV2 or G1	1090	1.6140249	0.6429438
upro	U. Protein (g/d) from SV2 or G1	1093	0.5328547	0.9412754
ratio	Urine Protein / Creatinine Ratio from SV2 or G1	1090	0.3264356	0.5195170
UUN	U. Urea Nitrogen (g/d) from SV2 or G1	1093	8.2865921	3.7864850
usod	U. Sodium (g/d) from SV2 or G1	1093	3.6832556	1.9793345
upot	U. Potassium (g/d) from SV2 or G1	1093	1.8168111	0.9820643

Table 4 col 1 (AASK Trial)

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
glgfr	Corrected G1 GFR (ml/min/1.73m ²)	1094	45.6402432	13.0484026	14.9520911	89.6958572
g2gfr	Corrected G2 GFR (ml/min/1.73m ²)	1094	47.1755398	15.2087000	11.9415207	98.8416724
MBGFR	Mean of Corrected G1 and G2 GFR (ml/min/1.73m ²)	1094	46.4078915	13.6267770	15.6529087	91.4297801
eGFR_trialbase		1094	46.7680442	14.1105756	15.0678204	93.5343833

Table 4 col 1 (AASK Trial)

The FREQ Procedure

LVH by EKG (at SV2; 0:no,1:yes, Fm14 Q8)				
lvh_ekg	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	494	56.14	494	56.14
1	386	43.86	880	100.00

Frequency Missing = 214

Table 1 part 1 (AASK Trial)

The FREQ Procedure

Table of ratio_cat by eGFR_trialbase_cat					
ratio_cat	eGFR_trialbase_cat				
Frequency Percent	1	2	3	4	Total
1	25 2.29	66 6.06	111 10.18	338 31.01	540 49.54
2	31 2.84	44 4.04	30 2.75	88 8.07	193 17.71
3	38 3.49	49 4.50	26 2.39	64 5.87	177 16.24
4	64 5.87	52 4.77	28 2.57	36 3.30	180 16.51
Total	158 14.50	211 19.36	195 17.89	526 48.26	1090 100.00
Frequency Missing = 4					

AASK Cohort analyses

to replicate Figure 1, Participant flow [al_evt_d=ESRD alone (0/1), al_evt_xd=ESRD or death (0/1)]

The FREQ Procedure

period=1

ESRD alone (0/1)				
al_evt_d	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	92	33.95	92	33.95
1	179	66.05	271	100.00

ESRD or death (0/1)				
al_evt_xd	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	7	2.58	7	2.58
1	264	97.42	271	100.00

ltfu_trans	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	271	100.00	271	100.00

al_evt_d	al_evt_xd	ltfu_trans	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	0	.	7	2.58	7	2.58
0	1	.	85	31.37	92	33.95
1	1	.	179	66.05	271	100.00

AASK Cohort analyses

to replicate Figure 1, Participant flow [al_evt_d=ESRD alone (0/1), al_evt_xd=ESRD or death (0/1)]

The FREQ Procedure

period=2

ESRD alone (0/1)				
al_evt_d	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	115	87.12	115	87.12
1	17	12.88	132	100.00

ESRD or death (0/1)				
al_evt_xd	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	97	73.48	97	73.48
1	35	26.52	132	100.00

ltfu_trans	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	35	26.52	35	26.52
0	73	55.30	108	81.82
1	24	18.18	132	100.00

al_evt_d	al_evt_xd	ltfu_trans	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	0	0	73	55.30	73	55.30
0	0	1	24	18.18	97	73.48
0	1	.	18	13.64	115	87.12
1	1	.	17	12.88	132	100.00

AASK Cohort analyses

to replicate Figure 1, Participant flow [al_evt_d=ESRD alone (0/1), al_evt_xd=ESRD or death (0/1)]

The FREQ Procedure

period=3

ESRD alone (0/1)				
al_evt_d	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	569	82.34	569	82.34
1	122	17.66	691	100.00

ESRD or death (0/1)				
al_evt_xd	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	495	71.64	495	71.64
1	196	28.36	691	100.00

ltfu_trans	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	691	100.00	691	100.00

al_evt_d	al_evt_xd	ltfu_trans	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	0	.	495	71.64	495	71.64
0	1	.	74	10.71	569	82.34
1	1	.	122	17.66	691	100.00

AASK Cohort analyses
to replicate Figure 1, Participant flow [al_evt_d=ESRD alone (0/1), al_evt_xd=ESRD or death (0/1)]

The FREQ Procedure

VISN				
VISN	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	673	98.25	673	98.25
3	3	0.44	676	98.69
6	3	0.44	679	99.12
9	2	0.29	681	99.42
12	1	0.15	682	99.56
27	3	0.44	685	100.00

AASK Cohort analyses
to compare visit number in urine mailing form versus visit number in data form

