



UvA-DARE (Digital Academic Repository)

Epidemiology of chronic kidney disease in Europe

Brück, K.

Publication date

2016

Document Version

Final published version

[Link to publication](#)

Citation for published version (APA):

Brück, K. (2016). *Epidemiology of chronic kidney disease in Europe*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

CHAPTER 5:

Quality of Reporting and Study Design of CKD Cohort Studies Assessing Mortality in the Elderly Before and After STROBE: A Systematic Review

Anirudh Rao, Katharina Brück, Shona Methven, Rebecca Evans, Vianda S. Stel, Kitty J. Jager, Lotty Hooft, Yoav Ben-Shlomo and Fergus Caskey

PLoS One. 2016 May 11; 11(5):e0155078.

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155078>

ABSTRACT

Background

The STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) statement was published in October 2007 to improve quality of reporting of observational studies. The aim of this review was to assess the impact of the STROBE statement on observational study reporting and study design quality in the nephrology literature.

Study Design

Systematic literature review.

Setting & Population

European and North American, Pre-dialysis Chronic Kidney Disease (CKD) cohort studies.

Selection Criteria for Studies

Studies assessing the association between CKD and mortality in the elderly (>65 years) published from 1st January 2002 to 31st December 2013 were included, following systematic searching of MEDLINE & EMBASE.

Predictor

Time period before and after the publication of the STROBE statement.

Outcome

Quality of study reporting using the STROBE statement and quality of study design using the Newcastle Ottawa Scale (NOS), Scottish Intercollegiate Guidelines Network (SIGN) and Critical Appraisal Skills Programme (CASP) tools.

Results

37 papers (11 Pre & 26 Post STROBE) were identified from 3621 potential articles. Only four of the 22 STROBE items and their sub-criteria (objectives reporting, choice of quantitative groups and description of and carrying out sensitivity analysis) showed improvements, with the majority of items showing little change between the period before and after publication of the STROBE statement. Pre- and post-period analysis revealed a Manuscript STROBE score increase (median score 77.8% (Inter-quartile range (IQR), 64.7–82.0) vs 83% (IQR, 78.4–84.9, $p=0.05$). There was no change in quality of study design with identical median scores in the two periods for NOS (Manuscript NOS score 88.9), SIGN (Manuscript SIGN score 83.3) and CASP (Manuscript CASP score 91.7) tools.

Limitations

Only 37 Studies from Europe and North America were included from one medical specialty. Assessment of study design largely reliant on good reporting.

Conclusions

This study highlights continuing deficiencies in the reporting of STROBE items and their sub-criteria in cohort studies in nephrology. There was weak evidence of improvement in the overall reporting quality, with no improvement in methodological quality of CKD cohort studies between the period before and after publication of the STROBE statement.

INTRODUCTION

Chronic kidney disease (CKD) is a complex chronic condition, and in recent years has emerged as a major public health problem (1, 2). CKD has been termed a “Geriatric Giant”, as this disproportionately affects the elderly and is assuming epidemic proportions. Also with increasing life expectancy, patients are surviving longer with chronic conditions including CKD (3). With the increasing burden of CKD, research of treatments developed to improve morbidity and mortality is vital (4). Randomised controlled trials (RCTs) indisputably hold many advantages over observational studies, but owing to ethical or other considerations, may be difficult or impossible to undertake (5–7). In nephrology there has not only been a lack of RCTs, but a large proportion of these RCTs have had negative or null findings (6). Observational studies can provide extremely valuable additional evidence, and when rigorously undertaken may yield similar results as RCTs at far lower expense (8–11).

Standardized reporting of cohort studies is crucial for the evaluation of the merits and flaws of observational research. Inadequate reporting is associated with potentially biased estimates of treatment effects and limits the assessment of a study’s strengths, weaknesses and generalizability (12). In order to address this, the STrengthening the Reporting of OBservational Studies in Epidemiology (STROBE) initiative developed recommendations on what should be incorporated in a precise and thorough report of an observational study. The STROBE statement and checklist were published in October 2007 (13, 14). These reporting guidelines were envisioned to make issues such as confounding, bias, and generalizability more ostensible. In the long term, this would improve the methodology of studies by increased awareness of these issues for researchers designing a new study (15, 16).

The scientific value and reliability of the conclusions drawn from a study are determined to a major extent by the quality of the study design (17). A variety of tools currently exist to assess the risk of bias (methodological quality) of observational studies, and are employed when undertaking a systematic review. These include quality scales, simple checklists, or checklists with a summary judgment for assessment of the risk of bias (18).

The objectives of this review were (a) to determine whether the publication of the STROBE statement is associated with an improvement in the reporting quality of cohort studies assessing mortality in elderly patients with CKD; and (b) to determine whether the publication of the STROBE statement is associated with a decrease in risk of bias (improvement in the methodological quality) of cohort studies assessing mortality in elderly patients with CKD.

METHODS

Data selection

A systematic literature search was performed in Medline and Embase using the OvidSP interface to identify all papers describing pre-dialysis CKD cohort studies in the elderly (> 65 years) where mortality was reported as an outcome. This systematic review was conducted as part of the background preparation for the EQUAL study which is an international (European) multicentre prospective observational cohort study looking at the timing of the start of dialysis in elderly patients (65 years) with estimated glomerular filtration rate (eGFR) of 20 ml/min and therefore the review is restricted to CKD cohort studies in the elderly (19). The search query is presented in Item A in the appendix.

Papers published between 1st January 2002 and 31st December 2013 were included, as the KDOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification were published in 2002 (20). Only articles published in English were considered for the purposes of the review. The initial search strategy yielded more than 10,000 hits, hence the number of studies were reduced by restricting the search to European and North American studies. Each article was double sifted at title, abstract and full text stage using predefined study inclusion and exclusion criteria. Any disagreements about inclusion were resolved by discussion.

The systematic review aimed to cover reporting and design of observational studies before and after the publication of the STROBE statement which was published in October 2007. We assessed reporting and methodological quality during two time periods: before STROBE between 1/1/2002-31/12/2007 and after STROBE 1/1/2008-31/12/2013, allowing a one-year run-in period. By excluding publications in the immediate twelve months post-STROBE we allowed a period of one year for submission, revision and publication of research adhering to the new guidelines.

Data extraction

The reporting of the selected studies was assessed using the STROBE checklist itself, and the methodological quality assessed using three tools. Thirteen of the 22 STROBE checklist items were assessed with 2 to 6 questions per item generating 55 questions. The STROBE checklist is presented in Table A in the appendix. These could be answered as “yes,” “partly,” “no,” “unclear,” or “not applicable”. We used similar methodology to that reported in the publication by Langan et al. (21).

To assess methodological quality, the articles were scored on the Newcastle Ottawa Scale (NOS). At the time this study was designed NOS was recommended by Cochrane for evaluating the risk of bias in observational studies for inclusion in systematic reviews (22,23).

The articles were also scored using the Scottish Intercollegiate Guidelines Network (SIGN) checklist for cohort studies (24), and Critical Appraisal Skills Programme (CASP) cohort studies checklist (25) to estimate concurrent validity of NOS tool. These three checklists were chosen because they were simple checklists without an additional summary judgement (26). The eligible papers that were identified by the sifting process were each scored using the STROBE, NOS, SIGN and CASP checklists by two reviewers. Where there was disagreement between reviewers, consensus was reached by discussion.

Outcome measure

Quality of study reporting was calculated by specific STROBE items and at a manuscript level. A STROBE question score (SQS) was calculated; the number of publications in a period that adequately reported a question divided by the number of publications in which this question was applicable, expressed as a percentage (item analysis). A Manuscript STROBE score (MSS) was calculated for every manuscript; the number of questions (maximum of 55 questions) adequately reported in the publication divided by the number of applicable questions, expressed as a percentage (manuscript analysis). Similarly, to assess the quality of study design the manuscript NOS score (MNOS), manuscript SIGN score (MSiS) and manuscript CASP score (MCAS) were calculated; the number of questions adequately addressed (in each appraisal tool) divided by the number of applicable items, expressed as a percentage in order to facilitate comparison.

Statistical analysis

Comparison between pre- and post-period SQS was performed by calculating the risk (proportion) difference between the two groups using the Wald test and respective 95% confidence intervals, with Benjamini and Hochberg adjusted p values (False Discovery Rate) to control for multiple testing (27). MSS, MNOS, MSiS and MCAS were reported as a median with respective interquartile range (IQR). Pre- and post-period median MSS, MNOS, MSiS and MCAS were compared using the Mann-Whitney (MW) test. Despite excluding articles published for a period of 1 year after introduction of STROBE, this could potentially have been insufficient for uptake and penetration of new information. Therefore a spline linear regression model was used to determine the impact of STROBE over time (28). Sub-group analyses of MSS were carried out restricting articles to those published in nephrology journals, STROBE endorsing and non-endorsing journals and by journal impact factor in the year that the article was published. Sensitivity analyses were carried out by excluding the outlying MSS if any data points were less than 1.5 interquartile ranges (IQRs) below the first quartile or above the third quartile ($< Q1 - 1.5 \times IQR$ or $> Q3 + 1.5 \times IQR$). Simple and weighted kappa statistics were used to compare agreement between reviewers for the NOS, SIGN and CASP checklists. All tests were twotailed, and p values, < 0.05 were considered statistically significant. Data were analysed using STATA v13.1 (College Station, TX, USA) and SAS v9.3 (SAS Institute, Cary, NC, USA) software.

Reporting

The study has been reported in accordance with PRISMA reporting guidelines.

RESULTS

Reporting Quality

Figure 1 shows the flow diagram of exclusions. Of the 3621 articles initially identified by the Medline and Embase search, 3584 (98.9%) were excluded after the sifting process. Only 37 articles met the pre-defined selection criteria for the scoring stage of the review during the inclusion period. Of these 37 articles, 11 were in the pre-STROBE era (1/1/2002-31/12/2007) & 26 in the post-STROBE period (1/1/2008-31/12/2013). Twenty-two of these articles were published in nephrology and 15 in other medical journals. The list of articles considered at the scoring stage of the study is provided in Item B in the appendix.

