The THAI Journal of SURGERY 2021;42(1):3-13.
Official Publication of The Royal College of Surgeons of Thailand

Special Article

Are Medical Doctors Scientists? Causal Inference Based on Observational Data

Panuwat Lertsithichai, MD

Department of Surgery, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

Abstract We present arguments for a view of what a clinician-scientist might be like, in the setting of everyday clinical practice. Clinical observations of good quality and completeness may, under suitable context or research framework, be viewed as evidence for claims of causation. Historical examples of clinicopathological observations in which causal claims were made and later confirmed, are briefly mentioned. We present in some detail a more recent causal inference framework in statistics, which is increasing accepted in practice, as well as a brief introduction to causal diagrams. Finally, we briefly present a view of a possible use of causal inference in the near future.

Keywords: Medical doctors, Clinician-scientist, Causal inference, Observational data

Introduction

To many medical doctors, the question of whether medicine is scientific has an obvious answer: of course, it is! Is not medicine, at least in the common, or conventional, form ("allopathic medicine"), based on anatomy, physiology, pathology, biochemistry, molecular biology, biostatistics, etc., all of which are well-known sciences? So the question seems settled. But what if the question is posed to clinicians, and specifically phrased as, "is *medical practice* scientific?" – should the answer still be yes?

Now the question seems a bit more difficult to answer, as there may be several ways to look at it. Some will say that, yes, medical practice is scientific as it uses the "scientific method" in the diagnosis and treatment of patients. It is difficult to define scientific method in a way which satisfies most people, but even if such a method can be satisfactorily defined, is it relevant to medical practice? After all, the aim of medical practice is the diagnosis and treatment of the patient, or prevention of disease, not on obtaining generalizable knowledge. Also, using a methodology similar to that of scientific research does not guarantee valid clinical reasoning, and some practitioners still claim that medical practice is more of an Art. Therefore, I think this argument is erroneous and perhaps misses the point.

Some will say that medical practice is scientific because the basis of medicine is scientific, mirroring the argument in the first paragraph. And general questions of diagnosis, treatment and prevention (in terms of generalizable knowledge, not focused on any particular patient) are often resolved using scientific methods (i.e., through scientific research). This seems valid as an argument if one considers, for example, teaching, engineering, and even coffee roasting as being scientific, since these activities all have some scientific basis.

If we accept medical practice as being scientific, then are medical doctors scientists?

Received for publication 10 March 2021; Accepted 15 March 2021

Correspondence address: *Panuwat Lertsithichai, MD, Department of Surgery, Ramathibodi Hospital, Mahidol University, Bangkok,Thailand; E-mail: panuwat.ler@mahidol.ac.th*

Some might say yes. Then they must say yes to teachers, engineers and coffee roasters as being scientists. Some might not go so far (my apologies to teachers, engineers and coffee roasters who might be offended by this). The point is, doctors are not scientists (as teachers, engineers, and coffee roasters are not) if the main aim of their profession is not to obtain novel, or new, generalizable knowledge. The latter is what scientists do. And – further apologies to some people I will now refer to – "scientists" working in the industries as laboratory technicians are not scientists, despite using scientific instruments – they are of course technicians (and doctors can be viewed as technicians in this sense, albeit one requiring long training and certification).

Why the fuss about all this naming, you might ask? Is it ever productive to argue about names? Well, no. Unless the argument is actually about a new way of looking at and doing things. Unless naming is just the first step towards that goal. Indeed, I will now argue that in most of what we do, in whatever profession, we can be scientists. Doctors (teachers, engineers, and coffee roasters) are scientists if they do certain things and look at themselves in ways that I will now describe. And if the medical profession and society accept certain agreements, the profession can be truly scientific and doctors will be scientists in the full sense of the word.[1]

How Can Doctors Be Scientists?

Medical doctors can make and record accurate clinical observations. With a supporting system of disease registries or specialized data capture protocols, data so generated can be used for rigorous observational research. With prespecified questions and a wide-ranging knowledge-based agenda, specific clinical observations suitably analyzed can lead to new knowledge made generalizable to other institutions or future patients. Newer ideas can be added to the agenda as required, and progress reports should be written on a regular basis, and published when important findings arise. Office staff should be available as research assistants. Thus, all clinicians can be scientists in this sense. If clinicians see themselves as being obligated to do all this, they are scientists.

With new frameworks for viewing certain observational data as having causal value, observational data can lead to reliable causal inferences (subject to reliable statistical analysis).[2] This causal framework, for example, based on the relatively recent idea of Bayesian causal networks (see later, below), can be readily applied. If it be agreed with the community in which the medical institute is embedded that medical research is essential, then, given the minimal risk (to patients) of observational research subject to applicable privacy laws, such research can be done without routine ethics committee oversight, not dissimilar to quality assurance surveys and reviews. The community must acknowledge that clinical observations made during clinical encounters for diagnostic and treatment purposes are also possibly subject to analysis with the aim of generating generalizable knowledge. Thus, informed consent specifically for research is not always required.

