Quantifying the Clinical Implications of Imprecision in Kidney Function Estimates

Morgan Sanchez, PhD[1,2], Jesse Cummings, MEng[1], James A. Diao, MD, MPhil[1-3], Ben Y. Reis, PhD[1,2,4], Ran D. Balicer, MD, MPH[2,5,6], Noa Dagan, MD, PhD[2,5,7], Arjun K. Manrai, PhD[1,2]

[1] *Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA*

[2] *Ivan and Francesca Berkowitz Family Living Laboratory, Harvard Medical School and Clalit Research Institute, Boston, MA, USA and Tel Aviv, Israel*

[3] *Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA*

[4] *Predictive Medicine Group, Boston Children's Hospital, Boston, MA, USA*

[5] *Clalit Research Institute, Innovation Division, Clalit Health Services, Tel Aviv, Israel*

[6] *School of Public Health, Faculty of Health Sciences, Ben Gurion University of the Negev, Be'er Sheva, Israel*

[7] *Faculty of Computer and Information Science, Ben Gurion University, Be'er Sheva, Israel*

Precision medicine requires personalizing not only a patient's estimated risk, but also the uncertainty surrounding decisions based on that estimate. We illustrate this challenge by using creatinine-based estimated glomerular filtration rate (eGFRcr), a measure of kidney function widely used to guide diagnosis and treatment in patients with chronic kidney disease. Using NHANES 2001–2018 data from 44,360 adults, we applied published distributions of measured GFR (mGFR) conditional on eGFRcr to estimate how often measured and estimated GFR yield different classifications at key clinical thresholds. We also compared prioritizing confirmatory testing by threshold-proximity with using an individual's probability of discordant classification. Among 74.5 million U.S. adults with eGFRcr <90 mL/min/1.73 m², 6.6 million had clinical characteristics recognized by guidelines as potential indications for confirmation. However, estimated and measured GFR are expected to fall on opposite sides of a single common threshold, 60 mL/min/1.73 m², in more than 14 million adults. Proximity-based testing closely approximated probabilistic prioritization for GFR thresholds, but rankings diverged for multivariable kidney failure risk estimates. These findings illustrate the population-scale implications of uncertainty in a routine clinical estimate. Accounting for how that uncertainty affects individual decisions could help target confirmatory testing and extend precision medicine beyond personalized point estimates.

Keywords: Clinical Decision Support; Precision Uncertainty; Estimated Glomerular Filtration Rate; Chronic Kidney Disease; Confirmatory Testing.

1. Introduction

Precision medicine has largely focused on personalizing risk estimates and treatments as point estimates while often ignoring the uncertainty surrounding these estimates and the clinical implications of this uncertainty. Population-level measures of model performance, such as bias or accuracy within a fixed percentage, do not directly answer the question often of interest to the physician of how likely the patient is to be on one side of the decision boundary versus the other. A

© 2026 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.

more complete precision-medicine framework would personalize not only the estimate, but also its uncertainty and its implications for care, including when to order confirmatory testing.

Kidney function assessment provides a useful case study. Glomerular filtration rate (GFR) is used to diagnose and stage chronic kidney disease, guide medication use, determine clinical-trial eligibility, and estimate kidney failure risk, often through thresholds on GFR itself or on estimated kidney failure risk. Though GFR can be measured, in routine care, it is often estimated via creatinine-based eGFR (eGFRcr). Guidelines recommend confirmatory testing when greater accuracy may affect a clinical decision, as alternative filtration markers can be costly and measured GFR can be difficult to obtain.[1] The need for confirmatory testing can arise because clinical conditions, medications, or exposures alter serum creatinine independently of GFR, or because ordinary estimation uncertainty obscures classification at a clinically meaningful threshold. The first setting is partly addressed through guideline-recognized indications for alternative filtration-marker testing or measured GFR. The second is less easily captured by categorical clinical indications because its importance depends jointly on the estimated value and the decision being considered.

As with many prediction tools, most evaluations of eGFR equations emphasize population-level bias, precision, and accuracy,[2–8] but recent work has characterized distributions of measured GFR conditional on eGFRcr.[9] Such distributions can be propagated through clinical thresholds to estimate the probability of decision-discordant classification, which reframes eGFR uncertainty as a decision problem. The same eGFRcr may be sufficiently informative for one clinical question but inadequate for another, depending on the relevant threshold, the patient's other risk factors, and the consequences of classification error. For thresholds on GFR itself, decision uncertainty may depend largely on distance from the boundary. For thresholds on quantities that rely on multiple inputs, including GFR, such as the Kidney Failure Risk Equation (KFRE), however, the GFR value corresponding to a given risk threshold is patient-specific because it also depends on age, sex, and albuminuria.[10] This distinction provides a setting in which to test when simple proximity is sufficient and when probabilistic uncertainty modeling may add information.

In this study, we used nationally-representative data from the National Health and Nutrition Examination Survey (NHANES) data to evaluate the population-level implications of decision-relevant eGFR uncertainty.[11] We first estimated the prevalence of guideline-recognized indications for confirmatory testing. We then integrated published conditional distributions of measured GFR given eGFRcr to estimate expected classification discordance at common GFR and KFRE thresholds. Finally, we compared simple proximity-based testing strategies with a probabilistic reference based on participant-level decision-discordance probabilities. Through these analyses, we assess when simple threshold proximity adequately represents decision uncertainty, when propagation through patient-specific decision boundaries may benefit from probabilistic modeling, and how uncertainty estimates could support more individualized and actionable kidney function assessment. Beyond the potential for improving clinical classification, such individualized assessments may have utility in new biomarker and therapeutic development, where separating true physiological change in kidney function from marker-specific measurement error is crucial. Such measurement error may arise from non-GFR biological determinants such as muscle mass and creatinine generation, tubular secretion, and inflammation. A first step in detecting biologically

meaningful signals and defining the mechanistic underpinnings of kidney function is identifying individuals with clinically relevant discordance between estimated and measured GFR.