Figure 1: Prisma flow diagram.

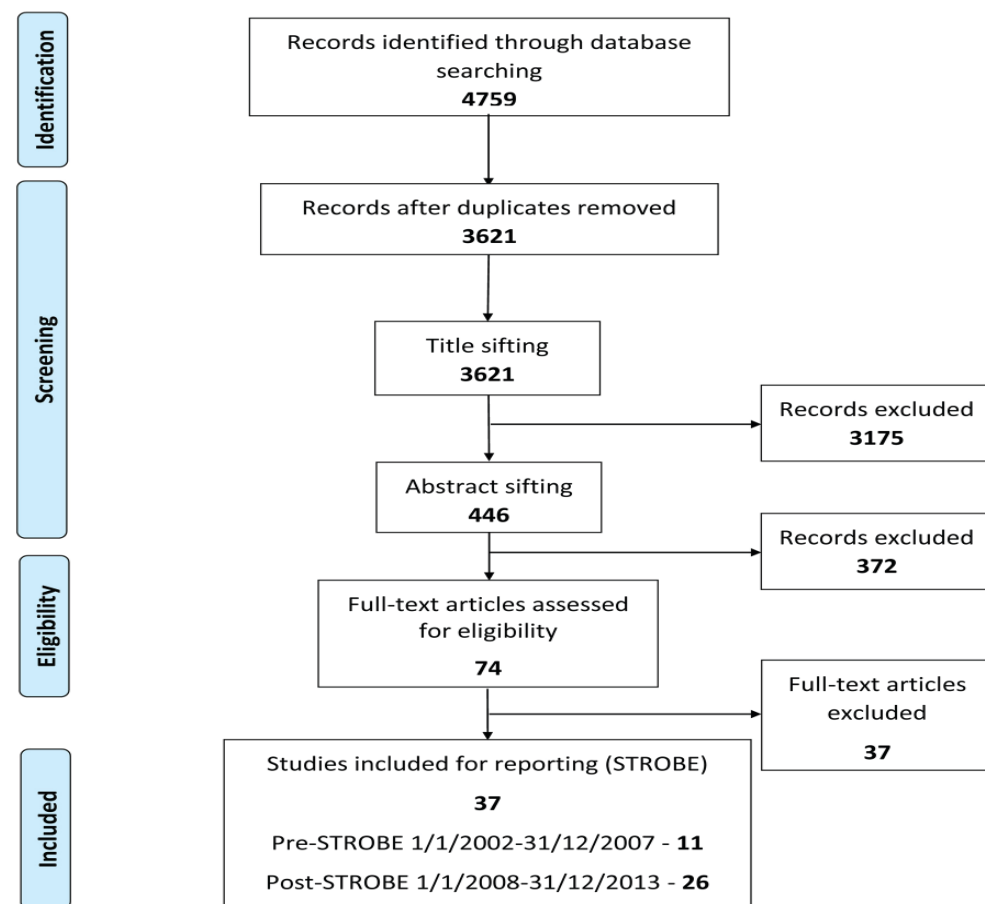


Table 1 summarises the STROBE, NOS, SIGN and CASP scores for each of the articles in the pre- and post-STROBE period. In most cases, reporting quality (STROBE) and methodological quality (NOS, SIGN and CASP) correlated well. However, in some articles methodological quality scored highly with a low score for reporting and vice versa.

Table 1: Summary of pre- and post-STROBE period Manuscript STROBE score (MSS), Manuscript NOS score (MNOS), Manuscript SIGN score (MSiS) & Manuscript CASP score (MCAS) by article.

Publication date	Journal	Study Reporting	Study Design		
		MSS	MNOS	MSiS	MCAS
Pre-STROBE					
Dec-02	Journal of American College of Cardiology	76.5	55.6	22.2	33.3
Jun-03	American Journal of Kidney Diseases	77.8	88.9	91.7	100.0
Oct-04	JASN	66.7	66.7	50.0	33.3
Apr-05	The Journal of the American Medical Association	88.7	88.9	100.0	100.0
Sep-05	JASN	64.7	88.9	80.0	83.3
Dec-05	JASN	84.6	66.7	61.5	83.3
Nov-06	British Medical Journal	82.0	100.0	100.0	91.7
Jul-07	Renal Failure	49.1	100.0	88.9	100.0
Jul-07	JASN	80.4	100.0	100.0	100.0
Nov-07	Nephrology Dialysis Transplantation	51.1	100.0	80.0	91.7
Dec-07	Archives of Internal Medicine	77.8	88.9	83.3	91.7
Post-STROBE					
Nov-08	Nephrology Dialysis Transplantation	83.0	88.9	83.3	100.0
Dec-08	Nephrology Dialysis Transplantation	84.9	77.8	90.0	83.3
Feb-09	JASN	72.9	100.0	83.3	91.7
Apr-09	American Journal of Kidney Diseases	90.0	100.0	69.2	100.0
Jul-09	American Journal of Kidney Diseases	83.7	100.0	100.0	91.7
Jul-09	Clinical JASN	75.0	37.5	25.0	33.3
Dec-09	Journal of American Geriatric Society	78.4	100.0	87.5	100.0
Jul-10	Nephrology Dialysis Transplantation	77.6	100.0	90.9	91.7
Oct-10	Journal of Nephrology	39.6	100.0	100.0	100.0
Sep-10	Journal of General Internal Medicine	83.0	77.8	54.5	41.7
Nov-10	Rejuvenation Research	73.6	88.9	80.0	91.7

Table 1: (continued).

Publication date	Journal	Study Reporting	Study Design		
		MSS	MNOS	MSiS	MCAS
Post-STROBE (continued)					
Sep-11	Clinical JASN	92.2	66.7	90.9	75.0
Jan-12	Nefrologia	50.0	77.8	90.9	91.7
Feb-12	Age and Ageing	83.0	100.0	91.7	100.0
Apr-12	Nephrology Dialysis Transplantation	87.8	88.9	91.7	100.0
Apr-12	The American Journal of Medicine	79.6	100.0	83.3	83.3
May-12	Journal of American Geriatric Society	83.0	88.9	91.7	100.0
May-12	Nephrology Dialysis Transplantation	88.0	100.0	66.7	75.0
Jun-12	The Journal of the American Medical Association	84.9	88.9	66.7	66.7
Jun-12	Clinical JASN	83.0	88.9	76.9	83.3
Jul-12	Journal of American Geriatric Society	79.6	88.9	100.0	91.7
Dec-12	Family Practice	84.6	77.8	63.6	83.3
Feb-13	American Journal of medicine	83.0	88.9	75.0	100.0
Apr-13	BMC Nephrology	84.6	100.0	70.0	100.0
May-13	BMC Nephrology	83.0	100.0	88.9	83.3
Sep-13	Clinical JASN	90.2	100.0	80.0	66.7

JASN= Journal of the American Society of Nephrology.

Some of the STROBE question scores showed a ceiling effect as they were already at a maximum level in the pre-STROBE period and could therefore only remain static or decline. Others saw improvements over the period such as “choice of quantitative groups” (30% vs 71%, $p=0.02$), “addressing of losses to follow up” (0% vs 36%, $p<0.001$), “description of and carrying out sensitivity analysis” (18% vs 58%, $p=0.01$ & 18% vs 65%, $p=0.002$) and “usage of flow diagram” (0% vs 19%, $p=0.01$). However, after adjusting for multiple testing, the change in only two items’ scores remained unlikely to be due to chance; “addressing of losses to follow up” ($p=0.02$) and “carrying out sensitivity analysis” ($p=0.04$). The majority of STROBE questions showed little improvement between the two periods. Some critical questions, such as hypothesis specification and those important to interpretation of study validity such as sample size estimation, addressing missing data, addressing loss to follow up, reason for non-participation and usage of flow diagram continue to be under reported with less than 50% reporting these items in both periods. Details regarding the reporting of the 55 STROBE items in the 37 included cohort studies are shown in Table 2.

Pre- and post-period analyses revealed an increase in MSS (median score 77.8 (IQR, 64.7–82.0) vs 83 (IQR, 78.4–84.9), $p=0.04$) (see Table 3). Any pre-STROBE period articles with MSS scores less than 47.4 and post-STROBE period less than 69 were considered to be outliers. Excluding outliers, the improvement in the MSS between the two periods showed a stronger statistical relationship ($p=0.01$). The results were essentially unchanged when restricted to nephrology journals or stratified by STROBE endorsing or non-endorsing journals, though there was less statistical power to test for differences. Journals with impact factor <5 saw greater change over the two periods when compared to journals with impact factor >5 but given the overlap in the confidence intervals this may have occurred by chance.

Time series analysis of MSS showed that there was a significant improvement in the quality of reporting in the latter three years (1/1/11 to 31/12/13) when compared to the first three years (1/1/2008 to 31/12/2010) after the introduction of the STROBE statement (Table 4). Longitudinal analysis of the MSS using a spline linear regression model (Figure 2), having excluded outliers, suggested a turning point in 2008 with a slight negative trend in the pre-STROBE period (coefficient -0.06 , SE 0.11) and a positive slope in the post-STROBE period (coefficient 0.21, SE 0.05) but this may have occurred by chance (Slope change coefficient 0.27, SE 0.16; p value= 0.10).

