REVIEW ARTICLE

NEPHROLOGY · APSN · WILEY

# Evidence-based medicine—When observational studies are better than randomized controlled trials

**Jizzo R. Bosdriesz[1]** | **Vianda S. Stel[1]** | **Merel van Diepen[2]** | **Yvette Meuleman[2]** | **Friedo W. Dekker[2]** | **Carmine Zoccali[3]** | **Kitty J. Jager[1]**

[1]ERA-EDTA Registry, Department of Medical Informatics, Amsterdam Public Health Research Institute, Amsterdam UMC-location AMC, University of Amsterdam, Amsterdam, The Netherlands

[2]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

[3]CNR-IFC, Clinical Epidemiology and Pathophysiology of Renal Diseases and Hypertension, Reggio Calabria, Italy

**Correspondence**
Dr Vianda S. Stel, ERA-EDTA Registry, Department of Medical Informatics, Amsterdam Public Health Research Institute, Amsterdam UMC-location AMC, University of Amsterdam, Room J1b-108, P.O. Box 22700, 1100 DE Amsterdam, The Netherlands.
Email: v.s.stel@amsterdamumc.nl

**Abstract**

In evidence-based medicine, clinical research questions may be addressed by different study designs. This article describes when randomized controlled trials (RCT) are needed and when observational studies are more suitable. According to the Centre for Evidence-Based Medicine, study designs can be divided into analytic and non-analytic (descriptive) study designs. Analytic studies aim to quantify the association of an intervention (eg, treatment) or a naturally occurring exposure with an outcome. They can be subdivided into experimental (ie, RCT) and observational studies. The RCT is the best study design to evaluate the *intended* effect of an intervention, because the randomization procedure breaks the link between the allocation of the intervention and patient prognosis. If the randomization of the intervention or exposure is not possible, one needs to depend on observational analytic studies, but these studies usually suffer from bias and confounding. If the study focuses on *unintended* effects of interventions (ie, effects of an intervention that are not intended or foreseen), observational analytic studies are the most suitable study designs, provided that there is no link between the allocation of the intervention and the unintended effect. Furthermore, non-analytic studies (ie, descriptive studies) also rely on observational study designs. In summary, RCTs and observational study designs are inherently different, and depending on the study aim, they each have their own strengths and weaknesses.

**KEYWORDS**

confounding, epidemiology, methodology, observational research, randomized controlled trial

## 1 | INTRODUCTION

A lead editorial in the British Medical Journal by Dave Sackett defined evidence-based medicine (EBM) as "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients."[1] In EBM, clinical research questions may be addressed by different study designs, each with their own merits and limitations. In the traditional hierarchy of study designs, the randomized controlled trial (RCT) is placed on top, followed by cohort studies, case-control studies, case reports and case series.[2] However, the foremost consideration for the choice of study design should be the research question. For some research questions, an RCT might be

*Jizzo R. Bosdriesz and Vianda S. Stel joint first authors.*

the most suitable design, whereas for other research questions observational study designs are to be preferred.

According to the Centre for Evidence-Based Medicine, clinical studies can be divided into analytic and non-analytic studies.[3] The types of clinical studies (analytic vs non-analytic studies) along with specific study designs and examples of research questions are given in Table 1. An analytic study aims to quantify the causal relationship between an intervention (eg, treatment) or a naturally occurring exposure (hereafter indicated as exposure) (eg, the presence of a disease) and an outcome.[3,4] To quantify the effect, it is necessary to compare the rates of the outcome in the intervention or exposed group, with that in a control group. Within analytic research, a further distinction can be made between experimental and observational analytic studies.[3] In experimental studies, that is, RCTs, the investigator intentionally manipulates the intervention by randomly allocating participants to the intervention or control group.[4] In contrast, in observational analytic studies, the intervention or exposure as well as the control group are simply measured (observed) without manipulation by the researcher. Non-analytic or descriptive studies on the other hand aim to describe what is happening in a population (eg, the incidence or prevalence of a disease), without quantifying a causal relationship, using observational study designs.[3,4]

This article aims to show when RCT are needed, and when observational studies are more suitable than RCT.

**SUMMARY AT A GLANCE**

In evidence-based medicine, clinical research questions may be addressed by different study designs. In this article, we explain that randomized controlled trials and observational study designs are inherently different, and depending on the study aim, they each have their own strengths and weaknesses.

