



Points of attention when conducting etiological research

Vianda S. Stel¹  | Nicholas C. Chesnaye¹ | Giovanni Tripepi²  |
Friedo W. Dekker³ | Carmine Zoccali² | Kitty J. Jager¹

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

²CNR-IFC, Center of Clinical Physiology, Clinical Epidemiology of Renal Diseases and Hypertension, Reggio Calabria, Italy

³Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

Correspondence

Vianda S. Stel, European Renal Association–European Dialysis and Transplant Association Registry, Department of Medical Informatics, Amsterdam UMC, location AMC, Room J1b-108, PO Box 22700, 1100DE Amsterdam, The Netherlands.

Email: v.s.stel@amsterdamumc.nl

Abstract

Epidemiological studies often aim to investigate the causal contribution of a risk factor to a disease or other outcome. In etiological research, one is usually interested in the (biological) mechanism(s) underlying the studied relationship. Inappropriate conduct of an etiological study may have major implications for the correctness of the results and interpretation of the findings. Therefore, in this paper, we aim to describe step by step how etiological research should be carried out, together with its common pitfalls. These steps involve finding and formulating a well-defined etiological research question, choosing an appropriate study design including a suitable comparison group, adequate modelling, and adequate reporting and interpretation of the results.

KEYWORDS

aetiology, epidemiology, etiological research, pitfalls, research methodology, statistical modelling

INTRODUCTION

Epidemiological studies often aim to investigate the causal contribution of a risk factor to a disease or other outcome. These risk factors may consist of an intervention (e.g., a treatment) or a naturally occurring exposure (e.g., the presence of a disease). In etiological studies, the ultimate goal of identifying risk factors is either to cure the patient, to prevent the occurrence of an outcome, or to prevent disease progression by preventing or intervening on these risk factors. This kind of research typically addresses the aetiology—or cause—of the outcome. To this end, in etiological research one is usually interested in the (biological) mechanism(s) underlying the studied relationship.

Within nephrology, there are numerous examples of etiological research. For example, Schuett et al.¹ studied the effect of clot density on all-cause and cardiovascular mortality in a cohort of 117 patients on haemodialysis. The researchers hypothesized a possible biological mechanism behind the studied relationship, by stating that fibrin clot structure is crucially involved in the development of cardiovascular

events. Their rationale was that if clot density is a risk factor for (cardiovascular) mortality in haemodialysis patients, the risk of mortality could be mitigated by reducing clot density.

When conducting etiological research, several steps must be carefully followed. These steps involve finding and formulating a well-defined etiological research question, choosing an appropriate study design including a suitable comparison group, adequate modelling, and adequate reporting and interpretation of the results. Inappropriate conduct of an etiological study can have major implications for the correctness of its conclusions. That is why in this paper we will address points of attention when conducting such research. We will do this by describing step by step how etiological research should be carried out.

1 | STEP 1 FORMULATING A HYPOTHESIS

Etiological research should start with a hypothesis on the (suspected) causal risk factors of a disease or other outcome. This hypothesis

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Nephrology* published by John Wiley & Sons Australia, Ltd on behalf of Asian Pacific Society of Nephrology.

must always be placed in context. The introduction typically includes a brief overview of the published literature and a description of why the current study is needed. In many studies, the hypothesis is not (entirely) innovative, and the hypothesis may add value to the current literature by, for example, studying the effects of a newer treatment or studying another outcome of interest. Please note that it may be valid to perform a study that is not innovative. One may for example repeat a study in another country to confirm previous findings. Indeed, a meta-analysis includes studies that have addressed more or less the same hypothesis.

2 | STEP 2 PHRASING A RESEARCH QUESTION

The formulation of a research question must express precisely what the study covers. The PICO format is a helpful approach to formulate a clearly defined research question. The acronym stands for Population, Intervention, Comparison and Outcome²:

- The *P (Population)* refers to a clear description of the individuals enrolled in the study, such as information on their age (e.g., children, adults, older population), sex, setting (e.g., hospitalized, community-dwelling), and health condition (e.g., healthy, having diabetes mellitus). The choice of the population has implications for the generalizability of the study results.
- The *I (Intervention)* refers to the intervention or exposure. This could for example be a therapeutic intervention (e.g., a drug), preventive intervention (e.g., diet), the presence of a disease (e.g., diabetes mellitus), or an environmental exposure (e.g., nephrotoxins).
- The *C (Comparison)* represents the comparison (or control) group of the intervention (e.g., usual/current treatment or placebo) or the non-exposed (e.g., those without diabetes mellitus). To quantify the effect, the occurrence of the outcome in the intervention or exposed group is compared with that in the control group.
- The *O (Outcome)* represents the pertinent disease or other outcomes (e.g., disease progression, mortality) being investigated.

