

MS KATERINA KIBURG (Orcid ID : 0000-0002-3894-5793)

DR ELIF I EKINCI (Orcid ID : 0000-0003-2372-395X)

Article type : Systematic Review or Meta-analysis

Article number: DME-2020-00226

Accepted date: 12 August 2020

Article type: Systematic Review

Copyeditor: Maria Hale

Proofs to: elif.ekinci@unimelb.edu.au

Figs 2 Tables 0, References 52

Author RH: N. Zafari *et al.*

Short title RH: CKD-EPI in people with diabetes

Systematic Review or Meta-analysis

Evaluation of the diagnostic performance of the creatinine-based Chronic Kidney Disease Epidemiology Collaboration equation in people with diabetes: a systematic review

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/DME.14391](https://doi.org/10.1111/DME.14391)

This article is protected by copyright. All rights reserved

N. Zafari¹, L. Churilov¹, L. Y.-L. Wong¹, M. Lotfaliany², M. Hachem¹, K. V. Kiburg⁴, L. Kong¹, N. Torkamani^{1,3}, H. Baxter⁵, R. J. MacIsaac⁴ and E. I. Ekinci^{1,3}

¹Department of Medicine, University of Melbourne, Austin Health, ²Department of Endocrinology and ³Austin Health Sciences Library, Austin Health, ⁴School of Population and Global Health, University of Melbourne, and ⁵Department of Endocrinology & Diabetes, University of Melbourne, St Vincent's Hospital, Melbourne, Victoria, Australia

Correspondence to: Elif I. Ekinci. E-mail: elif.ekinci@unimelb.edu.au

What's new?

- As a common complication of diabetes, diabetic kidney disease needs to be diagnosed at its earliest stages to be managed effectively.
- The Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation for estimating GFR (eGFR) is commonly used to diagnose diabetic kidney disease. However, scientific evidence regarding its diagnostic performance in people with diabetes shows controversy.
- This study showed that both the process and outcome of evaluating CKD-EPI in estimating GFR in adults with diabetes are inaccurate.
- We urge clinicians to interpret CKD-EPI eGFR with caution in people with diabetes and highlight the need to improve both the equation and its assessment methods.

Abstract

Aims GFR estimated with the creatinine-based Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI_{Cr}) equation is used to screen for diabetic kidney disease and

assess its severity. We systematically reviewed the process and outcome of evaluating CKD-EPI_{Cr} in estimating point GFR or GFR decline over time in adults with type 1 or type 2 diabetes.

Methods In this systematic review, MEDLINE, Embase and Cochrane Central Register of Controlled Trials were searched up to August 2019. Observational studies comparing CKD-EPI_{Cr} with measured GFR (mGFR) in adults with diabetes were included. Studies on people with kidney transplant, non-diabetes related kidney disease, pregnancy, potential kidney donors, and those with critical or other systematic illnesses were excluded. Two independent reviewers extracted data from published papers and disagreements were resolved by consensus. Risk-of-bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 tool. (PROSPERO registration number: CRD42018108776)

Results From the 2820 records identified, 29 studies (14 704 participants) were included. All studies were at risk of bias. Bias (eight different forms) ranged from -26 to 35 ml min⁻¹ 1.73 m⁻²; precision (five different forms) ranged between 9 and 63 ml min⁻¹ 1.73 m⁻²; accuracy (five different forms) ranged between 16% and 96%; the correlation coefficient between CKD-EPI_{Cr} and mGFR (four different forms) ranged between 0.38 and 0.86; and the reduced major axis regression slope ranged between 0.8 and 1.8.

Conclusions Qualitative synthesis of data suggested CKD-EPI_{Cr} was inaccurate in estimating point GFR or GFR decline over time. Furthermore, a lack of consistency in the methods and processes of evaluating the diagnostic performance of CKD-EPI_{Cr} limits reliable quantitative assessment. The equation needs to be improved in adults with diabetes.

[H1]Introduction

Given the high burden of diabetic kidney disease [1], it is crucial to have an accurate and precise diagnostic tool for the early detection of individuals at greatest risk of kidney damage. Promising interventions exist to halt the progression of kidney function decline, improve people's quality of life and reduce both cardiovascular and kidney mortality risk in

people with diabetic kidney disease [2,3]. Therefore, it is recommended that clinicians screen for diabetic kidney disease in people with diabetes at the time of diagnosis and annually thereafter [4,5].

GFR measured by plasma or kidney clearance of exogenous molecules (mGFR; Table S1) has the highest precision and accuracy in assessing GFR [6]. However, the invasive, laborious and expensive nature of these methods makes them inconvenient in the clinical setting [7]. Estimation of GFR from an individual's serum creatinine, on the other hand, is cheaper, more feasible and less invasive, and hence allows assessment of far larger numbers of individuals. Therefore, despite all the caveats, eGFR still is commonly used in the diagnosis and prediction of end-stage kidney disease. Incorporation of the CKD-EPI in the Kidney Function Risk Equation and validation in predicting end-stage kidney disease in those with diabetes are important [8,9].

The Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation has become one of the most widely accepted equations since its introduction in 2009 [8]. Although the formula was developed in 8254 participants and validated in 3896 (mean GFR $68 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$), only ~30% of both the development ($n = 2406$) and validation data sets ($n = 1089$) had diabetes. Further assessment of CKD-EPI_{Cr} in subgroups of the CKD-EPI study showed that the mean difference between CKD-EPI_{Cr} and mGFR was higher in people with diabetes than in those without [10]. Moreover, evaluation of CKD-EPI_{Cr} diagnostic performance in people with diabetes in other studies gave controversial results [11–17]. To the best of our knowledge, there is no systematically aggregated evidence on whether CKD-EPI_{Cr} is an appropriate tool with which to assess kidney function in people with diabetes. Therefore, we aimed to systematically review the process and outcomes of evaluation of diagnostic performance of CKD-EPI_{Cr} in estimating point GFR and GFR decline over time in adults with type 1 or type 2 diabetes.

[H1]Methods

[H2]Data sources and searches

A detailed protocol of this aggregated data systematic review has been published previously [18]. In summary, MEDLINE, Embase and Cochrane Central Register of Controlled Trials (CENTRAL) were searched from January 2009 (when the CKD-EPI_{Cr} was introduced) to January 2019 using Medical Subject Headings (MeSH) and text words related to diabetes mellitus, chronic kidney disease, estimated glomerular filtration rate using CKD-EPI_{Cr} and mGFR [18]. The International Clinical Trials Registry Platform Search Portal and ClinicalTrials.gov were searched for ongoing or recently completed trials, and PROSPERO was searched for ongoing or recently completed systematic reviews. Database searching was conducted by an expert clinical librarian.

