OBSERVATIONAL STUDIES VS. RANDOMIZED CONTROLLED TRIALS: AVENUES TO CAUSAL INFERENCE IN NEPHROLOGY

Csaba P Kovesdy, MD[1,2] and **Kamyar Kalantar-Zadeh, MD, PhD, MPH**[3,4]

[1]Division of Nephrology, Salem Veterans Affairs Medical Center, Salem Virginia

[2]Division of Nephrology, University of Virginia, Charlottesville, Virginia

[3]Harold Simmons Center for Kidney Disease Research and Epidemiology, Division of Nephrology and Hypertension, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA

[4]David Geffen School of Medicine at UCLA, Los Angeles, CA

Abstract

A common frustration for practicing Nephrologists is the adage that the lack of randomized controlled trials (RCTs) does not allow us to establish causality, but merely associations. The field of Nephrology, like many other disciplines, has been suffering from a lack of RCTs. The view that short of RCTs there is no reliable evidence has hampered our ability to ascertain the best course of action for our patients. However, many clinically important questions in medicine and public health such as the association of smoking and lung cancer are not amenable to RCTs due to ethical or other considerations. Whereas RCTs unquestionably hold many advantages over observational studies, it should be recognized that they also have many flaws that render them fallible under certain circumstances. We provide a description of the various pros and cons of RCTs and of observational studies using examples from the Nephrology literature, and argue that it is simplistic to rank them solely based on pre-conceived notions about the superiority of one over the other. We also discuss methods whereby observational studies can become acceptable tools for causal inferences. Such approaches are especially important in a field like Nephrology where there are myriads of potential interventions based on complex pathophysiologic states, but where properly designed and conducted RCTs for all of these will probably never materialize.

Keywords

observational studies; randomized controlled trials; causal inference

Introduction

A frequent lament in Nephrology is the lack of randomized controlled trials (RCTs), and thus the inability to conclusively establish a cause-and-effect relationship between an exposure and an outcome. This has led to some professional guidelines concluding that no firm recommendations can be made about the utility of many therapeutic interventions, [1]

Correspondence: Csaba P Kovesdy, MD FASN, Division of Nephrology, Salem VA Medical Center, 1970 Roanoke Blvd., Salem, VA 24153, Phone: 540-982 2463, Fax: 540-224 1963, csaba.kovesdy@va.gov.

causing significant confusion and disappointment among the rank and file. The prevailing view is that observational studies are in many ways flawed, and thus inferior; and their main utility is to generate hypotheses which can then be tested in RCTs, the crown jewels of investigative science. This mentality has resulted in a frustrating stalemate in many disciplines including Nephrology, where the low number of RCTs has limited the scope of many therapeutic interventions to short term goals of correcting biochemical abnormalities without targeting clinical outcomes such as morbidity or mortality. The negative results of several recent large RCTs[2–9] could result in an even more cautious attitude towards designing and conducting such investigations; hence it is likely that Nephrologists' frustrations with the level of evidence regarding the efficacy of many therapeutic interventions will probably persist for some time. Under such circumstances it is worthwhile re-examining our present paradigm which declares that observational studies and to some extent laboratory experiments are merely laying the groundwork for RCTs (Figure 1). We shall thus examine the pros and cons of observational studies vis a vis to RCTs along with various ways that could allow the former to be regarded as high level evidence.

Association vs. causation

All exploration starts with careful observation of the world around us, and scientists have used observations for hundreds of years to explain why events such as diseases occur. Rickets for example, which is now believed to be caused by vitamin D deficiency, became common during the industrial revolution as a result of the "modern" lifestyle of new city dwellers that resulted in significantly decreased exposure to sunlight. In 1822 Jedrzej Sniadecki, a Polish physician working at the University of Wilno proposed that the new lifestyle was linked to rickets and recommended exposure to open air and sunshine for the cure of the "English disease". [10] He thus established an association between the observed lifestyle and a disease, and proposed an experiment (which could have taken the form of an RCT) to test the hypothesis that the lifestyle caused the disease. Almost 200 years later we now know how ultraviolet radiation generates vitamin D which in turn affects bone health; but purely based on the association between lifestyle and rickets one cannot necessarily claim that the lack of sunshine was the proximal cause of the disease. An equally plausible hypothesis could have been that the disease itself caused a change in lifestyle, as patients with rickets were likely to be less mobile and hence more likely to be homebound; a phenomenon called reverse causation. One could also (maybe less plausibly) hypothesize that both rickets and reclusive lifestyle have a common source, which would than confound the association between the two and create a false sense that they are somehow linked. It is thus clear that one has to be careful when interpreting associations, as they are not always due to causal relationships. It is also true, that in the above example an RCT to test the hypothesis about sunshine's role in the disease process can provide more conclusive evidence, especially given knowledge of the condition's pathophysiology. But what if we are faced with a condition for which designing an RCT is not feasible, or is unethical, such as examining the association of tobacco smoking and lung cancer? Can we then turn to other types of investigations to establish causality? Are there certain rules that determine to what extent an observation can be considered valid and unbiased?

