Prospects for improved glomerular filtration rate estimation based on creatinine—results from a transnational multicentre study

Jonas Björk1,2,*, Ulf Nyman3,*, Marie Courbebaisse4, Lionel Couzi5, R. Neil Dalton6, Laurence Dubourg7, Natalie Ebert8, Björn O. Eriksen9, Francois Gaillard10, Cyril Garrouste11, Anders Grubb12, Magnus Hansson13,14, Lola Jacquemont15, Ian Jones16, Nassim Kamar17, Edmund J. Lamb18, Christophe Legendre19, Karin Littmann13,14, Christophe Mariat20, Toralf Melsom9, Lionel Rostaing21, Andrew D. Rule22, Elke Schaeffner8, Per-Ola Sundin16, Stephen Turner22, Anna Åkesson1,2, Pierre Delanaye23,** and Hans Pottel24,**

1Division of Occupational and Environmental Medicine, Lund University, Lund, Sweden, 2Clinical Studies Sweden, Forum South, Skåne University Hospital, Lund, Sweden, 3Department of Translational Medicine, Division of Medical Radiology, Lund University, Malmö, Sweden, 4Physiology Department, Georges Pompidou European Hospital, Assistance Publique Hôpitaux de Paris, Paris Descartes University, INSERM U1151-CNRS UMR8253, Paris, France, 5CHU de Bordeaux, Nephrologie–Transplantation–Dialyse, Université de Bordeaux, CNRS-UMR 5164 Immuno ConcEpT, Bordeaux, France, 6The Wellchild Laboratory, Evelina London Children’s Hospital, London, UK, 7Néphrologie, Dialyse, Hypertension et Exploration Fonctionnelle Rénale, Hôpital Edouard Herriot, Hospices Civils de Lyon, Lyon, France, 8Charité Universitätsmedizin Berlin, Institute of Public Health, Berlin, Germany, 9Metabolic and Renal Research Group, UiT The Arctic University of Norway, Tromsø, Norway, 10Renal Transplantation Department, Necker Hospital, Assistance Publique–Hôpitaux de Paris (AP-HP), Paris, France, 11Department of Nephrology, Clermont-Ferrand University Hospital, Clermont-Ferrand, France, 12Department of Clinical Chemistry, Skåne University Hospital Lund, Lund University, Lund, Sweden, 13Function area Clinical Chemistry, Karolinska University Laboratory, Karolinska University Hospital Huddinge, Stockholm, Sweden, 14Department of Laboratory Medicine, Karolinska Institute, Karolinska University Hospital Huddinge, Stockholm, Sweden, 15Renal Transplantation Department, CHU Nantes, Nantes University, Nantes, France, 16Department of Clinical Epidemiology and Biostatistics, School of Medical Sciences, Örebro University, Örebro, Sweden, 17Department of Nephrology, Dialysis and Organ Transplantation, CHU Rangueil, INSERM U1043, IFR–BMT, University Paul Sabatier, Toulouse, France, 18Clinical Biochemistry, East Kent Hospitals University NHS Foundation Trust, Canterbury, UK,
Hôpital Necker, AP-HP & Université Paris Descartes, Paris, France, Service de Néphrologie, Dialyse et Transplantation Rénale, Hôpital Nord, CHU de Saint-Etienne, France, Service de Néphrologie, Hémodialyse, Aphereses et Transplantation Rénale, Hôpital Michallon, CHU Grenoble-Alpes, La Tronche, France, Division of Nephrology and Hypertension, Mayo Clinic, Rochester, MN, USA, Department of Nephrology-Dialysis-Transplantation, University of Liége (ULg CHU), CHU Sart Tilman, Liége, Belgium and Department of Public Health and Primary Care, KU Leuven Campus Kulak Kortrijk, Kortrijk, Belgium

**P. Delanaye and H. Pottel shared last/senior authorship.**
*J. Björk and U. Nyman shared first authorship.
Correspondence to: Ulf Nyman; E-mail: ulf.nyman@bredband.net; Twitter handle: @PstoneOftheYear

ABSTRACT

Background. The Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine equation is routinely used to assess renal function but exhibits varying accuracy depending on patient characteristics and clinical presentation. The overall aim of the present study was to assess if and to what extent glomerular filtration rate (GFR) estimation based on creatinine can be improved.

Methods. In a cross-sectional analysis covering the years 2003–17, CKD-EPI was validated against measured GFR (mGFR; using various tracer methods) in patients with high likelihood of chronic kidney disease (CKD; five CKD cohorts, n = 8365) and in patients with low likelihood of CKD (six community cohorts, n = 6759). Comparisons were made with the Lund–Malmö revised equation (LMR) and the Full Age Spectrum equation.

Results. 78% patients aged 18–39 years old, CKD-EPI overestimated GFR with 5.0–16 mL/min/1.73 m² in median in both cohort types at mGFR levels <120 mL/min/1.73 m². LMR had greater accuracy than CKD-EPI in the CKD cohorts (P < 30%, the percentage of estimated GFR within 30% of mGFR, 83.5% versus 76.6%). CKD-EPI was generally the most accurate equation in the community cohorts, but all three equations reached P < 30 above the Kidney Disease Outcomes Quality Initiative benchmark of 90%.