[H3]Sample search query in MEDLINE search: Ovid interface

- 1 Diabetes Mellitus, Type 1/or Diabetes Mellitus, Type 2/or Diabetes Mellitus/or Diabetic Nephropathies/or Diabetes Complications/or (chronic kidney disease or CKD or diabetic or diabetes or T2DM).mp.
- 2 (eGFR or estimating equation* or CKD-EPI or Chronic Kidney Disease Epidemiology Collaboration or cystatin* or (estimat* and (GFR or glomerular filtration))). mp.
- 3 (mGFR or (measure* adj5 (GFR or glomerular filtration))).mp.
- 4 1 and 2 and 3.
- 5 Limit four to yr='2009-Current'.

A bridging search was conducted in August 2019 to ensure inclusion of any relevant studies published after January 2019. The electronic database search was supplemented by searching the grey literature, Google Advanced, and contacting authors where insufficient data were available in published reports. Of seven authors contacted, four provided further information. To ensure literature saturation, we scanned the reference lists of included studies and any relevant reviews.

[H2]Study selection

As per the published protocol, cross-sectional and cohort studies in English, Farsi, Dutch and Chinese were evaluated based on the following inclusion criteria: (1) study participants with type 1 or type 2 diabetes aged ≥ 18 years; (2) using CKD-EPI_{Cr} equation as the index test; (3) using the urinary clearance/plasma disappearance of inulin, iohexol, ¹²⁵I-iothalamate, Technetium-99m diethylenetriamine-penta-acetic acid (^{99m}Tc-DTPA) or chromium-51 ethylenediaminetetraacetic acid (⁵¹Cr-EDTA) as the reference test; and (4) report of at least one of the diagnostic performance specifiers of the CKD-EPI_{Cr} compared with the reference test. Studies in people with kidney transplant, chronic inflammatory diseases, non-diabetes related kidney disease, thalassaemia, heart failure, chronic use of corticosteroids, pregnancy, as well as potential kidney donors and critically ill patients were excluded because these conditions can interfere with diabetic kidney disease assessment.

[H2]Definition of outcome measurements

Dissimilarity between CKD-EPI_{Cr} eGFR and mGFR is assessed using several performance measures.

Bias (systematic error) is defined as the difference between the outcome reported by the index test vs. the reference test and can be reported as mean (SD) or median (IQR) on either an absolute or relative (proportion) scale. Examples of relative scale are mean percentage error [(mean difference between mGFR and eGFR relative to mGFR) $\times 100\%$], median percentage difference [(median difference between mGFR and eGFR relative to mGFR) $\times 100\%$], median percentage absolute difference [(median absolute difference between mGFR and eGFR relative to mGFR) $\times 100\%$]. Favourable bias for an index test is close to zero meaning that its estimates are not different from the reference test [19].

Precision is a measure of the statistical variance or random error of the index test from the reference test [20]. It is usually reported as SD or IQR of bias. It can also be assessed by the root mean square error between the eGFR and mGFR. Precision shows how close repeated measures would be to each other. The higher the variance of an index test, the lower its precision because this indicates repeating the test might result in numbers that are far from

each other. An unbiased test that is not highly precise indicates that although the difference between the index and reference tests is low, the chance of getting the same results when repeating the index test is low.

Accuracy is defined as the proportion of index test results that lie within a pre-specified range, e.g. 10% (P_{10}) or 50% (P_{50}), of the reference test reports [20].

Total deviation index is 'A boundary such that a majority, $100(1 - p)\%$, of the observations are within the boundary (measurement unit and/or percent) from their target values' [21].

Correlation coefficients range from -1 to $+1$, where values closer to ± 1 indicate strong agreement and numbers close to zero show poor agreement. According to Landis classification, an intraclass correlation coefficient of 0 = poor agreement, $0-0.2$ = slight agreement, $0.21-0.4$ = fair agreement, $0.41-0.6$ = moderate agreement, $0.61-0.8$ = substantial agreement and $0.81-1$ = almost perfect agreement [22].

The Bland-Altman plot illustrates the difference of the two tests being compared (i.e. $eGFR - mGFR$) against their means $[(eGFR + mGFR)/2]$ [23]. For normally distributed data, the expectation is that 95% of values will lie within two standard deviations of the mean bias. This range is called the limit of agreement. A serious drawback of the Bland-Altman plot is that despite assessing fixed bias well, it does not show proportional bias. Therefore, the degree of bias cannot be examined relative to the GFR spectrum [19].

Reduced major axis regression (RMAR) is a tool to evaluate disagreement between the two tests from two separate aspects, namely fixed and proportional bias via RMAR y -intercept and slope, respectively. An RMAR slope parallel to the line of perfect agreement ($y = x$, slope = 1) but with an intercept other than 0 indicates a fixed bias. By contrast, a slope other than 1 but with an intercept of 0 demonstrates the presence of proportional bias [17]. Despite being introduced more than 50 years ago, RMAR has not been applied widely in diagnostic performance evaluations in medicine and it is unclear why [19].

[H2]Data extraction and quality assessment

Covidence was used by two independent reviewers for abstract and full-text screening. Discrepancies were solved by discussion and seeking third-party opinion. The extracted data included:

- *study-related data*: type of study (cross-sectional or longitudinal), period of study (if longitudinal), year of publication, country of origin and sample size;
- *participant-related data*: type of diabetes (type 1 or type 2 diabetes), age, sex distribution, BMI, HbA_{1c}, albumin excretion rate, urine albumin to creatinine ratio and serum creatinine;
- *outcome-related data*: creatinine measurement method, values of CKD-EPI_{Cr} eGFR, exogenous filtration marker used to measure GFR, and mGFR value, as well as measures of estimation bias, precision, accuracy, total deviation index, Bland–Altman plot, correlation coefficient, reduced major axis regression, and any reports on change in mGFR and eGFR in time. Laboratory units were converted to SI units, if indicated, using the ScyMed MediCalc[®] online convertor.

Risk of study quality bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool (Table S2 and Fig. 1).

[H2]Data synthesis and analysis

Data were synthesized quantitatively. Results of the systematic review are reported in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) statement [24]. The study was registered in PROSPERO (registration number: CRD42018108776). The study's funder had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

[H1]Results

[H2]Study selection

In total, 2820 were identified. After removing duplicates ($n = 973$) in EndNote, 1847 records were screened via title/abstract, 218 of which remained for full-text screening. Twenty-nine papers were eligible for inclusion in the qualitative data synthesis. One further study was added during the bridging search (Fig. 1). Using QUADAS-2, all included studies were at risk of study quality bias, mainly (28 studies) due to measuring GFR with methods other than renal clearance of inulin, which is considered the gold standard of GFR measurement. Concern regarding applicability was low in all the included studies (Fig. S1).

