

**eTable 1.** Baseline characteristics of the Community Health Network (CHN) and Harborview Medical Center (HMC) subcohorts.

Characteristics	CHN		HMC	
	n = 16,656		n = 12,123	
<b>Demographics</b>				
Age (years) (mean, sd)	58.7	13.7	62.4	14.7
Male (n, %)	7,780	46.7	6,123	50.5
Race or ethnicity (n, %)				
Non-Hispanic White	4,675	28.1	7,082	58.4
Non-Hispanic Black	3,468	20.8	1,462	12.1
Hispanic	3,106	18.7	360	3.0
Asian	4,973	29.9	1,274	10.5
Other race or ethnicity	434	2.6	1,945	16.0
<b>Health insurance coverage</b>				
Uninsured (n, %)	6,001	36.0	2,339	19.3
Medicaid (n, %)	3,968	23.8	2,292	18.9
Medicare (n, %)	6,522	39.2	4,010	33.1
Private or commercial (n, %)	165	1.0	3,482	28.7
<b>Comorbidities</b>				
Diabetes (n, %)	3,721	22.3	2,848	23.5
Hypertension (n, %)	7,857	47.2	5,668	46.8
Cardiovascular disease (n, %)	3,162	19.0	4,785	39.5
Substance abuse (n, %)	3,127	18.8	3,467	28.6
Chronic viral disease (n, %)	3,978	23.9	1,941	16.0
<b>Laboratory data</b>				
eGFR mL/min/1.73m <sup>2</sup> (n, %)				
45 - 59	13,179	79.1	8,763	72.3
30 - 44	2,380	14.3	2,349	19.4
15 - 29	814	4.9	760	6.3
<15	283	1.7	251	2.1
Albumin (dipstick) categorical (n, %)				
None or trace	8,327	61.5	4,382	61.7
1+	2,603	19.2	1,272	17.9
2+	1,725	12.7	758	10.7
≥3+	885	6.5	696	9.8
Albumin (g/dL) (mean, sd)	3.9	0.7	3.6	0.6
Calcium (mg/dL) (mean, sd)	9.0	0.8	9.4	0.6
Hemoglobin (g/dL) (mean, sd)	12.8	2.1	13.1	2.0
Serum creatinine (mg/dL) (mean, sd)	1.5	1.1	1.51	1.1
Cholesterol (mg/dL) (mean, sd)	195.4	53.7	196.3	54.4

Missing values in the CHN cohort were distributed as follows: dipstick proteinuria 18.7%, serum cholesterol 12.4%, serum albumin 1.3%, and hemoglobin 0.4%; and the distribution of missing values in the HMC cohort was: dipstick proteinuria 41.4%, serum cholesterol 19.0%, serum calcium 12.9%, hemoglobin 3.8% and serum albumin 2.7%.

**eTable 2A.** Estimates (standard errors) of prediction performance measures for models 1-4 based on the Community Health Network cohort (n = 11,065 and n = 5,591 in the training and validation subsets).

<b>Year 1</b>				
	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>
<b>AUC</b>	0.92 (0.02)	0.95 (0.02)	0.95 (0.01)	0.95 (0.01)
<b>PE(x10)</b>	0.08 (<0.01)	0.08 (0.01)	0.08 (0.01)	0.08 (0.01)
<b>PCF(0.1)</b>	0.83 (0.05)	0.86 (0.05)	0.83 (0.04)	0.84 (0.05)
<b>PCF(0.2)</b>	0.93 (0.05)	0.98 (0.04)	0.94 (0.04)	0.98 (0.04)
<b>PNF(0.8)</b>	0.08 (0.09)	0.07 (0.03)	0.09 (0.03)	0.07 (0.03)
<b>PNF(0.9)</b>	0.19 (0.12)	0.12 (0.07)	0.14 (0.07)	0.12 (0.07)

<b>Year 3</b>				
	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>
<b>AUC</b>	0.86 (0.02)	0.93 (0.01)	0.93 (0.01)	0.93 (0.01)
<b>PE(x10)</b>	0.25 (0.01)	0.22 (0.02)	0.22 (0.03)	0.22 (0.03)
<b>PCF(0.1)</b>	0.60 (0.03)	0.74 (0.03)	0.75 (0.03)	0.81 (0.03)
<b>PCF(0.2)</b>	0.77 (0.03)	0.91 (0.03)	0.91 (0.03)	0.93 (0.02)
<b>PNF(0.8)</b>	0.24 (0.05)	0.12 (0.03)	0.12 (0.02)	0.10 (0.02)
<b>PNF(0.9)</b>	0.35 (0.06)	0.17 (0.06)	0.19 (0.05)	0.17 (0.05)

