



Beyond standardization: Reproducible approaches to deriving clinically meaningful variables for several measures of renal function

Amy Goodwin Davies, Evanette Burrows, Ashley Batugo, Kimberley Dickinson, Sara Deakyne Davies, Richard Hoyt, Janet Zahner, Vikas R. Dharnidharka, Bradley P. Dixon, Joseph T. Flynn, Mark M. Mitsnefes, William E. Smoyer, L. Charles Bailey, Hanieh Razzaghi, Michelle R. Denburg

Correspondence: goodwindaa@chop.edu



Introduction

Standardization to the CDM is often followed by additional steps to derive clinically meaningful variables. Input from collaborators with domain knowledge is critical.

With input from nephrologists, we developed reproducible approaches to deriving variables for 3 measures of renal function: github.com/PEDSnet/Renal_Function_Measures

1. Estimated glomerular filtration rate
2. Levels of proteinuria
3. Presence of hematuria

To illustrate these measures of renal function, we include distributions for 2 cohorts

- **Longer-term** nephrology (N = 38,751): Patients with ≥2 nephrology encounters separated by ≥90 days
- **Short-term** nephrology (N = 37,809): Patients with ≥1 nephrology encounter who do not meet criteria for longer-term nephrology cohort

Patients have ≥1 year of follow-up (2 face-to-face encounters separated by ≥1 year). Data are from 6 institutions* and PEDSnet¹ data Jan. 2009 to Dec. 2020.

Characteristic**	Short-term	Longer-term
Follow-up (years, any specialty)	7.9 (4.3, 12.0)	7.6 (3.9, 12.1)
Age at first visit (years)	2.6 (0.2, 8.3)	4.2 (0.3, 10.0)
Female	17,820 (47.1%)	17,287 (44.6%)
Nephrology encounters per person-year	0.2 (0.1, 0.3)	0.9 (0.4, 1.9)

Estimated glomerular filtration rate

Estimated glomerular filtration rate (eGFR) provides an estimate of kidney function and is used to classify chronic kidney disease (CKD) stage and monitor CKD progression.