Commented [1]:

[H2]Study characteristics

Cross-sectional analysis of data was conducted in 27 studies (including 10964 participants) and six studies (including 3740 participants) involved longitudinal analysis of data (Table S3). Five studies contained only participants with type 1 diabetes ($n = 4445$), 13 studies had only participants with type 2 diabetes ($n = 3148$), five studies reported both type 1 and type 2 diabetes ($n = 3248$), and six studies did not report the type of diabetes. Sample sizes for individual studies ranged from 15 to 3492 people with diabetes. Mean participant age ranged from 27 to 78 years; sex distribution, defined as proportion (%) of women participants, was 26% to 63% across the studies. Mean HbA_{1c} varied from 44 (6.2%) to 87 (10.1%) and mean (SD) serum creatinine from 61 (11) to 230 (203) $\mu\text{mol/l}$. Serum creatinine was measured using either the Jaffe kinetic method traceable to the isotope dilution mass spectrometry (IDMS) or the enzymatic method. Mean (SD) CKD-EPI_{Cr} estimated GFR varied from 53 (27) to 120 (17) $\text{ml min}^{-1} 1.73 \text{ m}^{-2}$. Reference test of measuring GFR varied in different studies and the mean (SD) of mGFR was between 48 (26) and 128 (19) $\text{ml min}^{-1} 1.73 \text{ m}^{-2}$. Some studies did not report baseline characteristics, detailed creatinine measurement method (Jaffe or enzymatic), eGFR or mGFR levels in participants with diabetes (Table S3).

[H2]Outcome measurements

As shown in Table S4, a variety of measures have been used in the literature to evaluate the diagnostic performance of CKD-EPI_{Cr} as the index test compared with the mGFR as the

reference test. Some 27 of the 29 studies reported at least one form of estimation bias, which ranged from -26 to $35 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$. Sixteen studies reported mean (SD) difference, ranging from -24 (24) to 35 (12) $\text{ml min}^{-1} 1.73 \text{ m}^{-2}$ (Fig. 2). In addition, estimation bias was reported as mean absolute difference (SD), mean percentage error (MPE) (SD), median difference (IQR), median percentage difference (IQR), median absolute difference (IQR) or median percentage absolute difference (IQR). Estimation precision, defined as one standard deviation of estimation bias, ranged from 9 to $25 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$. Estimation precision, defined as the interquartile range of estimation bias, ranged from 16 to $63 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$. Of the included 29 studies, 15 had a report of precision, 10 of which were reported as one standard deviation of estimation bias. P_{30} (the percentage of CKD-EPI_{Cr} eGFRs that fell within 30% of the mGFR), ranged from 35% to 96%. Of 29 studies, 20 reported P_{30} as a measure of estimation accuracy (Fig. 2). One study used total deviation index as a tool to compare performance of the estimation equation [17].

We identified four different correlation coefficients, i.e. Pearson, Spearman, intraclass and concordance correlation coefficients. The intraclass and concordance correlation coefficients between eGFR and mGFR ranged from 0.43 to 0.86. Whereas 10 studies reported Pearson or Spearman correlation coefficients, three presented the intraclass or the Lin's concordance correlation coefficient. The Bland-Altman plot was presented in 11 studies. Accordingly, limit of agreement between CKD-EPI_{Cr} and mGFR ranged from 57 to $158 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$. The RMAR slope ranged from 0.8 to 1.72 and the RMAR intercept from 13.4 to 29.6. Three studies presented the RMAR slope and intercept (Table S5).

The GFR change slope was reported in four of the six longitudinal studies, three of which showed a significant difference between the mGFR and eGFR change slope. Accordingly, the CKD-EPI_{Cr} underestimated the decline in mGFR (1.6 vs. 2.6, 3.2 vs. 6 and 0.9 vs. $3.4 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$) (Table S8).

[H1]Discussion

In this systematic review, we found that CKD-EPI_{Cr} diagnostic performance compared with mGFR was evaluated using a variety of measures, namely estimation bias, precision,

Commented [2]: TS: Insert Fig 2 near here

accuracy, Bland–Altman plot, correlation and reduced major axis. The outcomes of these measures ranged widely from being acceptable to unacceptable. Lack of consistency in how dissimilarity between CKD-EPI_{Cr} and mGFR is assessed hindered meta-analysis of the data and reaching a conclusion regarding CKD-EPI_{Cr} diagnostic performance. However, qualitative synthesis of the available data showed that CKD-EPI_{Cr} could either underestimate or overestimate point GFR and GFR decline over time in adults with diabetes. These findings are important to consider in a context in which: (1) eGFR is the recommended tool for routine assessment of kidney function in people with diabetes by many guidelines; (2) in recent times, there has been a focus on following early patterns of GFR decline to predict risk of progression to end-stage kidney disease; and (3) many pharmaceutical clinical trials consider eGFR decline as a surrogate endpoint to assess progressive kidney function loss in people with diabetic kidney disease. Knowing that CKD-EPI_{Cr} might not accurately represent measured GFR in people with diabetes; clinicians and researchers may need to interpret the test results cautiously.

The difference between CKD-EPI_{Cr} and mGFR point estimates (estimation bias) was reported most commonly as mean (SD), and was $> 10 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$ in 11 of 16 studies. Although some studies indicated that CKD-EPI_{Cr} overestimated mGFR [25–29] by as much as $35.4 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$ [30], others reported that the equation underestimated mGFR [12,14,15,28,31–35] by as much as $26 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$ [36]. This is of particular concern in hyperfiltration, which occurs commonly in diabetes [37]. A systematic review of the current literature showed that CKD-EPI_{Cr} estimation bias was higher in people with GFR $> 90 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$ [12,14–17,31–34]; and the equation fails to perform satisfactorily over higher ranges of GFR. These results are especially important for randomized clinical trials (RCTs) in which participants are selected based on their CKD-EPI_{Cr} levels or treatment effect is assessed by change in eGFR as the end point. In the former situation, the actual GFR in study participants may be higher or lower than the reported CKD-EPI_{Cr} and this might affect an individual's response to treatment. In the latter scenario, estimation bias can result in an overestimated or underestimated treatment effect because the slope of change presented by CKD-EPI_{Cr} might be more or less than the mGFR change slope. Hence, there may be a bias-by-treatment interaction effect that affects interpretation of the

effectiveness of the proposed treatment for diabetic kidney disease even when it produces random misclassification [38].