The FREQ Procedure

visitnum1	visitnum	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	0	582	19.00	582	19.00
3	3	43	1.40	625	20.40
6	6	37	1.21	662	21.61
9	6	1	0.03	663	21.65
9	9	6	0.20	669	21.84
12	9	3	0.10	672	21.94
12	12	522	17.04	1194	38.98
15	15	57	1.86	1251	40.84
18	15	1	0.03	1252	40.87
18	18	9	0.29	1261	41.17
21	21	3	0.10	1264	41.27
24	21	1	0.03	1265	41.30
24	24	472	15.41	1737	56.71
24	27	1	0.03	1738	56.74
27	27	41	1.34	1779	58.08
30	30	3	0.10	1782	58.18
33	33	3	0.10	1785	58.28
36	36	430	14.04	2215	72.31
39	39	31	1.01	2246	73.33
42	42	4	0.13	2250	73.46
45	45	4	0.13	2254	73.59
48	48	375	12.24	2629	85.83
51	51	41	1.34	2670	87.17
54	54	19	0.62	2689	87.79
57	57	133	4.34	2822	92.13
60	60	240	7.84	3062	99.97
63	63	1	0.03	3063	100.00

AASK Cohort analyses
to compare LVH at trial bsln (lvh_ekg) to LVH at cohort bsln (lvh)

The FREQ Procedure

LVH by EKG (at SV2; 0:no,1:yes, Fm14 Q8)				
lvh_ekg	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	141	20.41	141	20.41
0	317	45.88	458	66.28
1	233	33.72	691	100.00

LVH				
LVH	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	230	33.29	230	33.29
0	365	52.82	595	86.11
1	96	13.89	691	100.00

ANYLVH	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	56	8.10	56	8.10
0	355	51.37	411	59.48
1	280	40.52	691	100.00

lvh_ekg	LVH	ANYLVH	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	.	.	56	8.10	56	8.10
.	0	0	80	11.58	136	19.68
.	1	1	5	0.72	141	20.41
0	.	0	96	13.89	237	34.30
0	0	0	179	25.90	416	60.20
0	1	1	42	6.08	458	66.28
1	.	1	78	11.29	536	77.57
1	0	1	106	15.34	642	92.91
1	1	1	49	7.09	691	100.00

ANYLVH	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	56	8.10	56	8.10
0	355	51.37	411	59.48
1	280	40.52	691	100.00

AASK Cohort analyses
for table 1 col 3

The MEANS Procedure

Analysis Variable : age_cohort_base				
N	Mean	Std Dev	Minimum	Maximum
680	60.1975000	10.1960434	27.6000000	77.1000000

AASK Cohort analyses
for table 1 col 3

The FREQ Procedure

Gender (1=Male, 2=Female)				
GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	418	60.49	418	60.49
2	273	39.51	691	100.00

MARITAL				
MARITAL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	123	18.09	123	18.09
2	227	33.38	350	51.47
3	25	3.68	375	55.15
4	202	29.71	577	84.85
5	103	15.15	680	100.00

Frequency Missing = 11

1=never 2=married/married like 3=div/sep 4=wid				
mari_4	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	123	18.09	123	18.09
2	252	37.06	375	55.15
3	202	29.71	577	84.85
4	103	15.15	680	100.00

Frequency Missing = 11

ALONE				
ALONE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	475	69.85	475	69.85
1	205	30.15	680	100.00

Frequency Missing = 11

AASK Cohort analyses
for table 1 col 3

The FREQ Procedure

EMPLOY_STATS				
EMPLOY_STATS	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	184	27.06	184	27.06
2	61	8.97	245	36.03
3	19	2.79	264	38.82
4	244	35.88	508	74.71
5	24	3.53	532	78.24
6	79	11.62	611	89.85
7	1	0.15	612	90.00
8	1	0.15	613	90.15
9	1	0.15	614	90.29
10	66	9.71	680	100.00

Frequency Missing = 11

1=working 2=homemaker 3=retired 4=unemployed 5=student 6=other				
employ_6	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	246	36.18	246	36.18
2	19	2.79	265	38.97
3	244	35.88	509	74.85
4	103	15.15	612	90.00
5	2	0.29	614	90.29
6	66	9.71	680	100.00

Frequency Missing = 11

TOT_INCOME				
TOT_INCOME	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	52	7.65	52	7.65
2	134	19.71	186	27.35
3	93	13.68	279	41.03
4	52	7.65	331	48.68
5	82	12.06	413	60.74
6	44	6.47	457	67.21
7	25	3.68	482	70.88
8	27	3.97	509	74.85
9	13	1.91	522	76.76
10	8	1.18	530	77.94
11	150	22.06	680	100.00