2.1 | Common pitfall

The scope of etiologic research is often mistakenly considered similar to that of prognostic research. In medicine, it is quite common that etiologic questions are erroneously tested with instruments meant for prognostic research (and vice versa).³ Prognostic research focuses on the course and outcomes of disease and factors that predict this outcome regardless of the underlying biological mechanisms. Table 1 clarifies the difference between etiological and prognostic research by describing the different research questions and ideas behind them. Please note that besides the research question, also the modelling, reporting and interpretation of results is completely different for etiological and prognostic research.

SUMMARY AT A GLANCE

A summary of steps to carry out etiological research, including main points of attention. [Correction added on 7 April, after first online publication: Summary at a Glance statement has been updated.]

3 | STEP 3 CHOOSING A STUDY DESIGN

3.1 | Study designs

Questions about the causal effect of interventions should preferably be addressed by a randomized controlled trial (RCT). In an RCT, the investigator intentionally manipulates the comparison by randomly allocating participants to the intervention or control group. By randomization, one aims to prevent confounding by indication with the main advantage that the randomization procedure ensures that any remaining differences between the intervention and control group are determined by chance.⁴ RCTs are not always feasible due to ethical reasons and their high cost. In observational studies, the intervention or exposure group, as well as the control group, are simply measured without manipulation by the researcher. Different types of observational study designs are used to answer etiological research questions, such as cohort studies and case-

TABLE 1 The difference between an etiological and prognostic research aim

	Etiological research aim	Prognostic research aim
What is it about	Aetiology is about causal relationships and biological mechanisms	Prognosis is about prediction of a clinical outcome given the specific values of patient characteristics
Example of research question	Does a higher body mass index increase the risk of chronic kidney disease?	If we would like to determine a person's risk of chronic kidney disease, should we take the body mass index into account?
The idea behind it	If a higher body mass index has a causal relationship with the occurrence of chronic kidney disease, then the risk of chronic kidney disease may be reduced by losing weight.	If body mass index predicts chronic kidney disease, people at high risk of chronic kidney disease may be more easily identified if body mass index is added to the prediction model. This is only about risk calculation/identification (to identify people with a high risk of chronic kidney disease), and not about the intention to advise people to reduce weight.

control studies. The choice of the best study design depends on the research aim as well as on the merits and limitations of the study design.⁵

3.2 | Bias

In designing a study, the researcher should always attempt to reduce both random and systematic errors. Random error is nothing more than variability in the data. Random error can be high if, for example, one takes just one blood pressure measurement in each of 10 individuals to estimate the average blood pressure in the reference population. However, if one takes 10 blood pressure measurements in each (which reduces the within subjects error) of 100 different individuals (which reduces the between individuals error), the random error of the average blood pressure would decrease dramatically. Potential solutions to limit or reduce random error are to increase the number of measurements at individual level and/or the sample size.

Systematic error, on the other hand, occurs when the measurements are consistently wrong. The two main types of systematic error, or bias, are selection bias (i.e., errors in the selection procedure of study participants, and from factors affecting study participation) and information bias (i.e., errors in the measurement, collection or interpretation of the exposure, of the disease, or both).⁸ A systematic error can underestimate or overestimate the true frequency of an exposure or an outcome or the true relationship between an exposure and outcome. This type of error does not decrease when the number of measurements or sample size increases, because it systematically affects all individuals in the sample.

4 | STEP 4 STATISTICAL MODELLING

4.1 | Checking the data

One first needs to check the data on, for example, outliers and missing values, and then correct these data in the best possible way (e.g., by checking and sometimes correcting outliers or by imputation of missing values). In addition, one may need to check the distribution of variables (e.g., normal/non-normal distribution) as this helps in making correct decisions during data analysis.

