

Supplementary Information for

**Surveillance of CKD epidemiology in the US – a joint analysis of NHANES and KEEP**

Myers OB<sup>1</sup>, Pankratz VS<sup>1</sup>, Norris KC<sup>2</sup>, Vassalotti JA<sup>3</sup>, Unruh M<sup>1</sup>, Argyropoulos C<sup>1</sup>

<sup>1</sup>Internal Medicine, University of New Mexico, Albuquerque, NM 87131

<sup>2</sup>David Geffen School of Medicine, UCLA, Los Angeles, CA 90024

<sup>3</sup>National Kidney Foundation Inc., New York, New York

## Supplemental Methods

### **Propensity scores using logistic regression on reconstructed NHANES total population frequencies and with NHANES mobile examination center weights**

We conducted preliminary analyses to estimate propensity scores using the expected distribution of study covariates in the general NHANES population by forming cross-tabulations of NHANES participants for all combinations of covariate values for each survey period and using NHANES participants and their sampling weights. For the total population frequency approach we used sampling weights, primary sampling units (PSU) and strata provided with the NHANES data for participation in the mobile examination centers (MEC) in each two year survey period to estimate the population frequency for each cross-tabulation cell (SURVEYFREQ, SAS v9.4). MEC weights were used instead of interview weights because we estimate CKD prevalence measures based on MEC results. The Taylor series linearization method was used for variance estimation. A separate subgroup (domain) was used to exclude NHANES participants <20 years old, those receiving dialysis in the past year or those without valid data required to estimate eGFR (age, sex, race/ethnicity, serum creatinine). The selected domain for estimation of self-selection weights included 28,517 NHANES participants.

The reconstructed NHANES population was formed by cross-tabulation of participant characteristics resulted into subgroups having non-zero frequency weights. KEEP participants received a weighted frequency of 1.0 and were combined with NHANES frequency weighted data into a master analytic file for further analyses (TOTPOP file). A second analytic file contained selected 28,517 NHANES participants appended to KEEP (WTDPOP). In the WTDPOP analytic file MEC weights were designated as sampling weights for NHANES and KEEP participants received a sampling weight equal to 1.0. Selection models were estimated for

each analytic file to determine if propensity scores were affected by how we developed the NHANES subset of the analytic file. Weighted logistic regression was used to estimate the probability of KEEP participation separately for each NHANES survey periods from 2001-2002 to 2011-2012. Selection models included demographic variables (age, sex, race/ethnicity), recruitment variables (patient reported diabetes, hypertension, chronic kidney disease, and family history of CKD), other health-related factors (obesity status, smoking, family history of heart attack, and participant cardiovascular disease) and two socioeconomic variables (education and health insurance). Age was modeled as a restricted cubic spine (14) with interactions between sex and with race/ethnicity, and the remaining variables were entered as categorical. SAS v9.4 (SURVEYLOGISTIC) was used to estimate probability of being in the KEEP sample for each selection model and analytic file.

Propensity score estimates for KEEP participants were not sensitive to the method used to estimate the weights used in their estimation. This finding allowed us to use individual NHANES records and MEC sampling weights to model self-selection in KEEP. The WTDPOP approach did not include NHANES PSU and strata information because these are not available for KEEP, which would affect variance estimates for inference using the combined data. Our goal was to obtain point estimates for self-selection probabilities for each KEEP participant, and these estimates were not sensitive to the weighting approach.

Supplemental Table 1. Prevalence of CKD Stage 3 and Stage 4-5 by sex and age from NHANES and KEEP (2001-2012) without and with inverse probability weighting (% and 95% CI).

NHANES estimates accounted for sampling design. KEEP WTD-GEE estimates accounted for participant self-selection and clustering within regional affiliates. CKD – chronic kidney disease; KEEP – Kidney Early Evaluation Program; NHANES - National Health and Nutrition Examination Survey

