Item S1. Supplemental Methods

Covariate Measurement

The CRIC Study Central Laboratory measured calcium and phosphate via standard assays and plasma parathyroid hormone (PTH) using the Scantibodies total intact assay (Santee, California) from the annual study visits, specimen volume permitting. The University of Washington Laboratory measured the 1,25-dihydroxyvitamin D levels at the second and fifth annual CRIC study visits with a multiplex HPLC-mass spectrometry assay.[1] Urine protein to creatinine ratio (UPCR) was measured using a combination of 24 hour and spot urine collections using standard assays. Twenty four hour urinary phosphate and spot urine phosphate were measured using standard assays. Baseline CRP was measured in EDTA plasma samples using laser-based immunonephelometric methods on BNII (Siemens Healthcare Diagnostics, Deerfield, IL) and baseline IL-6 was measured using high-sensitivity ELISAs (Quantikine HS; R&D systems, Minneapolis, MN).[2]

Age, eGFR, UPCR, serum albumin, hemoglobin, CRP, IL-6, systolic blood pressure, body mass index, calcium, phosphate, PTH, 24 hour urinary phosphate, spot urine phosphate, and 1,25 dihydroxyvitamin D were all treated as continuous variables in our models. UPCR, PTH, CRP, and IL-6 were natural log–transformed. Sex, race, ethnicity, diabetes, current smoking, history of heart failure, stroke, peripheral vascular disease, number of hypertension meds, and ace/arb use were treated as categorical variables in our models. UPCR was categorized as missing, < 30 mg, 30-300 mg and > 300 mg for our FGF23 trajectory models.

Time-Varying Inverse Probability Weighting (IPW) Discrete Time Failure Models

Time-varying IPW discrete time failure models are performed in two steps. The first step employs calculation of the FGF23 exposure weights and in the second step the fitting of a discrete time failure model for the outcomes (ESRD, ESRD or death) by applying the final weight derived in the first step.[3] The data structure was set up in the longitudinal format, so each observation, in 1 year increments, could contribute up to 5 records in the dataset.

To calculate the exposure weight in the first step, we modeled FGF23 at each visit in quartiles because it was not-normally distributed.[3-5] We performed multinomial logistic regression on FGF23 quartiles with adjustment for age, sex, race, and ethnicity, estimated glomerular filtration rate, urine protein-to-creatinine ratio, serum albumin, hemoglobin, CRP, IL-6, diabetes, smoking, systolic blood pressure, body mass index, history of coronary artery disease, congestive heart failure, stroke, peripheral vascular disease, number of blood pressure medications, ACE/ARB use, calcium, phosphate, and PTH, and FGF23 from the previous 4 study visits, if available. The weight for a specific study visit was calculated as one over the cumulative probabilities of the observed FGF23 history up to that visit (i.e. the product of the probabilities of the observed FGF23 categories up to that visit). We stabilized the weights by multiplying the estimated probability of observed FGF23 history conditional on baseline-only predictors, including age, sex, race, and ethnicity, estimated glomerular filtration rate, urine protein-to-creatinine ratio, serum albumin, hemoglobin, CRP, and IL-6, diabetes, smoking, systolic blood pressure, body mass index, history of coronary artery disease, congestive heart failure, stroke, and peripheral vascular disease, number of blood pressure medications, ACE/ARB use, calcium, phosphate, and PTH. This was done by fitting a second model of FGF23 quartiles using the baseline-only predictors.

Given the small number of loss to follow-up (91 of 1597 patients [5.7%]), we did not calculate the censoring weight.[6] For the case-cohort design, the inverse probability weights are calculated using the subcohort and cases outside the subcohort. Contributions to the models for the inverse probability weight were themselves weighted inversely to the sampling probability for the case-cohort design. In our study, cases (inside or outside the subcohort) were given a weight of 1, while non-cases inside the subcohort were given a weight of 3.47 (=3939/1135). Non-cases outside the subcohort were not used to estimate the inverse probability weights. The final weight for each observation is the product of the exposure weight multiplied by the weight from the case-cohort design.