4.2 | Regression models

Within etiological research, the strength of the association between the risk factor and the outcome is shown by effect estimates, usually a relative risk, an odds ratio or a hazard ratio. These effect estimates are usually the result of regression analysis. The choice of the type of regression analysis depends on the type of outcome variable. Linear regression can be used for continuous outcome variables (e.g., systolic pressure),⁶ logistic regression for dichotomous outcome variables (e.g., being employed or not),⁶ and Cox regression for endpoints

including the time when an event occurs (e.g., time to death).⁷ When having time to event data, one usually also performs Kaplan Meier analyses to show the survival curves in a graph. As explained below, regression modelling (i.e., linear, logistic and Cox regression, but *not* the Kaplan Meier analyses) allows for the adjustment of potential confounders. Each type of regression analysis has (its own) assumptions and it is important to check if these assumptions are met to avoid misinterpretation of the results.

4.3 | Confounders

In observational studies, when studying the effect of a risk factor on an outcome, one should be aware that the association might actually depend on variables that are a common cause for both the intervention/exposure and the outcome, a so-called confounding variable. This is undesirable because a confounder may obscure the true effect of the risk factor on the outcome (negative confounding) or may generate/overestimate a relationship which does not exist (positive confounding). Adjustment for potential confounders in the regression analysis aims to take away the effect of confounding variables. In this case, one examines the association between a risk factor and outcome if all confounding variables have similar values in the comparison groups.

In order for a variable to qualify as a confounder it needs to satisfy the following three criteria⁸:

1. The variable must have an association with the outcome, that is, it should be a risk factor for the outcome
2. It must be associated with the exposure, that is, it must be unequally distributed between the exposed and non-exposed groups
3. It must not be an effect of the exposure or—linked to this—it should not be a factor in the causal pathway between the exposure and the outcome

Figure 1 gives an explanation about the criteria for confounding using two examples. Figure 1A shows that age is a potential confounder in the association between body mass index (BMI) (exposure) and chronic kidney disease (CKD) (outcome), because age fulfils the three criteria for confounding: age is associated with the CKD (criterion 1); age is associated with BMI (criterion 2); and age is not an effect of BMI (a higher or lower BMI does not make one older or younger) (criterion 3). Figure 1B shows that, in the same association, diabetes mellitus does not fulfil the three criteria for confounding (violation of the 3rd criterion) because diabetes mellitus is in the causal pathway (i.e., higher BMI may cause diabetes mellitus which may cause CKD). Diabetes mellitus is therefore a so-called mediator, and *not* a potential confounder, in the association between BMI and CKD. Literature research and expert opinion can be used to check these criteria for confounding, rather than checking the criteria using the data set of the study.

Several other methods can be used aiming to take away the effect of confounding variables, such as restriction (exclusion of a subgroup), stratification (analysing the data by subgroup), inverse probability

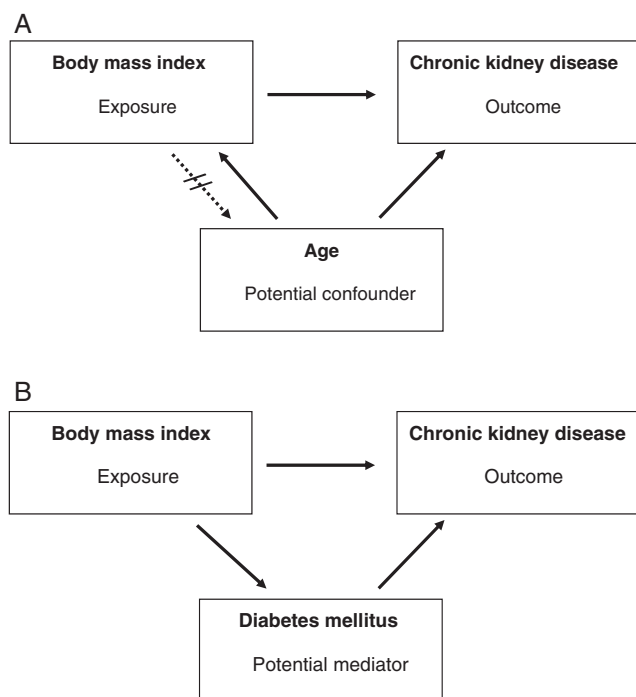


FIGURE 1 Schematic representation and explanation about the criteria for confounding: two examples. (A) Schematic representation of age as a potential confounder in the association between body mass index (exposure) and chronic kidney disease (outcome). Age fulfils the three criteria for confounding: (1) Age is associated with chronic kidney disease; (2) Age is associated with body mass index; and (3) Body mass index does *not* have an effect on age (a higher or lower body mass index does not make someone older or younger), and age is therefore not in the causal pathway. (B) Schematic representation of diabetes mellitus as mediator in the association between body mass index (exposure) and chronic kidney disease (outcome). Diabetes mellitus does not fulfil the three criteria for confounding (violation of the 3rd criterion), because body mass index has an effect on diabetes mellitus (i.e., higher body mass index may cause diabetes mellitus). Diabetes mellitus is therefore in the causal pathway in the association between body mass index and chronic kidney disease and therefore a so-called mediator (and *no* potential confounder)