Conclusions. None of the evaluated equations made optimal use of available data. Prospects for improved GFR estimation procedures based on creatinine exist, particularly in young adults and in settings where patients with suspected or manifest CKD are investigated.

Keywords: chronic kidney disease, creatinine, glomerular filtration rate, kidney function tests, renal failure

INTRODUCTION

In 2013, the Kidney Disease: Improving Global Outcomes work group recommended clinical laboratories to report estimated glomerular filtration rate (eGFR) in adults using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine equation [1], unless an alternative equation improves accuracy of glomerular filtration rate (GFR) estimates [2]. This recommendation has made the use of CKD-EPI common practice in many hospitals. However, there is evidence suggesting that GFR estimations based on creatinine can be improved in patients with known or suspected chronic kidney disease (CKD) [3–9]. Moreover, the CKD-EPI equation is not intended for the entire life span, as the present formulation of the equation severely overestimates GFR in children [10, 11]. This restricted scope of the CKD-EPI equation also most likely explains its substantial overestimation and low accuracy among young adults [6, 11, 12], and the implausible changes in eGFR when switching from the recommended creatinine-based Chronic Kidney Disease in Children equation (‘Schwartz bedside’) in paediatric care to CKD-EPI in adults [13]. Thus, although GFR estimation based on creatinine has a long history [14], it seems to still be an area where improvements can be made. Guidance in future refinement of GFR equations and estimation approaches are warranted from more solid empirical evidence from external validation studies.

Conclusions are important to help guide clinical practice and research. It is essential to distinguish between ‘diagnostic correctness’ and ‘diagnostic predictiveness’ when investigating GFR estimation accuracy [15]. Diagnostic correctness represents GFR estimation accuracy stratified by measured GFR (clearance of external tracers, mGFR) and is analogous to sensitivity and specificity of a binary test. For binary tests, there is a trade-off between sensitivity and specificity, which for GFR equations corresponds to a trade-off between accuracy of eGFR among patients with and without renal impairment [15, 16]. However, when a diagnostic test is used in clinical practice typically, the index test (eGFR) but not the reference method (mGFR) is available. Diagnostic predictiveness represents GFR estimation accuracy stratified by eGFR, analogous to presenting predictive values of a binary test for use in the clinical situation [15]. Diagnostic predictiveness is dependent not only on sensitivity (accuracy at low mGFR) and specificity (accuracy at normal mGFR), but also on the clinical setting, e.g. primary versus specialized healthcare and the resulting distribution of patient characteristics, symptomatology and severity of renal disease [15, 17]. Most validation studies of GFR equations have focused on diagnostic correctness [15, 18] and have often pooled data from different populations. However, overall validation in pooled data adds little insight into diagnostic predictiveness, as the pooling makes the study population and underlying clinical setting poorly defined [19]. Additionally, few studies have
studied diagnostic accuracy at two or more dimensions simultaneously, e.g. mGFR and age [9, 13, 20].

The overall aim of the present cross-sectional multicentre study was to assess if and to what extent diagnostic predictiveness of GFR estimation based on creatinine can be improved in two distinct clinical settings: (i) patients with high prior likelihood of CKD and (ii) patients with low prior likelihood of CKD. We compared CKD-EPI with one GFR equation developed specifically for use in patients with suspected or confirmed CKD [the Lund–Malmö revised equation (LMR)] [21] and one primarily intended for use in patients with no prior suspicion of CKD [the Full Age Spectrum (FAS) equation] [7, 22].

**MATERIALS AND METHODS**

**Patient data**

The European Kidney Function Consortium (EKFC), a new working group under the umbrella of the ERA-EDTA, has taken the initiative to pool and structure data on mGFR, plasma/serum creatinine, age, sex, height and weight of Europeans and non-black North Americans aged ≥18 years. Data were obtained from 11 cohorts (Supplementary data, Table S1) used in ongoing or published cross-sectional and longitudinal studies in France [7, 8, 23, 24], the UK [7–9, 25], Germany [7–9, 26], Norway [7, 8, 27], Sweden [4, 6, 9, 28, 29] and the USA [7, 8, 30, 31]. We classified the individual cohorts according to the clinical setting: (i) patients with high prior likelihood of CKD (labelled ‘CKD cohorts’; n = 5) and (ii) patients with low prior likelihood of CKD (labelled ‘community cohorts’; n = 6). Common causes for referral in the CKD cohorts were manifest or suspected diabetic nephropathy, interstitial nephritis, glomerulonephritis, nephrotic syndrome, haematuria, proteinuria, reflux nephropathy, myeloma, vasculitis, consideration of initiation of haemodialysis, control after organ transplantation and to dose drugs cleared by the kidneys. The community cohorts included studies of general populations, healthy older people and potential kidney donors.