There have been numerous attempts throughout history to codify the rules that establish the validity of a scientific investigation. A formal set of rules to describe valid experimentation and drug testing was established as early as 1025 AD by the Persian physician and philosopher Ibn Sina (Avicenna) in his work The Canon of Medicine. [11;12] More relevant to the topic discussed herein was a later effort by Sir Austin Bradford Hill, who established a number of criteria, which, if fulfilled would strengthen the possibility of a causal relationship. [13] As shown in Table 1, these criteria can by no means be applied as a simple checklist in order to achieve logical proof of causality. Short of maybe the criterion of

temporal sequence (i.e. the exposure has to precede the disease), none of the others (including experimentation) can be construed as both necessary and sufficient to prove causality beyond any doubt.

Experiments such as RCTs are in fact themselves merely observations; what distinguishes them is the apparent control that we believe we have over the circumstances of the observations. However, RCTs too are subject to a number of fallacies (vide infra) which could render their results invalid or questionable. It is thus naive to look at RCTs as a universal panacea that will always tell us what is right or what is wrong; as it is also simplistic to question the validity of observational studies merely because of their non-experimental and non-randomized nature.

Are RCTs infallible?

Observational studies could be biased in many ways; be it selection bias including survivor bias, information bias including from use of missing data or inappropriate categories, confounding bias, or any residual sources of bias, not to mention new sources or errors and bias upon analyses such as those due to over-adjustment, inappropriate modeling and assumption-violations. Due to the constant suspicion that the results of observational studies are affected by inadequate control of some or all of these biases preference is given to RCTs in order to have an unbiased assessment of causality. RCTs can, however, also suffer from considerable bias; as discussed below some of these can be controlled with careful planning, but others are inherent shortcomings of all RCTs that have to be considered when interpreting their results.

Non-adherence to the assigned intervention (Intention-to-Treat vs. As-Treated)

An important limitation of some RCTs is non-adherence to an assigned intervention that can render the planned comparison difficult. One can choose to maintain the validity of the randomization sequence by comparing the groups according to their original assignment, the so-called "intention-to-treat" analysis; the problem in this case will be that the observed effect may not be the result of the originally assigned treatment(s), rendering any causal inference about the intervention(s) invalid or biased towards the null. Choosing to compare patients treated with one vs. the other intervention irrespective of their original assignment (the so-called "as-treated" analysis) is not any better, as such a comparison will render the randomization moot and the RCT becomes an observational study riddled with its potential biases. A possible solution to this dilemma is the application of G-estimation to RCTs with significant non-compliance, which considers both assigned and received treatment simultaneously in a structural nested model.[14]

The challenge of retention

Inability to follow-up on enrolled participants, also known as censoring, can likewise hamper the interpretation of RCTs. The number of participants to be enrolled in an RCT is established based on assumptions that a certain number of patients will complete the study and a certain number of events will be observed which will then assure that a given minimum difference between the outcomes in the intervention arms can be detected. Deviations from these assumptions, whether due to larger-than-expected dropouts or fewer-than-planned events can result in a study that does not have the power to detect meaningful differences.

A recent example for several of the above pitfalls was the Dialysis Clinical Outcomes Revisited (DCOR) study,[8] which compared clinical outcomes between patients randomized to receive sevelamer hydrochloride or calcium-based phosphorus binders. Only 52% of patients assigned to sevelamer and 49% of patients assigned to calcium completed

the study, and the original duration of the study had to be extended due to lower-than-expected event rates. Due to these shortcomings the DCOR study should be regarded as inconclusive and not one that refutes the original hypothesis of the study.

Suboptimal randomization and heterogeneity

Another potential problem with RCTs is unsuccessful randomization, whereby the two comparison groups differ from one another in one or more important characteristic, thus creating confounding. This problem can be addressed by adjustment for the unbalanced variables, provided that we know what those are; but it is always possible that there may be some unmeasured variable which could end up confounding the results of the RCT. This can be a significant problem with small RCTs, especially if the mechanisms of action of the studied intervention are not clearly defined, making the identification of key covariates difficult. A frequently applied "remedy" for small RCTs is the technique of meta-analysis, whereby the results of many smaller studies are pooled together, thus increasing their statistical power. The problem that is often overlooked is that pooling together many poor-quality studies suffering from significant bias may alleviate the shortcoming stemming from their sample size, but does not render the aggregate results qualitatively any more valid than those of the individual studies themselves. Furthermore, significant heterogeneity among the individual studies included in the meta-analysis can make it difficult to unify their findings under a single umbrella. A recent example to these effects was a meta-analysis that examined the impact of vitamin D therapy on various biochemical end points.[15] One of the main results of this study (that treatment with active vitamin D does not lower parathyroid hormone levels) has been widely cited as proof towards the ineffectiveness of active vitamin D, ignoring the fact that many of the individual studies included in this meta-analysis had significant flaws individually and that there was a significant heterogeneity among them.[16]