weighting (estimating the probability of the exposure observed for a particular individual, and using the inverse predicted probability as a weight in subsequent analyses) and matching (participants in each group are matched regarding characteristics that might affect the outcome).

4.4 | Mediators

In addition to describing the effect size of risk factors on the outcome, researchers may want to unravel biological mechanisms. To this end, they may wish to explore to what extent the association between the risk factor and outcome is explained by a third variable that lies in the causal pathway (the mediator).⁹ For example, diabetes mellitus may be a mediator in the association between BMI and the development of

CKD (Figure 1B). The indirect effect (or mediation effect) corresponds to the effect of BMI on CKD that passes through diabetes, whereas the direct effect is the sole effect of BMI on CKD. Together, these effects add up to form the total effect of BMI on CKD. An often-used method to estimate the indirect effect, is to simply adjust for the mediator, and compare the effect estimate with that of the unadjusted model and/or the fully adjusted model. The difference then gives the effect of BMI on CKD that is explained by diabetes (i.e., the indirect effect). In causal mediation analysis confounders should be included to adjust for the associations between (a) the exposure and outcome; (b) the exposure and mediator; and (c) the mediator and outcome.⁹ Of note, etiological research questions often aim to determine the causal effect of an exposure (i.e., the total effect), in which case adjustment for mediators is unnecessary. If researchers wish to unravel biological mechanisms using mediator analyses, this should be clearly explained in the methods of the paper.

4.5 | Effect modifiers

Effect modification, also known as the heterogeneity of effect or interaction, is present when the magnitude of the association between the exposure and the outcome depends on a third variable.¹⁰ For example, one may wish to investigate whether the effect of BMI (exposure) on the risk of CKD (outcome) differs by sex (effect modifier). If the effect of BMI on CKD is strong in men, but not in women, the effect is heterogeneous, and presenting only the overall effect estimate would be misleading. Assessing effect modification therefore allows researchers to identify patient subgroups that may be at high or low risk, or in RCTs to identify subgroups in which a given treatment would be more or less beneficial. Effect modifiers can be identified by either performing a stratified analysis or by using an interaction term between the exposure and the effect modifier in regression models. In addition to estimating the magnitude of effect in each subgroup, the latter also allows for testing whether the effect modification is statistically significant.

4.6 | Common pitfalls

A common pitfall consists of failure to check for the criteria for confounding. For instance, without checking the criteria one may incorrectly adjust the association between a risk factor and an outcome for variables in the causal pathway (violation of the 3rd criterion). This results in overadjustment of the effect of that particular risk factor, leading to a real effect to go undetected. Causal pathways to an outcome may differ from one risk factor to another. Therefore, when studying the associations of more than one risk factor with the same outcome one should check the criteria for confounding for each association studied, because the confounders may be different for each association. For example, in the association between CRP and left ventricular mass in pre-dialysis patients, smoking is a potential

TABLE 2 Results of an etiological study on the association between body mass index and the development of chronic kidney disease in apparently healthy men^a

Baseline body mass index (kg/m ²)	Model 1 ^b unadjusted Odds ratio (95% confidence interval)	Model 2 adjusted for age at baseline	Model 3 ^c adjusted for all potential measured confounders	Model 4 ^d adjusted for all potential measured confounders + mediators
Continuous variable	1.09 (1.05–1.12)	1.05 (1.02–1.07)	1.05 (1.03–1.07)	1.03 (1.01–1.05)
Quintiles				
1. <22.7	1.00	1.00	1.00	1.00
2. 22.7–23.7	1.23 (1.01–1.46)	1.19 (0.98–1.45)	1.17 (0.96–1.43)	1.13 (0.96–1.43)
3. 23.8–25.0	1.25 (1.02–1.48)	1.23 (1.01–1.50)	1.24 (1.02–1.52)	1.15 (1.02–1.52)
4. 25.1–26.6	1.46 (1.21–1.79)	1.44 (1.19–1.75)	1.44 (1.19–1.75)	1.32 (1.19–1.75)
5. >26.6	1.46 (1.17–1.72)	1.43 (1.18–1.73)	1.45 (1.19–1.76)	1.26 (1.19–1.76)

^aThe results are based on the study of Gelber et al.,¹¹ with exception of the results of Model 1.

