PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Association between non-registration of chronic kidney disease and mortality and cardiovascular outcome: a time-to-event analysis of retrospective primary care data
AUTHORS	Van den Wyngaert, Ine; Mamouris, Pavlos; Ali, Endale Alemayehu; Vaes, Bert; Van Pottelbergh, Gijs

VERSION 1 – REVIEW

REVIEWER	Michele Provenzano Università di Bologna
REVIEW RETURNED	20-Jan-2024

GENERAL COMMENTS	Van den Wyngaert and Colleagues, in this original article, evaluated the association between non-diagnosed CKD and CV events and mortality in GP patients. The idea is of interest despite the fact that the evidence of late referral and impaired prognosi in CKD patients has already been demonstrated. The two concepts are not equal but similar. My comments: - In methods: please note the sub-distribution hazards derive from the Fine and Gray model. This should be specified - Statistical analysis is well described and complete. Also the sample size is huge and these are two major points. Please only better specify what is the method used for including the covariates in the Cox: backward selection? This is imporant since some findings in the Cox models are really hard to understand such as the protective factor of hypertension. I would prefer if authors run a stepwise automated method with a p value of 0.100 which retain only the associated variables and exclude the collinearity per se - Please cite and include in discussion the following papers: doi: 10.1016/j.numecd.2015.04.001. AND doi: 10.3390/life12081202. which are strictly related to the aim of the topic - Please implement in discussion the strategies on how to improve communication to GP from nephrologists. This is al current matter of discussion.

REVIEWER	Maria Eriksson Svensson Uppsala University, Renal medicine
REVIEW RETURNED	15-Feb-2024

GENERAL COMMENTS	In this paper “"Impact of under-registration of chronic kidney disease on mortality and cardiovascular outcome: a time-to-event analysis"” by Van den Wyngaert et al the authors aim to study the association between under-registration of chronic kidney disease and/with mortality and cardiovascular outcome. This is an interesting

	research question and a well written paper but there are some issues to be addressed. Please explain why 17% of subjects from GPs that did not meet the criteria for optimal registration were excluded because this introduces a selection basis that impacts generalizability of the study results. Prevalence of patients diagnosed with CKD using laboratory measurements (5.5%) is somewhat lower than previous studies. Is this a low-risk population? Two-thirds of these patients had no CKD diagnosis code which is in line with previous studies. Please discuss and refer to other studies. The authors use the term under-registration and, in the figures, unregistration instead of the non-registration, could you please explain the reason for this. A basic descriptive with clinical and biochemical characteristics of subjects with CKD total, registered CKD and non-registered CKD for comparison would be helpful for the reader. Please add. It is unclear to me how long the follow-up time in this study was but I may have missed this but if not. Please add All cause death is presented in table 1 not cardiovascular vs non-cardiovascular death. If this data is available, it would be of interest by registration status. Please add. Figure 3 is pivotal in this study and should be highlighted since non-registration may be a risk factor comparable to diabetes and hypertension but outweighed by age and stage of CKD by far. Please comment. In the discussions section, there is no discussion on reason for non-registration some that could be GP related, some patient-related or both. Please consider adding. In the discussions section, it would be important to discuss association vs causality, potential explanations for the findings in the study and potential implications and also suggest future studies. The large proportion of missingness even though it is handled with imputation should be mentioned as a limitation. Please add. A minor point is that the format of ref 25 should be checked and corrected.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: comments

- In methods: please note the sub-distribution hazards derive from the Fine and Gray model. This should be specified

Thank you for this comment. We specified in the manuscript the derivation of the sub-distribution hazards from the Fine Gray model.

*Line 139-140: ‘We estimated the hazard ratios (HRs) and derived the sub-distribution hazard ratios (sHRs) from the Fine and Gray model.’

- Statistical analysis is well described and complete. Also the sample size is huge and these are two major points. Please only better specify what is the method used for including the covariates in the Cox: backward selection? This is important since some findings in the Cox models are really hard to understand such as the protective factor of hypertension. I would prefer if authors run a stepwise automated method with a p value of 0.100 which retain only the associated variables and exclude the collinearity per se
*Thank you for the suggestion to use a stepwise automated method. However, we decided not to use backward selection. We selected the covariates based on literature. While step-wise variable selection methods can be useful for simplifying models and identifying potential predictors, they have several limitations when applied to Cox prediction hazard models [1-4]:
• Multiple testing issue: step-wise involves testing multiple hypothesis, one for each potential variable to include or exclude from the model. This increase the likelihood of finding false positive (Type I error) purely by chance.
• Loss of Power: Step-wise selection can result in the exclusion of potentially important variables from the model if they do not meet the arbitrary significance thresholds for inclusion. This can lead to a loss of statistical power to detect true associations between predictors and the outcome.
• Ignoring model assumptions: Cox regression assumes that the hazard ratios associated with each predictor are constant over time (proportional hazards assumption). Variable selection methods like backward selection do not directly address violations of this assumption, so it's crucial to assess the proportional hazards assumption in the final model.
• Selection Bias: step-wise selection methods tend to select variables that are strongly associated with the outcome in the dataset being analysed. This can lead to overfitting and biased estimates of the regression coefficients.

1) Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. Psychosomatic Medicine, 66:411–421.
2) Huberty, C. J. (1989). Problems with stepwise methods—better alternatives. Advances in Social Science Methodology, 1:43–70.
3) Malek, M. H. and Coburn, D. E. B. J. W. (2007). On the inappropriateness of stepwise regression analysis for model building and testing. European Journal of Applied Physiology, 101(2):263–264.
4) Sribney, B., Harrell, F., and Conroy, R. (2011). Problems with stepwise regression.

- Please cite and include in discussion the following papers: doi: 10.1016/j.numecd.2015.04.001. AND doi: 10.3390/life12081202. which are strictly related to the aim of the topic
*Thank you for these suggestions. We added the first reference in the manuscript (line 292-293). The second reference you mentioned is also very valuable. However, the focus of this article includes more than diabetic nephropathy and its treatment. We chose not to elaborate on this to keep the discussion as concise as possible.
- Please implement in discussion the strategies on how to improve communication to GP from nephrologists. This is al current matter of discussion.
Thank you for this suggestion. We added this in the discussion.
*Line 289-290: ‘Good mutual communication between GP and nephrologist through referral letters and clear consultation reports can contribute to this.’

3) Reviewer 2: comments

-Please explain why 17% of subjects from GPs that did not meet the criteria for optimal registration were excluded because this introduces a selection basis that impacts generalizability of the study results.
*Thank you for this comment. Once the GP data were inserted, registration performance was audited using a number of algorithms that compared their results with those of all other applicants. Only the data of the practices with an optimal registration performance were included in the database since

data completeness depends on the quality of registration of the participating general practitioner. With these criteria, we intend to minimize the risk of recording bias, where general practitioners only register certain, e.g. serious, diagnoses. The design, selection process, quality control procedures and comparability with other (inter)national registration networks were described in detail previously (Truyers C, Goderis G, Dewitte H, et al. The Intego database: background, methods and basic results of a Flemish General practice-based continuous morbidity registration project. BMC Med Inform Decis Mak 2014;14:48).

-Prevalence of patients diagnosed with CKD using laboratory measurements (5.5%) is somewhat lower than previous studies. Is this a low-risk population? Two-thirds of these patients had no CKD diagnosis code which is in line with previous studies. Please discuss and refer to other studies.

*Thank you for this remark. As mentioned in the method section, this study was conducted following on from previous work. [Van den Wyngaert I, et al. An exploration of under-registration of chronic kidney disease in Belgian general practices using logistic regression. Plos One. 2022;17(12):e0279291] In that research, the prevalence of non-registered CKD, the diagnostic delay (time between abnormal eGFR and diagnosis) and the baseline characteristics of the non-registered patient group were examined in a Belgian GP population. The same study population was used. A similar high percentage of nonregistered patients as in our study was described by Ryan et al (74%) in 2007. [Ryan TP, Sloand JA, Winters PC, Corsetti JP, Fisher SG. Chronic kidney disease prevalence and rate of diagnosis. Am J Med. 2007; 120(11):981–6.]

The prevalence of possible CKD in the current study (5.5%) was much lower than earlier published. Previous research in our own INTEGO-database also showed a higher prevalence of CKD (13%). [Van Pottelbergh G, Bartholomeeusen S, Buntinx F, Degryse J. The prevalence of chronic kidney disease in a Flemish primary care morbidity register. Age and Ageing. 2012;41(2):231-3] An explanation for this difference can be found in the study design and selected population. The current study included all patients ≥18 years old, while other studies only used data about patients ≥45 years old. In some patients only one eGFR <60 mL/min/1.73m2 or even no eGFR measurement was available. These patients may also belong to the CKD group, which could result in a higher prevalence. Besides, we did not take the presence of proteinuria into account in the detection of CKD patients. Mainly due to the lack of data on proteinuria, which brings us straight to the problem of underdetection of proteinuria in the Flemish general practice. As a consequence, we underestimated the prevalence among those patients. Both overestimation and underestimation due to single time-point detection of abnormal kidney function and selection of high-risk groups were previously described to explain the difference in prevalence. [Hill NR, Fatoba ST, Oke JL, Hirst JA, O'Callaghan CA, Lasserson DS, et al. Global Prevalence of Chronic Kidney Disease - A Systematic Review and Meta-Analysis. Plos One. 2016;11(7).][Glassock RJ, Warnock DG, Delanaye P. The global burden of chronic kidney disease: estimates, variability and pitfalls. Nat Rev Nephrol. 2017;13(2):104-14.]