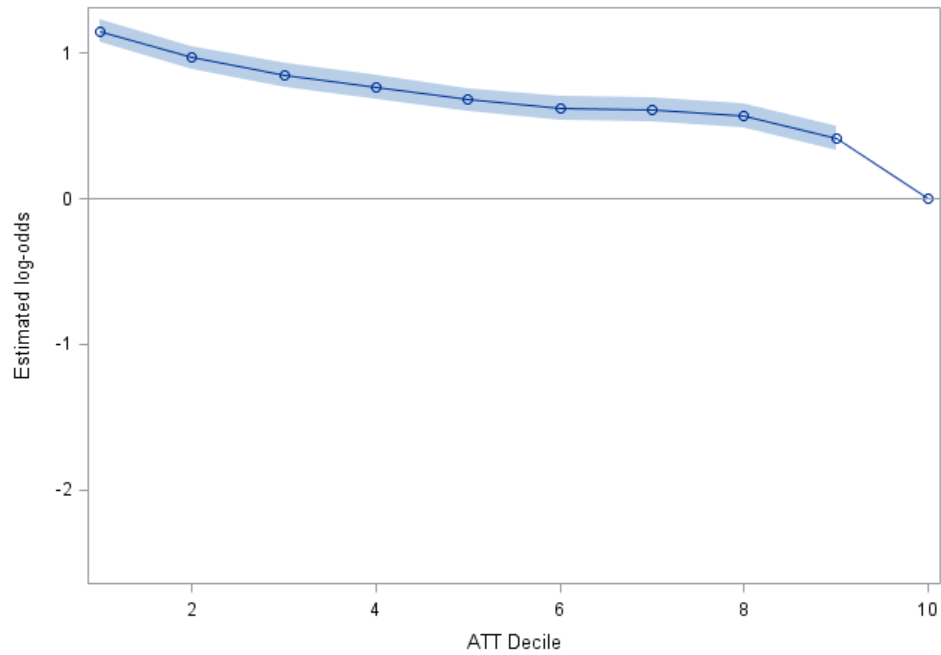
Sex	Age	Stage 3			Stage 4-5		
		NHANES	KEEP	KEEP WTD-GEE	NHANES	KEEP	KEEP WTD-GEE
Female	20-39	0.06 (0.02-0.18)	0.84 (0.70-1.00)	0.55 (0.34-0.89)	0.07 (0.02-0.20)	0.12 (0.08-0.19)	0.01 (0.00-0.02)
	40-59	2.52 (1.95-3.25)	5.44 (5.21-5.67)	4.46 (3.80-5.22)	0.12 (0.07-0.21)	0.34 (0.29-0.41)	0.17 (0.11-0.25)
	60-69	11.49 (9.68-13.59)	17.06 (16.53-17.61)	15.19 (13.54-17.01)	0.52 (0.33-0.82)	0.89 (0.77-1.04)	0.41 (0.27-0.63)
	70+	36.25 (34.23-38.32)	39.50 (38.73-40.28)	39.24 (37.17-41.34)	3.74 (3.05-4.58)	3.02 (2.77-3.31)	3.42 (2.66-4.40)
	20-39	0.24 (0.12-0.49)	1.64 (1.36-1.97)	0.88 (0.63-1.23)	0.01 (0.00-0.10)	0.29 (0.19-0.45)	0.07 (0.03-0.14)
Male	40-59	1.74 (1.30-2.32)	5.40 (5.08-5.75)	4.01 (3.25-4.96)	0.04 (0.02-0.11)	0.53 (0.43-0.65)	0.13 (0.07-0.21)
	60-69	9.84 (8.55-11.31)	16.09 (15.34-16.87)	13.67 (12.29-15.17)	0.65 (0.34-1.22)	1.11 (0.91-1.35)	0.74 (0.51-1.06)
	70+	31.74 (29.86-33.69)	36.24 (35.18-37.31)	35.66 (33.21-38.17)	2.72 (2.14-3.45)	3.56 (3.17-4.00)	3.41 (2.81-4.15)

Supplemental Table 2. Linear trend estimates for CKD Stage: NHANES and KEEP WTD-GEE. Coefficient estimates are for a two-year change. NHANES estimates account for sampling strata and clusters using two year sampling weights for mobile clinics, and are estimates for the domain that met selection criteria. KEEP estimates employed inverse probability weighting for self-selection and used GEE to account for affiliate sampling clusters.