## 2 | EXPERIMENTAL STUDIES

When the aim is to evaluate the *intended* effect of an intervention, the RCT is the gold standard (Figure 1A). An intended effect is the outcome the person who prescribes the intervention intends to achieve. For example, a physician may prescribe a certain drug with the intention to prevent mortality. This is different from *unintended* effects, which will be discussed later in this article.[5] By the strength of the design, the evidence produced by a sufficiently powered RCT is highly convincing in determining the presence or absence of a causal relationship between an intervention and its intended effect on the outcome.[6] In an RCT, randomization is used to allocate participants to the intervention group or the control group (eg, without intervention, with a placebo or an alternative treatment). By randomization, one

**TABLE 1** Types of clinical studies with related study designs and examples

| Types of clinical studies | Study designs | Studied association | Examples |
|---|---|---|---|
| *Analytic* | | | |
| Experimental | Randomized controlled trial | Between an intervention and an intended outcome | What is the effect of benazepril plus amlodipine vs benazepril plus hydrochlorothiazide on the progression of CKD?[10] |
| Observational analytic | Cohort study, case-control study, some cross-sectional studies | Between an intervention and an intended outcome | What is the difference in survival among ESRD patients on dialysis, on a transplant waiting list, and after receiving a renal transplant?[27] |
| | | Between an exposure[a] and outcome | What is the risk of all-cause mortality in each stage of CKD?[17] |
| | | Between an intervention and unintended outcome | What are the effects of ACE/AII inhibitors use vs no use, on peritoneal membrane transport characteristic in long-term PD patients?[21] |
| *Non-analytic* | | | |
| Descriptive | Case reports, case series, cross-sectional studies (surveys) | Not applicable | What is the prevalence of CKD in individual European countries?[25] |
| | | | What are the international time trends in the incidence of RRT for ESRD by PRD from 2005 to 2014?[26] |

Abbreviations: ACE, angiotensin-converting enzyme; AII, angiotensin-II; CKD, chronic kidney disease; CRF, chronic renal failure; ESRD, end-stage renal disease; PD, peritoneal dialysis; PRD, primary renal disease; RRT, renal replacement therapy.
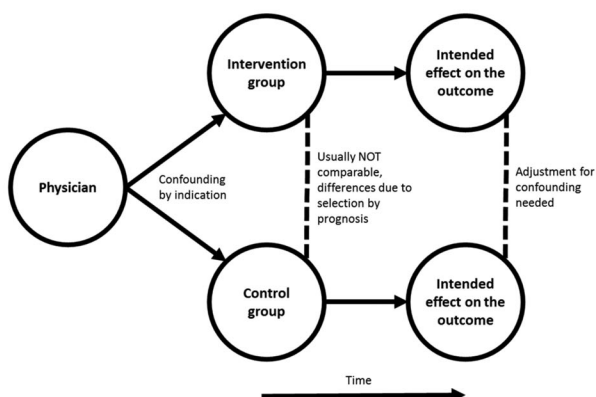[a]Here exposure is defined as a naturally occurring exposure.

(A) **Experimental study (RCT) – the association between an intervention and an intended effect on the outcome**
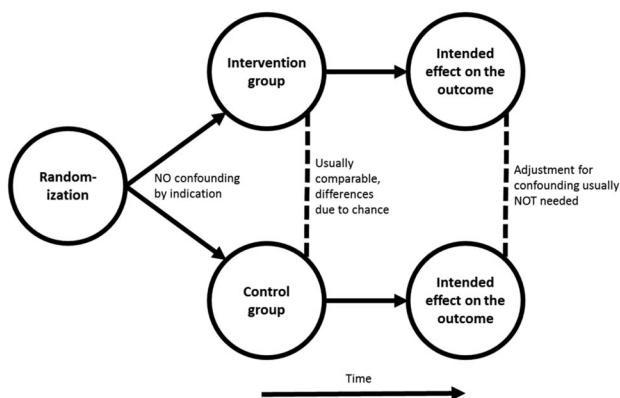
(C) **Observational analytic study – the association between a naturally occurring exposure and an outcome**
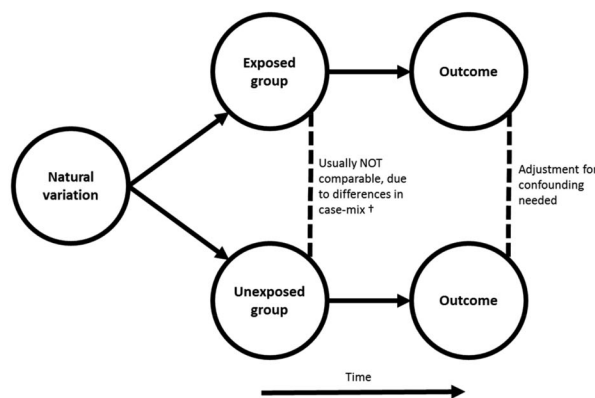
(B) **Observational analytic study – the association between an intervention and an intended effect on the outcome**