External validity

Besides the above pitfalls that can at times be difficult to predict when planning clinical trials, there are universal flaws of RCTs that need to be kept in mind before canonizing them. Perhaps the most important such problem is their limited external validity. A well designed and executed RCT (i.e. one with excellent internal validity) can indeed provide conclusions about the studied group of patients which was selected based on strict inclusion and exclusion criteria, and such conclusions can be regarded valid with great degree of certainty. They, however, do not necessarily inform us about how the same or similar interventions would fare under different circumstances; patients of different age, race, and gender, or those with different kidney or liver function may respond differently to a treatment due to myriads of potential reasons. Nephrologists are very often faced with this problem, as it is very tempting to fill the void of RCTs in patients with CKD with results from studies performed in patients with normal kidney function. One example to this problem is the treatment of hypertension; the Kidney Disease Outcome Quality Improvement (K-DOQI) guidelines made recommendations about the treatment of hypertension in CKD based on extrapolations from studies in patient populations without CKD, due to lack of studies in patients with CKD and end stage renal disease (ESRD). [17] This conjecture can be challenged both based on biological differences in how patients with CKD respond to changes in blood pressure, [18] and also based on observational studies suggesting that the association between blood pressure and clinical outcomes is different in patients with ESRD[19] and CKD[20] than that seen in the general population. [21] Nevertheless, the conviction that the lack of RCTs should prevent even speculations about causal inferences that contradict mainstream paradigms seems to trump these arguments at the moment.

When RCTs are not possible

Another practical shortcoming of RCTs is the ethical concern raised by some interventions. In order to completely eliminate bias one would have to assign potentially harmful interventions to, and thus intentionally jeopardize the wellbeing of study participants. It is easy to see how one cannot ethically conduct an RCT examining the harmful effects of smoking, for example. The causal role of smoking in engendering lung cancer is now universally accepted, even though it is based exclusively on observational data, and one could potentially construct alternative hypotheses how an association between smoking and lung cancer could be biased by confounders like for example inhalation of phosphorus from the fires of the matches used to light cigarettes. In multivariate analyses, adjusting for the status of "carrying matches" may indeed nullify the statistical association between smoking and lung cancer. Nevertheless, the causal association between tobacco smoking and lung cancer has hardly been questioned despite a lack of RCTs.

An example from the field of Nephrology where concerns for the wellbeing of participants made the completion of an RCT difficult was a recent clinical trial that compared early vs. late initiation of dialysis.[22] This study failed to achieve the pre-specified difference in GFR level at dialysis initiation between the two intervention arms, as 76% of patients in the late start arm initiated dialysis earlier then planned largely because of the investigators' concerns about the wellbeing of those patients who were supposed to delay dialysis initiation. In cases like these observational studies employing various techniques to minimize bias may be the only feasible way to gain better understanding of a disease process or an intervention (vide infra).

Cost and duration of RCTs

Finally, the most mundane deficiencies of RCTs are their high cost and the long time required for their completion. Compared to observational studies RCTs carry a price tag that is several magnitudes higher and typically they take many more years to complete. A significant proportion of RCTs are sponsored by Pharmaceutical companies to whom the investment is justified by future returns due to a new indication for their marketed drug or to higher sales boosted by the benefits shown by their agent in the clinical trial. The recent spate of large Pharmaceutical-sponsored RCTs showing neutral or unexpectedly deleterious effects of the applied interventions in ESRD/CKD patients[2–8] along with the economic recession and a change in the payment structure for medications in dialysis patients will foreseeably dampen the enthusiasm for corporate sponsorship of RCTs in this patient population, making it even more important to explore alternative ways to make causal inferences.

Can observational Studies Allow Causal Inference?

What are clinicians to do if there is no RCTs for most major problems faced in clinical practice of Nephrology? If we adhere to the "no RCT equals no conclusive proof" principle (Figure 1) we will never have all the answers we need, especially if we consider the above limitations of RCTs. It is thus worthwhile to explore alternative ways to try and make causal inferences using methods other than RCTs. The principle advantage of RCTs is their controlled nature which eliminates many of the biases that hamper the interpretation of observational studies. There are, however, techniques that can be applied to observational data in order to eliminate bias, which could theoretically allow causal inference. The simplest and most widely applied one is adjustment, whereby analyses can account for the bias introduced by a known confounder. A variant of adjustment is the use of propensity scores, where the likelihood of a certain intervention or treatment is quantified using known

determinants of the intervention, and analyses are then either conditioned (stratified) on or adjusted for the propensity scores.