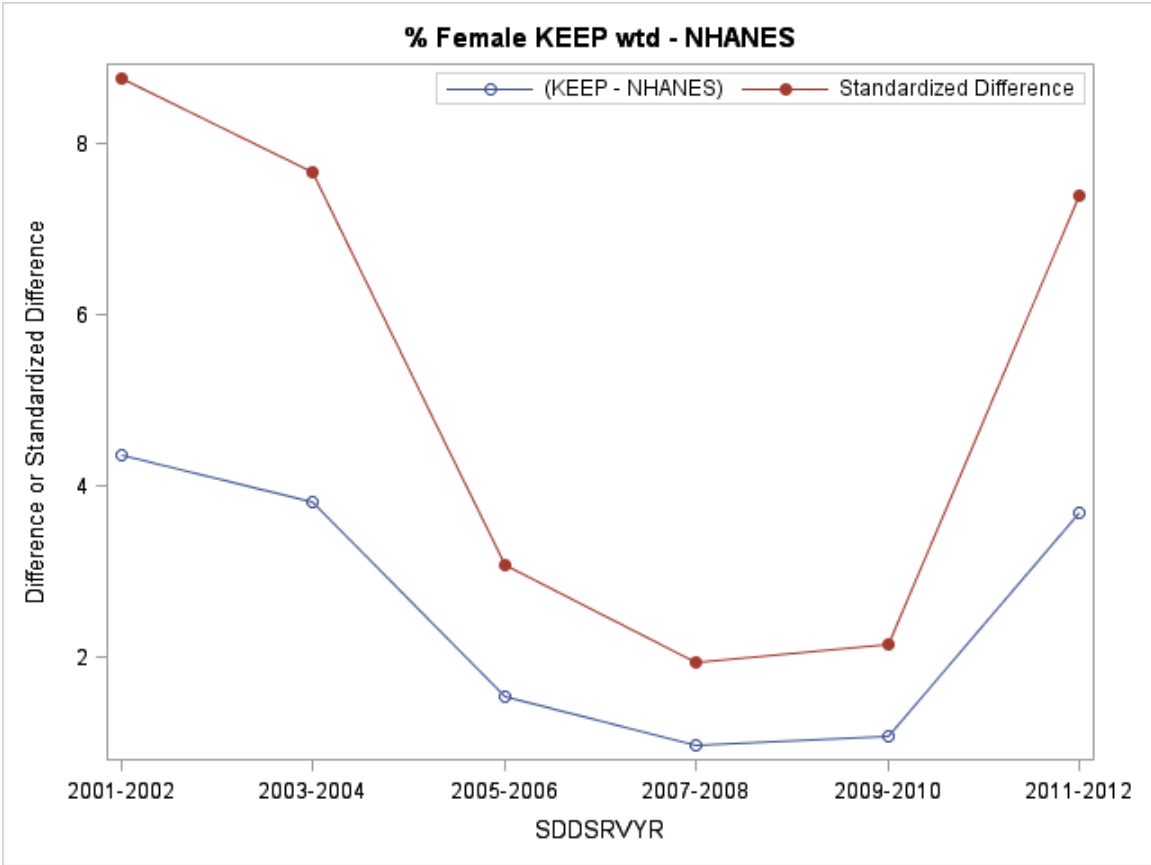
	Coefficient (95% CI) for 2-yr Change			
Stage	NHANES	P-value	KEEP WTD-GEE	P-value
3	-0.00085 (-0.0210, 0.0193)	0.93	-0.0035 (-0.0591, 0.0522)	0.90
3-5	0.000192 (-0.0194, 0.0198)	0.98	-0.0038 (-0.0591, 0.0515)	0.89
4-5	0.0122 (-0.0376, 0.0620)	0.63	-0.0075 (-0.1303, 0.1153)	0.90