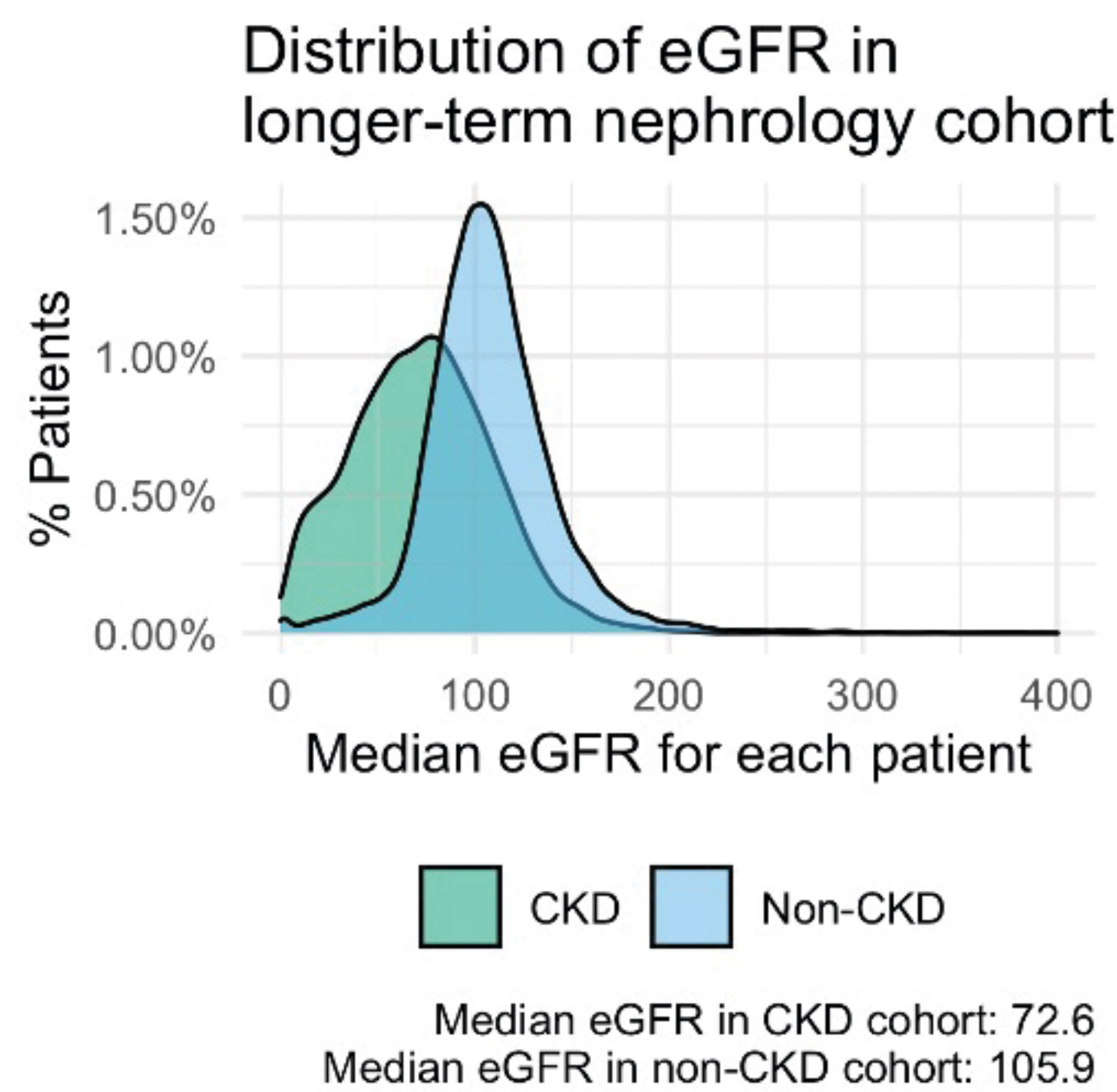
CDM → meaningful variable: eGFR is not always directly reported in the CDM and choice of eGFR equation varies. Calculating eGFR from serum creatinine measurements and height increases the number of available measurements and ensures consistent equation use for the calculation.

Implementation: Revised Bedside Schwartz Formula² (ages 1-17).

For each serum creatinine measurement, the closest available height within the specified time window is used in the calculation.

Evaluation: This approach leads to a large increase in the number of patients with eGFR measurements available for both cohorts. In the longer-term cohort, patients with ≥2 CKD diagnoses separated by ≥90 days have lower eGFRs, as expected.

Future work: Parameterize plausible bounds for serum creatinine and height, include options for various eGFR equations.



Characteristic**	Short-term	Longer-term
Directly-reported eGFR available	1,894 (5.0%)	5,205 (13.4%)
eGFR available	21,710 (57.4%)	32,453 (83.8%)
eGFR	106.2 (88.4, 125.9)	100.1 (79.8, 119.8)

Levels of proteinuria

Level of proteinuria (elevated protein in the urine) can indicate kidney injury and provide information about kidney disease progression.

