Supplementary Methods

Study Population

The Seattle Kidney Study (SKS) is a clinic-based, prospective cohort study of CKD based in Seattle, Washington.[37, 38] The SKS began recruiting participants in 2004 from outpatient Nephrology clinics at Harborview Medical Center and the Veterans' Affairs Medical Center, affiliated hospitals of the University of Washington. Eligibility criteria are age greater than 18 years and CKD of any stage not requiring dialysis. Exclusion criteria are current or prior kidney transplantation, dementia, institutionalization, expected to start renal replacement therapy or leave the area within 3 months, participation in a clinical trial, non-English speaking, or inability to undergo the informed consent process.

We restricted our study population to persons with moderate-to-severe CKD (stage III-IV; estimated GFR 15-59 ml/min/1.73m^2) by excluding 71 participants who had an eGFR $\geq$ 60ml/min/m^2 and 18 who had an eGFR <15ml/min/m^2 at baseline. Each site's institutional review board approved the study, and all participants provided informed consent. To focus on participants who had the capacity to exercise, we further excluded 33 patients who were unable to ambulate and required the use of a wheelchair. Finally, we excluded 118 subjects who had fewer than the two eGFR measurements necessary to calculate a slope, leaving a total of 256 participants for analysis (Figure 1). The 118 subjects who were excluded due to lack of follow-up eGFR measurements were of similar age, had a similar baseline eGFR, had a higher prevalence of diabetes and

coronary artery disease at baseline and had similar amounts of physical activity compared to included participants. Among them, 23 (19%) died prior to scheduled follow-up.

Measurement of physical activity

The Four Week Physical Activity History Questionnaire (FWH) was administered at baseline to estimate each participant's self-reported leisure-time physical activity level. The FWH queries participants regarding the frequency and duration in which they engaged in each of the following activities during the prior month: walking for exercise, jogging, biking, aerobics, golf, tennis, swimming, weight training, treadmill or aerobic machine.[39] To calculate minutes per week of total leisure-time activity, for each activity performed, we multiplied the frequency by the duration in the prior month, divided by 4 weeks, and summed across all activities.

The FWH questionnaire has been evaluated in the general population against doubly labeled water, heart rate monitoring, changes in maximal oxygen uptake, and accelerometry.[40-44] We previously reported significant associations of FWH with objectively measured energy expenditure using accelerometry (intra-class correlation coefficient (ICC): 0.28, p=0.037) and with percent time spent in moderate to vigorous activities (ICC=0.38, p=0.018) in a subset of 48 SKS participants.[37]

Estimated GFR

Serum samples are collected annually in SKS. We decided, prior to analysis, to evaluate cystatin C-based estimates of GFR because serum creatinine levels depend on

muscle mass, which declines with older age and may be influenced by exercise.[45] In secondary analyses, we also assessed GFR using the serum-creatinine-based CKD-EPI equation.[46] Cystatin C and creatinine-based estimates of GFR provide similar precision and validity compared to gold-standard radionucleotide measurement of GFR.[55] Serum cystatin C levels were measured from frozen serum samples stored at −70°C using either a particle-enhanced immunonephelometric assay (Gentian A/S, Norway) with a clinical chemistry analyzer (DXC600, Beckman Coulter Miami, Florida) or a nephelometer (BNII, Siemens Healthcare Diagnostics Inc., Deerfield, Illinois). A calibration equation to convert nephelometer-based cystatin C concentrations to DXC-based concentrations (gold-standard) was developed in a subset of 40 SKS participants who had identical serum samples measured with both assays (r^2=0.98). The calibration of cystatin C was performed by reconstitution of the cystatin C reference material (ERM-DAY7/IFCC) per its Certificate of Analysis and corrected by weight to yield a cystatin C concentration of 5.48 mg/L (uncertainty of 0.15mg/L). The assays are stable through several freeze-thaw cycles.[56] We calculated estimated GFR at each SKS study exam using the equation:[47]

$$\text{Estimated GFR}_{\text{cystatin C}} = 127.7 \cdot \text{cystatin C}^{-1.17} \cdot \text{age}^{-0.13} \cdot (0.91 \text{ if female; } 1.06 \text{ if black})$$

Incident end stage renal disease (ESRD), defined as the first occurrence of initiation of chronic dialysis or kidney transplantation, was identified during twice yearly surveillance. Study coordinators also specifically inquired about initiation of dialysis and fistula placement since the previous encounter. All self-reported initiations of dialysis or kidney transplantations were subsequently verified through medical record review.