2. Methods

2.1. *Study Population*

We used nine cycles of NHANES spanning 2001-2018.[11] The primary population included 44,360 nonpregnant adults aged ≥18 years with serum creatinine measurements. A secondary population of 43,782 participants with both serum creatinine and urinary albumin-to-creatinine ratio (UACR) measurements was used for kidney failure risk analyses. Survey weights were adjusted for missing creatinine in the primary population and for missing creatinine and UACR in the secondary population. All NHANES participants provided written consent under protocols approved by the National Center for Health Statistics Research Ethics Review Board.

2.2. *Demographic and Clinical Characteristics*

We extracted demographic characteristics, body mass index, smoking status, diabetes, hypertension, serum creatinine, UACR, and prescription medication use from NHANES files. Diabetes and hypertension were defined using self-report, medication use, laboratory measurements, and examination data. Serum creatinine was corrected according to NHANES analytic guidance where applicable.[12] eGFRcr was calculated using the 2021 CKD-EPI creatinine equation,[13] and albuminuria was categorized using standard UACR thresholds. Detailed variable definitions are provided in the Appendix.

2.3. *Clinical Indications for Confirmatory GFR Testing*

Kidney Disease: Improving Global Outcomes (KDIGO) guidelines recommend confirmatory testing when eGFRcr may not be sufficiently accurate for a GFR-dependent decision.[1] We grouped potential indications according to whether they primarily affected creatinine, cystatin C, or both markers. NHANES-observable creatinine-specific limitations included BMI <18.5 kg/m² as a proxy for low muscle mass, creatine supplement use, and medications affecting creatinine secretion. Cystatin C-specific limitations included current smoking and systemic corticosteroid use. Limitations affecting both markers included cirrhosis or advanced liver disease and class III obesity. Participants were classified as potentially suitable for cystatin C confirmation, as having a potential limitation of cystatin C confirmation, or as potentially requiring measured GFR based on these categories.

2.4. *Interventional Thresholds Using Kidney Function and Kidney Failure Risk Estimates*

We evaluated GFR thresholds of 30, 45, and 60 mL/min/1.73 m² because they recur across kidney disease staging, medication guidance, and other GFR-dependent decisions.[14–18] For each threshold, participants were classified according to whether eGFRcr fell below the boundary and whether modeled mGFR would yield a discordant classification.

Among participants with UACR measurements, we also evaluated 2-year KFRE thresholds of 10% and 40% and a 5-year threshold of 5%, which are used in kidney disease management.[1] eGFRcr-based KFRE classification was compared with classification after substituting modeled mGFR while holding age, sex, and UACR fixed.

2.5. *Estimating the National Burden of Indications for Confirmatory Testing*

We estimated the prevalence of each indication for confirmatory testing as well as combined confirmatory testing categories in the primary NHANES population and associated survey-adjusted U.S. adult population, indicating counts and proportion of the population with each individual indication and each category. Because most GFR-based decision thresholds fall below the normal GFR range, we also estimated these quantities among adults with eGFR < 90 mL/min/1.73 m^2.

2.6. *Estimating Measured GFR Distributions Conditional on Creatinine-Based Estimates*

We used published conditional distributions of mGFR given eGFRcr calculated with the 2021 CKD-EPI equation.[9,13] For each participant, mGFR quantiles were obtained from the published calculator and Figure 1 of Shafi et al.[9] A conditional cumulative distribution function was approximated by linear interpolation between quantiles. For each decision threshold, the participant-level probability of discordance was defined as the probability that mGFR and eGFRcr were on opposite sides of the GFR threshold. (Fig. 1)

2.7. *Estimating Expected Discordance at KFRE Risk Thresholds*

For each participant and KFRE threshold, we obtained the GFR value that placed the participant exactly at the threshold while holding age, sex, and UACR fixed. Because KFRE risk decreases monotonically with GFR, this value defined a participant-specific decision boundary. The conditional mGFR distribution was then used to estimate the probability that mGFR-based and eGFRcr-based KFRE classifications would differ.

2.8. *Survey-weighted National Estimates and Statistical Uncertainty*

Participant-level discordance probabilities were summed using NHANES survey weights to estimate the expected number and proportion of U.S. adults with discordant classification at each threshold. Survey weights, strata, and primary sampling units were incorporated into all analyses. 95% confidence intervals were calculated using the appropriate NHANES survey design.