Time-dependent confounding and causal models

A special case of confounding is the so called time-dependent confounding and when a confounder also acts simultaneously as an intermediate confounder; an example is the effect of active vitamin D on mortality: in clinical practice active vitamin D is initiated based on a patient's serum parathyroid hormone (PTH) level, which in itself can be associated with mortality (thus it is a confounder). Once, however, active vitamin D is started it will usually lower the PTH level (thus PTH is also an intermediate), but the lower PTH level will then often result in the discontinuation or in a dose-adjustment for active vitamin D (thus it is a time-dependent confounder and may also be in the causal pathway to mortality). Adjustment for baseline or even time-dependent values of PTH and/or active vitamin D doses in Cox models (as was done in numerous studies that examined the association of active vitamin D with mortality[23–27]) does not resolve the bias resulting from the complicated situation depicted above; more complex methods such as marginal structural models (also called "causal models") have been proposed to achieve this, [28] as was applied in some of the studies examining the effect of active vitamin D on survivial. [29;30] The above statistical methods are applied to address bias resulting from known confounders under different circumstances, but none of them can resolve bias from unknown or residual confounders. If, for example some patients are materially better off they could presumably afford better insurance which could allow them to receive a more expensive version of active vitamin D which has a more beneficial side effect profile. A study that compared mortality in patients who received a newer (paricalcitol) vs. an older version (calcitriol) of the active vitamin D found that the former was associated with lower mortality. [31] While it is possible that the better outcomes were caused by a lower incidence of hypercalcemia or hyperphosphatemia or by some other biological benefit of paricalcitol, it is just as plausible to hypothesize that a better socioeconomic situation in those who received paricalcitol was also linked to many unrelated benefits that could have impacted their survival. Without the ability to quantify all the possible ancillary benefits that may accompany paricalcitol use one cannot with certainty make causal inferences about the link between paricalcitol and survival in this observational study.

Instrumental variable

A technique that can potentially address unmeasured confounders in observational studies is the instrumental variable approach. [32] This method consists of identifying a so called instrument, which is a variable that (a) defines the intervention of interest (e.g. active vitamin D use); (b) affects the outcome (e.g. mortality) exclusively through the intervention of interest and (c) shares no common causes with the outcome (i.e. it lacks confounders). Association of an instrument with the outcome can be interpreted as unconfounded proof of a causal relationship between the actual intervention of interest and the outcome.

The simplest example of a valid instrumental variable is the random treatment allocation in a clinical trial, since (a) randomization determines the exposure to the intervention, (b) if randomization is successful than mortality is only affected through the treatment intervention, since all other relevant characteristics of exposed and unexposed patients are identical and (c) randomization by definition has no common source with the outcome. A similar framework can be applied to observational studies to allow for unbiased estimates of causation. Good examples are cases where a natural "randomization" of a characteristic such as a genetic trait at birth will allow the unconfounded examination of an association with future outcomes; these studies are also called Mendelian randomization studies. [33] A recent example was a study that examined the association of a certain genotype that

determines elevated C-reactive protein (CRP) levels with future ischemic heart disease. [34] Numerous observational studies have previously described an association between higher CRP and increased incidence of ischemic heart disease, but since atherosclerosis itself can cause higher CRP levels it is possible that such associations were due to reverse causation. Since persons who are born with a certain genotype that determines higher CRP levels would experience these higher levels prior to the development of heart disease (which is presumably not present at birth), a lack of association between the genotype and ischemic heart disease contradicted the hypothesis that CRP is causally responsible for the development of this condition. [34] Such Mendelian randomization studies can offer a suitable observational alternative to RCTs, even though one has to be careful to consider potential sources of bias even for such studies. It is possible for example to imagine an unidentified gene in the above study of CRP which is in linkage disequilibrium with the CRP gene but is involved in some other pathologic process linked to heart disease, and thus act as an unmeasured confounder.