Measurement of covariates

Weight was measured using calibrated scales, height with a wall-mounted tape measure, and waist circumference using a constant-tension tape. Prevalent conditions were determined based on participant responses to questionnaires and hospitalizations that occurred after initial SKS enrollment but prior to the initial assessment for this study. Medications were assessed by inventory assessment and missing medication data were completed by chart review.[48] We defined diabetes by any of the following: use of an oral hypoglycemic medication or insulin, fasting blood sugar ≥126mg/dL, non-fasting blood sugar ≥200mg/dL, or hemoglobin A1c ≥6.5%. Hypertension was defined by the use of any antihypertensive medication, systolic blood pressure ≥ 140 mmHg, or diastolic blood pressure ≥ 90 mmHg.[49] Blood samples were collected after an overnight fast and urine was collected as timed overnight voids. Except for hemoglobin A1c, which was measured on fresh blood, serum, plasma, and urine were stored at -80 °C until analysis. Concentrations of C-reactive protein (CRP) were measured in serum. Urine albumin and creatinine were measured in spot morning or overnight urine collections. Smoking status (current, former, or never) and alcohol use (current use *versus* none) were determined via baseline lifestyle questionnaires. The Short Physical Performance Battery (SPPB) was performed and the score was calculated as recommended.[52] A summary score of 0 to 12 (higher score indicating better function) is based on performance on three tasks: gait speed, chair rise, and three standing positions to assess balance.

Statistical analyses

We tabulated baseline participant characteristics according to physical activity category. The minutes per week of physical activity were categorized as none, 1 to 60, 60

to 150 and more than 150 minutes per week. The uppermost physical activity category was defined so as to correspond to an adherence to the American Heart Association Physical Activity Guidelines of 150 minutes per week of moderate physical activity.[50]

We used generalized estimating equations (GEE), accounting for within-participant clustering across time, to determine if the annualized relative change in eGFR differed across physical activity categories, after adjusting for potential confounding variables.[51] Based on ongoing work demonstrating minimal dependence of the relative change on the baseline eGFR value, using graphical methods and residual analyses, we selected relative (log eGFR) versus absolute change in eGFR as our primary outcome.

We chose covariates as potential confounding factors *a priori* based on plausibility that they could confound the association of physical activity level with change in estimated GFR. We investigated groups of potential confounding factors by constructing three multivariable models. We conducted subgroup analyses to evaluate whether associations of both the presence and the duration of physical activity with change in eGFR were robust after excluding participants who had prevalent coronary artery disease, those with prevalent diabetes and those who scored below 8 on the SPPB, which is associated with disability in lower extremity functioning.[52]

For the analysis of incident treated ESRD as the outcome, participants were considered at risk from the date of their baseline SKS study visit until the first occurrence of dialysis initiation, kidney transplantation or their data were censored due to death, loss to follow-up or end of study data collection period (January 1, 2012). We used Cox proportional hazards regression to estimate the relative hazard of ESRD after adjustment

for factors selected *a priori* based on evidence that they might confound the association between physical activity and ESRD. The exposure definition and regression models were similar to those constructed to evaluate change in estimated GFR, with additional adjustment for baseline eGFR, using cystatin C, in the final model. To assess the importance of competing risks, we performed sensitivity analyses using end-stage renal disease or death as the failure variable.

For the 27 participants who did not check any boxes on the physical activity questionnaire but answered all other questions, we used a multiple imputation procedure to replace missing values. Multiple imputation is an efficient and practical method to deal with missing data, in which each missing value is replaced by a list of *m* (in our case, m=10) simulated values.[53] For the regression analyses, these subjects' values were multiply imputed using chained equations.[53] The multiple analyses over the imputations were combined using Rubin's rules to account for the variability in the imputation procedure.[54]

Walking is the most commonly performed physical activity in part because it is highly accessible and rarely associated with injury. In an attempt to characterize the association of walking time alone and eGFR decline, we secondarily examined the association of minutes of walking per week with annualized relative change in eGFR. For these analyses, we additionally adjusted for time spent in other leisure-time physical activities.