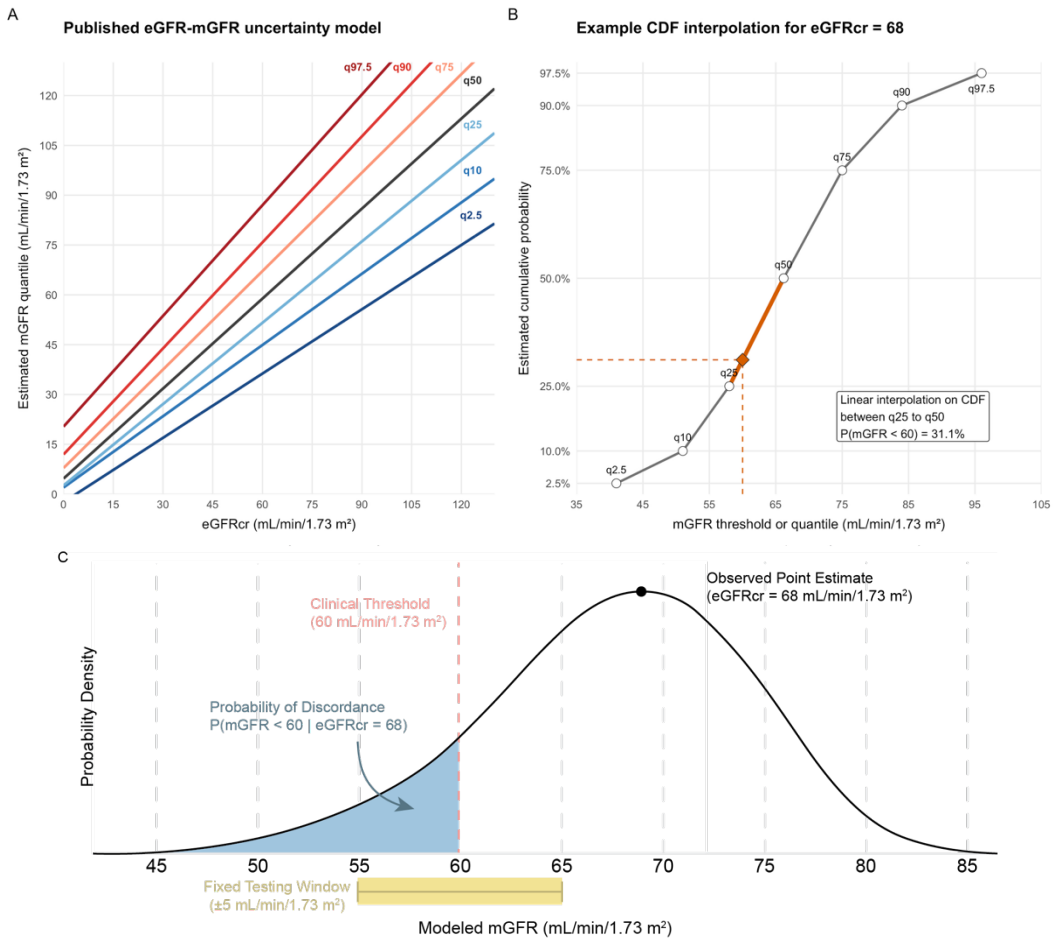


Fig. 1. Estimating decision-relevant GFR uncertainty from eGFR. (a) Conditional distributions of measured GFR across observed eGFRcr values. (b) Example conditional cumulative distribution for an individual with eGFRcr = 68 mL/min/1.73 m². (c) Decision-discordance probability, defined as the probability that measured GFR falls on the opposite side of a clinical threshold from eGFRcr.

2.9. *Proximity-Based Confirmatory Testing*

We evaluated proximity-based strategies for prioritizing confirmatory testing. For direct GFR thresholds, proximity was defined as the absolute distance between eGFRcr and the threshold; for KFRE thresholds, it was the absolute distance between eGFRcr and the participant-specific GFR value corresponding to the risk threshold. We assessed fixed windows of ±5, ±8, and ±10 mL/min/1.73 m² and continuous ranking by boundary distance. Performance was evaluated under the conditional mGFR model using testing burden, expected discordant classifications captured, percentage of total expected discordance captured, and expected discordant classifications per 100 tests. To contextualize the yield of proximity-based ranking under this model, we also compared it with a reference ranking based directly on estimated discordance probability.

3. Results

3.1. *Study Population Characteristics*

The primary population included 44,360 nonpregnant adults with serum creatinine measurements, and the secondary population included 43,782 adults with both serum creatinine and UACR measurements. Each represented approximately 230.1 million U.S. adults after survey weighting. Most participants had normal or mildly reduced kidney function. Median eGFRcr was 99.6 mL/min/1.73 m², and 95.7% had eGFRcr ≥60 mL/min/1.73 m². Additional characteristics are shown in Table 1.

3.2. *National Prevalence of Clinical Indications for Confirmatory GFR Testing*

NHANES-observable potential indications for confirmatory GFR testing varied substantially by filtration marker and indication category (Table 2). In the survey-weighted U.S. adult population, 5.3 million adults (2.29%) had a creatinine-specific limitation suggesting potential utility of cystatin C confirmation, 51.6 million adults (22.44%) had a potential limitation of cystatin C-based estimation only, and 16.6 million adults (7.22%) had a potential indication for mGFR when greater precision is required. Among adults with eGFRcr <90 mL/min/1.73 m² comprising 74.5 million individuals, these estimates were 1.4 million (1.88%), 13.0 million (17.49%), and 5.2 million (6.97%), respectively. Current smoking was the primary contributor to potential limitations of cystatin C-based estimation, while class III obesity was the primary contributor to potential indications for mGFR.

Table 1. Characteristics of the NHANES analytic populations and corresponding survey-weighted U.S. adult populations. Values are unweighted counts and percentages for NHANES participants and survey-weighted estimates for the U.S. population, unless otherwise indicated.