Methodological quality (study design); comparison in the pre- and post-STROBE period

We found no evidence for any change in the methodological quality of studies in the pre and post-STROBE period using the Newcastle Ottawa Scale (NOS) (median MNOS 88.9% (IQR, 66.7–100) vs 88.9% (IQR, 88.9–100), $p=0.51$), Scottish Intercollegiate Guidelines Network (SIGN) (median MSiS 83.3% (IQR, 61.5–100) vs 83.3% (IQR, 70–90.9), $p=0.93$) and Critical Appraisal Skills Programme (CASP) (median MCAS 91.7% (IQR, 83.3–100) vs 91.7% (IQR, 83.3–100), $p=0.93$) (Figure 3).

Inter-rater agreement

Agreement between raters for the NOS, SIGN and CASP tools was calculated using the simple or weighted Kappa coefficient. These were assessed at three levels: raters’ agreement on applicability, clarity (can’t say) and yes/no. The inter-rater agreement for each of the tools was overall inadequate, with the NOS tool having poor agreement between the three pairs of raters’. The CASP tool fared slightly better compared to the SIGN tool in raters’ assessment of clarity. A summary table of Kappa coefficients is included in Table B in the appendix.

Table 2: Median STROBE QUESTION SCORE (SQS), Difference (95% CI) with p value of the 55 data items (22 items were further subdivided to 55 questions in total) in 37 CKD cohort studies, by publication period.

Item number	Data Items	Pre-STROBE SQS	Post-STROBE SQS	Difference	LCI	UCI	p value	FDR	FDR*
Title and Abstract									
1A	Is the design described adequately in the title or abstract?	0.73	0.69	-0.04	-0.35	0.28	0.83	0.91	0.93
1B	Does the abstract provide an informative summary of what was done and found?	1.00	1.00	0.00	-	-	-	-	-
Introduction									
2	Is the scientific background and rationale for the investigation reported?	1.00	1.00	0.00	-	-	-	-	-
3A	Are any pre specified hypotheses reported?	0.18	0.23	0.05	-0.23	0.33	0.73	0.91	0.93
3B	Are the objectives reported?	0.73	0.96	0.23	-0.04	0.51	0.09	0.62	0.64
Methods									
4	Are the key elements (ie, retrospective/prospective, cohort/cross-sectional) of the study design presented early in the paper?	1.00	0.92	-0.08	-0.18	0.03	0.14	0.62	.
5A	Are the settings reported?	0.91	1.00	0.09	-0.08	0.26	0.29	0.62	0.74
5B	Are the locations reported?	1.00	0.88	-0.12	-0.24	0.01	0.07	0.52	.
5C	Are relevant dates including periods of recruitment reported?	1.00	0.96	-0.04	-0.11	0.04	0.31	0.62	.
5D	Are relevant dates including periods of exposure reported?	1.00	0.93	-0.07	-0.21	0.06	0.30	0.62	.

Table 2: (continued).

5E	Are relevant dates including periods of follow-up reported?	0.91	0.96	0.05	-0.13	0.24	0.58	0.81	0.87
5F	Are relevant dates including periods of data collection reported?	0.73	0.92	0.20	-0.09	0.48	0.17	0.62	0.74
6A	Are the eligibility criteria for participants described?	1.00	0.92	-0.08	-0.18	0.03	0.14	0.62	.
6B	Are the sources of participants described?	1.00	0.92	-0.08	-0.18	0.03	0.14	0.62	.
6C	Are the methods of selection described?	1.00	0.96	-0.04	-0.11	0.04	0.31	0.62	.
6D	Are the methods of follow-up described?	0.91	0.96	0.05	-0.13	0.24	0.579	0.81	0.87
6E	If it is a matched study, are the matching criteria and the numbers of exposed and unexposed described?	0.00	1.00	1.00	-	-	-	-	-
7A	Are all outcomes described if applicable?	0.91	0.96	0.05	-0.13	0.24	0.58	0.81	0.87
7B	Are all exposures described if applicable?	1.00	1.00	0.00	-	-	-	-	-
7C	Are all predictors described if applicable?	0.91	1.00	0.09	-0.08	0.26	0.29	0.62	0.74
7D	Are potential confounders described?	0.91	0.92	0.01	-0.18	0.21	0.89	0.93	0.94
7E	Are all effect modifiers described?	0.73	0.80	0.07	-0.23	0.38	0.64	0.86	0.91
7F	Are diagnostic criteria described if applicable?	0.80	0.92	0.12	-0.15	0.39	0.40	0.73	0.84
8A	Are the sources of data and details of methods of measurement given for each variable of interest?	0.91	0.85	-0.06	-0.28	0.16	0.57	0.81	0.87
8B	If there is more than 1 group, are the measurement methods comparable?	0.75	1.00	0.25	-0.17	0.67	0.25	0.62	0.74
9	Was there any effort to address potential sources of bias?	0.91	0.92	0.01	-0.18	0.21	0.89	0.93	0.94

* False Discovery Rate (FDR) calculated excluding questions which had 100% completeness in the pre-STROBE phase.

Table 2: (continued).

Item number	Data Items	Pre-STROBE SQS	Post-STROBE SQS	Difference	LCI	UCI	p value	FDR	FDR*
10	Did they describe how the study size was determined?	0.00	0.04	0.04	-0.04	0.13	0.31	0.62	0.74
11A	Did they describe how quantitative variables were handled in the analysis?	0.80	0.80	0.00	-0.29	0.29	1	1	1
11B	Did they describe which groupings were chosen for quantitative variables?	0.90	0.83	-0.07	-0.31	0.17	0.58	0.81	0.87
11C	Did they describe why quantitative groups were chosen?	0.30	0.71	0.41	0.07	0.75	0.02	0.17	0.74
12A	Did they describe all statistical methods including those to deal with confounding?	0.91	1.00	0.09	-0.08	0.26	0.29	0.62	0.99
12B	Did they describe methods to examine subgroups and interactions?	0.73	0.72	-0.01	-0.32	0.31	0.96	0.98	0.87
12C	Did they explain how missing data was addressed?	0.27	0.38	0.11	-0.21	0.43	0.50	0.81	.
12D	Did they explain if applicable how losses to follow-up were addressed?	0.00	0.36	0.36	0.16	0.56	<0.001	0.02	0.74
12E	Did they describe any sensitivity analysis?	0.18	0.58	0.40	0.10	0.69	0.01	0.14	0.12
Results									
13A	Did they report the numbers of individuals at each stage of the study numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, and completed follow-up and were analysed?	0.64	0.81	0.17	-0.15	0.49	0.30	0.62	0.74
13B	Did they give reasons for nonparticipation at each stage? ^a	0.30	0.48	0.18	-0.18	0.53	0.33	0.64	0.75
13C	Did they use a flow diagram if appropriate?	0.00	0.19	0.19	0.04	0.34	0.01	0.15	0.13

Table 2: (continued).

14A	Did they give the characteristics of study participants (eg, demographic, clinical, and social) and information on exposures and potential confounders?	1.00	1.00	0.00	-	-	-	-	-
14B	Did they indicate the number of participants with missing data for each variable of interest?	0.27	0.15	-0.12	-0.42	0.18	0.43	0.74	0.85
14C	Did they summarize follow-up time (average and total amount)?	0.91	1.00	0.09	-0.08	0.26	0.29	0.62	0.74
15A	Did they report numbers of outcome measures over time?	1.00	1.00	0.00	-	-	-	-	-
15B	Did they report summary measures over time?	0.82	0.92	0.10	-0.15	0.35	0.41	0.73	0.84
16A	Did they give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval)?	0.82	0.92	0.03	-0.24	0.29	0.84	0.91	0.93
16B	Did they detail which confounders were adjusted for and why they were included?	0.90	0.92	0.02	-0.19	0.24	0.83	0.91	0.93
16C	Did they report category boundaries when continuous variables were categorized?	0.90	0.88	-0.03	-0.25	0.20	0.83	0.91	0.93
16D	Did they, if relevant, consider translating estimates of relative risk into absolute risk for a meaningful period?	0.33	0.26	-0.07	-0.43	0.28	0.69	0.90	0.93
17A	Did they report on other analyses done, eg, analysis of subgroups or interactions?	0.64	0.85	0.21	-0.11	0.53	0.19	0.62	0.74
17B	Did they do a sensitivity analysis?	0.18	0.65	0.47	0.18	0.76	0.002	0.04	0.03
Discussion									
18	Did they summarize key results with reference to study objectives?	0.82	1.00	0.18	-0.05	0.41	0.12	0.62	0.70

* False Discovery Rate (FDR) calculated excluding questions which had 100% completeness in the pre-STROBE phase.

Table 2: (continued).

Item number	Data Items	Pre-STROBE SQS	Post-STROBE SQS	Difference	LCI	UCI	p value	FDR	FDR*
19	Did they discuss the limitations of the study taking into account potential sources of bias or imprecision (including discussion of the magnitude of any potential sources of bias)?	0.91	1.00	0.09	-0.08	0.26	0.29	0.62	0.74
20	Did they give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence?	1.00	1.00	0.00	-	-	-	-	-
21	Did they discuss the generalizability (external validity) of the study results?	0.73	0.76	0.04	-0.27	0.35	0.80	0.91	0.93
Other Information									
22A	Did they give the source of the funding in the present study and, if applicable, for the original study on which the present article is based?	0.73	0.77	0.03	-0.28	0.34	0.84	0.91	0.93
22B	Did they give the role of the funders in the present study and, if applicable, for the original study on which the present article is based?	0.36	0.46	0.09	-0.25	0.44	0.60	0.81	0.87

* False Discovery Rate (FDR) calculated excluding questions which had 100% completeness in the pre-STROBE phase.

Table 3: Summary of quality of reporting as assessed using the Manuscript STROBE Score (MSS).