Among 15 studies that reported estimation precision, 10 used one standard deviation of estimation bias, which was shown to be $> 10 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$ in nine studies [12,14,25,28,29,31–33,35]. The number was reported to be as high as $24.6 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$ in a study of 210 people with diabetes in South Korea (mean CKD-EPI_{Cr} 72.9, mean mGFR $93.1 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$) [35]. In other words, if the distribution of GFR is normal, in 95% of cases the mean difference between CKD-EPI_{Cr} and mGFR would be between -69.4 and $29 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$, which is noticeable considering that GFR usually ranges from 0 to $150 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$ [7]. Such variance can easily change an individual's kidney disease stage [39]. This becomes more important in clinical decision-making where reno-protective agents are considered for use in specific diabetic kidney disease stages [2,3,40–42].

Estimation accuracy, reported as P_{30} in 20 of 29 papers, widely varied from 48% [27] to 96% [43]. With such variation in P_{30} , it is difficult to consider CKD-EPI_{Cr} an accurate estimate of GFR. Moreover, the percentage approach is 'relative' in nature and allows for larger absolute under- or overestimations for larger mGFR values as long as these under- or overestimated values lie within $\pm 30\%$. In essence, very different absolute differences will be judged to be similar using this relative metric.

Among 13 studies reporting correlation coefficients, only three reported either intraclass or Lin's concordance correlation coefficients, which were 0.43 [17], 0.65 [22] and 0.86 [30]. Simple correlation coefficients are not appropriate because a high correlation is necessary, but not sufficient for similarity. For example, if eGFR always produces a value twice as high as mGFR, the two will be perfectly correlated, but very dissimilar [23]. Therefore, Pearson or Spearman correlation coefficients might not be an appropriate tool with which to assess diagnostic performance of the CKD-EPI_{Cr}. However, intraclass and Lin's concordance correlation coefficients assess the agreement between two different measurement methods for the same variable and hence are appropriate tools for statistical judgement [44,45]. It is critical that this issue is resolved to prevent further misuse and

misinterpretation of such coefficients in future studies of diagnostic performance assessment for clinical tests.

Only 38% ($n = 11$) of the included studies reported the Bland–Altman plot of mGFR and CKD-EPI_{Cr}, with a limit of agreement as wide as $158 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$ in a study with 209 people with type 2 diabetes in China [27]. Again, this is noticeable because GFR usually ranges from 0 to $150 \text{ ml min}^{-1} 1.73 \text{ m}^{-2}$ [7]. RMAR was used by only three studies to assess the agreement between CKD-EPI_{Cr} and mGFR, all pointed to a considerable proportional and fixed estimation bias of the equation [17,25,33].

Four of six studies in this review highlighted a significant underestimation of mGFR decline over time by CKD-EPI_{Cr} [15,17,33,46]. This is particularly important in interpreting the outcomes of RCTs where a 30–40% decline in eGFR is considered a surrogate endpoint of progressive kidney function loss [47]. Such underestimation of the true decline in GFR might lead to underestimation of treatment effects for the desirable medications studied.

This systematic review highlights considerable limitations in the original studies that aimed to evaluate diagnostic performance of CKD-EPI in people with diabetes. First, studies rarely conducted analysis in subgroups based on sex, age, BMI and race (only three of 29 studies reported age, sex and BMI, and only one reported race-split results). This is a considerable limitation because the diagnostic performance of CKD-EPI is under question in elderly people and those with obesity, as well as where race is not considered in estimating eGFR [48–50]. Second, only two of the included studies measured GFR by inulin clearance, which is considered the gold standard. Other techniques of measuring GFR (e.g. iohexol, ¹²⁵I-iothalamate, ^{99m}Tc-DTPA and ⁵¹Cr-EDTA) show bias and inaccuracy compared with inulin clearance [6]. Third, many of these GFR measurement techniques use the Bröchner–Mortensen correction factor which cannot obtain absolute clearance values for mGFR higher than $\sim 200 \text{ ml/min}$. This would be a potential source of bias in individuals with hyperfiltrating kidneys, especially younger adults with type 1 and type 2 diabetes. This issue was addressed recently by Jødal *et al.* [51]. However, the majority of the conducted studies addressing the question in this systematic review were

published before introduction of the newer Jødal–Brøchner–Mortensen correction factor and hence did not use this correction coefficient. Fourth, the included studies compared mGFR and eGFR levels corrected for body surface area and the formula used to correct for this differed between studies (Table 3). Furthermore, indexation of GFR for body surface area could induce errors in GFR assessment [52]. Therefore, it would be best to compare absolute mGFR and eGFR values (without correction for body surface area) to prevent introducing an extra source of bias. However, lack of reports based on absolute GFR in the included studies precluded such comparison in this systematic review. Lastly, because an individual's glycaemic status can affect mGFR levels, it is very important to evaluate mGFR values considering data regarding participants' blood glucose level [30]. However, only two of the included studies reported glycaemic status at the time of GFR measurement [30,31]. eGFR might still be the preferred method to diagnose or predict end-stage kidney disease despite its caveats. However, to improve the equation's diagnostic performance, precise assessment of its potential pitfalls in different populations and aggregation of the results are required. Therefore, studies need to be more consistent in their methods of measuring GFR and creatinine, their study design and analysis, as well as outcome report to enable aggregated data analysis.

This review has several limitations. First, we assessed performance of the creatinine-based CKD-EPI equation. Relying on creatinine to estimate GFR has the issue of creatinine level variation between individuals and within each individual at different times. However, this has been partly resolved by considering the major factors affecting creatinine excretion (i.e. age, sex, body size and race) in the formulae. Second, this study reviewed only papers published in English, Farsi, Dutch and Chinese, and we acknowledge that relevant data from papers in other languages may have been inadvertently excluded. However, it should be noted that only one paper was excluded from this review due to language restriction.

[H2] Conclusion

As a serious and common complication of diabetes, diabetic kidney disease needs to be detected as early as possible so that appropriate management of the disease and prevention of its progression can take place. Our systematic review shows that although the CKD-

EPI_{Cr} equation has improved overall estimation of GFR, it still underperforms in people with diabetes, particularly those at the early stages of diabetic kidney disease. We encourage further attempts to improve eGFR equations or identify alternative diagnostic tools to assess kidney function in people with diabetes, particularly those with hyperfiltrating kidneys. An analogy could be drawn between the current tests to diagnose diabetes, which range from raised HbA_{1c}, to fasting, random or 2-h postprandial glucose load. Although there is an overlap in these measures identifying diabetes, it is understood that for several reasons the tests may define different populations with diabetes, however, the risk from diabetes remains regardless of how it is categorized.

This systematic review also reveals that there is no consistency in the literature regarding measures to evaluate diagnostic performance of a continuous index test (e.g. eGFR) compared with a continuous reference test (e.g. mGFR). We urge researchers to reach a consensus on which methods to use for such evaluation. A standard checklist of mandatory statistical tools to be reported by such papers might be of great benefit for future aggregated data meta-analysis of similar research questions.