Large weights can create imprecision from inflation of standard errors of the effect estimate.[3] To lessen imprecision, we truncated weights in the time-varying IPW discrete time failure models. When we pooled the estimated weights from all time points in the study, 93 observations were truncated at 99th percentiles of the estimated weights. The largest final weight for a particular observation in the second step was 20. We assume the model assumptions of exchangeability, positivity and correct specification of the model used to estimate weights hold true. We implicitly assumed the exchangeability assumption held true given the large number of joint predictors of exposure and outcome we incorporated. We considered the positivity assumption, or that is the probability of being in any of the 4 FGF23 quartiles conditional on any combination of the predictors is strictly positive, to be satisfied given the large sample size and the fact that FGF23 categories were defined based on the quartiles of the distribution. Additionally, the largest weight we observed for a particular observation was 1202. If the positivity assumption

was violated, the final calculated weight for an observation would go to infinity. We also assumed that the models used to estimate weights were correctly specified.

In the second step, we fit a discrete time failure model[7,8] for the ESRD outcome. We applied the final weight calculated in the first step to the study visit level data using the Proc Genmod in SAS, version 9.4. We included the mean of the natural log–transformed FGF23 up to each study visit. We included the baseline predictors from the stabilizing weights deriving the numerator weight in the first step. We replicated the time-varying IPW discrete time failure models approach for our composite endpoint of ESRD or death. For the composite outcome, 93 observations were truncated at 99th percentiles of the estimated weights with the largest final weight for a particular observation being 20.

Group-Based Trajectory Models

We previously published detailed methods of FGF23 trajectory analyses in the CRIC Study.[9] We calculated the maximum likelihood that the longitudinal FGF23 values would fit a discrete mixture of two or more trajectories using SAS Proc Traj.[10-13] We used a semiparametric group-based modeling strategy, which assumes that a given population is made up of multiple trajectories and simultaneously estimates probabilities for multiple trajectories. We tested for different numbers and forms of potential trajectory groups and tested for best model fit using the Bayesian Information Criterion. Similar to prior analyses, we identified three FGF23 trajectory groups based on visual inspection and what we considered clinically meaningful categories: stable, slowing rising and rapidly rising.[9] We modeled the stable FGF23 trajectory group in quadratic terms and the slowing and rapidly rising groups in linear terms based on best model fit.

We calculated the posterior predicted probability for each participant being a member of each of the 3 trajectory groups and assigned the participant to the trajectory group for which they had the highest posterior predicted probability. The mean predicted probability for each trajectory group is the mean of all the individual posterior probabilities of participants assigned to a given group. These probabilities were 0.92 (95%CI 0.89-0.94), 0.91 (95% CI 0.89-0.92) and 0.95 (95%CI 0.95-0.96) for the rapidly rising, slowly rising and stable FGF23 trajectory groups, respectively. We compared the slopes of the FGF23 trajectory groups using the "TRAJTEST" macro,[10] and we tested for heterogeneity. We derived eGFR trajectories using similar methods.

We used multiple imputations to account for missing covariate data using a multiple regression procedure in IVEware 2.0.[49] We created five imputed datasets through a sequence of multiple regressions and drew values from the corresponding predictive distributions. The type of regression models used varied by the type of variable being imputed. This relies on the assumption that missing data is at random.[14] Using weighted Cox models modified for the case-cohort design, we tested the association of membership in the three different FGF23 trajectory groups with risks of ESRD on each of the five completed datasets individually and combined the results across the imputed datasets using the rules of Rubin,[15] as previously done.[9]

Works Cited

1. de Boer IH, Sachs MC, Chonchol M, Himmelfarb J, Hoofnagle AN, Ix JH, Kremsdorf RA, Lin YS, Mehrotra R, Robinson-Cohen C, Siscovick DS, Steffes MW, Thummel KE, Tracy RP, Wang Z, Kestenbaum B. Estimated GFR and circulating 24,25-dihydroxyvitamin D3 concentration: a participant-level analysis of 5 cohort studies and clinical trials. *Am J Kidney Dis.* 2014;64(2):187-197.