	Pre-STROBE		Post-STROBE		p value
	N	median MSS (IQR)	N	median MSS (IQR)	
All Journals	11	77.8 (64.7-82.0)	26	83 (78.4-84.9)	0.05
All Journals (excluding outliers)*	11	77.8 (64.7-82.0)	24	83 (79.6-84.9)	0.01
Nephrology Journals	6	72.25 (64.7-80.4)	16	83.4 (76.3-87.9)	0.10
STROBE endorsing Journals (3)	2	79.9 (77.8-82)	3	83.7 (83-90)	0.15
Non-STROBE endorsing Journals (13)	9	76.5 (57.9-82.5)	23	83 (77.6-84.9)	0.10
IMPACT FACTOR < 5	3	51.1 (49.1-77.8)	16	83 (78-84.6)	0.03
IMPACT FACTOR ≥ 5	8	79.1 (71.6-83.3)	10	83.35 (79.6-90.0)	0.07

*Excluding articles that were less than 1.5 interquartile ranges (IQRs) below the first quartile (< Q1-1.5×IQR). Pre-STROBE = 47.4 & Post-STROBE = 69

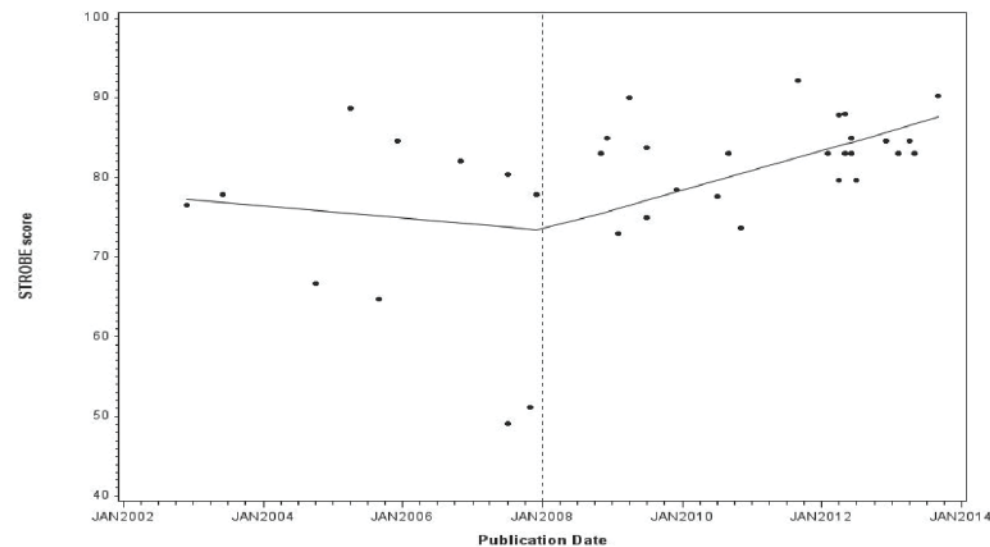
Table 4: Quality of the reporting of observational studies as assessed using the Manuscript STROBE score (MSS) over time.

	Pre-STROBE publication (period 1) 2002 - 2007	Immediate Post-STROBE publication (period 2) 2008 - 2010	Late Post-STROBE publication (period 3) 2011 - 2012	p value	
				period 1 to 2	period 1 to 3
N	11	10	13		
Median MSS	77.8	80.7	83.8	0.23	0.003
IQR	64.7-82.0	75.0-83.7	83.0-87.8		

DISCUSSION

This systematic review assessed the impact of the publication of the STROBE statement on quality of study design and reporting of methodology. It showed that, after publication of STROBE, a large proportion of the STROBE items and sub-criteria continue to be underreported in CKD cohort studies of mortality in elderly patients. Reporting rates were lowest for hypothesis specification, usage of flow diagrams and addressing missing data. There was evidence of improvement in the reporting quality of CKD cohort studies particularly in the latter three years of the post-STROBE period, which was also seen when looking at the temporal patterns but this may have occurred by chance. We found no evidence that the quality of study design as assessed by three different tools NOS, SIGN and CASP had improved. However, these quality assessment tools have poor to moderate inter-rater reliability and might not be suitable for use without consensus agreement between raters.

Figure 2: Time series of Manuscript STROBE scores (MSS) from spline linear regression models.



The publication of CKD guidelines in 2002 has potentially had an impact on the volume of CKD research with approximately 2.5 times the number of studies in the post-STROBE period compared to the pre-STROBE period (20).

Inadequate reporting not only hinders critical assessment by others of the strengths and weaknesses in study design, conduct, and analysis, it affects judgement of whether and how results can be included in systematic reviews and also impacts on the reader assessment of the studies generalizability (29). Our results are consistent with other studies assessing deficiencies in reporting of individual STROBE items such as sample size, use of flow diagram and reporting of missing data (21, 30–35).

A number of studies, including a Cochrane review, have demonstrated improvements in reporting quality of randomised control trials (RCTs) after the introduction of the Consolidated Standards of Reporting Trials (CONSORT) statement with a significant improvement in journals endorsing this guideline statement (36–40). An RCT has also shown that using reporting guidelines in the peer review process improves the quality of manuscripts (41). Our study showed weak evidence of improvement in the quality of reporting of CKD cohort studies over time following the introduction of the STROBE statement. The improvements unfortunately fell short of the intended expectations when compared to the impact the CONSORT statement had achieved upon the reporting quality of RCTs. These results were similar to the only other study looking at quality of reporting, published in the dermatology literature. Those authors attributed the lack of improvement

to the short follow up period after STROBE introduction (2008–10) (30). However, in our study the small improvement could be attributable to the fact that the reporting of nephrology literature in the pre-STROBE period was already of a higher standard (median MSS 77.8, IQR 64.7–82.0) in comparison to dermatology literature (median score 58, IQR 46–73).

Journal endorsement of reporting guidelines has been shown to improve reporting quality of manuscripts submitted to journals (41). Given that only two medical journals (British Medical Journal & Ageing) and one renal journal (American Journal of Kidney Diseases) included in this review had endorsed the STROBE statement, any evidence of improvement in reporting quality of cohort studies in nephrology literature is probably attributable to the penetration of STROBE statement over time rather than to its endorsement by journals (42). The lack of improvement of reporting standards seen in the STROBE endorsing journals is not an indictment of these journals but maybe attributable to the small sample size to accurately test for differences between the groups. An important observation that was made during the process of this review was that despite studies having similar reporting standards, reflected by their similar MSS, some studies had failed to adequately report essential criteria.

For most of the articles included in this study reporting and methodological quality were well correlated, however the assessment of the methodological quality of a study is largely dependent on adequate reporting of the research. Therefore, drawing any inferences about a study's design quality is made harder if the reporting quality is inadequate.

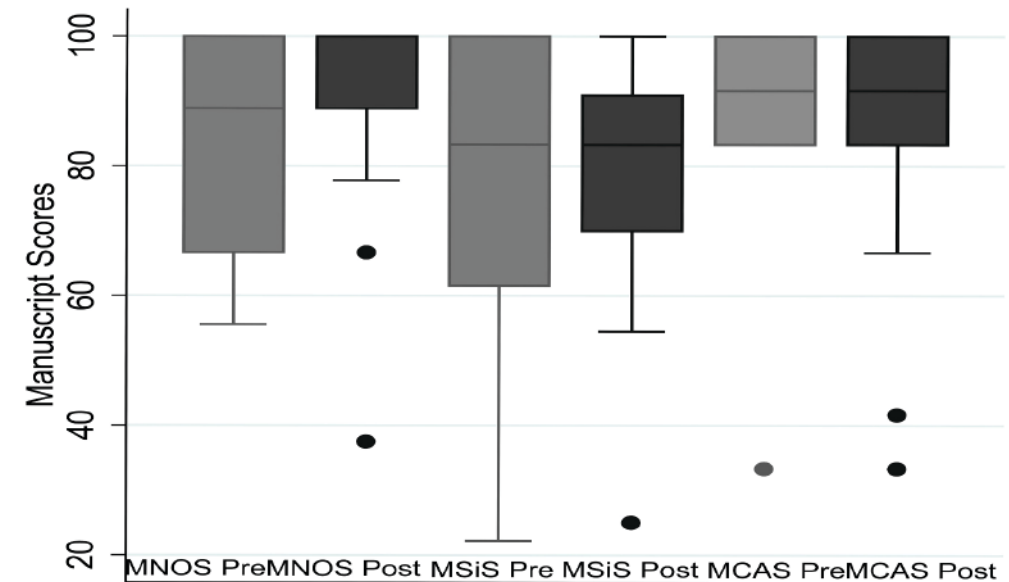
One of the main goals of reporting guidelines was to improve reporting clarity and not necessarily improve the quality of research, but in due course achieve it as an indirect effect. Due to interchangeable usage of the terminology 'reporting quality' and 'methodological quality', the STROBE statement has often been used inappropriately for the assessment of methodological quality of observational research (16). There are a number of assessment tools that have been developed to assess quality and susceptibility to bias in observational studies with only half of the identified tools have described their development or validity and reliability (26). The review by Sanderson et al. highlighted the lack of a single obvious tool for assessing quality of observational epidemiological studies (26). The bias assessment tools used in this study (NOS, SIGN and CASP) were subjective, differed by content, format and validity. The bias assessment tools identified deficiencies in the articles relating to consideration of participant's lost to follow up (attrition bias), exposure level or prognostic factor measured only once (detection bias), and inadequate methods of outcome assessment (detection bias). However, given that the assessment of methodological quality is largely reliant on the reporting of study design, one might therefore fail to detect differences in design quality if reporting is inadequate.

Also given the latency period of designing a new study, undertaking it and then publishing it, might have been simply too soon for the STROBE statement to have influenced the methodological quality of studies. The NOS tool was previously recommended by Cochrane for evaluating the risk of bias but published literature has demonstrated poor inter-rater reliability between individual reviewers (43, 44). The results of our study are consistent with these findings as all of the three tools (NOS, SIGN and CASP) showed poor agreement between individual reviewers. The usability of a tool depends on its clarity. Moreover, the tools contain items whose scoring is subjective and dependent on reviewers' perceptions and domain knowledge. Cochrane now recommends the ACROBAT-NRSI bias assessment tool for non-randomized studies which has been developed by members of the Cochrane Bias Methods Group and the Cochrane Non-Randomised Studies Methods Group (45, 46). However, at the time of drafting this manuscript, this tool remains yet to be tested for consistency between individual reviewers.