Patient data were pooled in an anonymous database for the present study at Lund University, Sweden. All procedures involving subjects and data followed the ethical principles for medical research involving human subjects established in the World Medical Association Declaration of Helsinki. For this type of retrospective study, all extracted data were fully anonymous without any personal information, therefore informed consent was not required according to the Regional Ethical Board approval in Lund, Sweden, which approved the study (Dnr 2018/220). The present study was limited to the first measurement of GFR in each patient, resulting in 8365 patients in the CKD cohorts (median age 60 years, 45% females, median mGFR 58 mL/min/1.73 m²) and 6759 patients in the community cohorts (median age 60 years, 55% females, median mGFR 88 mL/min/1.73 m²; Table 1).

**Laboratory methods**

Details of laboratory methods used are summarized in Supplementary data, Table S2. Clearance methods for determination of GFR (mGFR) included renal clearance of inulin, chromium-51 labelled ethylene-diamine-tetra-acetic-acid (51Cr-EDTA) and iothalamate, and plasma clearance of iohexol, all considered acceptable as reference tests [32]. Samples of creatinine were obtained on the day of GFR measurement in all cohorts but Stockholm, where samples within 48 h of mGFR were accepted. All centres used creatinine assays traceable to isotopic dilution mass spectrometry (IDMS) and standardized against primary reference material (National Institute of Standards and Technology Standard Reference Material 967), except for Kent, where it was measured directly with IDMS [25].

**GFR equations**

The CKD-EPI, FAS and LMR creatinine equations are presented in the Supplementary material. None of the included cohorts has been used for the development of any of the equations.

**Statistical evaluation**

Statistical evaluations were conducted using SPSS Statistics (version 25; IBM Corp.), STATA (version 14; StataCorp) and R (version 3.5.2), focussing on bias, precision and accuracy [33]. ‘Bias’ was defined as the median of the individual differences between eGFR and mGFR in mL/min/1.73 m². ‘Precision’ was assessed as the interquartile range (IQR) of the differences eGFR – mGFR. ‘Accuracy’ was assessed from the absolute error |eGFR – mGFR/mGFR and summarized as the median absolute percentage difference (absolute accuracy) and as the percentage of estimates within ±10% and ±30% of mGFR (P10 and P30). The complementary value 1 – P30 reflects the proportion of ‘large’ estimation errors [34]. The Kidney Disease Outcome Quality Initiative (K/DOQI) 2002 benchmark is to reach P30 accuracy of ≥90% [35, 36]. Equation performance was evaluated against this benchmark. We also assessed equation performance using pairwise comparisons with CKD-EPI equation as benchmark.

Non-parametric and asymptotic 95% confidence intervals (CIs) were calculated as measures of the statistical uncertainty in medians and proportions (P10/P30) of the overall results, respectively. CIs for IQR were estimated from the 2.5 to 97.5 percentiles of a simulated distribution obtained using a bootstrap method with 10 000 replications [37].

**Diagnostic correctness—stratification by mGFR**

Evaluation of diagnostic correctness implies analysis of equation
performance (bias, precision and accuracy) stratified by mGFR, analogous to reporting sensitivity and specificity of a binary test [15]. We collapsed the two cohort types to improve statistical precision and presented diagnostic correctness in a simultaneous stratification by mGFR (<30, 30–59, 60–89, 90–119 and ≥120 mL/min/1.73 m²) and age (18–39, 40–59, 60–69 and ≥70 years). We also added sex (female and male) as a third dimension to this stratification.

Diagnostic predictiveness—stratification by estimated GFR. Evaluation of diagnostic predictiveness implies analysis of equation performance (bias, precision and accuracy) stratified by estimated GFR, analogous to presenting predictive values of a binary test for use in the clinical situation [15]. ‘Accuracy diagrams’ were constructed using quantile regression with fractional polynomials (linear, logarithm and square) as input to illustrate how the estimation errors varied across eGFR for each equation in the two population types [15, 16]. In the diagrams, we expressed estimation errors in mL/min/1.73 m² using the quantiles (percentiles) Q10, Q50 (median bias) and Q90, where the accuracy interval (AI, Q10 to Q90), reflects the largest estimation error with 80% certainty. The presentation was limited to the range between 1% and 99% percentile of the estimated GFR values for each equation to limit the statistical uncertainty in the tails of the quantile curves. The constancy of bias stratified by eGFR in the accuracy diagrams is an indicator of how similarly an equation behaves in the validation compared with the original development cohort [16, 18].

Performance in a given eGFR range of each equation cannot be directly compared because it is unlikely that all evaluated equations produce GFR estimates within that range for the same set of patients [15]. Results in tables were therefore stratified by eGFR values calculated from the CKD-EPI equation (eGFRCKD-EPI: <30, 30–59, 60–89, 90–119 and ≥120 mL/min/1.73 m²) to permit direct comparison of diagnostic predictiveness between the equations in the same patients. Since diagnostic predictiveness is dependent on pretest likelihood of disease [17], performance was evaluated for each population type (CKD and community) separately. Quantile regression was used in multivariable models to investigate how the median bias was dependent not only on eGFR CKD-EPI in the intervals defined above, but also on age, sex and body mass index (BMI; <18.5, 18.5–24.9, 25.0–29.9 and ≥30.0 kg/m²).