Mendelian randomization studies need the presence of genetically determined characteristics; hence our ability to use this method is finite. Researchers have attempted to find instrumental variables that are not determined by nature or by randomized interventions for use in observational studies in order to bypass the need for randomized controlled trials. An example is a recent study of active vitamin D, where the percentage of active vitamin D use in a given hemodialysis center was proposed as the instrumental variable. [30] The authors of this study did not find a significant association between their instrument and mortality and concluded that such lack of association was proof that active vitamin D therapy is not causally linked to lower mortality. The criticism of the instrumental variable approach using instruments that are not created by randomization is that there is no telling if the chosen instrument is indeed conforming to the basic requirements that make it a valid tool; in the case of the above study[30] one could imagine that dialysis units with higher rates of active vitamin D use may provide different care in other (unmeasured) ways too, which in turn could also affect outcomes independently and thus act as a common source (an unmeasured confounder).[35]

Epilogue

What is thus the optimal method to allow for causal inferences? As discussed above, properly designed and conducted RCTs do provide a higher degree of certainty as they eliminate many of the biases that can be present in observational studies, but they are by no means infallible. A commonly cited example that is meant to prove that RCTs refute the results of observational studies is the one of studies examining the link between blood hemoglobin level and mortality in patients with CKD and ESRD. [2–5] These RCTs were designed based on older observational studies showing an association between higher hemoglobin and lower mortality, only to show that applying therapy with erythropoesis stimulating factors (ESF) to increase hemoglobin to certain predefined levels was either neutral or deleterious. [2–5] It is however simplistic to conclude from these studies that they completely refuted the results of all the observational studies linking higher hemoglobin to better survival; they merely established with a great degree of certainty that the application of ESF towards certain hemoglobin goals may be deleterious, without being able to discern if this was due to the (achieved or targeted) higher hemoglobin levels or to some other side effect of the ESF. Not surprisingly, due to this uncertainty the aforementioned studies have lead to a number of hypotheses which have or will be tested first in observational studies, [36] which may lead to yet more hypotheses. Furthermore, the notion that these RCTs were in discordance with all prior observational studies neglects the fact that some observational studies have in fact provided earlier evidence that hemoglobin levels above a certain threshold may be deleterious; [37] these were ignored either because of infatuation with

prevailing scientific paradigms at the time, or because of the a priori conviction that observational studies are always inferior. In retrospect the hemoglobin-ESF story in CKD provides a prime example for how we should not rank the various types of investigations based on prior convictions about their advantages or disadvantages, but rather consider all of them as pieces of the same puzzle (Figure 2), with all of them being able to provide useful information and increase our knowledge about a subject, but without any of them being infallible. We need to realize that absolute certainty cannot ever be attained in Medicine. Our goal should be to continuously broaden our knowledge to achieve increasing certainty about biological processes and disease therapies, while constantly keeping an open mind.

Reference List

1. KDIGO clinical practice guideline for the diagnosis, evaluation prevention, and treatment of Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD). Kidney Int Suppl. 2009; 76 (113):S1–S130.

2. Besarab A, Bolton WK, Browne JK, Egrie JC, Nissenson AR, Okamoto DM, et al. The effects of normal as compared with low hematocrit values in patients with cardiac disease who are receiving hemodialysis and epoetin. N Engl J Med. 1998; 339:584–590. [PubMed: 9718377]

3. Drueke TB, Locatelli F, Clyne N, Eckardt KU, Macdougall IC, Tsakiris D, et al. Normalization of hemoglobin level in patients with chronic kidney disease and anemia. N Engl J Med. 2006; 355:2071–2084. [PubMed: 17108342]

4. Pfeffer MA, Burdmann EA, Chen CY, Cooper ME, de Zeew D, Eckardt KU, et al. A trial of darbepoetin alfa in type 2 diabetes and chronic kidney disease. N Engl J Med. 2009; 361:2019–2032. [PubMed: 19880844]

5. Singh AK, Szczech L, Tang KL, Barnhart H, Sapp S, Wolfson M, et al. Correction of anemia with epoetin alfa in chronic kidney disease. N Engl J Med. 2006; 355:2085–2098. [PubMed: 17108343]

6. Fellstrom BC, Jardine AG, Schmieder RE, Holdaas H, Bannister K, Beutler J, et al. Rosuvastatin and cardiovascular events in patients undergoing hemodialysis. N Engl J Med. 2009; 360:1395–1407. [PubMed: 19332456]

7. Wanner C, Krane V, Marz W, Olschewski M, Mann JF, Ruf G, et al. Atorvastatin in patients with type 2 diabetes mellitus undergoing hemodialysis. N Engl J Med. 2005; 353:238–248. [PubMed: 16034009]

8. Suki WN, Zabaneh R, Cangiano JL, Reed J, Fischer D, Garrett L, et al. Effects of sevelamer and calcium-based phosphate binders on mortality in hemodialysis patients. Kidney Int. 2007; 72:1130–1137. [PubMed: 17728707]

9. Jamison RL, Hartigan P, Kaufman JS, Goldfarb DS, Warren SR, Guarino PD, et al. Effect of homocysteine lowering on mortality and vascular disease in advanced chronic kidney disease and end-stage renal disease: a randomized controlled trial. JAMA. 2007; 298:1163–1170. [PubMed: 17848650]

10. Sniadecki, J. Dziela. Warsaw: 1840. p. 273-274.

11. Brater DC, Daly WJ. Clinical pharmacology in the Middle Ages: principles that presage the 21st century. Clin Pharmacol Ther. 2000; 67:447–450. [PubMed: 10824622]