A strategy to improve inter-rater agreement would be tailoring and training of reviewers prior to implementation of the tools. Due to the poor reliability of the tools demonstrated here, it should be strongly considered that each study should be assessed by at least two reviewers prior to inclusion in a systematic review/meta-analysis.

One of the strengths of our study is that we studied the impact of STROBE upon both quality of reporting and study design. The study has good internal validity as the selection and evaluation processes were independently performed by two reviewers. However, as the articles were included from one field of medicine (CKD) we must be cautious in generalising our findings to other areas. The other limitation of the study was that it only covered articles from Europe and North America. There was also an imbalance in the number of studies assessed in the two periods probably due to the KDIGO CKD guidelines which were published in 2002. This imbalance could have potentially introduced a lack of power to detect difference in quality. It was also impossible to blind the reviewers to the publication date during the sifting stage of the review, and the journal name during the review of quality which could have biased the reviewers' assessment of quality of the study. Finally, whilst we examined a five-year period post-STROBE, it is possible that we failed to find any benefit for methodological quality due to the long latency period between designing a new study, obtaining funding, undertaking data collection, analysis and publication.

Figure 3: Box plot summarising methodological quality of the studies in the Pre- and Post-STROBE period as assessed using the NOS, CASP and SIGN.



CONCLUSION

This study highlights continuing deficiencies in the reporting of observational studies in the nephrology literature. However, the publication of the STROBE statement may have positively influenced the quality of some aspects of observational study reporting. There was no evidence, however, that methodological quality improved over this time period. With continued efforts from researchers and with particular focus on the domains identified as deficient by the STROBE statement and bias reporting tools, this presents an opportunity to improve the validity of observational research in nephrology. With increased awareness by authors and editors regarding compliance of manuscripts to the STROBE statement and journal endorsement of the STROBE statement, we hope that not only reporting but also the design of future studies will be improved.

Acknowledgments

We thank Ingeborg M. Nagel, medical information specialist and clinical librarian at Academic Medical Center, University of Amsterdam, Netherlands, for valuable input into search strategy for the systematic review

APPENDIX

Item A: Search strategy for systematic review.

Chronic Kidney Disease (CKD): exp renal insufficiency, chronic/; exp Cardiovascular Diseases/ep [Epidemiology]; kidney disease*.ti,ab.; renal disease*.ti,ab.; kidney insufficienc*.ti,ab.; renal insufficienc*.ti,ab.; kidney failure .ti,ab.; kidney dysfunction*.ti,ab.; renal dysfunction*.ti,ab.; kidney impairment*.ti,ab.; renal impairment*.ti,ab.; impaired kidney function*.ti,ab.; impaired renal function*.ti,ab.; decreased kidney function*.ti,ab.; decreased renal function*.ti,ab.; chronic kidney.ti,ab.; chronic renal.ti,ab.; CKD.ti,ab.; CRD.ti,ab.; ESRD.ti,ab.; ESKD.ti,ab.; CKF.ti,ab.

Cohort: exp cohort studies/; cohort*.pt,ti,ab.;Longitudinal*.pt,ti,ab.; Follow-up*.pt,ti,ab.; Follow up*.pt,ti,ab.; Prospective* .pt,ti,ab.; Retrospective* .pt,ti,ab.; Observational* .pt,ti,ab.

Elderly: exp Aged/; elder* .ti,ab.; Older .ti,ab.; Oldest .ti,ab.; old age .ti,ab.; old people .ti,ab.; Geriatric .ti,ab.; Aging .ti,ab.; Ageing .ti,ab.; Frail .ti,ab.; community dwelling .ti,ab.; nursing home .ti,ab.; home for the aged .ti,ab.; homes for the aged .ti,ab.; Residents .ti,ab.

Mortality: mortality/ or *"cause of death"/ or *fatal outcome/ or *hospital mortality/ or *mortality, premature/ or *survival rate/; exp Renal Insufficiency, Chronic/mo [Mortality]

Europe: 1. exp Europe/; 2. UK.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]; 3. United Kingdom.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]; 4. England.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]; 5. English.mp.; 6. (Austria or Austrian).mp.; 7. Belgium*.mp.; 8. Albania*.mp.; 9. Estonia*.mp.; 10. Latvia*.mp.; 11. Lithuania*.mp.; 12. Bosnia*-Herzegovina*.mp.; 13. Bulgaria*.mp.; 14. Croat*.mp.; 15. Hungarian*.mp.; 16. Hungary.mp.; 17. Macedonia*.mp.; 18. (Moldova or Moldavian* or Moldovan).mp.; 19. Montenegro*.mp.; 20. (Poland or Polish or Pole*).mp.; 21. Belarus*.mp.; 22. Romania*.mp.; 23. Russia*.mp.; 24. Serbia*.mp.; 25. Slovak*.mp.; 26. Sloven*.mp.; 27. Ukrain*.mp.; 28. Yugoslavia*.mp.; 29. (Finland or Finns or Finn or Finnish).mp.; 30. (France or French).mp.; 31. German*.mp.; 32. (Great Britain or Great British).mp.; 33. (Northern Ireland or Northern Irish).mp.; 34. (Greece or Greek).mp.; 35. Iceland*.mp.; 36. (Ireland or Irish).mp. 37. (Italy or Italian*).mp.; 38. Luxembourg*.mp.; 39. (Netherlands or Dutch).mp.; 40. (Portugal or Portuguese).mp.; 41. (Denmark or Danish).mp.; 42. (Norway or Norwegian*).mp.; 43. (Sweden or Swedish).mp.; 44. (Spain or Spanish).mp.; 45. (Switzerland or Swiss).mp.; 46. Armenia*.mp.; 47. Azerbaijan*.mp.; 48. Georgia*.mp.; 49. (Turkey or Turkish).mp.; 50. (Malta or Maltese).mp.; 51. Ulster.mp.; 52. Belgian*.mp.; 53. Andorra*.mp.; 54. (Cyprus or Cypriot*).mp.; 55. Czech*.mp.; 56. (Kazakhstan or Kazakh*).mp.; 57. Liechtenstein*.mp.; 58. (Monaco or Monacian* or Monegasue*).mp.; 59. (San Marino or Sammarinese*).mp.; 60. (Vatican or Vanticanian*).mp.

America: 1. exp North America/; 2. Canada.mp.; 3. (US or USA).mp.; 4. United States.mp.; 5. America*.mp.; 6. (Alabama or Alaska or Arizona or Arkansas or California or Colorado or Connecticut or Delaware or Florida or Georgia or Hawaii or Idaho or Illinois or Indiana or Iowa or Kansas or Kentucky or Louisiana or Maine or Maryland or Massachusetts or Michigan or Minnesota or Mississippi or Missouri or Montana or Nebraska or Nevada or New Hampshire or New Jersey or New Mexico or New York or North Carolina or North Dakota or Ohio or Oklahoma or Oregon or Pennsylvania or Rhode Island or South Carolina or South Dakota or Tennessee or Texas or Utah or Vermont or Virginia or Washington or West Virginia or Wisconsin or Wyoming).ti,ab.

Limits: English language

PY(Publication Year): Time period 1/1/2002 to 31/12/2013

PT (Publication Type): Deselect conference papers and other publication types which will not use STROBE guidance for reporting (e. letter to the editor) (autobiography or biography or case reports or clinical conference or clinical trial, all or clinical trial, phase i or clinical trial, phase ii or clinical trial, phase iii or clinical trial, phase iv or clinical trial or congresses or consensus development conference or consensus development conference, nih or controlled clinical trial or dictionary or directory or editorial or in vitro or interactive tutorial or interview or lectures or legal cases or legislation or letter or news or newspaper article or patient education handout or portraits or randomized controlled trial or video-audio media or webcasts)

Selection of human studies: Animals/ not humans/; Not previous

Item B: List of articles included for review by date of publication.

1. Fried LF, Shlipak MG, Crump C, et al. Renal insufficiency as a predictor of cardiovascular outcomes and mortality in elderly individuals. *Journal of the American College of Cardiology*. 2003;41:1364-1372.
2. Drey N, Roderick P, Mullee M, Rogerson M. A population-based study of the incidence and outcomes of diagnosed chronic kidney disease. *American Journal of Kidney Diseases*. 2003;42:677-684.
3. Foley RN, Murray AM, Li S, et al. Chronic Kidney Disease and the Risk for Cardiovascular Disease, Renal Replacement, and Death in the United States Medicare Population, 1998 to 1999. *Journal of the American Society of Nephrology*. 2005;16:489-495.
4. Shlipak MG, Fried LF, Cushman M, et al. Cardiovascular mortality risk in chronic kidney disease: Comparison of traditional and novel risk factors. *JAMA*. 2005;293:1737-1745.
5. Fried LF, Katz R, Sarnak MJ, et al. Kidney Function as a Predictor of Non-cardiovascular Mortality. *Journal of the American Society of Nephrology*. 2005; 16:3728-3735.
6. O'Hare AM, Bertenthal D, Covinsky KE, et al. Mortality Risk Stratification in Chronic Kidney Disease: One Size for All Ages? *Journal of the American Society of Nephrology*. 2006;17:846-853.
7. Hallan SI, Dahl K, Oien CM, et al. Screening strategies for CKD in the general