Funding sources

This study was supported by the Melbourne Research Scholarship. EIE was supported by the Sir Edwards Weary Dunlop Medical Research Foundation.

Competing interests

None declared.

Acknowledgements

We acknowledge Drs Andrew Levey, Lesley Inker and their colleagues from Chronic Kidney Disease Epidemiology (CKD-EPI), Drs Antonios Douros and Elke Schöffner from the Berlin initiative study, Drs Francisco Veríssimo Veronese and Sandra Pinho Silveiro for providing data to be included in this study.

This article is protected by copyright. All rights reserved

Authors' contributions

All authors contributed to the development of the study design and manuscript writing. NZ undertook title/abstract and full-text screening, assessed risk of study quality bias, extracted data and drafted the manuscript. HB developed the search strategy. KK undertook title/abstract screening. LW undertook full-text screening and risk of study quality bias assessment. MH undertook full-text screening. LK undertook data extraction. LC and ML provided statistical expertise. EE, NT and RM provided expertise on diabetic kidney disease and glomerular filtration rate assessment. All authors read, provided feedback and approved the final manuscript. EIE is the guarantor of this article.

References

- 1 GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017; **390**: 1211–1259.
- 2 Brenner BM, Cooper ME, De Zeeuw D, Keane WF, Mitch WE, Parving HH *et al.* Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. *N Engl J Med* 2001; **345**: 861–869.
- 3 Zelniker TA, Wiviott SD, Raz I, Im K, Goodrich EL, Bonaca MP *et al.* SGLT2 inhibitors for primary and secondary prevention of cardiovascular and renal outcomes in type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials. *Lancet* 2019; **393**: 31–39.
- 4 Chadban S, Howell M, Twigg S, Thomas M, Unit TD. *National Evidence Based Guideline for Diagnosis, Prevention and Management of Chronic Kidney Disease in Type 2 Diabetes*. Canberra: XXX, 2009.
- 5 Handelsman Y, Bloomgarden ZT, Grunberger G, Umpierrez G, Zimmerman RS, Bailey TS *et al.* American Association of Clinical Endocrinologists and American College of Endocrinology – Clinical practice guidelines for developing a diabetes

Commented [3]: Author: Please confirm that amended authors are correct

Commented [4]: Author: Is 'Unit' the name of an author?

- mellitus comprehensive care plan – 2015. *American Association of Clinical Endocrinologists* 2015; 413–437.
- 6 Soveri L, Berg UB, Björk J, Elinder CG, Grubb A, Mejare I *et al.* Measuring GFR: a systematic review. *Am J Kidney Dis* 2014; **64**: 411–424.
- 7 Bjornstad P, Cherney DZ, Maahs DM. Update on estimation of kidney function in diabetic kidney disease. *Curr Diabetes Rep* 2015; **15**: 57.
- 8 Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI *et al.* A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009; **150**: 604–612.
- 9 Major RW, Shepherd D, Medcalf JF, Xu G, Gray LJ, Brunskill NJ. The Kidney Failure Risk Equation for prediction of end stage renal disease in UK primary care: an external validation and clinical impact projection cohort study. *PLoS Med* 2019; **16**: e1002955.
- 10 Fan L, Inker LA, Rossert J, Froissart M, Rossing P, Mauer M *et al.* Glomerular filtration rate estimation using cystatin C alone or combined with creatinine as a confirmatory test. *Nephrol Dial Transplant* 2014; **29**: 1195–1203.
- 11 Porrini E, Ruggenenti P, Luis-Lima S, Carrara F, Jiménez A, de Vries APJ *et al.* Estimated GFR: time for a critical appraisal. *Nat Rev Nephrol* 2019; **15**: 177–190.
- 12 Camargo EG, Soares AA, Detanico AB, Weinert LS, Veronese FV, Gomes EC *et al.* The Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation is less accurate in patients with Type 2 diabetes when compared with healthy individuals. *Diabet Med* 2011; **28**: 90–95.
- 13 Rognant N, Lemoine S, Laville M, Hadj-Aïssa A, Dubourg L. Performance of the Chronic Kidney Disease Epidemiology Collaboration equation to estimate glomerular filtration rate in diabetic patients. *Diabetes Care* 2011; **34**: 1320–1322.
- 14 Silveiro SP, Araújo GN, Ferreira MN, Souza FDS, Yamaguchi HM, Camargo EG. Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation pronouncedly underestimates glomerular filtration rate in type 2 diabetes. *Diabetes Care* 2011; **34**: 2353–2355.

Commented [5]: Author: Are the publication details
Endocr Pract 2015 Apr;21 Suppl 1(Suppl 1):1-87.
doi: 10.4158/EP15672.GL

Commented [6]: Author: Are added publication details
correct

- 15 Gaspari F, Ruggenti P, Porrini E, Motterlini N, Cannata A, Carrara F *et al.* The GFR and GFR decline cannot be accurately estimated in type 2 diabetics. *Kidney Int* 2013; **84**: 164–173.
- 16 Maple-Brown LJ, Ekinci EI, Hughes JT, Chatfield M, Lawton PD, Jones GRD *et al.* Performance of formulas for estimating glomerular filtration rate in Indigenous Australians with and without Type 2 diabetes: The eGFR Study. *Diabet Med* 2014; **31**: 829–838.
- 17 Wood AJ, Churilov L, Perera N, Thomas D, Poon A, Macisaac RJ *et al.* Estimating glomerular filtration rate: performance of the CKD-EPI equation over time in patients with type 2 diabetes. *J Diabetes Complications* 2016; **30**: 49–54.
- 18 Zafari N, Churilov L, MacIsaac RJ, Torkamani N, Baxter H, Kiburg KV *et al.* Diagnostic performance of the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation at estimating glomerular filtration rate in adults with diabetes mellitus: a systematic review and meta-analysis protocol. *BMJ Open* 2019; **9**: e031558.
- 19 Ludbrook J. Statistical techniques for comparing measurers and methods of measurement: a critical review. *Clin Exp Pharmacol Physiol* 2002; **29**: 527–536.
- 20 Barnhart HX, Haber MJ, Lin LI. An overview on assessing agreement with continuous measurements. *J Biopharm Stat* 2007; **17**: 529–569.
- 21 Lin LIK. Total deviation index for measuring individual agreement with applications in laboratory performance and bioequivalence. *Stati Med* 2000; **19**: 255–270.
- 22 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159.
- 23 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **327**: 307–310.
- 24 Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; **151**: 264.
- 25 Zhao F, Zhang L, Lu J, Guo K, Wu M, Yu H *et al.* The Chronic Kidney Disease Epidemiology Collaboration equation improves the detection of hyperfiltration in Chinese diabetic patients. *Int J Clin Exp Med* 2015; **8**: 22084–22097.

