The Routledge Companion to Philosophy of Medicine

Miriam Solomon, Jeremy R. Simon, Harold Kincaid

Observational Research

Publication details

Olaf M. Dekkers, Jan P. Vandenbroucke
Published online on: 25 Oct 2016

How to cite :- Olaf M. Dekkers, Jan P. Vandenbroucke. 25 Oct 2016, Observational Research from: The Routledge Companion to Philosophy of Medicine Routledge
Accessed on: 09 Aug 2023

Please scroll down for document

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
What Is Observational Research?

In medical research, a fundamental distinction can be made between *observational* and *experimental* studies. In experimental studies, investigators perform an action intended to interfere with the course of the disease or condition under study. The prototype is the randomized experiment, in which patients are randomly allocated to an experimental intervention and a comparator or a placebo. For example, patients with subclinical hypothyroidism (a condition in which thyroid function is slightly diminished) might be randomized between treatment with thyroid hormone substitute or an identical looking but inert placebo. With such an experiment, investigators could aim to answer the question of whether treatment improves cardiovascular risk associated with this condition. The patients would be treated for the purpose of the study, and arguably, many patients would not have been treated in the same way had they not been study participants (e.g., because their doctors do not think that their subclinical hypothyroidism importantly affected their health). The outcome of such an experimental study is a priori uncertain, and besides examples in which the treatment helped or made no discernible difference, there are examples of experimental treatments that turned out to worsen the disease course (Besselink et al. 2008).

In what is commonly called “observational research,” the investigator does not intervene to allot the treatment (more generally, the exposure of interest). Some say that a better term for observational research would be “non-randomized research.” However, we will stick with the term observational research as it is commonly used, as not all experiments are randomized. Observational research may take several forms, some of which the patients will be aware of (when questionnaires are to be filled in or additional blood samples are taken). In other instances, patients will not notice that their disease course is studied (e.g., when routinely gathered data from electronic medical records or large administrative health care databases are analyzed). An observational study can even be conducted several years after disease has occurred. For example, investigators may want to know whether hypothyroidism is associated with an increased risk for mortality. For this purpose, they could select all patients with a hypothyroidism diagnosis from a medical database, link these data to data regarding vital status, and compare mortality risk between patients with hypothyroidism and a control group without the disease.

Although in observational studies investigators do not intervene in the treatment of patients, they are not passive recorders of data. In almost all observational studies, substantial preparation is required, and decisions have to be made on what to observe and how to observe it. These decisions fundamentally influence what can be found and how those findings are interpreted (Popper 1957).
Observational studies can be primarily descriptive or comparative. In descriptive studies the main goal is to determine risks or prognosis in specified groups, without further analytical exploration of underlying causes. The proportion of individuals with a certain chronic condition in a specified population can be estimated (e.g., the proportion of people with diabetes in the Netherlands), or the risk of an outcome can be assessed in a population that has a (disease) characteristic in common (e.g., mortality risk after a specific type of surgery). Such descriptive studies are important for informing policy makers, patients and doctors; they are usually not comparative but give a broad idea about the likely outcome of a disease in a particular patient.

In comparative studies, groups are compared to determine the effect of an “exposure”; such exposures can be external agents (e.g., smoking, air pollution) or internal agents (e.g., a gene variant). Medical interventions can also be compared observationally. The crucial difference from experiments is that the type of medical intervention is not allotted by the investigator to study the effect of the intervention; the investigator observes what happens in the course of usual medical care, where different doctors make different decisions. The purpose of these comparative studies is generally to disentangle causal effects. A comparative observational study could compare mortality risks between two types of surgery for the same indication, aiming to determine which is the best procedure. Although this is not an experiment, the aim is the same as that of an experiment: to learn from a comparison. A comparative study could also compare cardiovascular risk between patients with hypothyroidism and a population without this condition, thereby aiming to identify a potential causal role of the condition in disease occurrence. Of note, descriptive studies can be presented as an apparent comparison, if risks are presented for several populations, such as when diabetes risks are presented for different countries. As long as the studies do not aim to disentangle underlying causes for the differences, such studies are still descriptive—even if they can give clues to potential causes (Pearce 2011). The distinction between descriptive and comparative studies is important, as the considerations regarding validity of the study’s conclusions differ between these types of studies.

