Supplemental Methods

Laboratory Measurement Data

Manual examination of 15,794 unique test names (by author NB, a board-certified hematologist) reduced the number of distinct test names relevant to paraprotein testing to 4,133. We retrieved 63,696 laboratory records of assays performed on a urine or blood sample from cohort members from October 1, 1999 to September 30, 2002. Of these, we excluded 5,613 (8.8%) records likely representing cancelled or otherwise invalid results (e.g., hemolyzed sample, sample not received), leaving 58,083 records for further evaluation.
Paraprotein test records included results from one of several different assays (e.g., serum protein electrophoresis or immunofixation, urine protein electrophoresis or immunofixation), and we observed that results were potentially recorded in multiple fields within each record, including within a “comments” field. The comments field includes any free-from text pertaining to paraprotein results entered by the laboratory, which may include positive or negative phrasing left out of the main results fields.

Using a text-search algorithm, paraprotein test results meeting criteria indicating a positive result were provisionally classified as a positive. Next, records were again scanned for negative keywords and phrases (e.g., “absent”, “normal”, “not detected”) and classified as negative if any of these were present. A small subset of test results (5.3%, n= 1,165) remained “unclassified”, meaning there was insufficient information to report a result as positive or negative. Next, the positive and negative keywords and phrases were iteratively refined until the sensitivity (>70%) and specificity (>90%) of the algorithm results were acceptable when compared to manual review (training set) of a 2% (n=1,123) random sample of records (by author NB) and the number of unclassifiable tests could no longer be significantly reduced by successive refinements to the algorithm.

To validate our process, in the final step we drew a random validation sample of test records (2%, N=1,179). Among 1,074 of these validation records classified by the search algorithm and manual expert review, our criteria correctly classified 939 records (87%), with 74% sensitivity

(95% confidence interval, 69-78%) and 94% specificity (95% confidence interval, 92-96%) (Supplemental Table 5).

Finally, paraprotein test results within one year of cohort entry were considered for each veteran in the cohort. Results on the same date were considered positive if a veteran had one or more positive result on that date. For analysis, veterans were assigned the most recent paraprotein result prior to cohort entry, or for veterans with no paraprotein results within 1 year prior to entry, the first paraprotein result within one year of cohort entry was used. Of 2,156,317 veterans in the cohort, 21,898 (1%) had paraprotein test results meeting these criteria, and the remaining cohort members were classified as "untested" for analysis.

Statistical Analysis

We conducted several sensitivity analyses. First, we repeated the primary analysis after additionally adjusting for dipstick proteinuria (Negative; Trace/1+; 2+ and higher) among the veterans with information on proteinuria available from VA CDW (25). We chose to adjust for dipstick proteinuria, since total urine protein or protein to creatinine ratio are available for a much smaller number of patients and are more likely to be subject to confounding by indication. Urine dipstick was not used as an assessment of urinary paraprotein. Second, we repeated the primary analysis after excluding veterans with one or more ICD-9 diagnosis code for multiple myeloma (203.0, 203.00, 203.01, 203.02) in either VA or Medicare records within one year before or after cohort entry.

Supplemental Material

Supplemental Table 1

Keywords and LOINCs used to search VA laboratory records for assays plausibly related to paraprotein testing. Percentage sign indicates a wildcard which can be any number of characters including spaces. Searches were case-insensitive. For information on LOINC see https://loinc.org
Keywords
%protein%
%SPEP%
%UPEP%
%pep%
%electrophoresis%
%spike%
%monoclonal%
%immunofixation%
%immfix%
%spike%
%bence%jones%
%IFE%
%IEP%