Variable	NHANES SCr population 2001–2018	U.S. SCr population 2001–2018	NHANES SCr+UACR population 2001–2018	U.S. SCr+UACR population 2001–2018
Sample size, no.	44,360	230,108,444	43,782	230,108,444
Female, no. (%)	22,208 (50.06)	116,728,879 (50.73)	21,426 (50.09)	116,740,554 (50.73)
Median SCr (IQR), mg/dL	0.85 (0.29)	0.85 (0.28)	0.85 (0.29)	0.85 (0.28)
Median eGFR (IQR), mL/min/1.73 m^2	99.3 (29.8)	99.6 (27.5)	99.3 (29.7)	99.5 (27.4)
Median UACR (IQR), mg/g	6.78 (8.51)	6.3 (6.98)	6.78 (8.51)	6.3 (6.99)
Median BMI (IQR), kg/m^2	27.8 (8.29)	27.6 (8.21)	27.8 (8.28)	27.6 (8.21)
Race/ethnicity				
Black, no. (%)	9,522 (21.47)	26,603,289 (11.56)	9,363 (21.39)	26,603,289 (11.56)
White, no. (%)	18,598 (41.93)	154,773,674 (67.26)	18,329 (41.86)	154,725,248 (67.24)
Hispanic, no. (%)	11,812 (26.63)	32,379,855 (14.07)	11,707 (26.74)	32,440,478 (14.10)
Other, no. (%)	4,428 (9.98)	16,351,626 (7.11)	4,383 (10.01)	16,339,430 (7.10)
Age, years				
Aged 18–44, no. (%)	21,093 (47.55)	114,995,121 (49.97)	20,879 (47.69)	114,916,825 (49.94)
Aged 45–64, no. (%)	14,725 (33.19)	81,591,407 (35.46)	14,582 (33.31)	81,548,863 (35.44)
Aged 65 or older, no. (%)	8,542 (19.26)	33,521,916 (14.57)	8,321 (19.01)	33,642,757 (14.62)
BMI, kg/m^2				
<18.5, no. (%)	792 (1.81)	3,947,594 (1.74)	780 (1.80)	3,933,693 (1.73)
18.5–24.9, no. (%)	12,680 (29.04)	67,672,080 (29.80)	12,554 (29.05)	67,828,940 (29.80)
25.0–29.9, no. (%)	14,299 (32.75)	74,577,172 (32.84)	14,167 (32.78)	74,803,011 (32.86)
30.0–39.9, no. (%)	12,890 (29.52)	66,069,349 (29.09)	12,764 (29.53)	66,298,631 (29.12)
≥40, no. (%)	2,999 (6.87)	14,842,632 (6.54)	2,952 (6.83)	14,782,270 (6.49)
eGFR, mL/min/1.73 m^2				
<15, no. (%)	120 (0.27)	343,189 (0.15)	79 (0.18)	232,719 (0.10)
15–29, no. (%)	194 (0.44)	573,795 (0.25)	186 (0.42)	562,988 (0.24)
30–44, no. (%)	626 (1.41)	2,119,580 (0.92)	588 (1.34)	2,050,766 (0.89)
45–59, no. (%)	1,714 (3.86)	6,896,319 (3.00)	1,666 (3.81)	6,899,878 (3.00)
60–89, no. (%)	12,392 (27.94)	64,551,796 (28.05)	12,240 (27.96)	64,762,750 (28.14)
≥90, no. (%)	29,314 (66.08)	155,623,766 (67.63)	29,023 (66.29)	155,599,343 (67.62)
Albuminuria				
A1: <30 mg/g, no. (%)	38,860 (88.76)	207,645,663 (91.26)	38,860 (88.76)	209,923,305 (91.23)
A2: 30–299 mg/g, no. (%)	4,093 (9.35)	17,028,401 (7.48)	4,093 (9.35)	17,287,381 (7.51)
A3: ≥300 mg/g, no. (%)	829 (1.89)	2,855,614 (1.26)	829 (1.89)	2,897,758 (1.26)
Clinical characteristics				
Obesity, no. (%)	15,889 (36.39)	80,911,982 (35.63)	15,716 (36.37)	81,080,902 (35.62)
Diabetes, no. (%)	6,727 (15.16)	26,067,998 (11.33)	6,568 (15.00)	25,935,971 (11.27)
Hypertension, no. (%)	20,645 (46.54)	99,905,143 (43.42)	20,325 (46.42)	99,976,848 (43.45)
Current smoking, no. (%)	9,227 (20.80)	49,500,244 (21.51)	9,083 (20.75)	49,346,343 (21.44)

Table 2. NHANES-observable potential indications for confirmatory GFR testing in U.S. adults.

	NHANES SCr Population 2001-2018 no. (%)	**U.S. SCr Population 2001-2018 no. (%)**	**U.S. SCr Population w/ eGFR < 90 2001-2018 no. (%)**
Sample size	44,360 (100.00)	230,108,444 (100.00)	74,484,678 (32.37)
Creatinine-specific limitations			
Underweight proxy for low muscle mass	792 (1.79)	3,947,594 (1.72)	743,750 (1.00)
Creatine supplement	54 (0.12)	366,257 (0.16)	171,767 (0.23)
Creatinine secretion/interference medications	186 (0.42)	973,821 (0.42)	490,313 (0.66)
Cystatin C-specific limitations			
Current smoking	9,227 (20.80)	49,500,244 (21.51)	11,938,065 (16.03)
Systemic corticosteroids	530 (1.19)	2,545,842 (1.11)	1,296,867 (1.74)
Limitations affecting both markers			
Cirrhosis or advanced liver disease	20 (0.05)	99,325 (0.04)	33,183 (0.04)
Class III obesity (BMI >=40)	2,999 (6.76)	14,842,632 (6.45)	4,695,333 (6.30)
Joint confirmatory-test groups			
Potential indication for cystatin C confirmation	1,027 (2.32)	5,268,538 (2.29)	1,401,825 (1.88)
Potential limitation of cystatin C-based estimation	9,669 (21.80)	51,625,469 (22.44)	13,026,291 (17.49)
Potential indication for mGFR if precision is required	3,350 (7.55)	16,617,925 (7.22)	5,191,098 (6.97)