2. Amdur RL, Feldman HI, Gupta J, Yang W, Kanetsky P, Shlipak M, Rahman M, Lash JP, Townsend RR, Ojo A, Roy-Chaudhury A, Go AS, Joffe M, He J, Balakrishnan VS, Kimmel PL, Kusek JW, Raj DS. Inflammation and Progression of CKD: The CRIC Study. *Clin J Am Soc Nephrol.* 2016;11(9):1546-1556.
3. Anderson AH, Yang W, Townsend RR, Pan Q, Chertow GM, Kusek JW, Charleston J, He J, Kallem R, Lash JP, Miller ER, 3rd, Rahman M, Steigerwalt S, Weir M, Wright JT, Jr., Feldman HI. Time-updated systolic blood pressure and the progression of chronic kidney disease: a cohort study. *Ann Intern Med.* 2015;162(4):258-265.
4. Stephens-Shields AJ, Spieker AJ, Anderson A, Drawz P, Fischer M, Sozio SM, Feldman H, Joffe M, Yang W, Greene T. Blood pressure and the risk of chronic kidney disease progression using multistate marginal structural models in the CRIC Study. *Stat Med.* 2017;36(26):4167-4181.
5. Williamson T, Ravani P. Marginal structural models in clinical research: when and how to use them? *Nephrol Dial Transplant.* 2017;32(suppl_2):ii84-ii90.
6. Naimi AI, Moodie EE, Auger N, Kaufman JS. Constructing inverse probability weights for continuous exposures: a comparison of methods. *Epidemiology.* 2014;25(2):292-299.
7. Allison PD. Discrete-Time Methods for the Analysis of Event Histories. *Sociological Methodology.* 1982;13:61-98.
8. Joeng HK, Chen MH, Kang S. Proportional exponentiated link transformed hazards (ELTH) models for discrete time survival data with application. *Lifetime Data Anal.* 2016;22(1):38-62.
9. Isakova T, Cai X, Lee J, Xie D, Wang X, Mehta R, Allen NB, Scialla JJ, Pencina MJ, Anderson AH, Talierco J, Chen J, Fischer MJ, Steigerwalt SP, Leonard MB, Hsu CY, de

Boer IH, Kusek JW, Feldman HI, Wolf M. Longitudinal FGF23 Trajectories and Mortality in Patients with CKD. *J Am Soc Nephrol.* 2018;29(2):579-590.

10. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol.* 2010;6:109-138.
11. Cappola AR, O'Meara ES, Guo W, Bartz TM, Fried LP, Newman AB. Trajectories of dehydroepiandrosterone sulfate predict mortality in older adults: the cardiovascular health study. *J Gerontol A Biol Sci Med Sci.* 2009;64(12):1268-1274.
12. O'Hare AM, Batten A, Burrows NR, Pavkov ME, Taylor L, Gupta I, Todd-Stenberg J, Maynard C, Rodriguez RA, Murtagh FE, Larson EB, Williams DE. Trajectories of kidney function decline in the 2 years before initiation of long-term dialysis. *Am J Kidney Dis.* 2012;59(4):513-522.
13. Ravani P, Malberti F, Tripepi G, Pecchini P, Cutrupi S, Pizzini P, Mallamaci F, Zoccali C. Vitamin D levels and patient outcome in chronic kidney disease. *Kidney Int.* 2009;75(1):88-95.
14. Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey methodology.* 2001;27(1):85-96.
15. Rubin DB. *Multiple imputation for nonresponse in surveys.* Vol 81: John Wiley & Sons; 2004.