LOINCs										
15182-9	17007-6	55921-1	55924-5	11050-2	15189-4	49265-2	81231-3	48627-4	56768-5	44605-4
18299-8	18301-2	14002-0	29585-7	15188-6	33559-6	13169-8	12851-2	51436-4	32629-8	48771-0
18300-4	18302-0	15186-0	29586-5	25681-8	49908-7	13440-3	13438-7	51435-6	33647-9	10496-8
2458-8	2465-3	17009-2	34672-6	25682-6	6788-4	25700-6	14895-7	56766-9	42483-8	10497-6
25444-1	25445-8	18303-8	11219-3	27365-6	11051-0	33042-3	14896-5	35559-4	42484-6	48774-4
26871-4	30141-6	18304-6	14795-9	40635-5	15187-8	34440-8	32210-7	35561-0	72577-0	48777-7
30140-8	33415-1	2472-9	17793-1	40636-3	25683-4	49274-4	34435-8	51437-2	15181-1	48776-9
38487-5	38488-3	25446-6	1928-1	44394-5	25684-2	49275-1	34641-1	51438-0	21349-6	48773-6
44388-7	44595-7	26958-9	1929-9	6785-0	27394-6	49276-9	49296-7	66481-3	2457-0	48772-8
44389-5	47083-1	30142-4	31152-2	6786-8	40638-9	49277-7	49297-5	74637-0	44582-5	10499-2
55902-1	55922-9	44600-5	46084-0	6787-6	40639-7	49278-5	49298-3	44932-2	54900-6	10500-7
56121-7	55923-7	50992-7	49270-2	12777-9	44395-2	57457-4	49299-1	59800-3	6938-5	48775-1
58763-4	55974-0	55925-2	74773-3	36916-5	6789-2	74639-6	49300-7	50796-2	44589-0	
6778-5	56122-5	56123-3	43110-6	38176-4	6790-0	74641-2	49301-5	51439-8	15183-7	
6779-3	6780-1	58764-2	44792-0	38177-2	6791-8	74638-8	49302-3	51440-6	21350-4	
76487-8	6781-9	6783-5	44793-8	40637-1	12778-7	74640-4	38528-6	74636-2	2464-6	
74864-0	76490-2	6784-3	81632-2	43940-6	33944-0	13348-8	49587-9	50792-1	54901-4	
74869-9	77177-4	76491-0	14977-3	44555-1	38169-9	74665-1	50408-4	59801-1	55975-7	
74865-7	74862-4	74866-5	51406-7	44604-7	38178-0	74666-9	56154-8	50797-0	76489-4	
2459-6	74868-1	74870-7	55897-3	80515-0	40640-5	55295-0	33018-3	59563-7	44599-9	
27019-9	74863-2	74867-3	55898-1	40844-3	50978-6	48811-4	33358-3	75517-3	15185-2	
44587-4	26814-4	2473-7	55899-9	41759-2	53574-0	24351-9	33879-8	50793-9	21351-2	
44588-2	27018-1	27020-7	55900-5	48378-4	56481-5	24352-7	35560-2	59802-9	2471-1	
55901-3	44597-3	44601-3	56125-8	49553-1	80516-8	34539-7	40661-1	56767-7	54902-2	
15184-5	44598-1	44602-1	57778-3	80517-6	34550-4	77381-2	42482-0	57667-8	44603-9	

Supplemental Table 2. Adjusted hazard ratios for ESKD by paraprotein test status and eGFR category.

	eGFR (mL/min/1.73 m^2)							
	≥60		45-59		30-44		15-29	
Paraprotein	aHR[a]	(95% CI)	aHR[a]	(95% CI)	aHR[a]	(95% CI)	aHR[a]	(95% CI)
Untested	0.39	(0.33, 0.45)	0.43	(0.38, 0.50)	0.65	(0.58, 0.72)	0.79	(0.69, 0.89)
Negative	Ref.		Ref.		Ref.		Ref.	
Positive	1.48	(1.08, 2.01)	1.06	(0.78, 1.46)	1.03	(0.81, 1.32)	1.49	(1.16, 1.92)
Unclassified	1.11	(0.63, 1.96)	1.46	(0.93, 2.30)	1.11	(0.75, 1.63)	1.16	(0.75, 1.78)

Adjusted hazard ratios and 95% confidence intervals for end-stage kidney disease, excluding cohort members with a diagnosis code for multiple myeloma ± one year from cohort entry.