(D) **Observational analytic study – the association between an intervention and an unintended effect on the outcome**
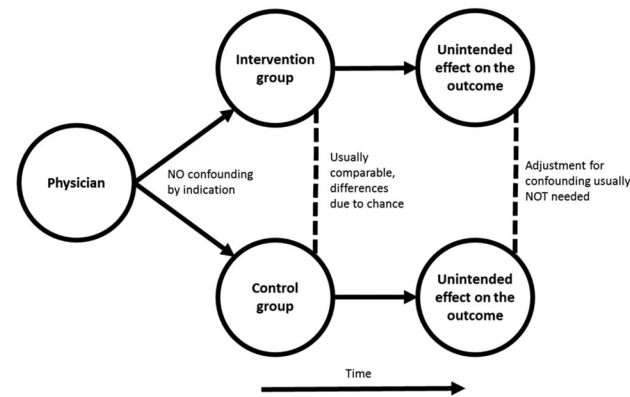
**FIGURE 1** Outline of different analytic studies using A, an experimental study (randomized controlled trial) and B, C, D, an observational analytic study. Case-mix* refers to differences in measured and unmeasured confounders between exposed and unexposed groups

aims to prevent confounding by indication, also known as selection by prognosis.[7] Confounding by indication usually occurs when clinicians decide who will receive the intervention, as their opinion about the patient's prognosis guides their decision on treatment allocation.[8] For example, patients with more severe symptoms usually receive treatment that is more intensive. As a result, the group receiving more intensive treatment may have worse outcomes due to their worse prognosis at baseline.[9] Without randomization, the intervention and the control group will usually be different with respect to their baseline characteristics and their prognosis. Of note, randomization does not guarantee that the intervention and control group will be exactly the same in terms of baseline characteristics. However, randomization does ensure that any remaining differences between the intervention and control groups are determined by chance.[8]

An example of an RCT is the study by Bakris et al in which 11 506 patients with hypertension who were at high risk of cardiovascular events were randomly assigned to receive benazepril plus amlodipine or benazepril plus hydrochlorothiazide.[10] These patients were followed over time to assess the effect of these different anti-

hypertensive treatments (intervention) with respect to slowing the progression of chronic kidney disease (CKD) (intended outcome). The investigators found that the risk of progression was lower using a combination of benazepril plus amlodipine than with benazepril plus hydrochlorothiazide (hazard ratio [HR]: 0.52) (95% confidence interval [95% CI] 0.41-0.65).

## 2.1 | Limitations of RCT

Randomized controlled trials have several important limitations. First, the generalizability of their results is often limited due to sampling bias. This occurs when the study sample or the groups resulting from randomization are not representative of the source population they were drawn from.[7,11] This can be due to strict inclusion criteria. RCTs outside the field of nephrology often routinely exclude CKD patients and therefore the generalizability of their results to patients with CKD may be questionable.[12] Second, RCT are generally expensive to perform,[13] and therefore the study samples are often relatively small and

their follow-up relatively short. As a result, there might be substantial baseline differences remaining in the measured and unmeasured confounders between the two groups after randomization (although these differences are determined by chance). Third, for some research questions on the intended effect of an intervention, an RCT may be impossible, unfeasible or unnecessary. For instance, an RCT might be too costly, it might be unethical to randomize the intervention, or the health benefit of the treatment may be so dramatic that observational studies can demonstrate their effectiveness. These and other reasons are described in more detail elsewhere.[2,14,15]

## 3 | OBSERVATIONAL ANALYTIC STUDIES

In the above-mentioned situations, when performing an RCT investigating the intended effect of an intervention is not possible or not justified, observational study designs, such as cohort studies (Figure 1B) or case-control studies are needed.[6] An observational study design may also be preferred for other reasons, such as the lack of generalizability in an RCT. Observational studies are also necessary to assess the effect of a naturally occurring exposure on an outcome (Figure 1C). Please note that in observational analytic studies investigating the effect of an exposure there are no intended or unintended effects, just outcomes.[16]