3.3. *Expected Decision Discordance at GFR and Kidney Failure Risk Thresholds*

Expected decision discordance diverged sharply depending on the specific threshold evaluated (Fig. 2A–B). For GFR decision thresholds, the expected number of U.S. adults whose classification would differ using modeled mGFR rather than eGFRcr increased at higher thresholds, from 591,423 adults at the GFR threshold of 30 mL/min/1.73 m² to 3,337,312 adults at 45 mL/min/1.73 m² and 14,248,035 adults at 60 mL/min/1.73 m² (Fig. 2A). These estimates corresponded to 0.26%, 1.45%, and 6.19% of the survey-weighted U.S. adult population, respectively.

Expected KFRE-based decision discordance was less common but remained evident across evaluated risk thresholds (Fig. 2B). Expected discordance was 168,045 adults for the 2-year KFRE ≥10% threshold, 74,471 adults for the 2-year KFRE ≥40% threshold, and 605,027 adults for the 5-year KFRE ≥5% threshold, corresponding to 0.07%, 0.03%, and 0.26% of the survey-weighted U.S. adult population, respectively.

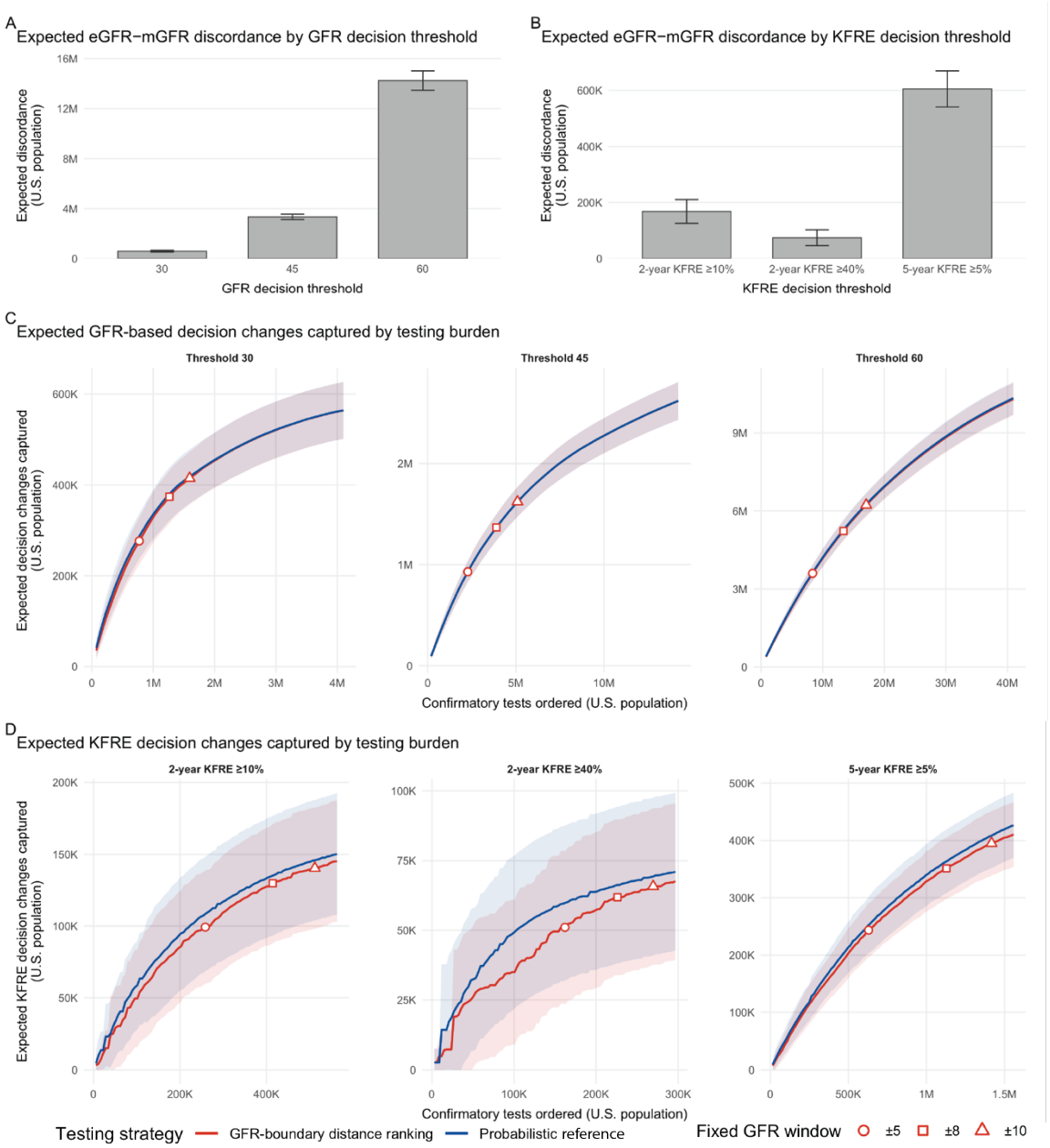


Fig. 2. Expected decision discordance and confirmatory testing trade-offs across GFR and KFRE thresholds. (a) Expected discordant classifications at direct GFR thresholds. (b) Expected discordant classifications at KFRE risk thresholds. (c–d) Expected decision changes captured as confirmatory-testing burden increases for GFR and KFRE thresholds, respectively. Proximity-based strategies are compared with probabilistic ranking; symbols indicate fixed testing windows of ±5, ±8, and ±10 mL/min/1.73 m². Error bars and ribbons show 95% confidence intervals.