- population: follow-up of cross sectional health survey. *BMJ*. 2006; 333:1047.
8. Wong CF, McCarthy M, Howse MLP, Williams PS. Factors Affecting Survival in Advanced Chronic Kidney Disease Patients Who Choose Not to Receive Dialysis. *Renal Failure*. 2007;29:653-659.
 9. O'Hare AM, Choi AI, Bertenthal D, et al. Age Affects Outcomes in Chronic Kidney Disease. *Journal of the American Society of Nephrology*. 2007;18:2758-2765.
 10. De Biase V, Tobaldini O, Boaretti C, et al. Prolonged conservative treatment for frail elderly patients with end-stage renal disease: the Verona experience. *Nephrology Dialysis Transplantation*. 2008;23:1313-1317.
 11. Hallan S, Astor B, Romundstad S, Aasarød K, Kvenild K, Coresh J. Association of kidney function and albuminuria with cardiovascular mortality in older vs younger individuals: The hunt ii study. *Archives of Internal Medicine*. 2007;167:2490-2496.
 12. Pizzarelli F, Lauretani F, Bandinelli S, et al. Predictivity of survival according to different equations for estimating renal function in community-dwelling elderly subjects. *Nephrology Dialysis Transplantation*. 2009;24:1197-1205.
 13. Nitsch D, Mylne A, Roderick PJ, Smeeth L, Hubbard R, Fletcher A. Chronic kidney disease and hip fracture-related mortality in older people in the UK. *Nephrology Dialysis Transplantation*. 2009;24:1539-1544.
 14. Snyder JJ, Collins AJ. Association of Preventive Health Care with Atherosclerotic Heart Disease and Mortality in CKD. *Journal of the American Society of Nephrology*. 2009;20:1614-1622.
 15. Roderick PJ, Atkins RJ, Smeeth L, et al. CKD and Mortality Risk in Older People: A Community-Based Population Study in the United Kingdom. *American Journal of Kidney Diseases*. 2009;53:950-960.
 16. Seyfarth M, Kastrati A, Mann JFE, et al. Prognostic Value of Kidney Function in Patients With ST-Elevation and Non-ST-Elevation Acute Myocardial Infarction Treated With Percutaneous Coronary Intervention. *American Journal of Kidney Diseases*. 2009;54:830-839.
 17. Carson RC, Juszczak M, Davenport A, Burns A. Is Maximum Conservative Management an Equivalent Treatment Option to Dialysis for Elderly Patients with Significant Comorbid Disease? *Clinical Journal of the American Society of Nephrology*. 2009;4:1611-1619.
 18. El-Ghoul B, Elie C, Sqalli T, et al. Nonprogressive Kidney Dysfunction and Outcomes in Older Adults with Chronic Kidney Disease. *Journal of the American Geriatrics Society*. 2009;57:2217-2223.
 19. Demoulin N, Beguin C, Labriola L, Jadoul M. Preparing renal replacement therapy in stage 4 CKD patients referred to nephrologists: a difficult balance between futility and insufficiency. A cohort study of 386 patients followed in Brussels. *Nephrology Dialysis Transplantation*. 2011;26:220-226.
 20. Pilotto A SD, Franceschi M, Aucella F, D'Ambrosio P, Scarcelli C, Ferrucci L. A multidimensional approach to the geriatric patient with chronic kidney disease. *Journal of Nephrology*. 2010;23:5-10.
 21. Dalrymple L, Katz R, Kestenbaum B, et al. Chronic Kidney Disease and the Risk of End-Stage Renal Disease versus Death. *J GEN INTERN MED*. 2011;26:379-385.
 22. Andrea Corsonello CP, Fabrizia Lattanzio, Sabrina Garasto, Francesco Corica, Silvia Bustacchini, Enrico E. Guffanti, Angela M. Abbatecola, Vincenzo Mari, Filippo L. Fimognari, and Raffaele Antonelli Incalzi. . Does Concealed Chronic Kidney Disease Predict Survival of Older Patients Discharged from Acute Care Hospitals? *Rejuvenation Research*. 2010;13:539-545.
 23. Muntner P, Bowling CB, Gao L, et al. Age-Specific Association of Reduced Estimated Glomerular Filtration Rate and Albuminuria with All-Cause Mortality. *Clinical Journal of the American Society of Nephrology*. 2011;6:2200-2207.
 24. Heras M F-RM, Sánchez R, Guerrero MT, Molina A, Rodríguez MA, Alvarez-Ude F. Elderly patients with chronic kidney disease: What happens after five years of follow-up? *Nefrologia*. 2012;32:300-305.
 25. Drion I, van Hateren KJJ, Joosten H, et al. CKD and mortality risk among older patients with type 2 DM (ZODIAC-24). *Age and Ageing*. 2012;41:345-350.
 26. Marks A, Black C, Fluck N, et al. Translating CKD epidemiology into patient care—the individual/public health risk paradox. *Nephrology Dialysis Transplantation*. 2012;27:iii65-iii72.
 27. Ahmed A, Fonarow GC, Zhang Y, et al. Renin-Angiotensin Inhibition in Systolic Heart Failure and Chronic Kidney Disease. *The American Journal of Medicine*. 2012;125:399-410.
 28. Binder EF, White HK, Resnick B, McClellan WM, Lei L, Ouslander JG. A Prospective Study of Outcomes of Nursing Home Residents with Chronic Kidney Disease with and without Anemia. *Journal of the American Geriatrics Society*. 2012;60:877-883.
 29. Navaneethan SD, Schold JD, Arrigain S, et al. Serum triglycerides and risk for death in Stage 3 and Stage 4 chronic kidney disease. *Nephrology Dialysis Transplantation*. 2012;27:3228-3234.
 30. Hemmelgarn BR, James MT, Manns BJ, et al. Rates of treated and untreated kidney failure in older vs younger adults. *JAMA*. 2012;307:2507-2515.
 31. Dobre M, Brateanu A, Rashidi A, Rahman M. Electrocardiogram Abnormalities and Cardiovascular Mortality in Elderly Patients with CKD. *Clinical Journal of the American Society of Nephrology*. 2012;7:949-956.
 32. Shastri S, Katz R, Rifkin DE, et al. Kidney Function and Mortality in Octogenarians: Cardiovascular Health Study All Stars. *Journal of the American Geriatrics Society*. 2012;60:1201-1207.
 33. Marks A, MacLeod C, McAteer A, et al. Chronic kidney disease, a useful trigger for proactive primary care? Mortality results from a large UK cohort. *Family Practice*. 2013;30:282-289.
 34. Ahmed A, Rich MW, Zile M, et al. Renin-Angiotensin Inhibition in Diastolic Heart Failure and Chronic Kidney Disease. *The American Journal of Medicine*. 2013;126:150-161.
 35. Westerberg P-A, Tivesten Å, Karlsson MK, et al. Fibroblast growth factor 23, mineral metabolism and mortality among elderly men (Swedish MrOs). *BMC nephrology*. 2013;14:85-85.
 36. Faller B, Beuscart J-B, Frimat L. Competing-risk analysis of death and dialysis initiation among elderly (≥80 years) newly referred to nephrologists: a French prospective study. *BMC Nephrology*. 2013;14:103-103.
 37. Huang X, Jiménez-Moleón JJ, Lindholm B, et al. Mediterranean Diet, Kidney Function, and Mortality in Men with CKD. *Clinical Journal of the American Society of Nephrology*. 2013;8:1548-1555.

Appendix Table A: STROBE scoring sheet.

Item No	Data Items	Rules/Explanatory notes	Answer choices			
			yes	partly	not	un-clear n/a
Title and Abstract						
1A	Is the design described adequately in the title or abstract?	If the study design was not specifically stated, this should be recorded as not being complete.				
1B	Does the abstract provide an informative summary of what was done and found?	The abstract provides key information that enables readers to understand a study and decide whether to read the article and should only present information that is provided in the article.				
Introduction						
2	Is the scientific background and rationale for the investigation reported?	Paper should give an overview of what is known on a topic and what gaps in current knowledge are addressed by the study.				
3A	Are any pre specified hypotheses reported?	Objectives are the detailed aims of the study. Well-crafted objectives specify populations, exposures and outcomes, and parameters that will be estimated.				
3B	Are the objectives reported?					
Methods						
4	Are the key elements (i.e. retrospective /prospective, cohort/cross-sectional) of the study design presented early in the paper?	For example, authors should indicate that the study was a cohort study, which followed people over a particular time period, and describe the group of persons that comprised the cohort and their exposure status. Authors should refrain from simply calling a study 'prospective' or 'retrospective' because these terms are ill defined.				

Appendix Table A: (continued).

5A	Are the settings reported?	Studies that did not report the setting or locations but referred readers to a previous publication should be considered as inconsistent with complete reporting. Readers need information on setting and locations to assess the context and generalizability of a study's results.				
5B	Are the locations reported?					
5C	Are relevant dates including periods of recruitment reported?	Authors should state dates rather than only describing the length of time periods. If the dates of recruitment were recorded anywhere in the article (not necessarily in the "Methods" section), the corresponding item should be rated as complete.				
5D	Are relevant dates including periods of exposure reported?					
5E	Are relevant dates including periods of follow-up reported?					
5F	Are relevant dates including periods of data collection reported?					
6A	Are the eligibility criteria for participants described?	Detailed descriptions of the study participants help readers understand the applicability of the results. Clinical, demographic and other characteristics of eligible participants should be described. Eligibility criteria may be presented as inclusion and exclusion criteria.				
6B	Are the sources of participants described?					
6C	Are the methods of selection described?					
6D	Are the methods of follow-up described?	Knowing details about follow-up procedures, including whether procedures minimized nonresponse and loss to follow-up and whether the procedures were similar for all participants, informs judgments about the validity of results.				
6E	If it is a matched study, are the matching criteria and the numbers of exposed and unexposed described?	Because matching can be done in various ways, with one or more controls per case, the rationale for the choice of matching variables and the details of the method used should be described. To allow readers to judge whether the matched design was appropriately taken into account in the analysis,				

Appendix Table A: (continued).