In cohort studies, participants are free of the outcome at study entry and are followed over time to assess who will develop the outcome and who will not. Cohort studies tend to be less costly to perform than RCTs, as they usually rely on less invasive and intensive methods of data collection. In addition, usually there are less ethical aspects to consider, since there is no intervention. However, a specific limitation of cohort studies is that they typically need to run for several years, and/or include many participants, in order to observe a sufficient number of occurrences of the (potentially rare) outcome. An example of a large cohort study, running for more than a decade, is the study by Wen et al.[17] They studied a large cohort of 462 293 adults to compare the all-cause mortality risk (outcome) between different stages of CKD (naturally occurring exposure). They found the HR for all-cause mortality to be lower in CKD stage 1 (HR: 1.8, 95% CI: 1.5-2.2), stage 2 (HR: 1.7, 95% CI: 1.5-1.9), and stage 3 (HR: 1.5, 95% CI: 1.4-1.6) compared with CKD stage 4 (HR: 5.3, 95% CI: 4.5-6.2) and stage 5 (HR: 9.1, 95% CI: 7.2-11.4).[17]

Case-control studies are usually more efficient than cohort studies, because instead of selecting individuals on the basis of exposure or intervention, the selection of individuals is based on the outcome. Patients with a certain outcome (cases) are compared with a *subset* of individuals who did not develop the outcome (controls).[6] As a result, the number of included persons in the control group can be limited. The researchers may then use retrospective data to find out to what extent cases and controls were exposed to the exposure or intervention of interest. The main limitations of case-control studies may include the difficulty in selecting an appropriate control group, and recall bias as data are always collected retrospectively. An example of a case-control study is the study by Fored et al on the association between socio-economic status (SES) (naturally occurring exposure) and chronic renal failure (CRF) (outcome).[18] They defined a source population from which they selected those with CRF as cases, and randomly selected a similar number of people without CRF as controls. They found that in families having a low SES, the risk of developing CRF was significantly higher (odds ratio [OR] = 1.6 [95% CI: 1.0-2.6] for men and OR = 2.1 [95% CI: 1.1-4.0] for women) than in families with a high SES.

The main drawback of observational analytic studies is that the intervention or exposure is not randomized, and therefore confounding by indication (in case of an intervention) or differences in case-mix between the exposed and unexposed groups (in case of a naturally occurring exposure) are likely to exist.[6] This means that there are usually differences in measured and unmeasured confounders between the comparison groups.[11] As a result, any observed effect of the intervention or the exposure on the outcome might be due to these baseline differences. A variety of methods exists aiming to address confounding in observational analytic studies. The most commonly used methods are given in Box 1. However, in most cases, residual confounding remains and as a result, it is often not possible to draw firm conclusions about causality from observational analytic studies.

## 4 | STUDY OF UNINTENDED EFFECTS

Unintended effects are all effects on outcomes—harmful, harmless or even beneficial—that are produced by an intervention or treatment, that were not originally intended by the person who prescribed the intervention.[4] In the field of medicine, unintended effects can be side effects, and particularly the side effects of pharmaceutical agents are widely studied. Unintended effects of interventions may be relatively uncommon and often hard to predict, and they are therefore usually not considered in clinicians' decision-making.[8] When clinicians do not know the unintended effects of an intervention, they cannot base their treatment allocation on these effects and therefore confounding by indication usually does not occur.[8] For this reason randomization is not needed, and thus observational analytic studies can be used to quantify the unintended effects of interventions (Figure 1D). Because unintended effects may occur less frequently, these studies may require a large sample size and sometimes extended follow-up.[8] In this perspective, RCTs are also less suitable as they normally include relatively small patient samples with relatively short follow-up times, and have limited generalizability as a result of rigid in- and exclusion criteria. It should be noted; however, that once the occurrence of an unintended effect of a treatment is well known, physicians may take this into account in their treatment allocation, and consequently confounding by indication may occur. Also, in the situation that the treatment allocation by the physician is related to a certain variable (eg, patients with more severe symptoms receive more intensive treatment) and this specific variable is related to the unintended outcome (eg, patients with more severe symptoms may suffer more from unintended effects) confounding by indication will also occur. In this