ESKD= end-stage kidney disease; aHR=adjusted hazard ratio; CI=confidence interval; eGFR=estimated glomerular filtration rate; Ref=reference category

[a]Adjusted for age, sex, Charlson, Diabetes (y/n), CHF (y/n), Vascular disease (y/n), race (African-American or other), and within category, continuous eGFR

Supplemental Table 3. Adjusted hazard ratios for ESKD by paraprotein test status and eGFR category.

	eGFR (mL/min /1.73 m^2)							
	≥60		45-59		30-44		15-29	
Paraprotein	aHR[a]	(95% CI)	aHR[a]	(95% CI)	aHR[a]	(95% CI)	aHR[a]	(95% CI)
Untested	0.39	(0.31, 0.50)	0.41	(0.32, 0.52)	0.62	(0.51, 0.75)	0.73	(0.58, 0.91)
Negative	Ref.		Ref.		Ref.		Ref.	
Positive	1.84	(1.18, 2.87)	1.18	(0.71, 1.98)	1.28	(0.86, 1.90)	1.59	(1.10, 2.30)
Unclassified	1.28	(0.51, 3.22)	0.28	(0.04, 2.09)	1.06	(0.59, 1.91)	0.69	(0.30, 1.59)

Adjusted hazard ratios and 95% confidence intervals for end-stage kidney disease among cohort members with proteinuria data.

ESKD= end-stage kidney disease; aHR=adjusted hazard ratio; CI=confidence interval; eGFR=estimated glomerular filtration rate; Ref=reference category

[a]Adjusted for age, sex, Charlson, Diabetes (y/n), CHF (y/n), Vascular disease (y/n), race (African-American or other), continuous eGFR within categories, and proteinuria (negative, T/1+, 2+ or more, missing)

Supplemental Table 4. Causes of ESKD by paraprotein test status and eGFR category.

	eGFR ≥60			eGFR 45-59			eGFR 30-44			eGFR 15-29		
Diagnosis (%)	Untested	Negative	Positive	Untested	Negative	Positive	Untested	Negative	Positive	Untested	Negative	Positive
Diabetes	51.3	55.6	42.1	48.0	53.4	49.2	45.9	48.5	32.5	44.2	52.7	48.3
Hypertension	20.4	12.2	7.9	29.1	22.2	26.2	33.7	30.2	39.8	34.7	29.6	26.4
MM/LC/Amyloidosis	2.0	6.9	27.6	1.1	2.3	9.2	0.4	2.0	9.0	0.2	1.2	4.0
Glomerulonephritis	6.0	8.0	6.6	4.8	8.1	4.6	5.3	7.1	4.9	6.8	6.2	6.9
Interstitial nephritis	3.2	2.7	4.0	3.1	1.8	4.6	3.2	3.4	1.6	3.3	2.5	4.6
Other	9.7	4.2	7.9	7.6	3.2	1.5	6.5	3.0	4.9	5.9	3.2	5.7
Uncertain etiology	4.1	4.8	2.6	4.4	6.3	4.6	3.9	3.7	7.3	3.7	4.3	4.0
Missing code	2.8	4.2	1.3	1.9	2.7	0.0	1.1	2.2	0.0	1.0	0.4	0.0
Total ESKD Cases	n=11179	n=189	n=76	n=7078	n=221	n=65	n=8811	n=408	n=123	n=7929	n=565	n=174

The percent of ESKD cases attributed to specific diagnoses by eGFR category is tabulated. Percentages are listed separately for patients not tested for monoclonal protein (untested) or tested with positive vs negative test results. Diagnoses were identified through the Centers for Medicare and Medicaid Services form 2728.

eGFR= estimated glomerular filtration rate in mL/min/1.73 m^2. ESKD= end-stage kidney disease, n= total number of ESKD cases for each eGFR category, MM= multiple myeloma, LC= light chain nephropathy.

Supplemental Table 5. Comparison of classification of paraprotein lab records by our criteria (algorithm) and "by-hand", i.e., by review of the record by study author (N.B.).

VALIDATION SET					
	Algorithm			Gold Standard (by-hand)	
	N	%		N	%
Negative	788	66.8		749	63.5
Positive	327	27.7		364	30.9
Unclassified	64	5.4		66	5.6
Total	1,179	100		1,179	100
	Gold Standard (by-hand)				
Algorithm	Negative	Positive	Unclassified	Total	
Negative	672	94	22	788	
Positive	41	267	19	327	
Unclassified	36	3	25	64	
Total	749	364	66	1,179	
Among records classified by both algorithm and gold standard					
	Value	95% CI			
Sensitivity	74.0%	(69.1 –78.4)			
Specificity	94.2%	(92.3 – 95.8)			

Supplemental Figure 1. Iterative method for classification of paraprotein laboratory records as positive, negative, or unclassified.