Threats to Validity in Observational Studies: Bias and Confounding

In this section, we describe the two most common threats to the validity of comparative observational research. The results of a study are considered to be valid if, under ideal circumstances with an infinite number of participants and infinite resources for assessment, the result of the study would yield “the truth.” A study is not valid if, even with infinite resources, its results will still be wrong. The two notions under which the possibilities for being wrong are described are bias and confounding.

All studies in medicine rely on measurement and classification of the disease (e.g., breast cancer) or disease-related states (e.g., quality of life in cancer patients, survival without metastases). However, diseases and other study endpoints are prone to inadequate measurement or misclassification, potentially leading to bias in study results. If such a misclassification depends on whether or not the person has a particular exposure, comparing exposed and non-exposed persons can give biased results. An example of potential misclassification might be a study where the risk of pneumonia is assessed and compared between patients treated with chemotherapy and a control group from the general population. In case of clinical suspicion, chemotherapy patients are more likely to be admitted to the hospital, where X-rays will be made, whereas in the control group the diagnosis is more often based on the clinical evaluation by a general practitioner, and diagnosis is not confirmed with an X-ray. Bias may occur if the X-ray is more sensitive for the diagnosis compared to clinical evaluation only. Misclassification bias can occur in descriptive as well as comparative observational studies. It is important to note that misclassification occurs in
different degrees, wherein some diseases, diagnostic tools, or classification systems are more prone to misclassification than others. For example, it is hardly possible to misclassify death. Of note, misclassification is also an issue in randomized trials, even though blinding aims to prevent the misclassification being different between the compared groups.

As mentioned, in comparative observational studies the aim of the study extends beyond the mere description of data and investigators want to study the effect of a treatment on a disease course, or the effect of a potential risk factor on the occurrence of a disease (so-called etiologic study). Potential causes of disease are often called risk factors. The comparison among two or more groups is central to comparative studies, and the validity of this comparison needs close attention when assessing validity in observational studies. The basic idea of a comparative study is that the group without the risk factor under study is used to determine the disease course in the absence of the risk factor under study. By comparing an outcome between the two groups (with and without the risk factors), investigators aim to ascribe any difference in outcome to a difference in the risk factor (or treatment). If there is a difference, the conclusion would then be that the risk factor under study is indeed a cause of the outcome under study—moreover, the study will also describe how strong the risk factor is.

This leads to the second major threat to validity of comparative observational studies: confounding. Causal inference by making the comparison between exposed and non-exposed is only valid under the assumption that all other risk factors (other than the risk factor under study) are balanced—i.e., on average similar, between the groups compared—the underlying idea is the notion of ceteris paribus (everything else being equal). Confounding is the term used to describe groups being different at baseline with regard to prognostic factors. This problem can occur by two mechanisms: in etiologic studies, risk factors tend to cluster (e.g., alcohol drinking is related to smoking), whereas in observational studies on therapeutic effects, treatment prescription is related to disease prognosis (this is called confounding by indication). For example, patients with newly diagnosed diabetes are more likely to be treated with drugs if their glucose levels are very high, but only with general lifestyle advice if their glucose levels are low. The opposite can also occur, such as when patients with advanced-stage cancer are no longer treated with chemotherapy. Of note, doctors are trained to treat patients according to the perceived prognosis of the patient. This is crucial for patients’ care, but it hampers the comparison of therapeutic interventions by observational studies. As such, confounding (by indication) is often held to be the central problem in observational studies of therapeutics.

However, the problem of confounding is also problematic in studies of general risk factors. Suppose that investigators want to study the effect of a vegetarian diet on mortality risk, and they compare a group of vegetarians to a group of non-vegetarians. In this context, no confounding would mean that all factors (other than vegetarianism) that influence mortality risk are similar in these two groups. How likely is that to be the case? Not very, as vegetarians will probably have a healthier lifestyle in general, being physically more active, are smoking less, possibly being higher educated, etc. So, numerous prognostic differences will exist between these groups, and it will be difficult to ascribe a difference in mortality risk between the groups to a difference in vegetarian diet only.