**BOX 1**    Methods used in observational analytic studies aiming to address confounding

| Method | Description | Potential limitations[a] | References |
|---|---|---|---|
| Restriction | Restricting the study sample by including only participants with equal or more similar values of a measured confounding variable, thereby reducing confounding. The association between exposure and outcome is studied in this restricted group only | Lower generalizability of results, reduced sample size and statistical power, less useful if there are many confounders | 28 |
| Stratification | Participants are divided into groups on the basis of a measured confounding variable. The association between exposure and outcome is studied in each group, and usually a weighted average of the association is calculated for the combined groups | Less useful if there are many confounders | 29,30 |
| Multi-variable regression | The effect of the exposure on the outcome is modelled together with all measured potential confounders (following the criteria for confounding), resulting in an estimated effect size that is adjusted for all confounders | Sample size and number of events determine how many confounders can be included in the model | 28,31 |
| Propensity score matching | The propensity score (PS) is the probability for any individual to be allocated to the treatment condition, calculated (usually by logistic regression) from their baseline characteristics. Thereafter each treated participant can be matched to an untreated participant, who had the same probability of receiving the treatment. Outcomes in both groups are then compared to determine the treatment effect | Finding matches can be difficult, leading to dropping of unmatched cases or accepting less-than-ideal matches | 32-34 |
| Inverse probability weighting | A probability is calculated for each individual, which is then used to weight the observations. These weights are calculated by taking the inverse of the PS (1/PS) for those in the exposed sample, and 1/1-PS for those in the unexposed sample. After weighting, regular statistical tests can be used, usually without further need for adjusting for observed confounders | Biased if the model to estimate weights is not specified correctly | 35,36 |
| Instrumental variables | An instrumental variable (IV) or *instrument* is first identified, this variable should meet three criteria: (a) the IV has a causal effect on the allocation of exposure, (b) the IV has no direct effect on the outcome (only through the actual exposure) and (c) the IV and the outcome do not share common causes. The second step is to estimate the proportion of variance in the allocation of the exposure attributable to the IV. When participants are grouped based on the IV, which is assumed to be pseudo-randomized, their results can be compared to determine the effect of the exposure | Finding a suitable IV can be difficult, potentially non-perfect correlation between IV and the exposure, assumed homogeneity in treatment effect may be violated | 37,38 |

[a]All these methods of addressing confounding in observational analytic studies share a common important limitation, namely residual confounding, as it is not possible to adjust for unknown, unmeasured or incorrectly assessed confounders.[31]

case, an RCT might be a more appropriate study design provided the unintended effect occurs with sufficient frequency.

Once a treatment has been introduced into clinical practice, one can study its unintended effects. To this end, pharmaceutical companies use so-called post-marketing surveillance, which is often mandated by the regulatory agencies, such as the American Food and Drug Administration, the European Medicines Agency and the Australian Therapeutic Goods Administration. To monitor the effectiveness and safety of drugs, and the occurrence of unintended effects, they rely on real-world observational study designs.[19] Most unintended effects, especially those that are adverse and require medical attention, are first brought to attention in case reports.[20] However, some

adverse effects might be very common, relatively benign, or only occur long after the treatment start, and not give rise to case reports, leading to underreporting. Therefore, more systematic surveillance systems that can link data from health-care insurers and government agencies, and observational cohort studies might be employed.[20]

An example of an observational study on unintended effects is the study by Kolesnyk et al, who assessed the unintended effects of using angiotensin-converting enzyme (ACE) and angiotensin-II (AII) inhibitors vs a control group without such medication on the peritoneal membrane function in a cohort of long-term peritoneal dialysis (PD) patients.[21] The investigators found that ACE/AII inhibitors had a protective effect on peritoneal membrane function (P =.037). Because the ACE/AII inhibitors were prescribed with the aim to treat hypertension or heart failure, and not intended to prevent peritoneal membrane damage in PD, this is considered to be an unintended effect, albeit a beneficial one.