Item No	Data Items	Rules/Explanatory notes	Answer choices			
			yes	partly	not	un-clear n/a
6E	continued	we recommend that authors describe in detail what statistical methods were used to analyse the data.				
7A	Are all outcomes described if applicable?	Disease outcomes require adequately detailed description of the diagnostic criteria.				
7B	Are all exposures described if applicable?					
7C	Are all predictors described if applicable?					
7D	Are potential confounders described?	If the confounders were recorded anywhere in the article (not necessarily in the "Methods" section), the corresponding item should be rated as complete.				
7E	Are all effect modifiers described?	Authors should declare all 'candidate variables' considered for statistical analysis, rather than selectively reporting only those included in the final models.				
7F	Are diagnostic criteria described if applicable?					
8A	or outcomes can make it more difficult to detect cause-effect relationships, or may produce spurious relationships.	The way in which exposures, confounders and outcomes were measured affects the reliability and validity of a study. Measurement error and misclassification of exposures or outcomes can make it more difficult to detect cause-effect relationships, or may produce spurious relationships.				
8B	If there is more than 1 group, are the measurement methods comparable?					
9	Was there any effort to address potential sources of bias?	Authors should have made attempts to address sources of bias, if they incorporated any tools to do this, eg, using standardized definitions or validated scoring systems should be rated as complete. Addressing sources of bias should never be considered "not applicable" in an observational study.				

Appendix Table A: (continued).

10	Did they describe how the study size was determined?	The importance of sample size determination in observational studies depends on the context. Investigators should report pertinent formal sample size calculations if they were done.				
11A	Did they describe how quantitative variables were handled in the analysis?	Authors should explain why and how they grouped quantitative data, including the number of categories, the cut-points, and category mean or median values.				
11B	Did they describe which groupings were chosen for quantitative variables?	Whenever data are reported in tabular form, the counts of cases, controls, persons at risk, person-time at risk, etc. should be given for each category. Tables should not consist solely of effect-measure estimates or results of model fitting.				
11C	Did they describe why quantitative groups were chosen?					
12A	Did they describe all statistical methods including those to deal with confounding?	In relation to statistical methods, unless authors state which confounders were adjusted for and why, should not be rated as complete. Authors should clarify reasons for particular analyses.				
12B	Did they describe methods to examine subgroups and interactions?	Readers need to know which subgroup analyses were planned in advance, and which arose while analysing the data.				
12C	Did they explain how missing data was addressed?	Authors should report the number of missing values for each variable of interest (exposures, outcomes, confounders) and for each step in the analysis. Authors should give reasons for missing values if possible, and indicate how many individuals were excluded because of missing data when describing the flow of participants through the study.				
12D	Did they explain if applicable how losses to follow-up were addressed?	Authors to report how many patients were lost to follow-up and what censoring strategies they used.				

Appendix Table A: (continued).

Item No	Data Items	Rules/Explanatory notes	Answer choices			
			yes	partly	not	un-clear n/a
12E	Did they describe any sensitivity analysis?	Sensitivity analyses are useful to investigate whether or not the main results are consistent with those obtained with alternative analysis strategies or assumptions.				
Results						
13A	Did they report the numbers of individuals at each stage of the study numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, and completed follow-up and were analysed?	Those included in a study often differ in relevant ways from the target population to which results are applied. This may result in estimates of prevalence or incidence that do not reflect the experience of the target population.				
13B	Did they give reasons for non-participation at each stage? ^a	Explaining the reasons why people no longer participated in a study or why they were excluded from statistical analyses helps readers judge whether the study population was representative of the target population and whether bias was possibly introduced.				
13C	Did they use a flow diagram if appropriate?					
14a	Did they give the characteristics of study participants (eg, demographic, clinical, and social) and information on exposures and potential confounders?	Readers need descriptions of study participants and their exposures to judge the generalizability of the findings. Information about potential confounders, including whether and how they were measured, influences judgments about study validity. Authors should give the mean and Standard deviation, or when the data have an asymmetrical distribution, as is often the case, the median and percentile range (eg, 25 th and 75 th percentiles).				

Appendix Table A: (continued).

14B	Did they indicate the number of participants with missing data for each variable of interest?	As missing data may bias or affect generalizability of results, authors should tell readers amounts of missing data for exposures, potential confounders, and other important characteristics of patients. Should also include the extent of loss to follow-up.				
14C	Did they summarize follow-up time (average and total amount)?	Readers need to know the duration and extent of follow-up for the available outcome data.				
15A	Did they report numbers of outcome measures over time?	Authors should report the numbers of events for each outcome of interest. Consider reporting the event rate per person-year of follow-up. If the risk of an event changes over follow-up time, present the numbers and rates of events in appropriate intervals of follow-up or as a Kaplan-Meier life table or plot.				
15B	Did they report summary measures over time?					
16A	Did they give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval)?	Readers can compare unadjusted measures of association with those adjusted for potential confounders and judge by how much, and in what direction, they changed.				
16B	Did they detail which confounders were adjusted for and why they were included?	Authors should explain all potential confounders considered, and the criteria for excluding or including variables in statistical models. Decisions about excluding or including variables should be guided by knowledge, or explicit assumptions, on causal relations.				
16C	Did they report category boundaries when continuous variables were categorized?	Authors should report the category boundaries; and report the range of the data and the mean or median values within categories.				
16D	Did they, if relevant, consider translating estimates of relative risk into absolute risk for a meaningful period?	It was only appropriate to translate relative risk into absolute risk if there was convincing evidence of a causal association.				

^a For studies based on disease registries/databases, a number of the checklist items are not applicable, eg, dates of recruitment, number eligible at each stage of the study, reasons for nonparticipation, or flow diagrams.

Appendix Table A: (continued).

Item No	Data Items	Rules/Explanatory notes	Answer choices			
			yes	partly	not	un-clear n/a
17A	Did they report on other analyses done, eg, analysis of subgroups or interactions?	Authors should report which analyses were planned, and which were not. This will allow readers to judge the implications of multiplicity, taking into account the study's position on the continuum from discovery to verification or refutation.				
17B	Authors should report which analyses were planned, and which were not. This will allow readers to judge the implications of multiplicity, taking into account the study's position on the continuum from discovery to verification or refutation.	Sensitivity analyses are helpful to investigate the influence of choices made in the statistical analysis, or to investigate the robustness of the findings to missing data or possible biases.				
Discussion						
18	Did they summarize key results with reference to study objectives?	The short summary reminds readers of the main findings and may help them assess whether the subsequent interpretation and implications offered by the authors are supported by the findings.				
19	Did they discuss the limitations of the study taking into account potential sources of bias or imprecision (including discussion of the magnitude of any potential sources of bias)?	Authors should identify the sources of bias and confounding that could have affected results, but also to discuss the relative importance of different biases, including the likely direction and magnitude of any potential bias. Authors may compare the study being presented with other studies in the literature in terms of validity, generalizability and precision. In this approach, each study can be viewed as contribution to the literature, not as a stand-alone basis for inference and action.				

Appendix Table A: (continued).

20	Did they give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence?	Authors should consider potential sources of bias, including loss to follow-up and non-participation. Due consideration should be given to confounding, the results of relevant sensitivity analyses, and to the issue of multiplicity and subgroup analyses. Authors should also consider residual confounding due to unmeasured variables or imprecise measurement of confounders. Authors should put their results in context with similar studies and explain how the new study affects the existing body of evidence, by referring to a systematic review.				
21	Did they discuss the generalizability (external validity) of the study results?	Can results be applied to an individual, groups or populations that differ from those enrolled in the study with regard to age, sex, ethnicity, severity of disease, and co-morbid conditions? Are the nature and level of exposures comparable, and the definitions of outcomes relevant to another setting or population? Are results from health services research in one country applicable to health systems in other countries?				
Other Information						
22A	Did they give the source of the funding in the present study and, if applicable, for the original study on which the present article is based?					
22B	Did they give the role of the funders in the present study and, if applicable, for the original study on which the present article is based?	Unless the role of the funders was specifically stated this should not be recorded as complete.				

The number and proportion of reported items (“yes” & “partly” responses) and not reported items (all responses except “yes”, “partly” or “not applicable”) will be analysed for each study.

Appendix table B: Summary of simple and weighted Kappa coefficient, measuring agreement between reviewers for the NOS, SIGN and CASP tool.

	Kappa	Agreement	P value
NOS (weighted)			
Reviewer 1 vs Reviewer 2	0.18	Poor	0.12
Reviewer 1 vs Reviewer 3	0.28	Fair	0.004
Reviewer 1 vs Reviewer 4	0.12	Poor	0.02
SIGN (unweighted)			
<i>Does not apply vs Everything else grouped</i>			
Reviewer 1 vs Reviewer 2	-0.02	poor	0.87
Reviewer 1 vs Reviewer 3	0.05	poor	0.51
Reviewer 1 vs Reviewer 4	0.19	poor	<0.001
<i>Can't say vs No/Yes grouped</i>			
Reviewer 1 vs Reviewer 2	-0.04	poor	0.62
Reviewer 1 vs Reviewer 3	0.47	moderate	<0.001
Reviewer 1 vs Reviewer 4	-0.04	poor	0.49
<i>No vs Yes</i>			
Reviewer 1 vs Reviewer 2	0.31	Fair	0.004
Reviewer 1 vs Reviewer 3	0.51	Moderate	<0.001
Reviewer 1 vs Reviewer 4	0.24	Fair	<0.001
CASP (unweighted)			
<i>Can't tell vs No/Yes grouped</i>			
Reviewer 1 vs Reviewer 2	0.90	Very good	0.30
Reviewer 1 vs Reviewer 3	0.72	good	<0.001
Reviewer 1 vs Reviewer 4	0.05	poor	0.05
<i>Yes vs No</i>			
Reviewer 1 vs Reviewer 2	0.69	good	<0.001
Reviewer 1 vs Reviewer 3	-0.06	poor	0.64
Reviewer 1 vs Reviewer 4	-0.07	poor	0.40