Table S1. Baseline characteristics of individuals within and outside the random subcohort

	Total CRIC Study Population	**Random Subcohort**	**Outside Random Subcohort**
	N=3939	**N=1135**	**N=2804**
Age, years	57.7 ± 11.0	57.8 ± 10.8	57.6 ± 11.1
Female, %	45.1	43.2	46.0
Black, %	42.1	42.5	41.9
Hispanic, %	12.6	11.9	12.9
Current smoking, %	13.1	10.6	14.2
Body mass index, kg/m2	32.1 ± 7.8	32.3 ± 8.1	32.0 ± 7.7
Systolic blood pressure, mmHg	128.5 ± 22.2	127.0 ± 21.4	129.1 ± 22.5
Hypertension, %	86.1	86.2	86.1
Diabetes, %	48.4	48.6	48.4
Heart failure, %	9.7	7.8	10.5
Stroke, %	10.0	9.0	10.3
Peripheral vascular disease, %	6.7	5.5	7.1
Coronary artery disease, %	21.9	20.5	22.4
Number of hypertension medications	2.6 ± 1.5	2.4 ± 1.3	2.6 ± 1.5
ACE/ARB Use, %	68.8	69.5	68.5
IL-6, pg/mL	1.9 (1.1 –3.1)	1.9 (1.1 – 3.2)	1.9 (1.1 – 3.1)
CRP, mg/L	2.5 (1.0 – 6.3)	2.5 (1.1 – 6.5)	2.4 (1.0 – 6.1)
eGFR, ml/min/1.73m2	44.3 ± 15.0	44.8 ± 14.6	44.1 ± 15.2
UPCR, g/g	0.15 (0.06 – 0.78)	0.13 (0.05 – 0.67)	0.16 (0.06 – 0.81)
Hemoglobin, g/dL	12.6 ± 1.8	12.6 ± 1.7	12.6 ± 1.8
Serum albumin, g/dL	3.9 ± 0.5	4.0 ± 0.4	3.9 ± 0.5
Calcium, mg/dL	9.2 ± 0.5	9.2 ± 0.5	9.2 ± 0.5
Phosphate, mg/dL	3.7 ± 0.7	3.7 ± 0.7	3.7 ± 0.7
PTH, pg/mL	54.0 (35.0 – 89.6)	54.1 (35.0 – 85.6)	54.0 (35.0 – 90.8)
FGF23, RU/mL	145.5 (95.8 – 239.2)	138.3 (93.8 –215.8)	148.7 (96.9 –247.3)

Results are reported as proportions or mean ± standard deviation, median (interquartile range).

Abbreviations: CRIC, Chronic Renal Insufficiency Cohort; ACE/ARB, angiotensin converting enzyme inhibitor/angiotensin receptor blocker; IL-6, interleukin 6; CRP; c-reactive protein; eGFR, estimated glomerular filtration rate; UPCR, urine protein-to-creatinine ratio; PTH, parathyroid hormone; FGF23, fibroblast growth factor 23

Table S2. Baseline characteristics of the total CRIC population, the case-cohort population and trajectory analyses population