AASK Cohort analyses
for table 1 col 3

The FREQ Procedure

Frequency Missing = 11

1=lt 15K 2=15 to <40K 3=40k+ 4=declined				
tot_inc_3	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	279	41.03	279	41.03
2	178	26.18	457	67.21
3	73	10.74	530	77.94
4	150	22.06	680	100.00

Frequency Missing = 11

1=private hmo other 2=medicaid medicare ONLY 3=none of above/missing				
insurancetype	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	206	30.29	206	30.29
2	274	40.29	480	70.59
3	200	29.41	680	100.00

Frequency Missing = 11

AASK Cohort analyses
for table 1 col 3

The FREQ Procedure

TOT_INCOME				
TOT_INCOME	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	52	7.65	52	7.65
2	134	19.71	186	27.35
3	93	13.68	279	41.03
4	52	7.65	331	48.68
5	82	12.06	413	60.74
6	44	6.47	457	67.21
7	25	3.68	482	70.88
8	27	3.97	509	74.85
9	13	1.91	522	76.76
10	8	1.18	530	77.94
11	150	22.06	680	100.00

Frequency Missing = 11

1=lt 15K 2=15 to <40K 3=40k+ 4=declined				
tot_inc_3	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	279	41.03	279	41.03
2	178	26.18	457	67.21
3	73	10.74	530	77.94
4	150	22.06	680	100.00

Frequency Missing = 11

AASK Cohort analyses
for table 1 col 3

The FREQ Procedure

1=private hmo other 2=medicaid medicare ONLY 3=none of above/missing				
insurancetype	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	206	30.29	206	30.29
2	274	40.29	480	70.59
3	200	29.41	680	100.00

Frequency Missing = 11

insurancetype	PRIVATE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	0	117	17.21	117	17.21
1	1	89	13.09	206	30.29
2	0	274	40.29	480	70.59
3	0	147	21.62	627	92.21
3	1	52	7.65	679	99.85
3	9	1	0.15	680	100.00

Frequency Missing = 11

insurancetype	HMO	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	0	122	17.94	122	17.94
1	1	84	12.35	206	30.29
2	0	274	40.29	480	70.59
3	0	154	22.65	634	93.24
3	1	44	6.47	678	99.71
3	9	2	0.29	680	100.00

Frequency Missing = 11

insurancetype	MEDICAID	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	0	206	30.29	206	30.29
2	0	125	18.38	331	48.68
2	1	149	21.91	480	70.59
3	0	172	25.29	652	95.88
3	1	27	3.97	679	99.85
3	9	1	0.15	680	100.00

Frequency Missing = 11

AASK Cohort analyses
for table 1 col 3

The FREQ Procedure

insurancetype	MEDICARE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	0	206	30.29	206	30.29
2	0	74	10.88	280	41.18
2	1	200	29.41	480	70.59
3	0	102	15.00	582	85.59
3	1	97	14.26	679	99.85
3	9	1	0.15	680	100.00

Frequency Missing = 11

insurancetype	OTH_INSURE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	0	150	22.06	150	22.06
1	1	56	8.24	206	30.29
2	0	274	40.29	480	70.59
3	0	180	26.47	660	97.06
3	1	19	2.79	679	99.85
3	9	1	0.15	680	100.00

Frequency Missing = 11

insurancetype	PRIVATE	HMO	MEDICAID	MEDICARE	OTH_INSURE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	0	0	0	0	1	46	6.76	46	6.76
1	0	1	0	0	0	65	9.56	111	16.32
1	0	1	0	0	1	6	0.88	117	17.21
1	1	0	0	0	0	72	10.59	189	27.79
1	1	0	0	0	1	4	0.59	193	28.38
1	1	1	0	0	0	13	1.91	206	30.29
2	0	0	0	1	0	125	18.38	331	48.68
2	0	0	1	0	0	74	10.88	405	59.56
2	0	0	1	1	0	75	11.03	480	70.59
3	0	0	0	0	0	93	13.68	573	84.26
3	0	0	0	1	1	11	1.62	584	85.88
3	0	0	1	0	1	1	0.15	585	86.03
3	0	0	1	1	1	3	0.44	588	86.47
3	0	1	0	1	0	27	3.97	615	90.44
3	0	1	0	1	1	1	0.15	616	90.59
3	0	1	1	0	0	5	0.74	621	91.32
3	0	1	1	1	0	5	0.74	626	92.06
3	0	9	1	0	0	1	0.15	627	92.21
3	1	0	0	1	0	32	4.71	659	96.91

AASK Cohort analyses
for table 1 col 3

The FREQ Procedure

insurancetype	PRIVATE	HMO	MEDICAID	MEDICARE	OTH_INSURE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
3	1	0	0	1	1	3	0.44	662	97.35
3	1	0	1	0	0	1	0.15	663	97.50
3	1	0	1	1	0	10	1.47	673	98.97
3	1	1	0	1	0	5	0.74	678	99.71
3	1	1	1	0	0	1	0.15	679	99.85
3	9	9	9	9	9	1	0.15	680	100.00