As we discussed these findings in the previous article, we decided to mention this again in the discussion.

-The authors use the term under-registration and, in the figures, unregistration instead of the non-registration, could you please explain the reason for this.

*Thank you for this comment. It seems to be a semantic confusion.
We used ‘unregistered’ to indicate that the diagnosis was ‘non-registered’. Patients with unregistered CKD were identified if they had no diagnostic CKD code for any time during the ≥12-month lookback period before the first eGFR measurement and up to 6 months post-index date. Those with a documented U99 during this time period were considered as having registered CKD. We changed ‘under-registration’ and ‘unregistered’ to ‘non-registration’ and ‘non-registered to avoid confusion.

-A basic descriptive with clinical and biochemical characteristics of subjects with CKD total, registered CKD and non-registered CKD for comparison would be helpful for the reader. Please add.

*Line 93-96: For these characteristics we would like to refer to our earlier article in which these data are discussed. [Van den Wyngaert I, et al. An exploration of under-registration of chronic kidney disease in Belgian general practices using logistic regression. Plos One. 2022;17(12):e0279291] Since this data has already been published previously, we cannot add these tables again. A reference to this article is given in the methods section (line 93-96).

-It is unclear to me how long the follow-up time in this study was but I may have missed this but if not. Please add

Thank you for this remark. As mentioned in the method section, this study was conducted following on from previous work. [Van den Wyngaert I, et al. An exploration of under-registration of chronic kidney disease in Belgian general practices using logistic regression. Plos One. 2022;17(12):e0279291]. The follow-up and diagnostic delay was discussed in this article. For clarity we mentioned this again in the results section.

*Line 168: 'The maximum follow-up was 3.97 years.'

-All cause death is presented in table 1 not cardiovascular vs non-cardiovascular death. If this data is available, it would be of interest by registration status. Please add.

*Thank you for this suggestion. Indeed it would be interesting to distinguish between cardiovascular and non-cardiovascular mortality. Unfortunately we do not have this information. General practitioners do register when the patient has died, but the cause of death usually cannot be determined from the EHR.

-Figure 3 is pivotal in this study and should be highlighted since non-registration may be a risk factor comparable to diabetes and hypertension but outweighed by age and stage of CKD by far. Please comment.

Thank you for this valuable comment. You correctly noted that the effect of non-registration was outweighed by age and stage of CKD. This was therefore stated more clearly in the discussion.

*Line 227-229: 'We must note that non-registration may be a risk factor to mortality comparable to diabetes, but outweighed by age and stage of CKD by far.'
-In the discussions section, there is no discussion on reason for non-registration some that could be GP related, some patient-related or both. Please consider adding.
We believe it would be useful to further investigate the causes of non-registration first. This was added to the discussion.
*Line 261-263: 'Moreover, it seems useful to investigate why the diagnosis was not registered in the EHR. Based on these results, the problem of non-registration could be addressed.'
-In the discussions section, it would be important to discuss association vs causality, potential explanations for the findings in the study and potential implications and also suggest future studies.
Thank you for these suggestions. We added information add the following lines:
*Line 227-230: 'We must note that non-registration may be a risk factor to mortality comparable to diabetes, but outweighed by age and stage of CKD by far. An association was found, but causality was not investigated. It is unclear whether better registration will lead to a better outcome, so this should be a topic for further research.'

-The large proportion of missingness even though it is handled with imputation should be mentioned as a limitation. Please add.
We added the missingness to the limitations.

*Line 307: ‘The large proportion of missingness is a limitation as well.’

-A minor point is that the format of ref 25 should be checked and corrected.
Thank you for this remark. The reference was corrected.

*Line 379-380: ‘Sundström J, et al. Prevalence, outcomes, and cost of chronic kidney disease in a contemporary population of 2,4 million patients from 11 countries: The CaReMe CKD study. Lancet Reg Health Eur. 2022;20:100438.’

VERSION 2 – REVIEW

REVIEWER	Michele Provenzano Università di Bologna
REVIEW RETURNED	26-Apr-2024

GENERAL COMMENTS	The manuscript has been improved.

REVIEWER	Maria Eriksson Svensson Uppsala University, Renal medicine
REVIEW RETURNED	21-Apr-2024

GENERAL COMMENTS	Thank you for letting me read your interesting work. I have now read your reviewer responses and the updated manuscript. All my concerns and questions have been addressed in a clear and stringent way.