	Total CRIC Study Population **N = 3939**	**Case-cohort Population** **N = 1597**	**Trajectory Analyses Population** **N = 1163**
Age, years	57.7 ± 11.0	57.1 ± 11.1	57.4 ± 10.9
Female, %	45.1	42.8	45.1
Black, %	42.1	45.4	43.2
Hispanic, %	12.6	13.3	11.4
Current smoking, %	13.1	12.5	11.3
Body mass index, kg/m2	32.1 ± 7.8	32.3 ± 7.9	32.2 ± 7.8
Systolic blood pressure, mmHg	128.5 ± 22.2	130.6 ± 23.0	127.0 ± 21.3
Hypertension, %	86.1	88.4	85.6
Diabetes, %	48.4	53.4	47.8
Heart failure, %	9.7	9.1	7.7
Stroke, %	10.0	9.8	9.1
Peripheral vascular disease, %	6.7	7.0	5.3
Coronary artery disease, %	21.9	21.2	19.8
Number of hypertension medications	2.6 ± 1.5	2.6 ± 1.3	2.4 ± 1.3
ACE/ARB Use, %	68.8	70.6	70.0
IL-6, pg/mL	1.9 (1.1 –3.1)	2.0 (1.2 – 3.2)	1.8 (1.1 – 3.0)
CRP, mg/L	2.5 (1.0 – 6.3)	2.5 (1.1 – 6.5)	2.6 (1.1 – 6.4)
eGFR, ml/min/1.73m2	44.3 ± 15.0	41.9 ± 14.4	45.2 ± 14.2
UPCR, g/g	0.15 (0.06 – 0.78)	0.29 (0.07 – 1.38)	0.15 (0.06 – 0.71)
Hemoglobin, g/dL	12.6 ± 1.8	12.4 ± 1.8	12.7 ± 1.7
Serum albumin, g/dL	3.9 ± 0.5	3.9 ± 0.5	4.0 ± 0.4
Calcium, mg/dL	9.2 ± 0.5	9.1 ± 0.5	9.2 ± 0.5
Phosphate, mg/dL	3.7 ± 0.7	3.8 ± 0.7	3.7 ± 0.7
PTH, pg/mL	54.0 (35.0 – 89.6)	60.8 (37.0 – 102.8)	53.0 (34.2 – 83.8)
FGF23, RU/mL	145.5 (95.8 – 239.2)	156.2 (101.2 –250.6)	137.7 (94.1 –217.9)

Results are reported as proportions or mean ± standard deviation, median (interquartile range).

Abbreviations: CRIC, Chronic Renal Insufficiency Cohort; ACE/ARB, angiotensin converting enzyme inhibitor/angiotensin receptor blocker; IL-6, interleukin 6; CRP; c-reactive protein; eGFR, estimated glomerular filtration rate; UPCR, urine protein-to-creatinine ratio; PTH, parathyroid hormone; FGF23, fibroblast growth factor 23

Table S3. FGF23 trajectories and risks of ESRD truncated to 3 time points

FGF23 trajectory group	Total N	ESRD N	Unadjusted HR	Model 1 HR	Model 2 HR	Model 3 HR	Model 4 HR	Model 5 HR
Stable ln FGF23/year = 0.05	769	124	Reference	Reference	Reference	Reference	Reference	Reference
Slowly rising ln FGF23/year = 0.29	544	356	7.15 (5.68–9.00)	7.34 (5.73–9.41)	3.19 (2.38–4.28)	4.55 (3.34–6.20)	3.14 (2.21–4.46)	3.01 (2.11–4.30)
Rapidly rising ln FGF23/year = 0.58	94	78	20.33 (12.69–32.57)	19.50 (11.65–32.65)	7.66 (4.79–12.26)	13.84 (8.10–23.63)	5.40 (2.84–10.29)	5.07 (2.67–9.66)

Up to 3 annual time points, median duration of subsequent follow-up time 4.6 years in 1407 total participants at risk

Results are reported as hazard ratios compared to the referent group.

Covariate adjustment are for covariate values at third annual study visit (time 0 for 3 time point analyses) except when stated.

Model 1: stratified by center, adjusted for age, sex, race, and ethnicity

Model 2: Model 1 plus eGFR, UPCR, serum albumin, hemoglobin, CRP (baseline), and IL-6 (baseline)

Model 3: Model 2 plus diabetes, smoking, systolic blood pressure, body mass index, history of coronary artery disease, history of heart failure, history of stroke, and history of peripheral vascular disease, number hypertension medications, and ACE/ARB use

Model 4: Model 3 plus calcium, phosphate, PTH, lnFGF23, and baseline eGFR

Model 5: Model 4 plus eGFR trajectory

Abbreviations: FGF23, fibroblast growth factor 23; ESRD, end-stage renal disease; HR, hazard ratio; eGFR, estimated glomerular filtration rate; UPCR, urine protein-to-creatinine ratio; CRP, c-reactive protein; IL-6, interleukin 6; ACE/ARB, angiotensin converting enzyme inhibitor/angiotensin receptor blocker; PTH, parathyroid hormone