^bModel 1: The unadjusted odds ratios were not provided in the original paper and for educational purposes self-invented unadjusted odds ratios were added to this table.

^cModel 3: Adjusted for age at baseline (40 to 44, 45 to 49, 50 to 54, 55 to 59, 60 to 64, 65 to 69, and 70 years), smoking at baseline (never, past, current), alcohol consumption at baseline (1 drink/day, 2 to 6 drinks/week, 1 drink/week), exercise (none, 1 time/month to 4 times/week, ≥5 times/week), and parental history of myocardial infarction before the age of 60 years.

^dModel 4: Adjusted for all variables in Model 3 plus diabetes (history at baseline or during follow-up), hypertension at baseline or follow-up (defined as self-reported systolic blood pressure 140 mm Hg or diastolic blood pressure 90 mm Hg or antihypertensive medication use regardless of blood pressure), elevated cholesterol level at baseline or during follow-up (defined as 240 mg/dl [6.21 mmol/L] or history of past or current use of cholesterol-lowering medication), and development of cardiovascular disease during follow-up (defined as self-reported myocardial infarction, stroke, angina, coronary artery bypass graft or percutaneous coronary angioplasty).

confounder (because smoking fulfils all criteria for confounding). On the other hand, in the same study population, the association between sex and left ventricular mass should not be adjusted for smoking because smoking may be in the causal pathway between sex and left ventricular mass (usually the prevalence of smoking is higher in men; violation of 3rd criterion).

It is not good to select confounders based on a *p*-value. The main reason for this is that the *p*-value depends on the size of the sample. An additional pitfall is the use of incorrect statistical methods, that are related to prognostic research (e.g., the use of backward selection to select covariates in the model or estimations of the area under the curve).⁶ These analyses are incorrect for etiological research as they predict outcomes without the need to formulate potential biological mechanisms and adjustment for potential confounders.

5 | STEP 5 REPORTING RESULTS

5.1 | Effect estimates

Preferably, the results section of an article presents both the unadjusted and adjusted effect estimates which makes it possible to compare them. If the unadjusted and adjusted effect estimates are different, we can conclude that the confounders obscured the true effect of the studied risk factor. Even if the unadjusted and adjusted effect estimates are similar, it is recommended to present both so this similarity becomes visible. Effect estimates should be presented with a 95% confidence interval (CI). The 95% CI gives information about the variability of the effect estimate, and defines a range of values

that contain, with 95% certainty, the true effect estimate. Please note that a larger study sample gives more precise effect estimates and narrower CIs.

Table 2 presents the study results based on a study by Gelber et al.¹¹ on the association between BMI and the development of CKD after 14 years of follow-up in a cohort of 11 104 initially healthy men. Logistic regression was used to analyse the data. Please note that Table 2 contains some differences with the results presented in the original paper (see footnote of Table 2).

Table 2 shows the unadjusted effect estimate (i.e., odds ratio) of BMI as a *continuous* exposure variable (per 1 kg/m²) on the development of CKD after 14 years (1.09, 95% CI 1.05–1.12, Model 1). This means that with each 1 kg/m² increase in BMI people are 9% more likely to develop CKD. After adjustment for age, which satisfies the criteria for confounding (Model 2), the odds ratio was 1.05 (95% CI 1.02–1.07). The odds ratio of Model 1 (1.09) and model 2 (1.05) are different. Therefore, age is a confounder in the association between BMI and CKD. In other words, in Model 1, age obscures the real effect of BMI on CKD, and in Model 2, adjustments have been made for age which eliminates the confounding effect of age on the association between BMI and the development of CKD. In Model 3 the researchers adjusted for all potential measured confounders in the study (age, smoking, alcohol consumption, and parental history of myocardial infarction before the age of 60 years). The odds ratio did not change in comparison to Model 2 and was still 1.05 (95% CI 1.03–1.07). In all cases 1.00 was not included in the confidence interval. This means that there was a statistically significant association between BMI and the development of CKD in all models.