<b>Year 5</b>				
	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>
<b>AUC</b>	0.84 (0.02)	0.92 (0.01)	0.92 (0.01)	0.93 (0.01)
<b>PE(x10)</b>	0.38 (0.02)	0.34 (0.04)	0.34 (0.05)	0.33 (0.05)
<b>PCF(0.1)</b>	0.54 (0.03)	0.68 (0.03)	0.69 (0.03)	0.73 (0.03)
<b>PCF(0.2)</b>	0.71 (0.03)	0.88 (0.02)	0.87 (0.02)	0.90 (0.02)
<b>PNF(0.8)</b>	0.28 (0.04)	0.14 (0.02)	0.14 (0.02)	0.13 (0.02)
<b>PNF(0.9)</b>	0.41 (0.04)	0.22 (0.05)	0.23 (0.05)	0.20 (0.05)

Model 1: adjusted for age, sex, race-ethnicity and eGFR

Model 2: model 1 covariates + dipstick proteinuria and an interaction between eGFR and dipstick proteinuria

Model 3: model 2 covariates + health insurance coverage, comorbidities (diabetes mellitus, CVD, hypertension, substance abuse and chronic viral disease)

Model 4: model 3 covariates + additional laboratory variables (albumin, calcium, hemoglobin and cholesterol)

AUC: area under the ROC curve;

PE(x10): prediction error multiplied by a factor of 10;

PCF(q): proportion of events captured if proportion  $q$  of the population at highest risk is followed;

PNF(p): proportion of the population at highest risk that needs to be followed to capture the proportion  $p$  of the events.

**eTable 2B.** Estimates (standard errors) of prediction performance measures for models 1-4 based on the Harborview Medical Center cohort (n = 8,029 and n = 4,094 in the training and validation subsets).

<b>Year 1</b>				
	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>
<b>AUC</b>	0.94 (0.02)	0.96 (0.02)	0.96 (0.01)	0.97 (0.01)
<b>PE(x10)</b>	0.12 (0.01)	0.11 (0.01)	0.11 (0.01)	0.11 (0.02)
<b>PCF(0.1)</b>	0.88 (0.05)	0.87 (0.05)	0.87 (0.05)	0.87 (0.04)
<b>PCF(0.2)</b>	0.93 (0.04)	0.95 (0.04)	0.95 (0.04)	0.97 (0.03)
<b>PNF(0.8)</b>	0.08 (0.06)	0.06 (0.03)	0.06 (0.03)	0.05 (0.03)
<b>PNF(0.9)</b>	0.13 (0.11)	0.13 (0.07)	0.16 (0.06)	0.13 (0.06)

<b>Year 3</b>				
	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>
<b>AUC</b>	0.91 (0.02)	0.95 (0.01)	0.95 (0.01)	0.96 (0.01)
<b>PE(x10)</b>	0.30 (0.04)	0.26 (0.04)	0.24 (0.04)	0.24 (0.05)
<b>PCF(0.1)</b>	0.72 (0.03)	0.83 (0.03)	0.81 (0.03)	0.82 (0.03)
<b>PCF(0.2)</b>	0.87 (0.03)	0.91 (0.03)	0.90 (0.03)	0.92 (0.03)
<b>PNF(0.8)</b>	0.14 (0.05)	0.09 (0.03)	0.10 (0.02)	0.09 (0.02)
<b>PNF(0.9)</b>	0.30 (0.08)	0.20 (0.06)	0.20 (0.05)	0.17 (0.05)

<b>Year 5</b>				
	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>
<b>AUC</b>	0.89 (0.02)	0.93 (0.01)	0.94 (0.01)	0.94 (0.01)
<b>PE(x10)</b>	0.66 (0.08)	0.56 (0.09)	0.55 (0.10)	0.54 (0.10)
<b>PCF(0.1)</b>	0.61 (0.03)	0.73 (0.03)	0.74 (0.03)	0.73 (0.03)
<b>PCF(0.2)</b>	0.80 (0.03)	0.87 (0.03)	0.86 (0.03)	0.89 (0.03)
<b>PNF(0.8)</b>	0.20 (0.05)	0.15 (0.03)	0.14 (0.02)	0.13 (0.02)
<b>PNF(0.9)</b>	0.30 (0.06)	0.22 (0.05)	0.24 (0.05)	0.23 (0.05)