RESULTS

Overall results by cohort type

All three equations showed no major bias overall but substantial imprecision in both the CKD and community cohorts. In the CKD cohorts, none of them reached the K/DOQI 2002 benchmark of an P30 accuracy ≥90%, while they all reached this benchmark in the community cohorts (Table 2). LMR had better bias, higher precision and greater accuracy than both CKD-EPI and FAS in the CKD cohorts. The P30 difference was seven percentage points (P30 = 83.5% for LMR versus 76.6% for CKD-EPI and 76.5% for FAS) in the community cohorts (Table 2).

<table>
<thead>
<tr>
<th>Equations</th>
<th>Bias (mL/min/1.73 m²)</th>
<th>Precision (mL/min/1.73 m²)</th>
<th>Absolute error (%)</th>
<th>P10 (%)</th>
<th>P30 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CKD cohorts (n = 8365)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CKD-EPI</td>
<td>3.8 (3.5–4.1)</td>
<td>16.6 (16.1–17.0)</td>
<td>16.0 (15.6–16.4)</td>
<td>33.4 (32.3–34.4)</td>
<td>76.6 (75.7–77.5)</td>
</tr>
<tr>
<td>FAS</td>
<td>3.3 (3.0–3.6)</td>
<td>15.5 (15.0–16.0)</td>
<td>15.3 (14.9–15.7)</td>
<td>34.3 (33.3–35.3)</td>
<td>76.5 (75.6–77.4)</td>
</tr>
<tr>
<td>LMR</td>
<td>−1.4 (−1.7, −1.1)</td>
<td>14.4 (14.0–14.8)</td>
<td>14.1 (13.8–14.5)</td>
<td>35.9 (34.9–37.0)</td>
<td>83.5 (82.7–84.3)</td>
</tr>
<tr>
<td>Community cohorts (n = 6759)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CKD-EPI</td>
<td>2.1 (1.8–2.6)</td>
<td>17.2 (16.7–17.7)</td>
<td>10.4 (10.1–10.7)</td>
<td>48.6 (47.4–49.8)</td>
<td>91.0 (90.4–91.7)</td>
</tr>
<tr>
<td>FAS</td>
<td>−1.4 (−1.8, −1.0)</td>
<td>17.8 (17.3–18.3)</td>
<td>10.8 (10.5–11.0)</td>
<td>46.9 (45.7–48.0)</td>
<td>91.9 (91.3–92.6)</td>
</tr>
<tr>
<td>LMR</td>
<td>−6.7 (−7.0, −6.4)</td>
<td>16.7 (16.2–17.1)</td>
<td>11.5 (11.2–11.8)</td>
<td>45.7 (42.5–44.9)</td>
<td>92.8 (92.1–93.4)</td>
</tr>
</tbody>
</table>

Median bias (eGFR − mGFR) and precision (IQR) expressed in mL/min/1.73 m², absolute error expressed as median absolute percentage error |eGFR − mGFR|, and P10 and P30 accuracy (percentage of GFR estimates within 10% and 30% of mGFR).
for FAS), which corresponds to seven fewer estimation errors exceeding 30% per 100 tested CKD patients if LMR is used. Differences between the equations in the P10–P50 accuracy range were larger in the CKD than in the community cohorts (Figure 1). CKD-EPI was generally the most accurate equation in the community cohorts, as reflected by lower absolute percentage error and higher P10 than the two other equations, but the difference versus FAS was smaller than versus LMR (Table 2). Statistical evaluations of the pairwise comparisons in performance with CKD-EPI as benchmark are presented stratified by cohort type in Supplementary data, Table S3. All comparisons had narrow CIs as a result of the large cohort sizes. Results for the 11 individual cohorts are presented in Supplementary data, Table S4.

Diagnostic correctness—results stratified by mGFR, age and sex

LMR was generally the least biased equation at all ages in patients with known GFR (mGFR < 90 mL/min/1.73 m²) (Table 3). It was generally also more precise than CKD-EPI and generally more accurate than both CKD-EPI and FAS at all mGFR levels < 90 mL/min/1.73 m² (Supplementary data, Table SSA-C). CKD-EPI and FAS showed marked overestimations at younger ages at all mGFR levels < 90 or 120 mL/min/1.73 m², respectively (Table 3). As an example, CKD-EPI overestimated GFR by 13 mL/min/1.73 m² (95% CI 11.5–15.5 mL/min/1.73 m²) on average in individuals < 40 years of age with mGFR 30–59 mL/min/1.73 m².

LMR exhibited noticeable underestimation in patients with mGFR > 90 mL/min/1.73 m² (Table 3). CKD-EPI and FAS also yielded underestimations in patients with high mGFR but mostly to a lesser degree than LMR. LMR and CKD-EPI were both more precise than FAS at mGFR > 90 mL/min/1.73 m² (Supplementary data, Table SSA), but CKD-EPI was more accurate due to its lower bias (Supplementary data, Tables SSB and SSC). None of the three equations showed any consistent differences in bias among females and males (Supplementary data, Table S6).