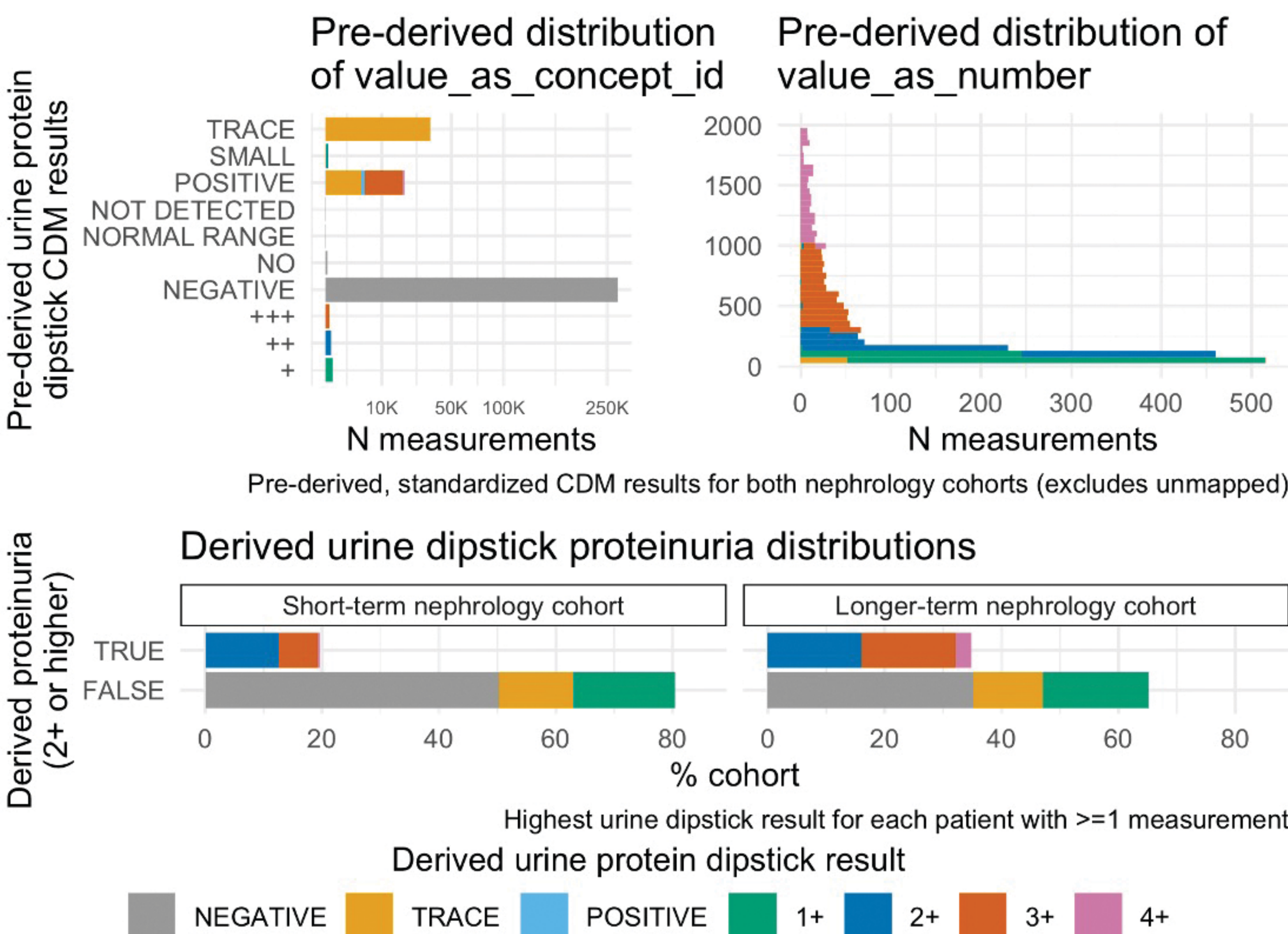
CDM → meaningful variable: Proteinuria is measured in several distinct ways and standardized representation in the CDM varies, e.g.,

- **Urine dipstick:** “1+” to “4+” scale (value_as_concept_id) or numeric mg/dL estimates (value_as_number)
- **Urine protein to creatinine ratios (UPCRs):** not always directly reported, can be calculated from urine protein and urine creatinine measurements

Implementation: Heterogenous urine dipstick results are classified. Where UPCRs are not directly reported, UPCRs are calculated from separate urine protein and urine creatinine labs within the specified time window.

Evaluation: Deriving UPCR from separate urine protein and urine creatinine measurements leads to >2x increase in the number of patients with available data for both cohorts. A greater proportion of patients in the longer-term nephrology cohort have urine protein measurements available and meet criteria for proteinuria, as expected. Periodic manual re-review is required as data is updated.

Future work: Further parameterize approach to incorporate flexibility but reduce variability in downstream processing decisions, e.g., thresholds for proteinuria, plausible bounds for quantitative measurements, time window for associating urine protein and urine creatinine measurements.