## 5 | DESCRIPTIVE STUDIES

Observational study designs are also needed for non-analytic studies (ie, descriptive studies). Such studies examine the frequency of risk factors, diseases or other health outcomes in a population, without assessing causal relationships.[4] Descriptive studies can for example be used to inform health-care professionals and policymakers on the amount of public health resources needed.[22] For descriptive studies, one can use population statistics, or draw a sample from the population. In this article, we mainly focus on cross-sectional descriptive studies and descriptive studies describing time trends. A more detailed description of types of descriptive studies can be found elsewhere.[23]

Cross-sectional descriptive studies can usually be implemented in a relatively short timeframe and with a reasonable budget. As the frequency of a disease may differ between groups and may depend on factors such as age and sex, it is common to standardize the results for these factors using a reference population.[24] An example is the study by Brück et al, which describes the prevalence of CKD in 13 European countries.[25] The CKD prevalence was age- and sex-standardized to the population of the 27 Member States of the European Union to enable comparison between countries. The results of that study suggested substantial international differences in the prevalence of CKD across European countries, varying from 3% to 17% for CKD stages 1 to 5, and from 1% to 6% for CKD stages 3 to 5.[25] Descriptive studies may suffer from sampling bias due to their sample selection methods, which may hamper the generalizability of the results.

Other descriptive studies can describe (long-term) time trends, for instance the incidence or prevalence of a disease. Insight into such trends may allow policymakers, health-care practitioners and researchers to identify newly emerging health threats, to optimize future allocation of resources, initiate preventive efforts and monitor changes over time.[2,22,25] However, even if trends indicate favourable changes following preventive efforts, descriptive studies alone cannot establish causal links between prevention and outcome. Another caveat of studies investigating long-term trends is that they need data collection over several years, preferably using the same methodology. An example of a descriptive study describing time trends is the one by Stel et al that made an international comparison of trends in the incidence of renal replacement therapy (RRT) for patients with end-stage kidney disease (ESKD) by primary renal disease.[26] The incidence of RRT for ESKD due to diabetes mellitus or hypertension was found to strongly increase in Asia between 2005 and 2014, whereas both declined in Europe. Conversely, the incidence of RRT for ESKD due to glomerulonephritis was stable or decreased in all included countries.

## 6 | CONCLUSIONS

The answer to the question "is an observational study better than an RCT?" depends on the research question at hand. There is not one overall gold standard study design for clinical research. With respect to analytic studies, the RCT is the best study design when it comes to evaluating the *intended* effect of an intervention. However, this type of research represents only a fraction of all clinical research. Observational analytic studies are most suitable if randomization of the intervention or exposure is not feasible or if the research question focuses on *unintended* effects of interventions. For non-analytic or descriptive studies, observational study designs are also needed. We conclude that RCT and observational studies are inherently different, and each have their own strengths and weaknesses depending on the study question.

### CONFLICT OF INTEREST
The authors declare no potential conflict of interest.

### ORCID
*Jizzo R. Bosdriesz* https://orcid.org/0000-0002-1038-7814

### REFERENCES
1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312 (7023):71-72.
2. Jager KJ, Stel VS, Wanner C, Zoccali C, Dekker FW. The valuable contribution of observational studies to nephrology. *Kidney Int*. 2007;72 (6):671-675.
3. Medicine TCfE-B. CEBM. Study Designs: University of Oxford; 2014 https://www.cebm.net/2014/04/study-designs/. Accessed April 1, 2020.
4. Porta MS, Greenland S, Hernán M, IdS S, Last JM. In: Porta M, ed. *A Dictionary of Epidemiology*. 6th ed. New York, NY: Oxford University Press; 2014:343.

5. Miettinen OS. The need for randomization in the study of intended effects. *Stat Med*. 1983;2(2):267-271.

6. Noordzij M, Dekker FW, Zoccali C, Jager KJ. Study designs in clinical research. *Nephron Clin Pract*. 2009;113(3):8-21.

7. Jager KJ, Tripepi G, Chesnaye NC, Dekker FW, Zoccali C, Stel VS. Where to look for the most frequent biases? *Nephrology*. 2020;25(6):435-441.

8. Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet*. 2004;363(9422):1728-1731.

9. Kyriacou DN, Lewis RJ. Confounding by indication in clinical research. *JAMA*. 2016;316(17):1818-1819.

10. Bakris GL, Sarafidis PA, Weir MR, et al. Renal outcomes with different fixed-dose combination therapies in patients with hypertension at high risk for cardiovascular events (ACCOMPLISH): a prespecified secondary analysis of a randomised controlled trial. *Lancet*. 2010;375(9721):1173-1181.

11. Tripepi G, Jager KJ, Dekker FW, Wanner C, Zoccali C. Bias in clinical research. *Kidney Int*. 2008;73(2):148-153.

12. Zoccali C, Blankestijn PJ, Bruchfeld A, et al. Children of a lesser god: exclusion of chronic kidney disease patients from clinical trials. *Nephrol Dial Transplant*. 2019;34(7):1112-1114.