Model 1: adjusted for age, sex, race-ethnicity and eGFR

Model 2: model 1 covariates + dipstick proteinuria and an interaction between eGFR and dipstick proteinuria

Model 3: model 2 covariates + health insurance coverage, comorbidities (diabetes mellitus, CVD, hypertension, substance abuse and chronic viral disease)

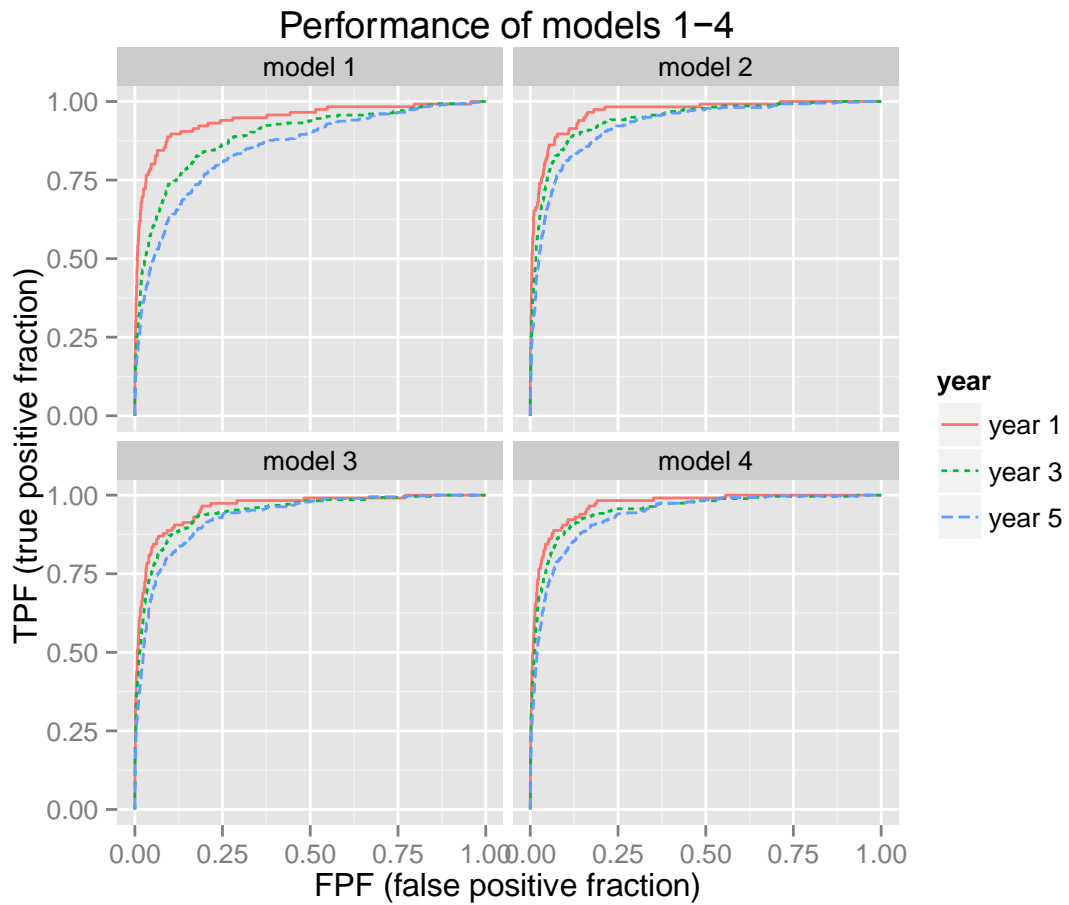
Model 4: model 3 covariates + additional laboratory variables (albumin, calcium, hemoglobin and cholesterol)

AUC: area under the ROC curve;

PE(x10): prediction error multiplied by a factor of 10;