*The following PEDSnet (pedsnet.org) institutions were included: Children’s Hospital of Philadelphia (CHOP); Children’s Hospital of Colorado; Nationwide Children’s Hospital; Nemours Children’s Health System (a Delaware and Florida health system); Seattle Children’s Hospital; and St. Louis Children’s Hospital. PEDSnet database v4.1.**Categorical reported as N patients (% cohort) and continuous reported as median (IQR). Results calculated across all available data for patients. We gratefully acknowledge feedback and assistance from the PEDSnet Data Coordinating Center at the Children’s Hospital of Philadelphia, as well as two anonymous OHDSI reviewers. This project was funded in part by a grant (PCRnet CRN-2020-007) from the Patient-Centered Outcomes Research Institute (PCORI). Research reported in this publication was funded by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under awards P50DK114786 (Children’s Hospital of Philadelphia Pediatric Center of Excellence in Nephrology) and R21DK116151. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
1. Forrest C, Margolis P, Bailey LC, et al. PEDSnet: a national pediatric learning health system. J Am Med Inform Assoc. 2014; 21(4):602-606. 2. Schwartz GJ, Munoz A, Schneider MF, et al. New equations to estimate GFR in children with CKD. J Am Soc Nephrol. 2009;20(3):629-637.

Characteristic**	Short-term	Longer-term
Urine dipstick protein measurement available	20,044 (53.0%)	28,823 (74.4%)
Proteinuria via urine dipstick (≥2+)		
within entire cohort	3,924 (10.4%)	10,034 (25.9%)
within cohort with available measurement(s)	3,924 (19.6%)	10,034 (34.8%)
Directly-reported UPCR available	1,954 (5.2%)	6,850 (17.7%)
Directly-reported UPCR	0.14 (0.07, 0.40)	0.22 (0.09, 0.80)
Proteinuria via directly-reported UPCR (≥0.2 mg:mg)		
within entire cohort	861 (2.3%)	4,230 (10.9%)
within cohort with available measurement(s)	861 (44.1%)	4,230 (61.8%)
Derived UPCR available	4,733 (12.5%)	13,933 (36.0%)
Derived UPCR	0.15 (0.07, 0.42)	0.21 (0.09, 0.76)
Proteinuria via derived UPCR (≥0.2 mg:mg)		
within entire cohort	2,277 (6.0%)	9,094 (23.5%)
within cohort with available measurement(s)	2,277 (48.1%)	9,094 (65.3%)
Proteinuria via combined definition (≥2+/ \geq 0.2 mg:mg)		
within entire cohort	5,153 (13.6%)	13,429 (34.7%)
within cohort with available measurement(s)	5,153 (24.3%)	13,429 (42.9%)

Presence of hematuria

Blood in the urine (hematuria) can be a sign of glomerular kidney disease.

Categorization of most frequent urine blood value_source_value results

CDM → meaningful variable: Dipstick and microscopy urine blood tests use the same lab codes. For correct interpretation, test type must be classified based on result. Microscopic test results take precedence. For both test types, CDM results are heterogenous and require classification, e.g.,

- **Dipstick:** Small-Large, 1+/2+/3+
- **Microscopy:** None, 0-2/2-5/5-10, TNTC (“Too numerous to count”)

Implementation: Measurement is identified as dipstick/microscopy based on result. Heterogenous results are classified. Hierarchy is applied so microscopy results take precedence.

Evaluation: A greater proportion of longer-term cohort have measurements available and evidence for hematuria, as expected. Periodic manual re-review is required as data is updated.

Future work: Parameterize timeframe for associating multiple urine blood tests, increase granularity of categorization (e.g., from NEGATIVE/POSITIVE to NEGATIVE/1+/2+ etc.).

Characteristic**	Short-term	Longer-term
Urine blood measurement available	21,215 (56.1%)	29,607 (76.4%)
Hematuria		
within entire cohort	7,480 (19.8%)	14,675 (37.9%)
within cohort with available measurement(s)	7,480 (35.3%)	14,675 (49.6%)