In the analysis of comparative studies, statistical methods are often used to tackle the problem of confounding, to try to statistically restore the “balance” between the groups. The basic principle of such statistical models is that any comparison is done between groups of patients with similar confounder characteristics (e.g., the comparison between vegetarians and non-vegetarians is made separately for smokers and non-smokers—as there will be more non-smokers among the vegetarians—and afterwards the two comparisons are averaged over the smokers and the non-smokers to yield an overall comparison). Statistical models can indeed compensate for baseline differences between the groups, but only under the assumption that all
differences are known and perfectly measured. This condition is rarely met. How likely is it that we have measured all prognostically important differences between vegetarians and non-vegetarians? The investigators might use statistical methods to account for differences in blood pressure, but more subtle and difficult-to-measure differences (such as differences in driving style, physical activity, or mood) will probably not be measured and therefore not be incorporated into the statistical model; these factors will remain unbalanced when comparing these two groups. The final verdict regarding the causality of the association remains uncertain if unmeasured confounding is an issue in a study. The same goes for confounding by indication in an observational study of therapeutic effects; some baseline prognostic factors might be known and measured (such as age or cancer staging), but other risk factors might be known only very crudely (e.g., if the presence of diabetes is known for all patients, but information about duration and severity of the diabetes is lacking).

It should be emphasized that confounding is not a yes/no phenomenon but a matter of degree depending on the association under study (see Figure 22.1). For some associations, such as the association between a vegetarian diet and mortality, the confounding might be very large. At the other end of the spectrum, when studying whether a specific genetic polymorphism is a risk factor for myocardial infarction, the association will likely not be confounded as the segregation of genes is a random process (there are exceptions to this general rule, for example when two genes are in linkage disequilibrium) (Smith & Ebrahim 2004).

In the above discussion, we have dealt with the typical situation in which two (or more) groups are compared. However, verdicts about causality are not always made in exactly the same way. In some instances a study compellingly shows a causal effect, even if the study does not involve a control group. An example is the use of pancreatic islet transplantation in patients with type 1 diabetes mellitus. Type 1 diabetes is characterized by destruction of pancreatic islets, and therefore the inability to produce insulin. In an experimental study, eight

![Figure 22.1 Degree of expected confounding in etiologic studies](https://example.com/figure22.1.png)

---

240
patients received pancreatic islet transplantation, and after a follow-up period of one year, five patients were still not in need of insulin treatment (Hering et al. 2005). The study showed a therapy to be effective, even though no control group was included.

Why is this a convincing study? The reason is that we know for sure what would have happened to the diabetes patients in the absence of an islet transplantation: all diabetes patients would with certainty have continued their insulin therapy (assuming no deaths), as insulin is by no means capable of curing the disease, and spontaneous recovery is not reported. More generally speaking, a study without a control group can be convincing if two conditions are met: (1) we have a perfectly clear and certain picture of what would have happened in the absence of the experimental intervention, and (2) the difference between the known course without the intervention and the actual course with the experimental therapy (such as islet transplantation) is clear-cut and large. This also means that the known disease course should not have large fluctuations, as it is otherwise difficult to distinguish a treatment effect from fluctuation in the disease course. Both conditions are met in the islet transplantation study: the difference in effect is clear-cut (insulin dependence versus insulin independence), and the natural course of type 1 diabetes is characterized by a continuous need for insulin, so the difference is also large. Moreover, we know about the pathophysiology of diabetes, and the intervention counters the mechanism of the disease.

Causal Inference in Observational Research—Counterfactual Reasoning

Over the past decade, the above-mentioned problems in causal inference in observational research have led to attempts by epidemiologists to structure observational research according to what is called “counterfactual reasoning.” There is at present an extensive literature about this counterfactual view on epidemiology (Hernan 2004). In fact, from a philosophical perspective, this newer counterfactual view is likely a mixture of different philosophical views on causation, such as counterfactual thinking, interventionism, and contrastivism (Vandenbroucke et al. 2016). We will introduce this counterfactual reasoning by examples.

Suppose there is a heavy storm. We see a tree falling down, and we witness how the tree hits and ruins a parked car. It is not risky to argue that the falling tree ruined the car, as the whole causal chain was observed: storm, tree falling, car ruined. We can then reason counterfactually that “if there had not been a storm, the tree would not have fallen, and the car would not have been damaged.” Unfortunately, such obvious causal inference is hardly possible in the study of risk factors for diseases, for the simple reason that we are not in a position to directly observe the causal chain that leads to a disease. There are several reasons for this (e.g., some risk factors may take years to lead to a disease), but the fundamental problem is that we have few tools to directly observe medical causes in action. We cannot see high cholesterol levels actually causing heart disease; we have no observations that show high glucose levels in patients with diabetes causing neuropathy—in marked contrast to the tree that we see falling and crashing onto the car. Causal inference (i.e., inferring that an association is causal) is therefore necessarily also based on judgment. So the central question is how to infer a causal relation from data when we cannot directly observe the causal process and how such inference can be justified. Here philosophy enters the picture.