- 26 Douros A, Ebert N, Jakob O, Martus P, Kreutz R, Schaeffner E. Estimating kidney function and use of oral antidiabetic drugs in elderly. *Fundam Clin Pharmacol* 2015; **29**: 321–328.
- 27 Liu X, Qiu X, Shi C, Huang H, Huang J, Li M *et al.* Modified glomerular filtration rate-estimating equations developed in Asiatic population for Chinese patients with type 2 diabetes. *Int J Endocrinol* 2014; **2014**: 521071.
- 28 Padala S, Tighiouart H, Inker LA, Contreras G, Beck GJ, Lewis J *et al.* Accuracy of a GFR estimating equation over time in people with a wide range of kidney function. *Am J Kidney Dis* 2012; **60**: 217–224.
- 29 Iliadis F, Didangelos T, Ntemka A, Makedou A, Moraliadis E, Gotzamani-Psarakou A, *et al.* Glomerular filtration rate estimation in patients with type 2 diabetes: creatinine- or cystatin C-based equations? *Diabetologia* 2011; **54**: 2987–2994.
- 30 Maahs DM, Bushman L, Kerr B, Ellis SL, Pyle L, McFann K *et al.* A practical method to measure GFR in people with type 1 diabetes. *J Diabetes Complications* 2014; **28**: 667–673.
- 31 Scarr D, Bjornstad P, Lovblom LE, Lovshin JA, Boulet G, Lytvyn Y *et al.* Estimating GFR by serum creatinine, cystatin C, and β 2-microglobulin in older adults: results from the Canadian study of longevity in type 1 diabetes. *Kidney Int Rep* 2019; **4**: 786–796.
- 32 Machado JD, Camargo EG, Boff R, da Silva Rodrigues L, Camargo JL, Soares AA *et al.* Combined creatinine–cystatin C CKD-EPI equation significantly underestimates measured glomerular filtration rate in people with type 2 diabetes mellitus. *Clin Biochem* 2018; **53**: 43–48.
- 33 MacIsaac RJ, Ekinci EI, Premaratne E, Lu ZX, Seah J-M, Li Y *et al.* The Chronic Kidney Disease-Epidemiology Collaboration (CKD-EPI) equation does not improve the underestimation of glomerular filtration rate (GFR) in people with diabetes and preserved renal function. *BMC Nephrol* 2015; **16**: 198.
- 34 Xu Q, Li X, Gao B, Xu Y, Wang Y, Zhang N *et al.* Comparative performance of four equations estimating glomerular filtration rate in adult Chinese diabetics. *J Endocrinol Invest* 2013; **36**: 293–297.

- 35 Oh SJ, Lee JI, Ha WC, Jeong SH, Yim HW, Son HS *et al.* Comparison of cystatin C- and creatinine-based estimation of glomerular filtration rate according to glycaemic status in type 2 diabetes. *Diabet Med* 2012; **29**: e121.
- 36 Cheuiche AV, Soares AA, Camargo EG, Weinert LS, Camargo JL, Silveiro SP. Comparison between IDMS-traceable Jaffe and enzymatic creatinine assays for estimation of glomerular filtration rate by the CKD-EPI equation in healthy and diabetic subjects. *Clin Biochem* 2013; **46**: 1423–1429.
- 37 Magee GM, Bilous RW, Cardwell CR, Hunter SJ, Kee F, Fogarty DG. Is hyperfiltration associated with the future risk of developing diabetic nephropathy? A meta-analysis. *Diabetologia* 2009; **52**: 691–697.
- 38 Page LA, Henderson M. Appraising the evidence: what is measurement bias? *Evid Based Ment Health* 2008; **11**: 36–37.
- 39 Levin A, Stevens PE, Bilous RW, Coresh J, De Francisco ALM, De Jong PE *et al.* Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Suppl* 2013; **3**: 1–150.
- 40 Lewis EJ, Hunsicker LG, Bain RP, Rohde RD. The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. *N Engl J Med* 1993; **329**: 1456–1462.
- 41 Lewis EJ, Hunsicker LG, Clarke WR, Berl T, Pohl MA, Lewis JB *et al.* Renoprotective effect of the angiotensin-receptor antagonist Irbesartan in patients with nephropathy due to type 2 diabetes. *N Engl J Med* 2001; **345**: 851–860.
- 42 Perkovic V, Jardine MJ, Neal B, Bompoint S, Heerspink HJL, Charytan DM *et al.* Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. *N Engl J Med* 2019; **380**: 2295–2306.
- 43 Inker LA, Schmid CH, Tighiouart H, Eckfeldt JH, Feldman HI, Greene T *et al.* Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med* 2012; **367**: 20–29.
- 44 Whitfield JW. Intra-class rank correlation. *Biometrika* 1949; **36**: 463.
- 45 Lin LIK. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989; **45**: 255.

Commented [7]: Author: Please confirm that added publication details are correct

- 46 van Rijn MHC, Metzger M, Flamant M, Houillier P, Haymann J-P, van den Brand JAJG *et al.* Performance of creatinine-based equations for estimating glomerular filtration rate changes over time. *Nephrol Dial Transplant* 2020; **35**: 819–827.
- 47 Levey AS, Gansevoort RT, Coresh J, Inker LA, Heerspink HL, Grams ME *et al.* Change in albuminuria and GFR as end points for clinical trials in early stages of CKD: a scientific workshop sponsored by the National Kidney Foundation in collaboration with the US Food and Drug Administration and European Medicines Agency. *Am J Kid Dis* 2020; **75**: 84–104.
- 48 Schaeffner ES, Ebert N, Delanaye P, Frei U, Gaedeke J, Jakob O *et al.* Two novel equations to estimate kidney function in persons aged 70 years or older. *Ann Intern Med* 2012; **157**: 471.
- 49 Nyman U, Grubb A, Larsson A, Hansson LO, Flodin M, Nordin G *et al.* The revised Lund–Malmö GFR estimating equation outperforms MDRD and CKD-EPI across GFR, age and BMI intervals in a large Swedish population. *Clin Chem Lab Med* 2014; **52**: 815–824.
- 50 Levey AS, Tighiouart H, Titan SM, Inker LA. Estimation of glomerular filtration rate with vs without including patient race. *JAMA Intern Med* 2020; **180**: 793–795.
- 51 Jodal L, Brochner-Mortensen J. Commentary to ‘Correction of the slope-intercept method for the measurement of glomerular filtration rate’. *Nucl Med Commun* 2015; **36**: 407–408.
- 52 Delanaye P, Mariat C, Cavalier E, Krzesinski J-M. Errors induced by indexing glomerular filtration rate for body surface area: reductio ad absurdum. *Nephrol Dial Transplant* 2009; **24**: 3593–3596.