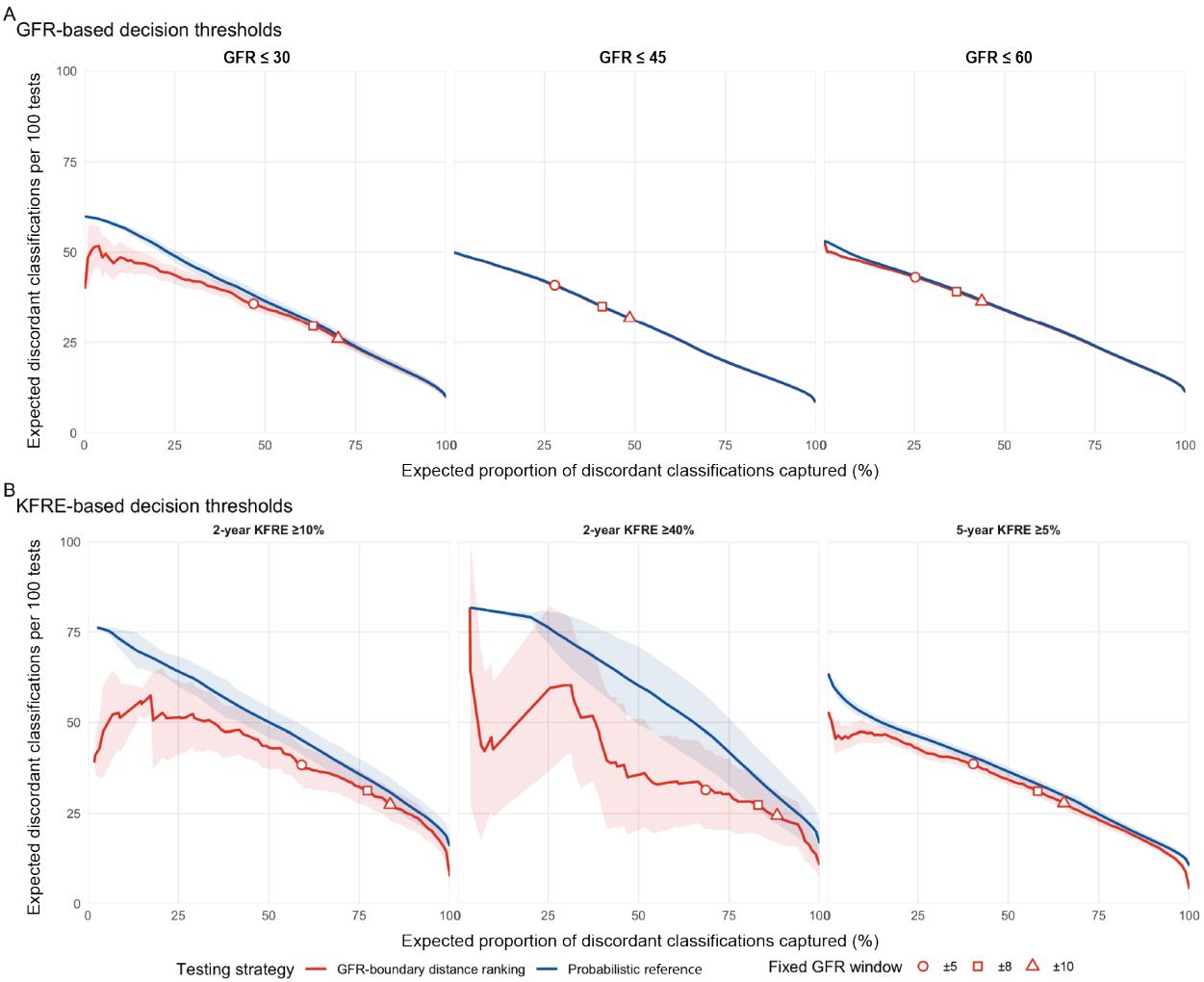


Fig. 3. Confirmatory testing efficiency across GFR and KFRE decision thresholds. Expected discordant classifications per 100 confirmatory tests are shown against the proportion of total expected discordant classifications captured for (a) GFR thresholds and (b) KFRE risk thresholds. The y- and x-axes are analogous to precision and recall, respectively, with both quantities calculated from model-implied discordance probabilities rather than observed outcomes. Proximity-based ranking is compared with probabilistic ranking. Symbols indicate fixed testing windows, and ribbons show 95% confidence intervals.

3.4. *Confirmatory Testing Burden and Captured Decision Discordance*

Increasing the testing window captured more expected decision changes but required substantially greater testing burden (Fig. 2C–D). For direct GFR thresholds, proximity-based testing closely approximated probabilistic ranking across most of the burden range, with modest separation at the 30 mL/min/1.73 m² threshold under highly restricted testing. For example, a ±5 mL/min/1.73 m² window would test approximately 772,000 adults and capture 277,000 expected discordant classifications at the threshold of 30, compared with 8.36 million tested and 3.61 million captured at the threshold of 60.