According to the idea of counterfactual reasoning in epidemiology, if we study the effect of a risk factor or treatment on a disease in a specific population, we need to know what would have happened in the absence of the risk factor/treatment, the so-called counterfactual situation. Think about the use of a control group in medical research (the non-exposed group as mentioned above). The control group is used to quantify what would have happened to a specific study population in the absence of the risk factor or experimental treatment under
consideration. Inferring causality would indicate that a disease occurs or at least is more likely to occur, in the presence of the risk factor, whereas it would not have occurred, or would have been less likely to occur, without the risk factor. Using control groups to define what would have happened in the absence of the risk factor or treatment under study resembles a counterfactual way of defining causality: “X causes Y because the counterfactual ‘if not X then not Y’ is true” (Menzies 2001). Intuitively, such a counterfactual theory has some appeal, as it resembles reasoning in ordinary life. A simple statement with causal connotations, such as “he was late (Y) because it was snowing (X),” implicitly assumes knowledge of what would have happened “if it had not been snowing”; it assumes that the counterfactual “if it had not been snowing, then he would not have been late” is true.

Despite its intuitive appeal, applying a counterfactual framework to observational studies has some drawbacks. First, a risk factor can be a cause of a disease even if the counterfactual “if not X then not Y” is not true. In other words, a risk factor can be a disease cause even if in the absence of the risk factor (the counterfactual situation) the disease still occurs. This is the case when, in the absence of the risk factor under study, a “back-up cause” makes the disease occur. For example, if in a patient a stroke is caused by smoking, then it might be that in the absence of smoking the stroke still would have occurred but be caused by the patient’s diabetes (as a back-up cause). This so-called problem of preemption shows that counterfactual dependence is not a necessary condition for causation. A second problem is that the counterfactual framework does not distinguish between meaningful causes and causes that fit the counterfactual definition but that are obviously irrelevant (Broadbent 2013). Think about causes as the presence of oxygen or the existence of a temperature on Earth that is compatible with life: these factors would be causes of a myocardial infarction within a counterfactual framework. It would make more sense to call these conditions “feasibility conditions” rather than causes, as they set the stage for all causes to act. A third problem is that the type of counterfactual thinking that is advocated in epidemiology is often explicitly limited to “humanly possible actions” and limited to external interventions. In contrast, “states,” such as being obese or having a certain genetic trait, fall outside of this type of counterfactual thinking in epidemiology. For a more detailed critique, see Glymour and Glymour 2014.

Fundamentally, a counterfactual theory provides no rules or tools for justifying causation in observational studies. More formally, the counterfactual framework is semantically appealing as it seems to give a well-described meaning to causation that fits with some types of observational studies, but it is hardly of epistemic value, as it does not add to justification of causal claims. It might be true that if we know the counterfactual outcome for certain, we can infer causality, but this only shifts the problem, as the new question arises how to know whether the counterfactual situation is in fact truly known. By definition, we cannot know the counterfactual outcome for certain, as this counterfactual situation does not exist.

Think again about the impact of a vegetarian diet on mortality. We do not know what the outcome in the vegetarians would have had they been carnivores, and the closest we can get is to perform a study including a comparison group. How can we know with certainty that the non-vegetarians represent the outcome the vegetarians would have had they not been vegetarians? We only have some certainty if there is no confounding or if confounding is adequately dealt with statistically. This means that a counterfactual reasoning cannot in itself justify causality in a specific observational study. Claiming that X is a cause of Y because the counterfactual statement “if not X then not Y” is true adds nothing to a claim that something is the cause of a disease. Such a claim can only be put forward if the central assumptions of epidemiologic assumptions are met—i.e., the assumptions of no bias and no (unmeasured) confounding.
In randomized trials the situation is slightly different, as the randomization produces a control group that can be seen as a tool to determine the counterfactual outcome. Why? Because the randomization procedure is designed to equally distribute prognostic factors between groups. As treatment decisions in a randomized trial are not influenced by prognosis, confounding by indication is thereby circumvented. However, randomization is relying on a chance procedure, and there is no guarantee that for a particular trial the groups still will be balanced. Randomization gives only an “asymptotic” reassurance (i.e., that in the long run, if an infinite number of trials is done, on average they will show truth). The difference between the actual outcome and the expected outcome can be explained by a game of throwing two dice. The prior probability of throwing any number below 12 is 35/36. Now suppose the dice are thrown and the outcome is 12. Referring to the prior probability and complaining that this was not expected is no longer meaningful. Unlikely events can occur by chance. This is similarly true for a randomized experiment: the randomization provides an expected balance but not a guarantee of balance in a particular study (Senn 2013).