TABLE 3 Summary of steps to carry out etiological research, including main points of attention

Steps to carry out etiological research	Main points of attention
Step 1: Formulating a hypothesis	<ul style="list-style-type: none"> Etiological research should start with a hypothesis on the causal risk factors of a disease or other outcome, and the hypothesis must always be placed in context
Step 2: Phrasing a research question	<ul style="list-style-type: none"> The formulation of a research question must express well enough what the study is about The PICO format can be used to formulate the question The aim of etiological research should not be formulated as prognostic research (and vice versa)
Step 3: Choosing a study design	<ul style="list-style-type: none"> The choice of the study design depends on the research aim as well as on the merits and limitations of the study design In designing a study, one should attempt to reduce random and systematic errors
Step 4: Statistical modelling	<ul style="list-style-type: none"> Before analysing the data, one first needs to check the data (e.g., outliers, missing data) The choice of the type of regression analysis depends on the outcome variable. One should check if the assumptions of the type of regression analysis used are satisfied In observational etiological research it is usually needed to adjust for confounders The criteria of confounding must be taken into account when choosing potential confounders When studying the association of more than one risk factor with the same outcome, one should check the criteria for confounding for each studied association separately (and not automatically use the same confounders for all associations studied) Literature research and expert opinion should be used to check the criteria for confounding, rather than checking the criteria with the data set at hand Adjustment for mediators is only correct in case of unravelling biological mechanisms whereas it is incorrect in determining the real effect of the risk factor. This should be clearly explained in the study methods If the effect of the risk factor on the outcome is heterogenous (e.g., different for males and females), it is usually needed to present the effect estimate for both groups separately as presenting the overall effect estimate alone would be misleading One should not use statistical methods that are related to prognostic research as these methods predict outcomes without the need to formulate potential biological mechanism and adjust for potential confounders One should not select confounders based on a <i>p</i>-value
Step 5: Reporting results	<ul style="list-style-type: none"> Only the effect estimate of the variable of interest, and not of the confounders, should be presented, including its 95% confidence interval Preferably, both unadjusted and adjusted estimates should be reported Both the absolute and relative risk should be presented Checklists, such as the STROBE or CONSORT statements, can be used for state-of-the-art reporting of the study methods and results
Step 6: Interpretation of the results	<ul style="list-style-type: none"> The article should discuss the etiological aspects of the study (e.g., potential biological mechanisms) To formulate a valid conclusion, possible limitations of the study such as residual confounding, bias and lack of statistical power must be taken into account

Table 2 also shows the association between BMI as a *categorical* variable and the development of CKD. Again, the results show the unadjusted odds ratios (Model 1), the odds ratios adjusted for age (Model 2) and the odds ratios additionally adjusted for all measured potential confounders (Model 3). These odds ratios reflect a comparison with the reference group, that is, patients with a BMI (< 22.7 kg/m²). For example, in Model 1, the unadjusted odds ratio for patients with a BMI greater than 26.6 kg/m² was 1.46 (95% CI 1.17–1.72). This means that patients with a BMI greater than 26.6 kg/m² are 46% more likely to develop CKD than patients with a BMI < 22.7 kg/m² (reference group).

In the same study, the authors also aimed to unravel the biological mechanism, and for that reason, described a second hypothesis in their study. To this end, they created Model 4 in which they not only added all potential measured confounders but also potential mediators (i.e., diabetes mellitus, hypertension, elevated cholesterol level, use of cholesterol-lowering medication, and development of cardiovascular disease during follow-up). In Model 4, the odds ratio for BMI

as a continuous variable was 1.03 (95% CI: 1.01–1.05). When comparing the odds ratios of Model 3 (1.05) and Model 4 (1.03), we can conclude that these mediators reduced the effect of BMI by 0.02. This means that slightly less than half of the effect of BMI on CKD may be explained by these mediators.

5.2 | Absolute risk

The size of an effect should be expressed in both relative and absolute terms. For instance, a study may present a relative risk estimate of 1.5, reflecting an impressive 50% increase in risk in the exposed group compared with the unexposed group. However, if the absolute risk of the outcome is only 1% in the unexposed group, this relative risk corresponds to a—somewhat less impressive—0.5% increase in absolute risk in the exposed group. The number needed to treat ($[1/\text{absolute risk difference}] \times 100$) in this example is 200, which means that one needs to treat 200 patients to prevent one outcome. Both absolute and relative

measures are therefore required to interpret whether this constitutes a clinically relevant effect in the context of the research question.