Diagnostic predictiveness—results stratified by estimated GFR and cohort type

The accuracy diagrams further illustrate the substantial imprecision across eGFR for all three equations and in both two population types (Figure 2). Low and constant bias were seen both for LMR in the CKD cohorts and for CKD-EPI in the community cohorts at eGFR < 120 mL/min/1.73 m².

In the CKD cohorts, LMR had lower bias and higher accuracy than CKD-EPI in patients with eGFR > 60 mL/min/1.73 m² according to the CKD-EPI equation (eGFRCKD-EPI; Table 4). Estimates of the CKD-EPI equation > 120 mL/min/1.73 m² occurred among 9% of all patients in the CKD cohorts, most of them young (median age 21 years). The overestimation exceeded 20 mL/min/1.73 m² on average for these patients when CKD-EPI or FAS was used, whereas the estimates from LMR were virtually unbiased for the same patients. In particular, FAS exhibited imprecision at high eGFRCKD-EPI in the community cohorts. The underestimation exceeded 20 mL/min/1.73 m² on average for these patients when CKD-EPI or LMR were used, whereas the estimates from LMR were virtually unbiased for the same patients. In particular, FAS exhibited imprecision at high eGFRCKD-EPI in the community cohorts.

In the community cohorts, changing equation from CKD-EPI to either FAS or LMR would not consistently improve accuracy in patients with eGFR < 120 mL/min/1.73 m² (98% of all patients; Table 5). FAS exhibited increasing overestimations at high eGFRCKD-EPI in the community cohorts. The underestimation of LMR varied between 4 and 7 mL/min/1.73 m² across all levels of eGFRCKD-EPI. None of the three equations was consistently more precise than the others across all levels of eGFRCKD-EPI.

Multivariable quantile regression based on eGFRCKD-EPI, age, sex and BMI confirmed the substantial overestimations for CKD-EPI and FAS at high levels of eGFRCKD-EPI in the CKD cohorts (Table 6). Additionally, overestimation due to underweight was noted for all three equations and with similar magnitude.

---

Table 3. Diagnostic correctness (bias stratified by mGFR) in the two cohort types combined (CKD and community; n = 15 124)

<table>
<thead>
<tr>
<th>mGFR</th>
<th>Age intervals (years)</th>
<th>18–39</th>
<th>40–59</th>
<th>60–69</th>
<th>≥70</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;30, number</td>
<td></td>
<td>101</td>
<td>270</td>
<td>338</td>
<td>1073</td>
</tr>
<tr>
<td>CKD-EPI</td>
<td></td>
<td>5.0 (3.6–7.0)</td>
<td>1.4 (0.7–2.8)</td>
<td>2.9 (2.1–3.9)</td>
<td>2.0 (1.5–2.5)</td>
</tr>
<tr>
<td>FAS</td>
<td></td>
<td>9.7 (7.8–12.7)</td>
<td>6.6 (5.5–7.6)</td>
<td>5.6 (4.8–6.4)</td>
<td>2.9 (2.5–3.3)</td>
</tr>
<tr>
<td>LMR</td>
<td></td>
<td>2.8 (1.6–3.8)</td>
<td>1.2 (0.4–1.9)</td>
<td>1.8 (1.1–2.5)</td>
<td>0.2 (–0.2, 0.7)</td>
</tr>
<tr>
<td>30–59, number</td>
<td></td>
<td>381</td>
<td>656</td>
<td>767</td>
<td>1443</td>
</tr>
<tr>
<td>CKD-EPI</td>
<td></td>
<td>13.4 (11.5–15.5)</td>
<td>3.7 (2.6–5.0)</td>
<td>3.4 (2.0–4.4)</td>
<td>4.0 (3.4–4.6)</td>
</tr>
<tr>
<td>FAS</td>
<td></td>
<td>14.5 (13.0–16.4)</td>
<td>7.1 (6.0–8.3)</td>
<td>2.7 (1.8–3.8)</td>
<td>–0.2 (–0.8, 0.2)</td>
</tr>
<tr>
<td>LMR</td>
<td></td>
<td>6.3 (4.1–8.6)</td>
<td>1.3 (0.3–2.6)</td>
<td>0.6 (–0.9, 1.9)</td>
<td>–0.3 (–0.8, 0.5)</td>
</tr>
<tr>
<td>60–69, number</td>
<td></td>
<td>736</td>
<td>1772</td>
<td>1556</td>
<td>1133</td>
</tr>
<tr>
<td>CKD-EPI</td>
<td></td>
<td>15.6 (13.6–17.3)</td>
<td>8.0 (7.3–8.9)</td>
<td>7.0 (6.2–7.6)</td>
<td>5.9 (5.1–5.6)</td>
</tr>
<tr>
<td>FAS</td>
<td></td>
<td>10.9 (9.5–12.2)</td>
<td>4.4 (3.8–5.1)</td>
<td>–0.4 (–1.0, 0.3)</td>
<td>–4.5 (–5.5, –3.8)</td>
</tr>
<tr>
<td>LMR</td>
<td></td>
<td>2.6 (1.4–3.8)</td>
<td>–0.1 (–0.7, 0.5)</td>
<td>–2.2 (–2.7, –1.6)</td>
<td>–4.4 (–5.2, –3.9)</td>
</tr>
<tr>
<td>90–119, number</td>
<td></td>
<td>1278</td>
<td>2088</td>
<td>798*</td>
<td>147*</td>
</tr>
<tr>
<td>CKD-EPI</td>
<td></td>
<td>10.4 (9.2–11.5)</td>
<td>–1.9 (–2.5, –1.4)</td>
<td>–5.3 (–6.1, –4.5)</td>
<td>–9.5 (–12.2, –7.0)</td>
</tr>
<tr>
<td>FAS</td>
<td></td>
<td>4.4 (3.4–6.0)</td>
<td>–2.8 (–3.4, –2.2)</td>
<td>–9.5 (–10.8, –8.5)</td>
<td>–14.2 (–16.8, –12.7)</td>
</tr>
<tr>
<td>≥120, number</td>
<td></td>
<td>338</td>
<td>249</td>
<td>“”</td>
<td>“”</td>
</tr>
<tr>
<td>CKD-EPI</td>
<td></td>
<td>–5.2 (–7.3, –2.8)</td>
<td>–23.9 (–26.4, –21.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAS</td>
<td></td>
<td>–5.8 (–9.3, –3.3)</td>
<td>–17.9 (–21.4, –15.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMR</td>
<td></td>
<td>–25.7 (–27.0, –24.2)</td>
<td>–32.4 (–35.7, –31.0)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Median bias (eGFR – mGFR in mL/min/1.73 m²; 95% CIs) of CKD-EPI, FAS and LMR stratified by mGFR and age (years). The lowest bias is marked with bold and italic in each stratum.