Confounding by indication is a central problem in observational studies of intended effects of interventions (i.e., effects that one hopes to see). The question is whether confounding plays a similar role in studies of unintended effects, such as side effects of treatments. Suppose two antibiotics are available for a specific disease (e.g., urinary tract infection). When studying treatment effects (e.g., whether symptoms of the infection recede more quickly or not), confounding by indication is an issue, as the one antibiotic might more often be given to patients with more severe symptoms. But what about side effects such as allergic reactions? As doctors cannot predict such side effects, for a study that compares the two drugs with regard to side effects, confounding is not an issue (Vandenbroucke 2004). This is graphically illustrated in Figure 22.2 (Schneeweiss 2007): an intended effect of Coxibs (fewer gastrointestinal adverse effects compared to standard painkillers) leads to selective prescribing to people with previous gastrointestinal problems and is strongly confounded in contrast to the unintended effect of causing cardiovascular disease, which was unknown at the time of marketing, so that cardiovascular risk factors were not taken into account on prescribing. To further guarantee the absence of confounding, patients with a known history of side effects (e.g., a known history of allergic reactions to drugs) can be excluded from the analysis (Schneeweiss et al. 2007). These theoretical considerations that confounding by indication is usually not an issue in the study of side effects are supported empirically by several studies showing that observational studies on side effects show results similar to results from randomized trials (Golder et al. 2011).

This idea of no confounding for unexpected events does not only apply to studies of drugs. In fact, most causes of disease that are investigated successfully in epidemiology could be called unintended effects: lung cancer is an unintended effect of smoking. One starts smoking, say at age 11, without any knowledge as to one’s probability to develop lung cancer, and for reasons that have nothing to do with one’s risk of lung cancer. In the same way, exposure to asbestos and its subsequent risk of mesotheliomas is due to a choice of work at a particular factory, at the time (in the 1950s) when the risk of exposure to asbestos was not known at all. Similar reasoning applies to leukemia in children being caused by prenatal X-ray exposure, which a long time ago was routinely applied to assess the width of the mother’s pelvis before birth (Vandenbroucke 2004). These are examples where not much confounding is expected, as the outcomes studied are unintended and unpredictable. Mind that confounding still needs consideration, as risk factors might cluster.
Should Only Interventions Be Studied?

Experimental studies are characterized by studying interventions (in a broad sense), whereas observational studies can also study risk factors. Some risk factors are characterized by an unclear onset (obesity, hypercholesterolemia) or represent a lifelong state (gender, genetic make-up). It should be emphasized that studying risk factors and not interventions poses one important restriction on the interpretation of study results: one might very well come to the verdict that some risk factor is causal, but that does not yet guide the type of intervention that is needed to reduce or normalize the risk, nor does it predict what will happen if the risk factor would be wiped out totally. Say that researchers have performed the perfect study (no confounding), with a valid estimate, of the effect of obesity on mortality, and they conclude that mortality risk is doubled in the obese compared to the non-obese. This does not automatically mean that intervening on obesity (e.g., by a highly effective diet) normalizes mortality risk among the obese. It might well be that, despite our spectacular diet being truly effective, obesity has induced irreversible organ damage, making it impossible to normalize mortality risk at all. Moreover, dieting might be less effective than exercise, or either might be less effective than a combination. Even more difficult to foresee, people might change their behavior when on a diet (e.g., by starting to smoke). Others might become depressed by the intervention and turn to alcohol. This means that, even assuming etiologic causality can be inferred from observations, such causal inference cannot directly be translated into an estimate of the effect of specific interventions on the risk factor (Greenland 2005).
Thus, from a health policy perspective, studying risk factors has disadvantages compared to studying interventions, as the latter can inform treatment policy directly. Randomized studies on interventions have the additional advantage that the justification of causal claims is more straightforward compared to observational studies. Does this mean that we should only study interventions and abstain from studying hypertension, hypercholesterolemia, obesity, gender, genes, and socio-economic status because these are not interventions? A view stating that only interventions can be studied neglects the fact that the well-conducted study of risk factors is an important step in identifying targets for interventions (Glymour & Glymour 2014). If we had no clue that obesity increases mortality risk, we would not have thought to perform studies that intervene on obesity. In a world where no observational studies of risk factors were performed, nor their potential causality assessed, researchers likely would have no idea what conditions to intervene on.