5.3 | Reporting checklists

To strengthen the reporting of studies in epidemiology one should follow the so-called reporting checklists, such as the STROBE statement¹² for observational studies and the CONSORT statement for RCTs.¹³

5.4 | Common pitfalls

A common pitfall is to report the full model, that is, showing the “effect estimates” of the confounders (these are usually given in the output of the statistical program). However, the relationship between a confounder and the outcome may have different confounders than the confounders used in the model. Moreover, one is usually not interested in the effect of the confounder on the outcome, as it is the effect of a confounder that one would like to eliminate.

6 | STEP 6: INTERPRETATION OF THE RESULTS

An etiological study discusses its findings in relation to the (biological) mechanism(s) that may underlie the associations that were found. In this type of studies, the conclusion typically includes an answer on the presence or absence of an association between a risk factor and an outcome. Although RCTs are the best study design to evaluate the causal effect of interventions, RCTs may also have limitations. For example, the generalizability of its results is often limited due to sampling bias. Although in observational studies the etiological association may be adjusted for potential confounders, it is not possible to adjust for unknown confounders, and also confounders may not have been properly assessed. This is the reason why after adjustment for potential confounding variables, residual confounding may still exist and causal inference is impossible to achieve. To draw a valid conclusion, possible study limitations such as residual confounding, bias, and lack of statistical power must be taken into account.

7 | SUMMARY AND CONCLUSION

This paper presents the steps to carry out an etiological study together with its common pitfalls (a summary is given in Table 3).

Inaccuracies in the conduct of such a type of studies may have major implications for the correct interpretation of the findings.

CONFLICT OF INTEREST

There is no conflict of interest.

ORCID

Vianda S. Stel  <https://orcid.org/0000-0002-0007-1947>

Giovanni Tripepi  <https://orcid.org/0000-0002-3580-4602>

REFERENCES

- Schuetz K, Savva A, Maxeiner S, et al. Clot structure: a potent mortality risk factor in patients on hemodialysis. *J Am Soc Nephrol*. 2017; 28:1622-1630.
- Cross NB, Craig JC, Webster AC. Asking the right question and finding the right answers. *Nephrology*. 2010;15(1):8-11.
- van Diepen M, Ramspek CL, Jager KJ, Zoccali C, Dekker FW. Prediction versus aetiology: common pitfalls and how to avoid them. *Nephrol Dial Transplant*. 2017 Apr 1;32(suppl_2):ii1-ii5.
- Stel VS, Jager KJ, Zoccali C, Wanner C, Dekker FW. The randomized clinical trial: an unbeatable standard in clinical research? *Kidney Int*. 2007;72:539-542.
- Bosdriesz JR, Stel VS, van Diepen M, et al. Evidence-based medicine—when observational studies are better than randomized controlled trials. *Nephrology*. 2020;25:737-743.
- Tripepi G, Jager KJ, Dekker FW, Zoccali C. Linear and logistic regression analysis. *Kidney Int*. 2008;73:806-810.
- van Dijk PC, Jager KJ, Zwinderman AH, Zoccali C, Dekker FW. The analysis of survival data in nephrology: basis concepts and methods of Cox regression. *Kidney Int*. 2008;74:705-709.
- Jager KJ, Zoccali C, MacLeod A, Dekker FW. Confounding: what it is and how to deal with it. *Kidney Int*. 2008;73:256-260.
- VanderWeele TJ. Mediation analysis: a Practitioner's guide. *Annu Rev Public Health*. 2016;37(1):17-32.
- Corraini P, Olsen M, Pedersen L, Dekkers OM, Vandembroucke JP. Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators. *Clin Epidemiol*. 2017;8(9):331-338.
- Gelber RP, Kurth T, Kausz AT, et al. Association between body mass index and CKD in apparently healthy men. *Am J Kidney Dis*. 2005;46:871-880.
- von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007; 370(9596):1453-1457.
- Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Ann Intern Med*. 2010;152(11):726-732.

How to cite this article: Stel VS, Chesnaye NC, Tripepi G, Dekker FW, Zoccali C, Jager KJ. Points of attention when conducting etiological research. *Nephrology*. 2021;26: 701–707. <https://doi.org/10.1111/nep.13875>