For KFRE thresholds, testing windows were centered on participant-specific GFR boundaries. A ±5 mL/min/1.73 m² window would capture approximately 99,000 of 258,000 tested adults at the 2-year 10% threshold, 51,000 of 162,000 at the 2-year 40% threshold, and 244,000 of 630,000 at the 5-year 5% threshold. Wider windows increased capture but reduced efficiency. Because KFRE risk depends jointly on age, sex, GFR, and UACR, proximity to the implied GFR boundary did not fully capture modeled discordance probability.

3.5. *Testing Efficiency and Yield*

Testing efficiency declined as strategies expanded to capture more expected discordance (Fig. 3). For direct GFR thresholds, distance-based ranking closely tracked the probabilistic reference strategy, although modeled probability modestly improved prioritization at the 30 mL/min/1.73 m² threshold when testing was highly restricted.

Separation was greater for KFRE thresholds, particularly the 2-year thresholds. Although proximity to the participant-specific GFR boundary identified relatively high-yield individuals, probabilistic ranking generally captured more expected decision changes per 100 tests. Thus, simple proximity was an effective proxy for direct GFR decisions but less complete when uncertainty was propagated through a multivariable risk model.

4. Discussion

Precision medicine has an understudied dimension, that clinical estimates are increasingly personalized, yet the uncertainty used to interpret and act on them remains understudied or generic. This problem is not specific to kidney function; it extends to any clinical prediction model whose output is acted upon at a meaningful threshold. A patient-specific point prediction may indicate which side of a threshold a patient occupies without conveying how likely plausible values are to cross that boundary. A decision-centered framework would instead propagate a conditional predictive distribution through the relevant threshold, yielding quantities such as the probability that kidney function lies below a medication cutoff or that underlying disease risk exceeds a treatment threshold. Without such tools, clinicians are left to apply population-level performance metrics, such as bias or P30, to individual threshold-based decisions for which those metrics are not directly actionable. Kidney function assessment offers a high-stakes and tractable setting in which to study this problem.

In this nationally representative analysis, guideline-recognized clinical contexts identified relatively limited subgroups with potential indications for confirmatory testing, whereas uncertainty near commonly used decision thresholds produced a larger population-level burden of expected discordance. For thresholds on GFR itself, simple proximity closely approximated probabilistic prioritization, suggesting that fixed testing windows may provide a practical heuristic. For KFRE thresholds, participant-specific decision boundaries introduced greater heterogeneity, and discordance probability conditioned on eGFR value provided more information than boundary distance alone.

Even probabilistic prioritization conditioned on eGFR could not eliminate the burden-yield trade-off created by overlap between eGFR- and mGFR-based classifications. Capturing most

expected discordance may require testing many individuals whose classifications would remain unchanged. Thus, uncertainty modeling conditioned only on eGFR may improve prioritization, but it cannot by itself make broad confirmatory testing uniformly efficient.

Because this spread partly reflects non-GFR determinants of creatinine, which include muscle mass, tubular secretion, and diet, the discordance we quantify is not just important for clinical classification but may have utility as a biological signal. Identifying patients whose estimated and measured GFR diverge may aid in biomarker and therapeutic development, where separating true changes in kidney function from marker-specific error is essential, linking decision-relevant uncertainty to the biological mechanisms that help shape it.

Our findings point to two complementary directions. First, uncertainty models should become more individualized. In kidney function assessment, this may mean conditioning on variables such as age, sex, comorbidities, body composition, albuminuria, medications, and clinical context. Second, those distributions must be translated into practical decision support: an estimate, the probability of crossing the relevant decision boundary, and guidance about whether confirmation is likely to change care. Guideline-identified clinical contexts represent an important attempt to personalize eGFRcr interpretation, but they are unlikely to identify all patients for whom confirmatory testing could alter a decision. Future tools should therefore personalize both the uncertainty distribution and the decision through which it is interpreted.

5. Limitations

This study used cross-sectional NHANES data, which do not include measured GFR or longitudinal clinical decisions. Accordingly, mGFR-based classification was modeled rather than observed. NHANES also incompletely captures muscle mass, frailty, diet, medication exposure, supplement use, and high-stakes clinical contexts. Guideline-recognized indications were therefore approximated using imperfect observable proxies and may be under- or overestimated.

Decision-discordance estimates relied on externally derived distributions of mGFR conditional on eGFRcr. These distributions may not transport fully to the NHANES population and do not account for variation in uncertainty by age, sex, body composition, comorbidities, medications, or other clinical characteristics. Interpolation between reported quantiles may also imperfectly represent distribution tails. The resulting estimates should therefore be interpreted as model-based expected discordance, not observed reclassification or fully individualized uncertainty.

The clinical importance of discordance depends on the decision and patient context. The evaluated GFR thresholds do not represent a uniform indication for testing among all nearby adults, and KFRE analyses additionally assume that the eGFRcr-conditioned mGFR distribution is applicable across age, sex, and UACR strata. KFRE results are therefore best interpreted as a sensitivity analysis showing how modeled GFR uncertainty propagates through patient-specific risk boundaries. Repeated eGFRcr measurements may also reduce uncertainty from transient biological and analytical variation, although persistent non-GFR determinants may still produce systematic error. Our single-measurement analysis could not assess how incorporating prior measurements would alter decision uncertainty and downstream interventional decisions.

Finally, the probabilistic reference strategy was generated and evaluated using the same conditional mGFR model. Its higher model-implied yield in some settings is therefore not independent evidence of superior real-world performance. The comparison instead characterizes how closely simple proximity-based strategies reproduce prioritization under the assumed uncertainty model. The strategies were not prospectively validated, and expected classification changes do not necessarily correspond to management changes, patient benefit, cost-effectiveness, or clinical outcomes.