Commented [8]: Author: Please confirm that added publication details are correct

FIGURE 1 Study selection flow chart. CKD-EPI_{Cr}; Creatinine-based Chronic Kidney Disease Epidemiology Collaboration equation.

FIGURE 2 Forest plot of the included studies characteristics and outcome reports. (a) Mean eGFR, (b) mean mGFR, (c) CKD-EPI_{Cr} estimation accuracy and (d) estimation bias

reported in the included studies. P30 is defined as the percentage of eGFR results that lie within 30% of the mGFR values. The estimation bias is reported as mean (eGFR-mGFR).

Commented [9]: Author: Is this eGFR 'minus' mGFR?

Supporting Information

Additional Supporting Information can be found in the online version of this article:

Figure S1. The Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-2 of the included studies.

Table S1. Methods of measuring GFR.

Table S2. The Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-2 of the included studies.

Table S3. Baseline characteristics of the included studies.

Table S4. Methods used to evaluate the diagnostic performance of the creatinine-based Chronic Kidney Disease Epidemiology Collaboration eGFR compared with the reference test.

Table S5. Performance of the creatinine-based Chronic Kidney Disease Epidemiology Collaboration eGFR compared with the reference test in showing point GFR.

Table S6. Performance of the creatinine-based Chronic Kidney Disease Epidemiology Collaboration eGFR compared with the reference test in showing point GFR in people with type 1 diabetes.

Table S7. Performance of the creatinine-based Chronic Kidney Disease Epidemiology Collaboration eGFR compared with the reference test in showing point GFR in people with type 2 diabetes.

Table S8. Performance of the creatinine-based Chronic Kidney Disease Epidemiology Collaboration eGFR compared with the reference test in showing GFR decline over time.

Appendix S1. Supplementary references.

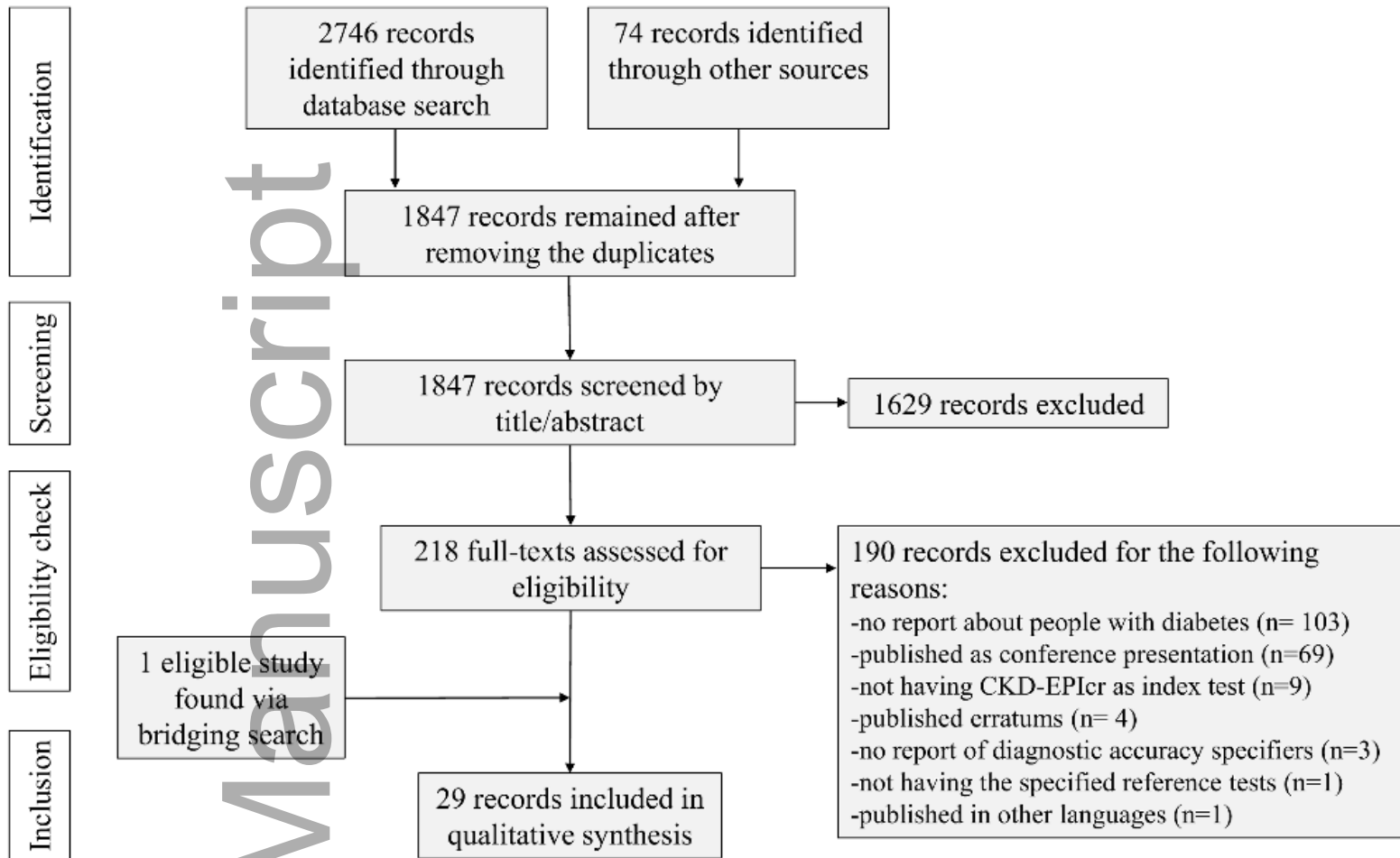
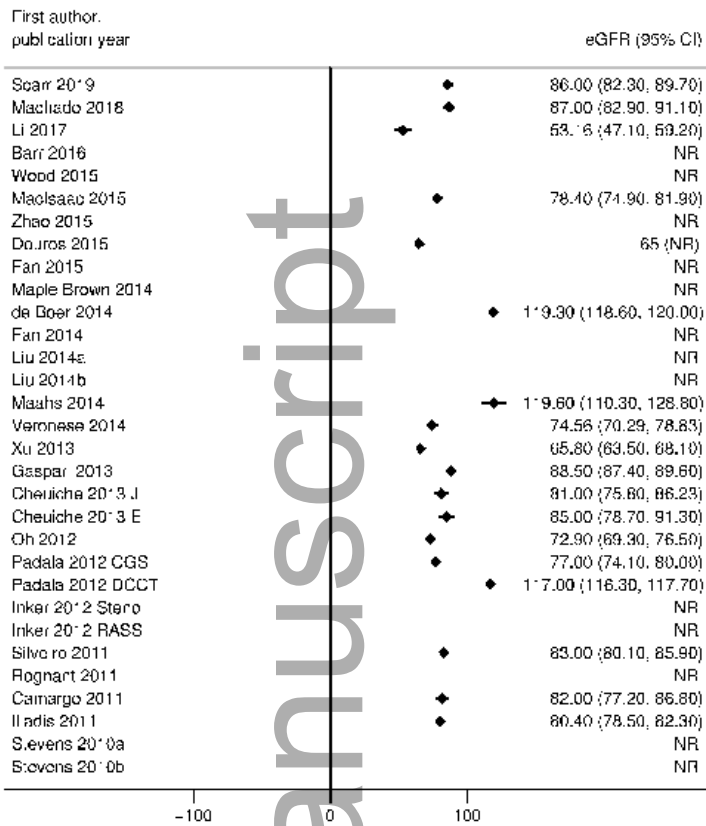
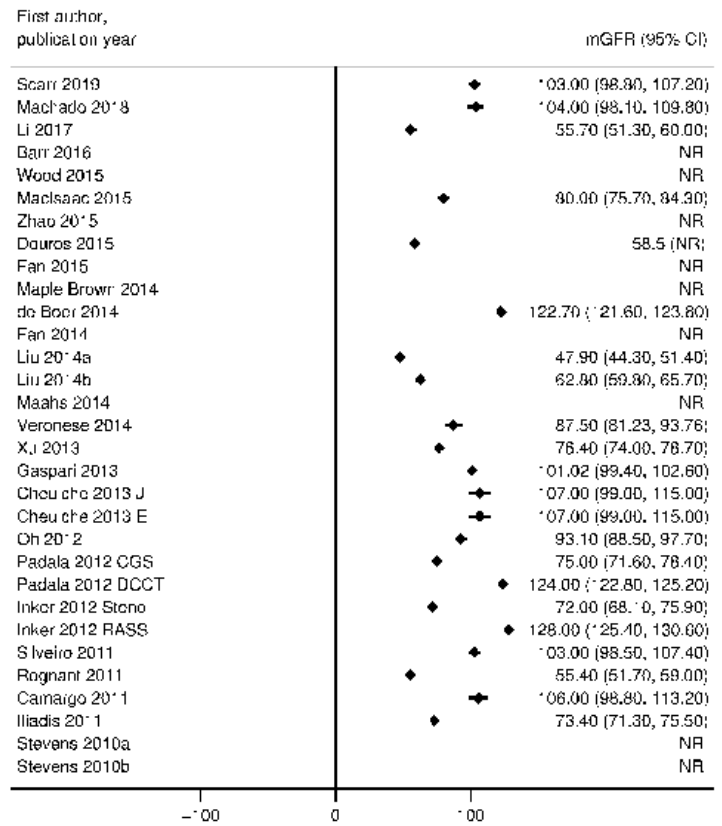