*mGFR intervals ≥120 were collapsed with 90–119 due to small numbers (n < 100).
Age and sex did not have strong independent effects on bias for any of the three equations in the CKD cohorts. Bias varied more noticeably with eGFR CKD-EPI for CKD-EPI and FAS than for LMR also in the multivariable quantile regression models for the community cohorts (Table 6). In addition, bias varied according to age for CKD-EPI and LMR. Being male increased the underestimation of the LMR equation. BMI at any level was not related to bias for any of the three equations in the community cohorts.

DISCUSSION

The salient finding of our comprehensive validation study is that prospects for improved GFR estimation based on creatinine still exist. The widely used CKD-EPI equation was generally sufficiently accurate with $P_{30}$ exceeding 90% only in patients with low likelihood of CKD, but it was neither the most accurate equation in patients with known or suspected renal impairment nor the most accurate in young adults irrespective of their renal status. The FAS equation shared similar weaknesses as CKD-EPI when applied in the CKD cohorts and in young adults. LMR, on the other hand, was the most accurate equation among patients with high likelihood of CKD but did not perform as well in the community cohorts.

The explanation for the superior accuracy of LMR in patients with known or suspected renal impairment may be that the equation was formulated with the explicit goal to improve sensitivity (estimations in CKD patients) [21], whereas the goal of developing CKD-EPI was to improve specificity (estimations in patients with normal mGFR) [1]. The development of FAS was established from a mathematical construction based on...
A major strength of the present study was the large sample size, which allowed for evaluation of diagnostic accuracy with sufficient statistical precision in three dimensions simultaneously (mGFR, age and sex). Another strength was the consistent stratification on cohort type in the evaluation of diagnostic predictiveness, as clinical presentation and related prevalence of CKD is fundamental for the interpretation of eGFR. This stratification, for example, highlighted how the expected error in eGFR as well as the influence of low BMI may differ importantly depending on clinical setting. Yet, another strength was that measurement of plasma/serum creatinine was based on enzymatic assays or directly measured with IDMS in all but two cohorts that only partly used Jaffe. A major limitation was that available data did not allow for separate validation in additional subgroups such as patients with diabetes, malignant disease or with organ transplantation, or other ethnicities than Europeans and non-black North Americans. The initial CKD-EPI study suggested that eGFR based on creatinine must be multiplied by a correction factor (1.159) to yield valid results for African-American patients, which means that a thorough assessment of algorithm fairness and accountability is warranted before implementation [39].

age-average GFR and population-normalized serum creatinine valid for a healthy population [7, 22], which may explain its poor performance in patients with severe CKD.

The noticeable differences in performance depending on cohort type suggest that incorporation of prior information about the patient’s disease history (e.g. pre-test probability of CKD) could improve estimation accuracy. More complex GFR estimation procedures, either rule-based choice of equation depending on clinical setting or machine learning algorithms [38], would open up the possibility of fine-grained estimations depending on, for example, age or pre-test probability of CKD. Such estimation procedures should ideally be applicable for the full age span of children, adults and older people, and use of cystatin C when available, and could also incorporate other patient characteristics such as height and weight to avoid overestimation of GFR in underweight patients. However, complex algorithms are often less transparent than explicitly formulated estimating equations, which means that a thorough assessment of algorithm fairness and accountability is warranted before implementation [39].