We also evaluated whether associations of physical activity and eGFR change differed according to stage of CKD, sex and race. A 3-way multiplicative interaction term

(e.g. race x time x physical activity) and the corresponding 2-way interaction terms were entered in the second regression model and the Wald test was used to assess statistical significance of the three-way interaction term. All p-values were two-tailed (α=0.05). All analyses were performed using STATA release 11.2 (College Station, TX).

Supplementary Table 1. Association of Walking Time and Annualized Relative Change in eGFR-cystatin C

Walking duration (minutes per week)	N	Percent Annual Change in eGFR-cystatin C							
		Unadjusted		Model 1		Model 2		Model 3	
		% change per year	95 % CI	% change per year	95 % CI	% change per year	95 % CI	% change per year	95 % CI
None	106	-10.9	(-13.5,-8.1)	-11.2	(-14.0,-8.1)	-10.8	(-12.7,-7.3)	-8.6	(-11.6,-5.5)
1-60	60	-4.7	(-9.1,+0.1)	-4.7	(-9.4,+0.3)	-4.2	(-11.2,+0.4)	-3.6	(-8.4,+1.3)
60-150	52	-3.5	(-7.7,+0.9)	-3.8	(-8.4,+0.1)	-4.5	(-8.7,-0.2)	-4.7	(-9.2,0.0)
≥150	41	-7.1	(-11.7,-2.4)	-7.5	(-12.4,-2.4)	-6.1	(-10.6,-1.4)	-5.3	(-10.6,+3.5)
p-value for trend		0.043		0.064		0.041		0.068	
Per 60 min/week increment		0.61 (0.14, 1.07)		0.52 (0.03, 1.00)		0.50 (0.02, 0.98)		0.46 (-0.05, 0.97)	

Model 1: age, race, gender, site, time spent in other leisure-time activities
Model 2: Model 1 + education, body mass index, diabetes, smoking status, alcohol, prevalent coronary artery disease
Model 3: Model 2 + hemoglobin A1c, systolic blood pressure, ACE-inhibitor use, ARB use, statin use, C-reactive protein
P-values for continuous association (per 60 minute/week increment in physical activity) were 0.027, 0.030 and 0.043, and 0.068 for unadjusted model and models 1, 2 and 3 respectively.

Supplementary Table 2. Association of Physical Activity and Annualized Relative Change in eGFR-CKD EPI

Leisure-time Physical	Percent Annual Change in eGFR-CKD EPI (95% CI)		
Activity Level	Model 1	Model 2	Model 3
None	-11.1 (-13.8, -8.3)	-10.9 (-13.7, -8.1)	-9.1 (-12.3, -5.9)
0-60 minutes per week	-8.0 (-10.4, -5.5)	-8.2 (-10.7, -5.8)	-7.8 (-10.4, -5.2)
60-150 minutes per week	-8.3 (-11.1, -5.6)	-8.5 (-11.2, -5.8)	-9.3 (-12.6, -6.5)
>150 minutes per week	-6.5 (-8.8, -4.1)	-7.1 (-9.5, -4.7)	-6.4 (-9.0, -3.7)
p-for-trend	0.010	0.032	0.061
Per 60 min/week increment	0.70 (0.17, 1.22)	0.55 (0.01, 1.09)	0.39 (0.02, 0.77)

Model 1: age, race, gender, site
Model 2: Model 1 + education, body mass index, diabetes, smoking status, alcohol, prevalent coronary artery disease
Model 3: Model 2 + hemoglobin A1c, systolic blood pressure, ACE-inhibitor use, ARB use, statin use, C-reactive protein, baseline GFR-cystatin C
P-values for continuous association (per 60 minute/week increment in physical activity) were 0.009, 0.047 and 0.037, and 0.068 for models 1, 2 and 3 respectively.