Frequency Missing = 11

AASK Cohort analyses
for table 2 col 3

The FREQ Procedure

SMOKE				
SMOKE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	289	42.50	289	42.50
2	114	16.76	403	59.26
3	277	40.74	680	100.00

Frequency Missing = 11

ALCOHOL				
ALCOHOL	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	254	37.35	254	37.35
1	322	47.35	576	84.71
2	104	15.29	680	100.00

Frequency Missing = 11

AASK Cohort analyses
for table 2 col 3

The FREQ Procedure

exercise	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	266	39.06	266	39.06
1	415	60.94	681	100.00

Frequency Missing = 10

AASK Cohort analyses
for table 2 col 3

The MEANS Procedure

Analysis Variable : htn_year_cohort				
N	Mean	Std Dev	Minimum	Maximum
674	19.6510386	9.8399226	4.0000000	51.6000000

AASK Cohort analyses
for table 3 col 3

The MEANS Procedure

Variable	Label	Mean	Std Dev
wt_cohort_base		92.1542525	22.3466703
bmi_cohort_base		31.4628635	7.1284635
SBP_AVG	SBP_AVG	135.7531847	21.9632411
DBP_AVG	DBP_AVG	80.8503185	12.2237206
PULS_SIT	PULS_SIT	68.6735669	12.1391508

AASK Cohort analyses
for table 4 col 3

The MEANS Procedure

Variable	Label	N	Mean	Std Dev
SCR	SCR	689	2.3158200	1.5072005
SUN	SUN	596	30.2432886	18.3026711
ALB	ALB	596	3.9890940	0.3207112
GLUC	GLUC	596	101.5167785	34.6308948
TOT_S_C	TOT_S_C	596	201.8137584	47.5975129
HDL_S_C	HDL_S_C	596	46.5587248	14.9484798
TRIG	TRIG	525	144.7238095	94.4917597
LDL_S_C	LDL_S_C	525	125.8076190	42.8385833
UACID	UACID	596	8.8907718	2.1803165
CAL	CAL	596	9.6013423	0.5918195
P	P	596	3.5567114	0.7787980
calp		596	34.0486074	7.0837995
HEMATOCRIT	HEMATOCRIT	676	38.2967456	5.4903083
K	K	595	4.2638655	0.6065483

AASK Cohort analyses
for table 4 col 3

The MEANS Procedure

Analysis Variable : uprocreat		
N	Mean	Std Dev
670	0.3786291	0.8173447

AASK Cohort analyses
for table 4 col 3

The MEANS Procedure

Variable	Label	N	Mean	Std Dev
VOLUME	VOLUME	670	2217.16	1081.00
up24	urinary protein, mg/24hr	670	531.6619746	1056.72
ucr24	urinary creatinine, mg/24hr	670	1589.08	849.9359218
uun24	urinary urea nitrogen, mg/24hr	670	724.6766418	442.2491917
una24mol	urinary sodium, g/24hr	670	3.7597429	2.8988342
uk24mol	urinary potassium, g/24hr	670	1.8774902	1.1144210

AASK Cohort analyses
for table 4 col 3

The MEANS Procedure

Analysis Variable : eGFR		
N	Mean	Std Dev
678	43.8831373	16.6986674

AASK Cohort analyses
for table 4 col 3

The FREQ Procedure

lvh_cohort	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	365	79.18	365	79.18
1	96	20.82	461	100.00

Frequency Missing = 230

AASK Cohort analyses
for table 4 col 3

The MEANS Procedure

Analysis Variable : uprocreat				
uprocreat_cat	N Obs	N	Minimum	Maximum
1	369	369	0.0046802	0.0797546
2	95	95	0.0802998	0.2181501
3	98	98	0.2215289	0.6530920
4	108	108	0.6631579	8.3415978

AASK Cohort analyses
for table 4 col 3

The MEANS Procedure

Analysis Variable : eGFR				
eGFR_cat	N Obs	N	Minimum	Maximum
1	147	147	3.8002320	29.8002592
2	131	131	30.0160768	39.9471885
3	108	108	40.2060283	47.9259858
4	292	292	48.0916038	93.1225821

Table of uprocreat_cat by eGFR_cat					
uprocreat_cat	eGFR_cat				
Frequency Percent	1	2	3	4	Total
1	37 5.58	65 9.80	66 9.95	199 30.02	367 55.35
2	19 2.87	21 3.17	18 2.71	36 5.43	94 14.18
3	32 4.83	22 3.32	11 1.66	32 4.83	97 14.63
4	56 8.45	20 3.02	12 1.81	17 2.56	105 15.84
Total	144 21.72	128 19.31	107 16.14	284 42.84	663 100.00
Frequency Missing = 28					