### Table 4. Diagnostic predictiveness (bias, precision and accuracy stratified by eGFR) in CKD cohorts

<table>
<thead>
<tr>
<th>eGFRCKD-EPI</th>
<th>Number</th>
<th>Age (years)</th>
<th>BMI (kg/m²)</th>
<th>Equation</th>
<th>Bias (mL/min/1.73 m²)</th>
<th>Precision (mL/min/1.73 m²)</th>
<th>Absolute error (%)</th>
<th>P30 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;30</td>
<td>1504</td>
<td>72</td>
<td>27</td>
<td>CKD-EPI</td>
<td>−0.3 (−0.7, 0.0)</td>
<td>6.8 (4.6–7.2)</td>
<td>18.1 (16.8–18.9)</td>
<td>74.5 (72.3–76.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FAS</td>
<td>2.0 (1.6–2.2)</td>
<td>7.0 (6.7–7.4)</td>
<td>19.6 (18.2–20.8)</td>
<td>68.4 (66.1–70.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LMR</td>
<td>−1.1 (−1.4, −0.8)</td>
<td>6.7 (4.4, 7.2)</td>
<td>17.8 (17.0–18.6)</td>
<td>77.1 (75.0–79.3)</td>
</tr>
<tr>
<td>30–59</td>
<td>2419</td>
<td>65</td>
<td>27</td>
<td>CKD-EPI</td>
<td>1.6 (1.2–2.1)</td>
<td>12.5 (11.9–13.3)</td>
<td>15.3 (14.7–16.0)</td>
<td>78.6 (77.0–80.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FAS</td>
<td>1.9 (1.5–2.4)</td>
<td>13.5 (13.0–14.0)</td>
<td>15.8 (15.2–16.6)</td>
<td>76.6 (74.9–78.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LMR</td>
<td>−1.8 (−2.4, −1.4)</td>
<td>12.6 (11.9–13.1)</td>
<td>15.7 (14.9–16.4)</td>
<td>81.1 (79.5–82.7)</td>
</tr>
<tr>
<td>60–89</td>
<td>2130</td>
<td>58</td>
<td>25</td>
<td>CKD-EPI</td>
<td>6.5 (5.6–7.2)</td>
<td>19.7 (18.9–20.8)</td>
<td>15.8 (14.8–16.6)</td>
<td>76.3 (74.5–78.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FAS</td>
<td>1.8 (1.0–2.7)</td>
<td>19.7 (18.7–20.7)</td>
<td>14.0 (13.3–14.5)</td>
<td>82.1 (80.5–83.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LMR</td>
<td>−0.7 (−1.4, 0.0)</td>
<td>18.8 (18.0–19.9)</td>
<td>13.8 (13.2–14.3)</td>
<td>84.0 (82.4–85.5)</td>
</tr>
<tr>
<td>90–119</td>
<td>1578</td>
<td>46</td>
<td>24</td>
<td>CKD-EPI</td>
<td>10.1 (9.2–11.2)</td>
<td>21.2 (20.2–22.4)</td>
<td>13.6 (12.7–14.6)</td>
<td>80.4 (78.5–82.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FAS</td>
<td>7.3 (6.4–8.0)</td>
<td>21.8 (20.5–23.2)</td>
<td>12.6 (12.1–13.4)</td>
<td>81.6 (79.6–83.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LMR</td>
<td>−2.8 (−3.7, −1.8)</td>
<td>20.8 (19.7–22.0)</td>
<td>11.7 (11.1–12.3)</td>
<td>87.1 (86.8–91.6)</td>
</tr>
<tr>
<td>≥120</td>
<td>734</td>
<td>21</td>
<td>21</td>
<td>CKD-EPI</td>
<td>22.5 (20.6–23.8)</td>
<td>24.3 (22.3–25.8)</td>
<td>21.1 (19.2–22.5)</td>
<td>67.0 (63.6–70.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FAS</td>
<td>20.4 (17.9–22.3)</td>
<td>33.7 (30.7–37.1)</td>
<td>19.6 (17.7–21.9)</td>
<td>65.8 (62.4–69.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LMR</td>
<td>−0.6 (−2.9, 1.1)</td>
<td>25.2 (22.9–27.4)</td>
<td>11.2 (10.6–12.4)</td>
<td>88.7 (86.4–91.0)</td>
</tr>
</tbody>
</table>

Estimation results for CKD-EPI, FAS and LMR (95% CI) stratified by eGFR (mL/min/1.73 m²) based on the CKD-EPI equation (eGFRCKD-EPI). The best result for each performance marker is marked with bold and italic in each stratum. Number of patients, median age (years) and BMI (kg/m²) are given in each stratum. Median bias (eGFR – mGFR) and precision (IQR) expressed in mL/min/1.73 m². Absolute error expressed as median absolute percentage error (eGFR – mGFR/mGFR, and P10 and P30 accuracy (percentage of GFR estimates within 10% and 30% of mGFR).
It seems logical to use the same correction factor for other creatinine-based equations such as LMR (or FAS) to extend their applicability to African Americans. However, we believe that differences in creatinine generation should be corrected at the creatinine level rather than at the GFR level to avoid misleading interpretations that GFR differs between ethnicities. This may be achieved using the approach implemented in the FAS equations [7, 40].