REFERENCES

- Mangione F, Dal Canton A. Chronic kidney disease epidemic: myth and reality. *Intern Emerg Med*. 2011; 6(1):69–76. doi: 10.1007/s11739-011-0686-4
- Jha V, Garcia-Garcia G, Iseki K, Li Z, Naicker S, Plattner B, et al. Chronic kidney disease: global dimension and perspectives. *The Lancet*. 382(9888):260–72. doi: 10.1016/S0140-6736(13)60687-X
- Ahmed Abdelhafiz SA, Flint K, El Nahas AM. Is Chronic Kidney Disease in Older People a New Geriatric Giant? *Aging Health*. 2011; 7(5):749–62.
- El Nahas AM, Bello AK. Chronic kidney disease: the global challenge. *The Lancet*. 365(9456):331–40. doi: 10.1016/S0140-6736(05)17789-7
- Kovesdy CP, Kalantar-Zadeh K. Observational Studies Versus Randomized Controlled Trials: Avenues to Causal Inference in Nephrology. *Advances in Chronic Kidney Disease*. 19(1):11–8. doi: 10.1053/j.ackd.2011.09.004
- Novak JE, Inrig JK, Patel UD, Califf RM, Szczech LA. Negative trials in nephrology: what can we learn? *Kidney Int*. 2008; 74(9):1121–7. doi: 10.1038/ki.2008.286
- Jager KJ, Stel VS, Wanner C, Zoccali C, Dekker FW. The valuable contribution of observational studies to nephrology. *Kidney Int*. 2007; 72(6):671–5. <http://www.nature.com/ki/journal/v72/n6/supinfo/5002397s1.html>.
- Concato J, Shah N, Horwitz RJ. Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs. *New England Journal of Medicine*. 2000; 342(25):1887–92. doi: 10.1056/NEJM200006223422507
- Thadhani R, Tonelli M. Cohort Studies: Marching Forward. *Clinical Journal of the American Society of Nephrology*. 2006; 1(5):1117–23. doi: 10.2215/cjn.00080106
- Benson K, Hartz AJ. A Comparison of Observational Studies and Randomized, Controlled Trials. *New England Journal of Medicine*. 2000; 342(25):1878–86. doi: 10.1056/NEJM200006223422506
- Johnston SC, Rootenberg JD, Katrak S, Smith WS, Elkins JS. Effect of a US National Institutes of Health programme of clinical trials on public health and costs. *The Lancet*. 367(9519):1319–27. doi: 10.1016/S0140-6736(06)68578-4
- Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of Reporting of Observational Longitudinal Research. *American Journal of Epidemiology*. 2005; 161(3):280–8. doi: 10.1093/aje/kwi042
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet*. 370(9596):1453–7. doi: 10.1016/S0140-6736(07)61602-x
- Nijsten T, Spuls P, Stern RS. Strobe: A beacon for observational studies. *Archives of Dermatology*. 2008; 144(9):1200–4. doi: 10.1001/archderm.144.9.1200
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *International Journal of Surgery*. 12(12):1495–9. doi: 10.1016/j.ijsu.2014.07.013
- da Costa BR, Cevallos M, Altman DG, Rutjes AWS, Egger M. Uses and misuses of the STROBE statement: bibliographic study. *BMJ Open*. 2011; 1(1). doi: 10.1136/bmjopen-2010-000048
- Carlson MDA, Morrison RS. Study Design, Precision, and Validity in Observational

- Studies. *Journal of Palliative Medicine*. 2009; 12(1):77–82. doi: 10.1089/jpm.2008.9690 PMC2920077.
18. Shamliyan T, Kane RL, Dickinson S. A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. *Journal of Clinical Epidemiology*. 2010; 63(10):1061–70. doi: 10.1016/j.jclinepi.2010.04.014
 19. Jager KJ, Ocak G, Drechsler C, Caskey FJ, Evans M, Postorino M, et al. The EQUAL study: a European study in chronic kidney disease stage 4 patients. *Nephrology Dialysis Transplantation*. 2012. doi:10.1093/ndt/gfs277
 20. National Kidney Foundation. K/DOQI Clinical practice guidelines for chronic kidney disease: evaluation, classification and stratification. *American Journal of Kidney Diseases*. 2002; 39(Suppl 2):S1–S246.
 21. Langan S, Schmitt J, Coenraads P, Svensson A, von Elm E, Williams H, et al. The reporting of observational research studies in dermatology journals: A literature-based study. *Archives of Dermatology*. 2010; 146(5):534–41. doi: 10.1001/archdermatol.2010.87
 22. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality if nonrandomized studies in meta-analyses. Available: http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm. Accessed July 28, 2015.
 23. Higgins JPT, Green S (Eds): *Cochrane handbook for systematic reviews of interventions* version 5.1.0 [Accessed March 2011]. Available: http://handbook.cochrane.org/chapter_13/13_5_2_3_tools_for_assessing_methodological_quality_or_risk_of.htm. Accessed July 28, 2015.
 24. Scottish Intercollegiate Guidelines Network (SIGN) checklist for cohort studies. <http://www.sign.ac.uk/methodology/checklists.html>. Accessed July 28, 2015. Available: <http://www.sign.ac.uk/methodology/checklists.html>.
 25. CASP UK. Critical Appraisal Skills Programme (CASP) checklist for cohort studies. <http://www.caspuk.net/#!casp-tools-checklists/c18f8>. Accessed July 28, 2015. Available: <http://www.casp-uk.net/#!casp-tools-checklists/c18f8>.
 26. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *International Journal of Epidemiology*. 2007; 36(3):666–76. doi: 10.1093/ije/dym018
 27. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 1995; 57(1):289–300.
 28. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics*. 2002; 27(4):299–309. doi: 10.1046/j.1365-2710.2002.00430.x
 29. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ: British Medical Journal*. 2007; 335(7624):806–8. doi: 10.1136/bmj.39335.541782.AD PMC2034723.
 30. Bastuji-Garin S, Sbidian E, Gaudy-Marqueste C, Ferrat E, Roujeau J-C, Richard M-A, et al. Impact of STROBE Statement Publication on Quality of Observational Study Reporting: Interrupted Time Series versus Before-After Analysis. *PLoS ONE*. 2013; 8(8):e64733. doi: 10.1371/journal.pone.0064733
 31. Müller M, Egger M. Strengthening the reporting of observational epidemiology (STROBE) in sexual health. *Sexually Transmitted Infections*. 2009; 85(3):162–4. doi: 10.1136/sti.2007.028985
 32. Poorolajal J, Cheraghi Z, Irani AD, Rezaeian S. Quality of Cohort Studies Reporting Post the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement. *Epidemiol Health*. 2011; 33(0):e2011005–0. doi: 10.4178/epih/e2011005
 33. Fung AE, Palanki R, Bakri SJ, Depperschmidt E, Gibson A. Applying the CONSORT and STROBE Statements to Evaluate the Reporting Quality of Neovascular Age-related Macular Degeneration Studies. *Ophthalmology*. 116(2):286–96.e4. doi: 10.1016/j.ophtha.2008.09.014
 34. Sorensen AA, Wojahn RD, Manske MC, Calfee RP. Using the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement to Assess Reporting of Observational Trials in Hand Surgery. *The Journal of Hand Surgery*. 2013; 38(8):1584–9.e2. doi: 10.1016/j.jhsa.2013.05.008
 35. Papatheanasiou AA, Zintzaras E. Assessing the Quality of Reporting of Observational Studies in Cancer. *Annals of Epidemiology*. 2010; 20(1):67–73. doi: 10.1016/j.annepidem.2009.09.007
 36. Plint Amy C M D, Morrison Andra, Schulz Kenneth, Altman Douglas G, Hill Catherine and Gaboury Isabelle. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust* 2006; 185(5):263–7.
 37. Moher D, Jones A, Lepage L, for the CG. Use of the consort statement and quality of reports of randomized trials: A comparative before-and-after evaluation. *JAMA*. 2001; 285(15):1992–5. doi: 10.1001/jama.285.15.1992
 38. Anna Partsinevelou EZ. Quality of reporting of randomized controlled trials in polycystic ovary syndrome. *Trials*. 2009; 10(106).
 39. Lucy Turner LS, Altman Douglas G, Schulz Kenneth F, Moher David. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Systematic Reviews*. 2012; 1(60).
 40. Kane RL, Wang J, Garrard J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *Journal of Clinical Epidemiology*. 2007; 60(3):241–9. doi: 10.1016/j.jclinepi. 2006.06.016
 41. Cobo E, Cortés J, Ribera JM, Cardellach F, Selva-O'Callaghan A, Kostov B, et al. Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical journal: masked randomised trial 2011-11-22 12:16:35.
 42. STROBE Statement. STROBE Endorsement. Available: <http://www.strobe-statement.org/index.php?id=strobe-endorsement>. Accessed October 29, 2015. Available: <http://www.strobe-statement.org/index.php?id=strobe-endorsement>.
 43. Hartling L, Milne A, Hamm MP, Vandermeer B, Ansari M, Tsertsvadze A, et al. Testing the Newcastle Ottawa Scale showed low reliability between individual

- reviewers. *Journal of Clinical Epidemiology*. 66 (9):982–93. doi: 10.1016/j.jclinepi.2013.03.003
44. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol*. 2010; 25(9):603–5. doi: 10.1007/s10654-010-9491-z
 45. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials. *BMJ*. 2011; 343. doi: 10.1136/bmj.d5928
 46. Sterne JAC HJ, Reeves BC on behalf of the development group for ACROBAT-NRSI. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI), Version 1.0.0. <http://www.riskofbias.info>. Accessed October 30, 2015. 24 September 2014. Available from: <http://www.riskofbias.info>.