### Causal Inference: The Role of Integrating Evidence

As discussed, arguments about the comparability of groups are important for causal inference. We can, however, never be certain that all measured and unmeasured prognostic factors are equally distributed—not even in a randomized trial, because chance baseline differences remain possible. This is to a certain extent a limitation of current knowledge, as we have no guarantee that we know all factors that are prognostically relevant to a specific outcome. Moreover, chance cannot be ruled out as a possible explanation for results in a single study. One could argue that given this inherent uncertainty, causality should never be judged based on a single observational study. One way forward is to place the results from studies in a much broader context. Until now, whether talking about randomized studies or about observational studies, we have mainly considered numerical results from individual studies. However, these data need integration for causal inference. In general, empirical studies are interpreted in the light of other knowledge. For example, it was found that certain compounds of tobacco smoke induce mutations at a p53 mutational hotspot in epithelial lung cells. Without epidemiological knowledge about smoking and lung cancer, searching for such carcinogenic mechanisms in tobacco smoking would be nonsensical. Even more to the point, if this particular mechanism would not have “worked,” the basic scientist would not have concluded that smoking is not a cause of lung cancer, but she would conclude that she has been looking at the wrong mechanism. Thus, basic science depends on epidemiology for interpretation of studies (Vandenbroucke 1998). In exactly the same way, epidemiology needs input from the outside: numbers and risks can only be collected and interpreted meaningfully in the light of reasoned arguments about explanations and mechanisms. One cannot set up a randomized trial and decide on a dosage and a particular type of patient without strong outside knowledge (e.g., from basic science, animal research, general physiology, or consistent observations on humans) that can guide our understanding about the likely mechanism of the disease and the likely mechanism of drug action. The same is true for observational research on risk factors; the epochal studies on cardiovascular risk factors like serum cholesterol were in part based in mice experiments about producing atherosclerosis. Thus, deciding how and what to study, and rendering the ultimate verdict that something is a cause, requires a judgment based on results from various types of research.

Ultimately, there are no rules for combining the results from different studies into one final judgment. Ruling out alternative explanations is one key way to help this judgment. For example, findings about the influence of smoking habits on pregnancy outcomes could be questioned, because it might be argued that perhaps women who smoke during pregnancy...
will also be different in a host of other factors (confounding by lifestyle, nutrition, or socio-economic class). However, when pregnancy outcomes were related to smoking by fathers, almost no effect was found. This is compatible with the idea of a direct effect of smoking in mothers (Howe et al. 2012). Such a way of reasoning is called triangulation of evidence. In other instances, a pathophysiologic mechanism tilts the scales: a contentious debate on the question of whether a new type of oral contraceptives led to more venous thrombosis than older types of contraceptives was strengthened by a finding of a stronger effect on coagulation by the newer contraceptives, showing the epidemiologic findings to be in line with basic science (Vandenbroucke et al. 2001).

This means that in observational research, and actually in all research, different types of reasoning play a role when coming to a final verdict: quantitative contrast-thinking, counterfactual thinking, thinking about mechanisms, and reasoning about the conditions under which an exposure might cause an outcome to occur. In observational studies it seems apparent that its conclusions rely on all of these types of reasoning. Randomized trials give the impression of standing on their own, given the single intervention that is studied experimentally. However, also in randomized trials, in their set-up and interpretation, ancillary knowledge from other science is crucially included. Lastly, all causal inference is ultimately “inference to the best explanation” and remains an inference under uncertainty. In the end, the judgment that something is a cause always remains a risk, as judgment is fallible. People try to escape this situation with diverse methodological strategies, making the rules for establishing causal inference ever stricter. Still, regardless of strategy, judgments about causality will, for the foreseeable future, remain grounded in the integration and weighting of different types of relevant research and theory.

References


OBSERVATIONAL RESEARCH


Senn, S., 2013. Seven myths of randomisation in clinical trials. Statistics in Medicine, 32(9), 1439–1450.


Further Reading


Senn, S., 2013. Seven myths of randomisation in clinical trials. Statistics in Medicine, 32(9), 1439–1450.