A potential limitation of the generalizability of the results is that the Swedish data constituted two-thirds of patients in the CKD cohorts. However, none of these cohorts has been used for the development of LMR. The results of the Swedish cohorts were consistent with those from another European centre, Lyon (France), insofar as LMR performed better than CKD-EPI, while the latter was more accurate in the North American Chronic Renal Insufficiency Cohort Study (CRIC) cohort (Supplementary data, Table S3). One possible explanation for these divergent data, Table S3 ). One possible explanation for these divergent results across studies may be the use of different methods when measuring GFR. However, apart from renal clearance of inulin, considered the ‘gold standard’ for measuring GFR, renal clearance of iohalamate and $^{51}$Cr-EDTA as well as plasma clearance of iohexol have all been considered sufficiently accurate methods to measure GFR [32]. In addition, the single plasma sample method is highly concordant with a multiple sample strategy [43–45] providing that sampling time is adjusted according to estimated renal function [46] as used in the present cohorts.

In conclusion, the present study provides evidence that the widespread CKD-EPI equation is generally sufficiently accurate according to K/DOQI benchmark only in patients with low likelihood of CKD. Accuracy of GFR estimation based on creatinine can be improved in patients with known or suspected renal impairment and in young adults. Caution is necessary when using any of the three evaluated equations in clinical practice, as all exhibited considerable imprecision. Incorporation of the pre-test probability of CKD in the GFR estimation procedure can be an important step towards improved accuracy of eGFR across the full spectrum of age and renal function.

### DATA SHARING

The EKFC dataset used in the present study is hosted by the Lund University Population Research Platform. Legal and ethical restrictions prevent public sharing of the dataset. Data can be made available for collaborations upon request to interested researchers but would generally require a new ethical permission and the permission of each of the data-owners. You can find contact information for the data host at https://www.lupop.lu.se/

### SUPPLEMENTARY DATA

Supplementary data are available at ckj online.

### ACKNOWLEDGEMENTS

The Chronic Renal Insufficiency Cohort Study (CRIC) was conducted by the CRIC Investigators and supported by the National Institute of Diabetes and Digestive and Kidney...
Diseases (NIDDK). The data from the CRIC Study reported here were supplied by the NIDDK Central Repositories. This manuscript was not prepared in collaboration with investigators of the CRIC study and does not necessarily reflect the opinions or views of the CRIC study, the NIDDK Central Repositories or the NIDDK. We also thank Physiologist Dr Marion Vallet and Physiologist Dr Renaud de la Faille for measuring glomerular filtration rate at Toulouse and Bordeaux University Hospital, respectively. J.B., U.N., M.C., L.D., N.E., B.O.E., F.G., A.G., E.J.L., T.M., A.D.R., E.S., P.D. and H.P. are members of the European Kidney Function Consortium.

**FUNDING**

Swedish Research Council (Vetenskapsrådet; grant no. 2019-00198).

**AUTHORS’ CONTRIBUTIONS**

J.B. and U.N. contributed to analysis and interpretation of data, drafting the article, provided intellectual content of critical importance to the work described and finally approved the version to be published. All other authors contributed with analysis and interpretation of data, revising the article, provided intellectual content of critical importance to the work described and finally approved the version to be published.

**CONFLICT OF INTEREST STATEMENT**

The results presented in this article have not been published previously in whole or part. U.N., J.B. and A.G. have developed the Lund-Malmö revised equation that was validated in the present study. H.P. has developed the Full Age Spectrum equation equation that was validated in the present study. U.N. and J.B. have received reimbursement for letting GE Healthcare AB distribute the computer programme OmniVis in radiology departments for estimation of glomerular filtration rate based on the creatinine equations presented in the current article, with no special preference of any of the included equations. U.N. has received lecture fees from GE Healthcare AB. M.C. has received grant support from BIOPAL, USA. R.N.D. is a Director of and minority shareholder in a University/NHS spin-out company, SpOtOn Clinical Diagnostics and has grant supports from NHS Health Technology Assessment and Juvenile Diabetes Research Foundation. N.E. has received lecture fees from Siemens Healthineers. B.O.E. has received lecture fees from Sanofi-Aventis. N.K. has received consulting fees or paid advisory boards, lecture fees and travel support from the following companies: Abbvie, Amgen, Astellas, Chiesi, Fresenius Medical Care, Gilead, Merck Sharp and Dohme, Neovii, Novartis, Roche, Sanofi and Shire. C.L. received consulting fees or paid advisory boards from CSL Behring and Novartis and lecture fees from Sandoz. E.S. has received lecture fees from Siemens Healthineers and Fresenius Kabi. All remaining authors declared no competing interests.

**REFERENCES**