## Supplemental Figures

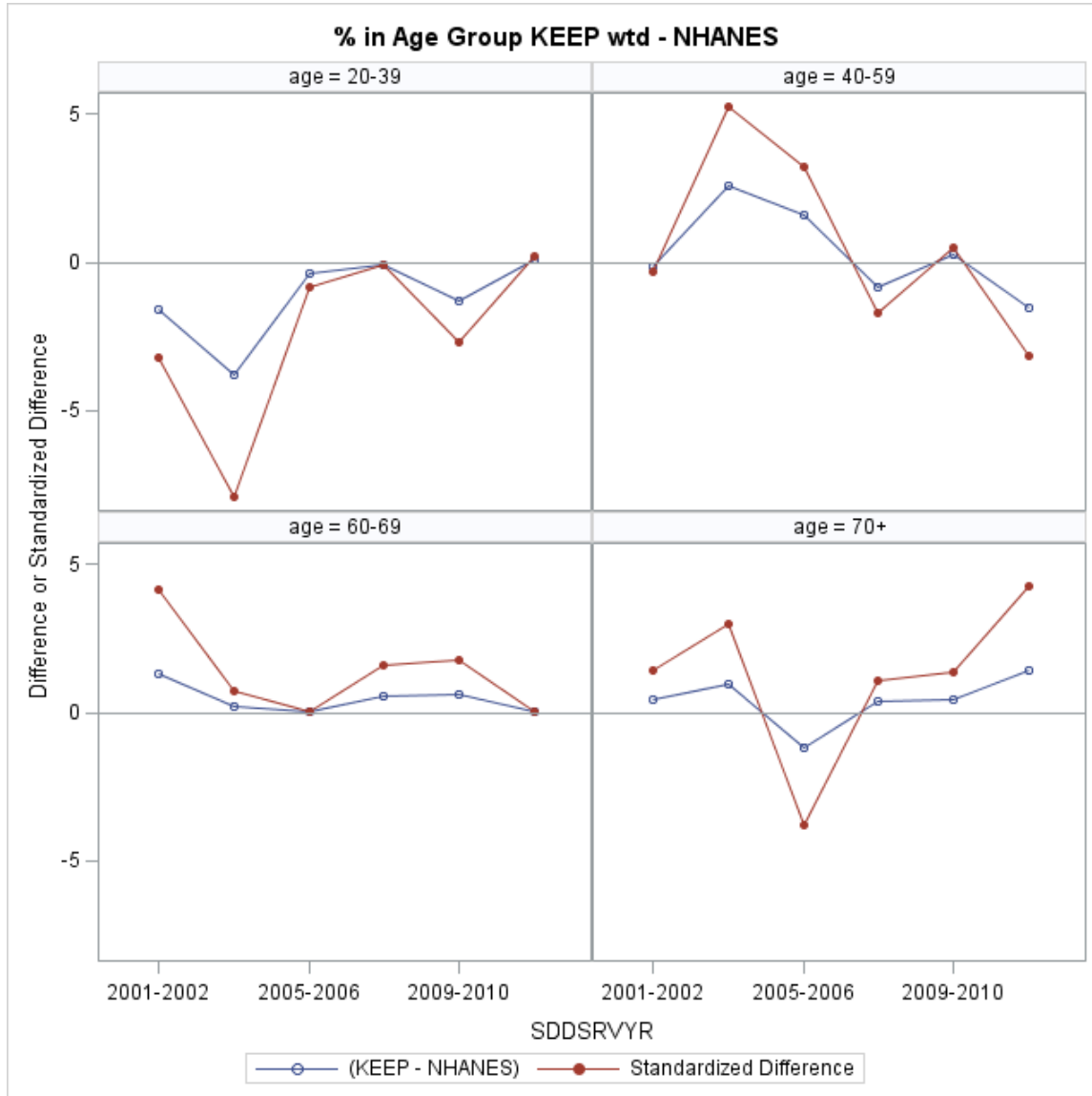
Supplemental Figure 1. Estimated log-odds and 95% confidence intervals for CKD Stage 3-5 within deciles of average treatment effect on the treated (ATT) weights (referent = decile 10). Smaller deciles have higher probability of being in KEEP relative to larger deciles ( $p < 0.001$ ).



Supplemental Figure 2. Balance by year for sex between NHANES general population and weighted KEEP. Balance by NHANES survey period assessed using unstandardized and standardized differences in proportions (Austin et al. 2007).



Supplemental Figure 3. Balance by year for age between NHANES general population and weighted KEEP. Balance by NHANES survey period assessed using unstandardized and standardized differences in proportions (Austin et al. 2007).





Supplemental Figure 4. Balance by year for race/ethnicity between NHANES general population and weighted KEEP. Balance by NHANES survey period assessed using unstandardized and standardized differences in proportions (Austin et al. 2007).