12. Kalantar-Zadeh K, Navab M. The importance of origins? Science. 2005; 309:1673–1675. [PubMed: 16150992]

13. Hill AB. THE ENVIRONMENT AND DISEASE: ASSOCIATION OR CAUSATION? Proc R Soc Med. 1965; 58:295–300. [PubMed: 14283879]

14. Mark SD, Robins JM. A method for the analysis of randomized trials with compliance information: an application to the Multiple Risk Factor Intervention Trial. Control Clin Trials. 1993; 14:79–97. [PubMed: 8500308]

15. Palmer SC, McGregor DO, Macaskill P, Craig JC, Elder GJ, Strippoli GF. Meta-analysis: vitamin D compounds in chronic kidney disease. Ann Intern Med. 2007; 147:840–853. [PubMed: 18087055]

16. Coyne DW. Vitamin D compounds in chronic kidney disease. Ann Intern Med. 2008; 148:969–970. [PubMed: 18559843]

17. K/DOQI clinical practice guidelines on hypertension and antihypertensive agents in chronic kidney disease. Am J Kidney Dis. 2004; 43:S1–290. [PubMed: 15114537]

18. Palmer BF. Renal dysfunction complicating the treatment of hypertension. N Engl J Med. 2002; 347:1256–1261. [PubMed: 12393824]

19. Kalantar-Zadeh K, Kilpatrick RD, McAllister CJ, Greenland S, Kopple JD. Reverse epidemiology of hypertension and cardiovascular death in the hemodialysis population: the 58th annual fall conference and scientific sessions. Hypertension. 2005; 45:811–817. [PubMed: 15699452]

20. Kovesdy CP, Trivedi BK, Kalantar-Zadeh K, Anderson JE. Association of low blood pressure with increased mortality in patients with moderate to severe chronic kidney disease. Nephrol Dial Transplant. 2006; 21:1257–1262. [PubMed: 16421161]

21. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet. 2002; 360:1903–1913. [PubMed: 12493255]

22. Cooper BA, Branley P, Bulfone L, Collins JF, Craig JC, Fraenkel MB, et al. A randomized, controlled trial of early versus late initiation of dialysis. N Engl J Med. 2010; 363:609–619. [PubMed: 20581422]

23. Kalantar-Zadeh K, Kuwae N, Regidor DL, Kovesdy CP, Kilpatrick RD, Shinaberger CS, et al. Survival predictability of time-varying indicators of bone disease in maintenance hemodialysis patients. Kidney Int. 2006; 70:771–780. [PubMed: 16820797]

24. Melamed ML, Eustace JA, Plantinga L, Jaar BG, Fink NE, Coresh J, et al. Changes in serum calcium, phosphate, and PTH and the risk of death in incident dialysis patients: a longitudinal study. Kidney Int. 2006; 70:351–357. [PubMed: 16738536]

25. Naves-Diaz M, Alvarez-Hernandez D, Passlick-Deetjen J, Guinsburg A, Marelli C, Rodriguez-Puyol D, et al. Oral active vitamin D is associated with improved survival in hemodialysis patients. Kidney Int. 2008; 74:1070–1078. [PubMed: 18633342]

26. Tentori F, Hunt WC, Stidley CA, Rohrscheib MR, Bedrick EJ, Meyer KB, et al. Mortality risk among hemodialysis patients receiving different vitamin D analogs. Kidney Int. 2006; 70:1858–1865. [PubMed: 17021609]

27. Kovesdy CP, Ahmadzadeh S, Anderson JE, Kalantar-Zadeh K. Association of activated vitamin D treatment and mortality in chronic kidney disease. Arch Intern Med. 2008; 168:397–403. [PubMed: 18299495]

28. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000; 11:550–560. [PubMed: 10955408]

29. Teng M, Wolf M, Ofsthun MN, Lazarus JM, Hernan MA, Camargo CA Jr, et al. Activated injectable vitamin D and hemodialysis survival: a historical cohort study. J Am Soc Nephrol. 2005; 16:1115–1125. [PubMed: 15728786]

30. Tentori F, Albert JM, Young EW, Blayney MJ, Robinson BM, Pisoni RL, et al. The survival advantage for haemodialysis patients taking vitamin D is questioned: findings from the Dialysis Outcomes and Practice Patterns Study. Nephrol Dial Transplant. 2009; 24:963–972. [PubMed: 19028748]

31. Teng M, Wolf M, Lowrie E, Ofsthun N, Lazarus JM, Thadhani R. Survival of patients undergoing hemodialysis with paricalcitol or calcitriol therapy. N Engl J Med. 2003; 349:446–456. [PubMed: 12890843]

32. Glymour, MM.; Greenland, S. Causal diagrams. In: Rothman, K.; Greenland, S.; Lash, TL., editors. Modern epidemiology. 3. Philadelphia: Lippincott Williams & Wilkins; 2008. p. 183-209.

33. Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. Int J Epidemiol. 2004; 33:30–42. [PubMed: 15075143]

34. Zacho J, Tybjaerg-Hansen A, Jensen JS, Grande P, Sillesen H, Nordestgaard BG. Genetically elevated C-reactive protein and ischemic vascular disease. N Engl J Med. 2008; 359:1897–1908. [PubMed: 18971492]

35. Kalantar-Zadeh K, Kovesdy CP. Clinical outcomes with active versus nutritional vitamin D compounds in chronic kidney disease. Clin J Am Soc Nephrol. 2009; 4:1529–1539. [PubMed: 19661219]

36. Streja E, Kovesdy CP, Greenland S, Kopple JD, McAllister CJ, Nissenson AR, et al. Erythropoietin, iron depletion, and relative thrombocytosis: a possible explanation for hemoglobin-survival paradox in hemodialysis. Am J Kidney Dis. 2008; 52:727–736. [PubMed: 18760517]

37. Regidor DL, Kopple JD, Kovesdy CP, Kilpatrick RD, McAllister CJ, Aronovitz J, et al. Associations between changes in hemoglobin and administered erythropoiesis-stimulating agent and survival in hemodialysis patients. J Am Soc Nephrol. 2006; 17:1181–1191. [PubMed: 16565261]

38. Shinaberger CS, Kopple JD, Kovesdy CP, McAllister CJ, van Wyck D, Greenland S, et al. Ratio of paricalcitol dosage to serum parathyroid hormone level and survival in maintenance hemodialysis patients. Clin J Am Soc Nephrol. 2008; 3:1769–1776. [PubMed: 18701614]
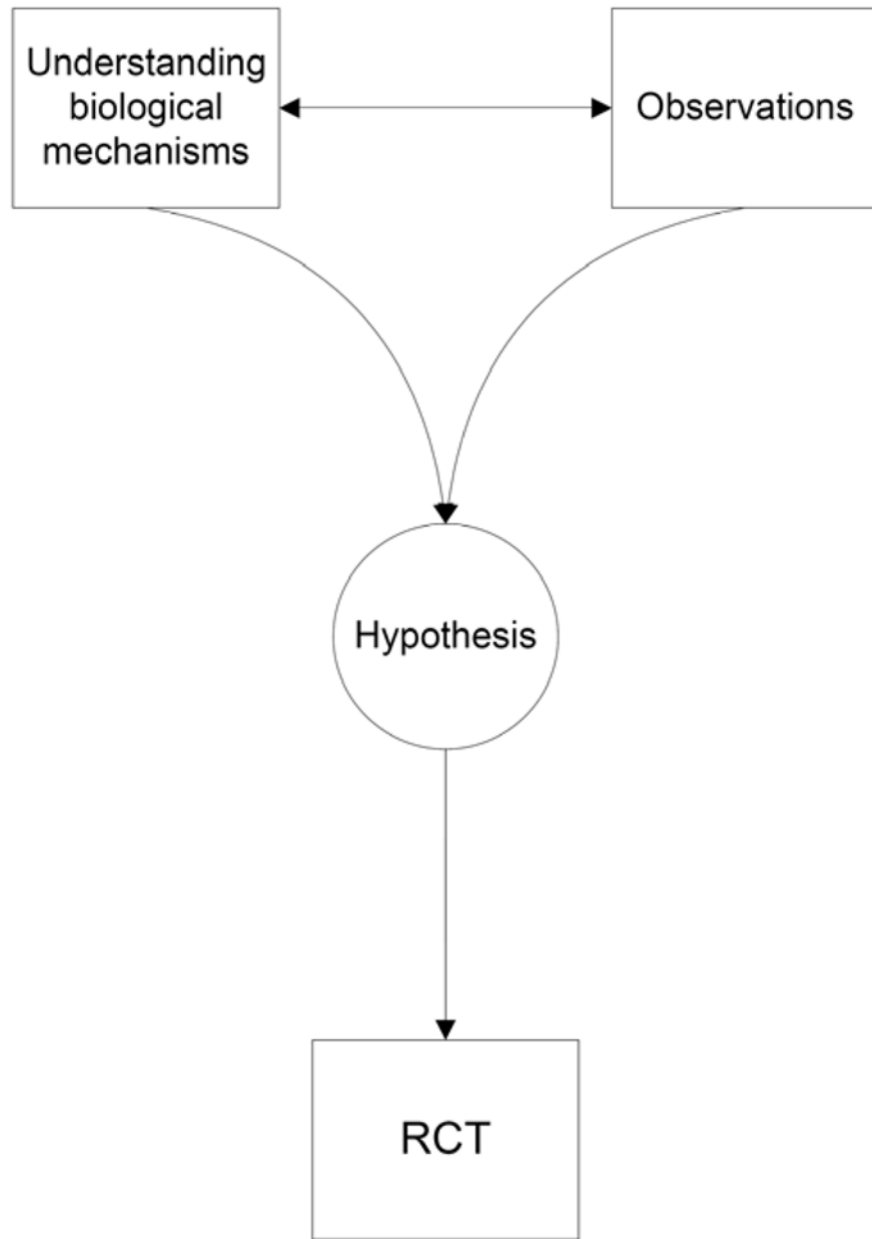
Figure 1.
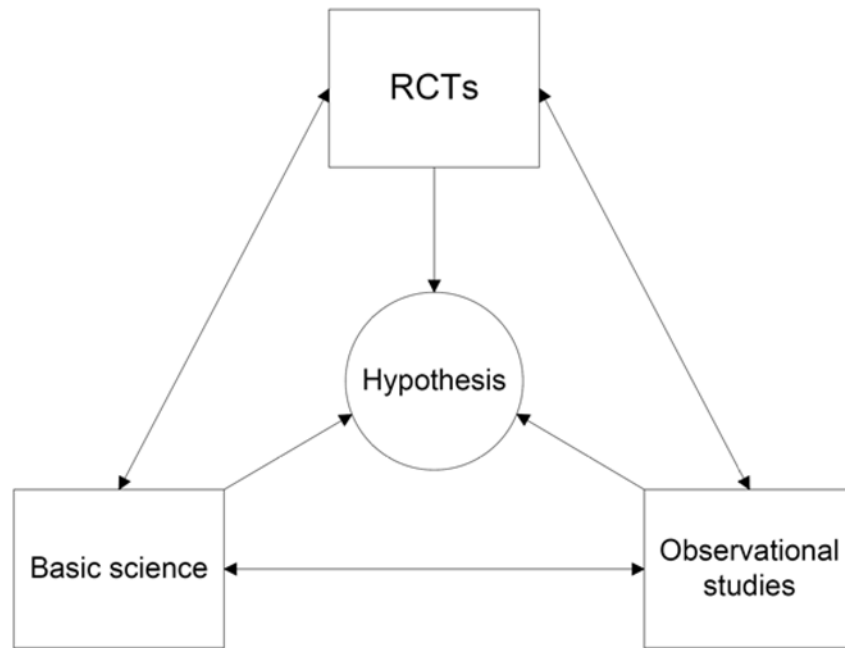Current ranking paradigm of scientific investigations.