13. Claiborne Johnston S, 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. *Lancet*. 2006;367(9519):1319-1327.

14. Grootendorst DC, Jager KJ, Zoccali C, Dekker FW. Observational studies are complementary to randomized controlled trials. *Nephron Clin Pract*. 2010;114(3):173-177.

15. Black N. Why we need observational studies to evaluate the effectiveness of health care. *Brit Med J*. 1996;312(7040):1215-1218.

16. Cross NB, Craig JC, Webster AC. Asking the right question and finding the right answers. *Nephrology*. 2010;15(1):8-11.

17. Wen CP, Cheng TY, Tsai MK, et al. All-cause mortality attributable to chronic kidney disease: a prospective cohort study based on 462 293 adults in Taiwan. *Lancet*. 2008;371(9631):2173-2182.

18. Fored CM, Ejerblad E, Fryzek JP, et al. Socio-economic status and chronic renal failure: a population-based case-control study in Sweden. *Nephrol Dial Transplant*. 2003;18(1):82-88.

19. Steinke DT. Essentials of pharmacoepidemiology. In: Thomas D, ed. *Clinical Pharmacy Education, Practice and Research*. 1st ed. Oxford, UK: Elsevier; 2019:203-214.

20. Brewer T, Colditz GA. Postmarketing surveillance and adverse drug reactions: current perspectives and future needs. *JAMA*. 1999;281(9):824-829.

21. Kolesnyk I, Dekker FW, Noordzij M, le Cessie S, Struijk DG, Krediet RT. Impact of ACE inhibitors and AII receptor blockers on peritoneal membrane transport characteristics in long-term peritoneal dialysis patients. *Perit Dial Int*. 2007;27(4):446-453.

22. Tripepi G, D'Arrigo G, Jager KJ, Stel VS, Dekker FW, Zoccali C. Do we still need cross-sectional studies in nephrology? Yes we do! *Nephrol Dial Transplant*. 2017;32(2):19-22.

23. Aggarwal R, Ranganathan P. Study designs: part 2-descriptive studies. *Perspect Clin Res*. 2019;10(1):34-36.

24. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Stratification for confounding–part 2: direct and indirect standardization. *Nephron Clin Pract*. 2010;116(4):322-325.

25. Bruck K, Stel VS, Gambaro G, et al. CKD prevalence varies across the European general population. *J Am Soc Nephrol*. 2016;27(7):2135-2147.

26. Stel VS, Awadhpersad R, Pippias M, et al. International comparison of trends in patients commencing renal replacement therapy by primary renal disease. *Nephrology*. 2019;24(10):1064-1076.

27. Wolfe RA, Ashby VB, Milford EL, et al. Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *N Engl J Med*. 1999;341(23):1725-1730.

28. Jager KJ, Zoccali C, Macleod A, Dekker FW. Confounding: what it is and how to deal with it. *Kidney Int*. 2008;73(3):256-260.

29. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Stratification for confounding–part 1: the Mantel-Haenszel formula. *Nephron Clin Pract*. 2010;116(4):c317-c321.

30. Fitzmaurice G. Confounding: regression adjustment. *Nutrition*. 2006;22(5):581-583.

31. Normand SL, Sykora K, Li P, Mamdani M, Rochon PA, Anderson GM. Readers guide to critical appraisal of cohort studies: 3. Analytical strategies to reduce confounding. *BMJ*. 2005;330(7498):1021-1023.

32. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar Behav Res*. 2011;46(3):399-424.

33. Benedetto U, Head SJ, Angelini GD, Blackstone EH. Statistical primer: propensity score matching and its alternatives. *Eur J Cardio-Thorac*. 2018;53(6):1112-1117.

34. Fu EL, Groenwold RHH, Zoccali C, Jager KJ, van Diepen M, Dekker FW. Merits and caveats of propensity scores to adjust for confounding. *Nephrol Dial Transplant*. 2019;34(10):1629-1635.

35. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168(6):656-664.

36. Mansournia MA, Altman DG. Inverse probability weighting. *BMJ*. 2016;352:i189.

37. Lousdal ML. An introduction to instrumental variable assumptions, validation and estimation. *Emerg Themes Epidemiol*. 2018;15:1.

38. Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. *Annu Rev Public Health*. 1998;19:17-34.