6. Conclusion

Precision medicine has largely personalized clinical estimates while treating the uncertainty around them as generic; our findings show why that uncertainty, and the decisions it feeds, warrant personalization too. In this nationally representative analysis, uncertainty in eGFRcr produced substantial expected discordance at common GFR and kidney failure risk thresholds. Simple threshold proximity provided a practical approximation for direct GFR decisions, whereas our results suggest that probabilistic modeling may add information when uncertainty is propagated through patient-specific risk boundaries. Improving kidney function assessment will therefore require more than better filtration-marker equations: it will require decision-centered models that personalize uncertainty, communicate the probability of clinically meaningful reclassification, and identify when confirmatory testing is likely to change care.

7. Acknowledgments

This research was supported by the Ivan and Francesca Berkowitz Family Living Laboratory Collaboration at Harvard Medical School and Clalit Research Institute and NIH/NIDDK Award R01DK137993.

8. Appendix

8.1. *Variable Definitions*

Obesity was defined as BMI ≥ 30 kg/m^2, and diabetes was defined by self-reported clinician diagnosis, self-reported use of insulin or diabetes medications, HbA1c ≥6.5%, or plasma glucose ≥126 mg/dL. Hypertension was defined as self-reported clinician diagnosis or clinician recommendation for antihypertensive medication, median systolic blood pressure >130 mm Hg, or median diastolic blood pressure >80 mm Hg. Albuminuria categories were defined by UACR.

References

1. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int*. 2024;105(4S):S117-S314.

2. Selistre L, Rabilloud M, Cochat P, et al. Comparison of the Schwartz and CKD-EPI equations for estimating glomerular filtration rate in children, adolescents, and adults: A retrospective cross-sectional study. *PLoS Med*. 2016;13(3):e1001979.

3. Fan L, Inker LA, Rossert J, et al. Glomerular filtration rate estimation using cystatin C alone or combined with creatinine as a confirmatory test. *Nephrol Dial Transplant*. 2014;29(6):1195-1203.

4. Evans M, van Stralen KJ, Schön S, et al. Glomerular filtration rate-estimating equations for patients with advanced chronic kidney disease. *Nephrol Dial Transplant*. 2013;28(10):2518-2526.

5. Murata K, Baumann NA, Saenger AK, Larson TS, Rule AD, Lieske JC. Relative performance of the MDRD and CKD-EPI equations for estimating glomerular filtration rate among patients with varied clinical presentations. *Clin J Am Soc Nephrol*. 2011;6(8):1963-1972.

6. Hojs R, Bevc S, Ekart R, Gorenjak M, Puklavec L. Kidney function estimating equations in patients with chronic kidney disease: GFR estimation and CKD patients. *Int J Clin Pract*. 2011;65(4):458-464.

7. Feng JF, Qiu L, Zhang L, et al. Multicenter study of creatinine- and/or cystatin C-based equations for estimation of glomerular filtration rates in Chinese patients with chronic kidney disease. *PLoS One*. 2013;8(3):e57240.

8. Inker LA, Tighiouart H, Coresh J, et al. GFR estimation using β-Trace protein and β2-microglobulin in CKD. *Am J Kidney Dis*. 2016;67(1):40-48.

9. Shafi T, Zhu X, Lirette ST, et al. Quantifying Individual-Level Inaccuracy in Glomerular Filtration Rate Estimation : A Cross-Sectional Study. *Ann Intern Med*. 2022;175(8):1073-1082.

10. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA*. 2011;305(15):1553-1559.

11. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Available at: http://www.cdc.gov/nchs/nhanes/

12. Selvin E, Manzi J, Stevens LA, et al. Calibration of serum creatinine in the National Health and Nutrition Examination Surveys (NHANES) 1988-1994, 1999-2004. *Am J Kidney Dis*. 2007;50(6):918-926.

13. Inker LA, Eneanya ND, Coresh J, et al. New Creatinine- and Cystatin C–Based Equations to Estimate GFR without Race. *N Engl J Med*. 2021;385(19):1737-1749.

14. Kidney Disease: Improving Global Outcomes (KDIGO) Diabetes Work Group. KDIGO 2022 Clinical Practice Guideline for Diabetes Management in Chronic Kidney Disease. *Kidney Int*. 2022;102(5S):S1-S127.

15. Kidney Disease: Improving Global Outcomes (KDIGO) Blood Pressure Work Group. KDIGO 2021 Clinical Practice Guideline for the Management of Blood Pressure in Chronic Kidney Disease. *Kidney Int*. 2021;99(3S):S1-S87.

16. Baker M, Perazella MA. NSAIDs in CKD: Are they safe? *Am J Kidney Dis*. 2020;76(4):546-557.

17. U.S. Food and Drug Administration. Glucophage (metformin hydrochloride) Tablets and Glucophage XR (metformin hydrochloride) Extended-Release Tablets: Prescribing Information. *US Food and Drug Administration*. Published online 2017. Accessed June 29, 2026. https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/020357s037s039,021202s021s023lbl.pdf

18. National Cancer Institute. Broadening/Modernizing Eligibility Criteria for National Cancer Institute (NCI) Sponsored Clinical Trials. Published online May 7, 2024. Accessed June 29, 2026. https://dctd.cancer.gov/research/ctep-trials/trial-development/eligibility-guidance.pdf