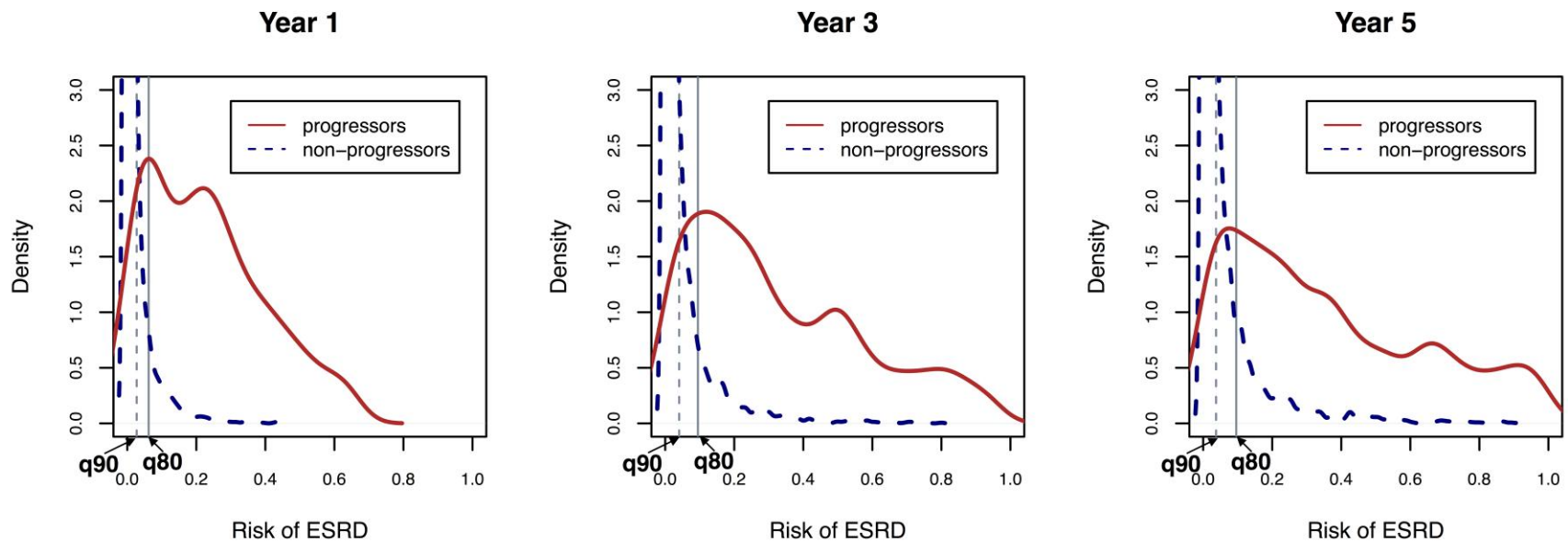
PCF(q): proportion of events captured if proportion  $q$  of the population at highest risk is followed;

PNF(p): proportion of the population at highest risk that needs to be followed to capture the proportion  $p$  of the events.



**eFigure 1.** The Receiver Operating Characteristics curve (ROC) summarizing the discrimination performance of models 1-4 for 1, 3, and 5-year time frames.

Model 2 showed a marked improvement in performance over model 1, especially for 3- and 5-year time frames. Models 3 and 4 offered minimal performance benefits over model 2. The performance of all models was highest for a 1-year time frame, and declined slightly over longer time frames.



	Year 1			Year 3			Year 5		
	N	≥q80	≥q90	Total	≥q80	≥q90	Total	≥q80	≥q90
<b>Progressors</b>	116	93	104	282	225	253	391	313	352
<b>Non-progressors</b>	9120	364	824	8298	503	1166	6847	622	1380

**eFigure 2.** The distributions of predicted risk of ESRD among subjects who did not develop ESRD (non-progressors) in a given time frame (dashed blue line) and subjects who progressed to ESRD (progressors) in that time frame (solid red). We considered three time frames - 1 year, 3 and 5 years. 80% of the ESRD progressors are to the right of the vertical solid grey line (q80), and 90% of them are to the right of the vertical dashed grey line (q90). Predicted risk at the q80 threshold is 0.062, 0.095, 0.096 for years 1, 3, and 5, respectively, and for q90 it is 0.027, 0.040, 0.037 for years 1, 3, and 5, respectively. The risk predictions are based on model 2. The density for the non-progressors (dashed blue) is ten times the height shown in the figure, reaching 30 (not shown). The table below the figure summarizes the numbers of progressors and non-progressors with predicted risk above q80 and q90 for years 1, 3 and 5 from the validation subset (n=9,564).