Figure 2.
Suggested alternative ranking paradigm of scientific investigations.

Table 1

Hill's criteria of causal inference in epidemiology.

Criterion	Definition	Counterargument
1. Temporal relationship	Exposure precedes the disease.	None
2. Strength of association	The stronger the association, the more likely it is that the relation is causal.	Weak associations do not rule out causality. Furthermore, the presence of a strong unmeasured confounder could result in strong associations without true causation. Strength of association should not be mistaken for statistical significance.
3. Dose response	Increasing amount of exposure increases the risk proportionally.	Causality can also be present with mechanisms that have threshold effects in which case increasing exposure would not lead to further increase in effect. It is also possible that the presence of an unmeasured confounder could change the shape of the dose-effect curve in spite of a true cause-effect relationship.[38]
4. Consistency	The association is consistent when results are replicated in studies in different settings using different methods.	Certain biologic processes may be different under different circumstances, hence the lack of an association in one population does not rule out a cause-effect relationship in another.
5. Biologic plausibility	The association agrees with currently accepted understanding of biologic processes.	It is always possible that new mechanisms of disease are discovered; hence studies that disagree with established understanding of biological processes may force a reevaluation of accepted beliefs. An association that is seemingly biologically implausible may be merely highlighting a yet unknown mechanism.
6. Experimentation	The condition can be altered (prevented or ameliorated) by an appropriate experimental intervention.	A negative experiment cannot always rule out a cause-effect relationship. A positive experiment in one group of patients cannot be construed as proof of causality in all groups. Flawed experiments can lead to erroneous assumptions of causality.
7. Specificity	A single putative cause produces a specific effect.	A complex mechanism of action does not rule out a cause effect relationship, it merely makes it more difficult to explain.
8. Biologic coherence	The association is consistent with the natural history of the disease.	There may be unknown aspects of any diseases' natural history. Seemingly anomalous associations could be causally explained by such unknown aspects.
9. Analogy	There are similar associations in other populations or under different settings.	Biological differences between populations make it possible that the same exposure has different results depending on the circumstances. This does not rule out the possibility that associations in certain populations are due to cause-effect relationships.