Figure 1. Study selection flow chart

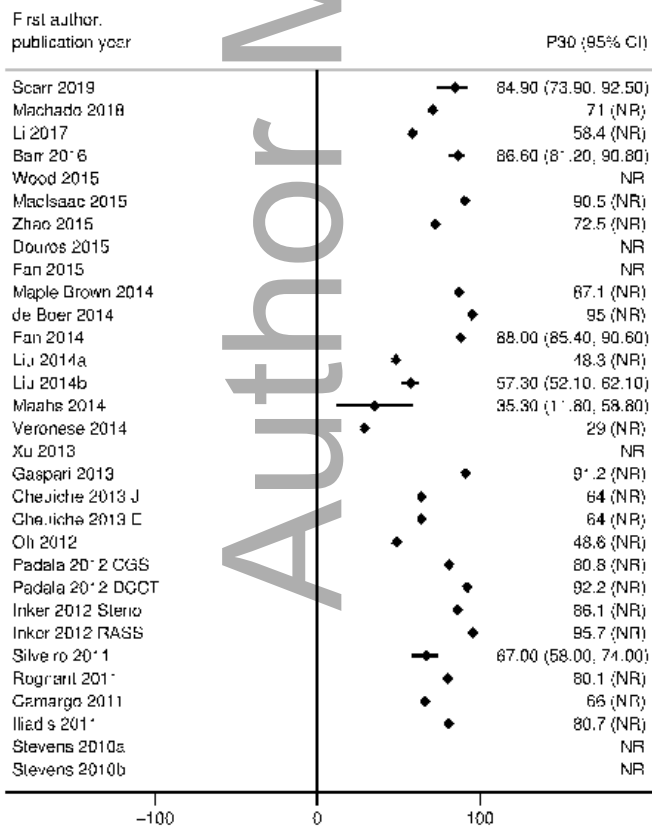
dme_14391_f1.eps



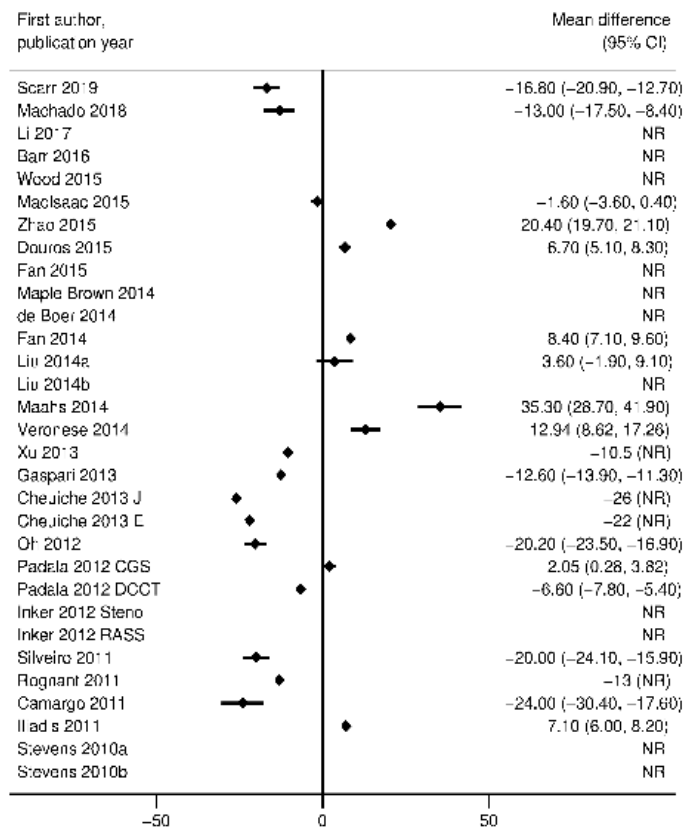
(a.) Estimated GFR



(b.) Measured GFR



(c.) Estimation accuracy



(d.) Estimation bias

Figure 2. Forest plot of the included studies characteristics and outcome reports.