Experimental clinical research, specifically randomized controlled trials (RCTs), can be institutionalized as well. Again, certain agreements and understanding with the embedding community is mandatory. Informed consent, with appropriate and full documentation, and ethics committee oversight are required only for research on new and unproven interventions. For research questions requiring randomization while not making a difference to clinical management and treatment efficacy as currently known, e.g., comparing established "equivalent" therapies, randomization can be done (without a bloated informed consent process as is presently required) through routine randomization software, although a protocol should be publicized or publicly approved beforehand. Once so institutionalized, RCTs should be easier to implement during routine clinical practice, taking no more time than the usual informing of risks & benefits of various established treatment options. But the doctor must mention to the patient that an RCT is going to be implemented. As an acceptable, or ethical, motivation, at least in the initial stages implementing publicly acknowledged institutionalized research, each treatment option can be made available at no cost (perhaps supported by the institute or state). If a patient prefers otherwise, he or she is excluded, and treatment is chosen based on patient preference. The only documentation needed may be as a note in the electronic registry. If further restrictions must be made during follow up, or if certain interventions must later be withheld or instituted, not based on standard clinical practice but on research mandate, then again full informed consent must be obtained, under ethics committee supervision, and any extraneous research-related interventions must be provided free of charge. All doctors can easily participate.

These are a part of the main ideas. There are difficulties with these proposals, but major obstacles include the feasibility of an institute and community-based long-term clinical research agreement, a trust in clinician responsibility, a good or ethical management of conflicts of interests, and public interest and acceptance.[3]

Historical Examples Of Clinician-Scientists

There are many important historical examples where clinicians are scientists (in the modern sense) in their own right, making valid causal inferences based on accurate observations. In fact, these observations had so much influence that later clinicians who fancied themselves as being scientists (perhaps not in the modern sense) modeled their scientific approach on these famous observational studies, without adequately understanding the associated biases and specific contextual features, and, equally important, not having the tools at hand for controlling these biases even if so recognized. As a consequence, clinical science did not advance as far as it should until relatively recently.

Some selected examples can be provided. The first two are well-known.[4] Hippocrates of Cos (~ 460 – 370 BC), to whom is attributed a collection of writings known as the Hippocratic corpus, showed in a seemingly haphazard series of cases or case vignettes that illnesses have repeatable and recognizable patterns, and that fairly accurate prognostication (or clinical prediction) can be made. There was no need to invoke supernatural, and thus unknowable or unobservable, explanations. Fevers are regularly observed to have ternary (every 3rd day) and quaternary (every 4th day) spiking patterns. Certain observable clinical manifestations can predict almost certain death (e.g., the Hippocratic Facies). Some diseases occur more frequently in one season, others in another. Galen of Pergamum (AD 129 – 200+) observed that section of the recurrent laryngeal nerve reliably produced abnormal phonation in animals, and bilateral section caused asphyxia. Galen, of course, also seemingly observed or inferred what later proved to be a great many errors due to mistaken assumptions, which were passed down for over a millennium before being corrected by later investigators.

One clinician-scientist I will focus on in a bit more detail is Giovanni Battista Morgagni (AD 1682 – 1771; Figure 1). Imagine spending, altogether, over 50 years observing and recording in detail almost 700 patients who later died, and correlating clinical observations with postmortem examination results. That, in essence, was what Morgagni did. This effort was probably the first major systematic clinicopathological investigation in the world. There had been several earlier reports, of smaller series of patients who were dissected after death, often observed inaccurately or reported for their extreme deviation from the normal, and had been collected together in an unsystematic way into a book (e.g., the *Sepulchretum sive anatomica practica*) which strongly influenced Morgagni. But Morgagni's work went far beyond anything existing before, and was truly monumental. His observations were accurate, meticulously analyzed given the time period, presented in a form of induction – deduction arguments in a series of letters, and published when the author was 80 years old, 10 years before his death.[5,6]

This work was "The seats and causes of disease investigated by anatomy" (*De sedibus et causis morborum per anatomen indagatis*), published in 1761[7] (Figure 2), written while Morgagni was a Professor of Medicine and later of Anatomy at the University of Padua, Italy. It is the true founding document of pathological anatomy, or gross pathology. It was over 1,300 pages in length, published in 5 volumes arranged by organ-systems. It emphasized, in a convincing fashion, that diseases are mainly localized in certain organs, correlating strongly with clinical symptoms and signs. It ushered in the age of organ-based disease pathology and practically ended the ancient humoral theory of diseases (as well as the idea of astrological influences on health!). Some sample observations from the books are in order.

In patients with cardiac and pulmonary symptoms, details of clinical findings were closely correlated with autopsy findings, in most cases describing without doubt conditions and diseases modern-day doctors would recognize.[5] Valvular heart diseases were described in great detail, some recognizable today as valvular endocarditis along with its symptoms and signs. Symptoms recognizable as angina pectoris was correlated with blocked coronary artery. The first description of cor pulmonale showed an understanding of the effect of the heart disease on the lungs. Patients who died of difficulty of breathing had autopsy findings described so accurately, in a first ever reported extensive pneumonia with hepatization, that those passages could be used in a modern-day textbook. Morgagni described cases of aortic arch aneurysm, one eroding through the sternum from which the patient bled to death.

Figure 1 Giovanni Battista Morgagni on the frontispiece of his book, *De sedibus et causis morborum per anatomen indagatis*, 1761

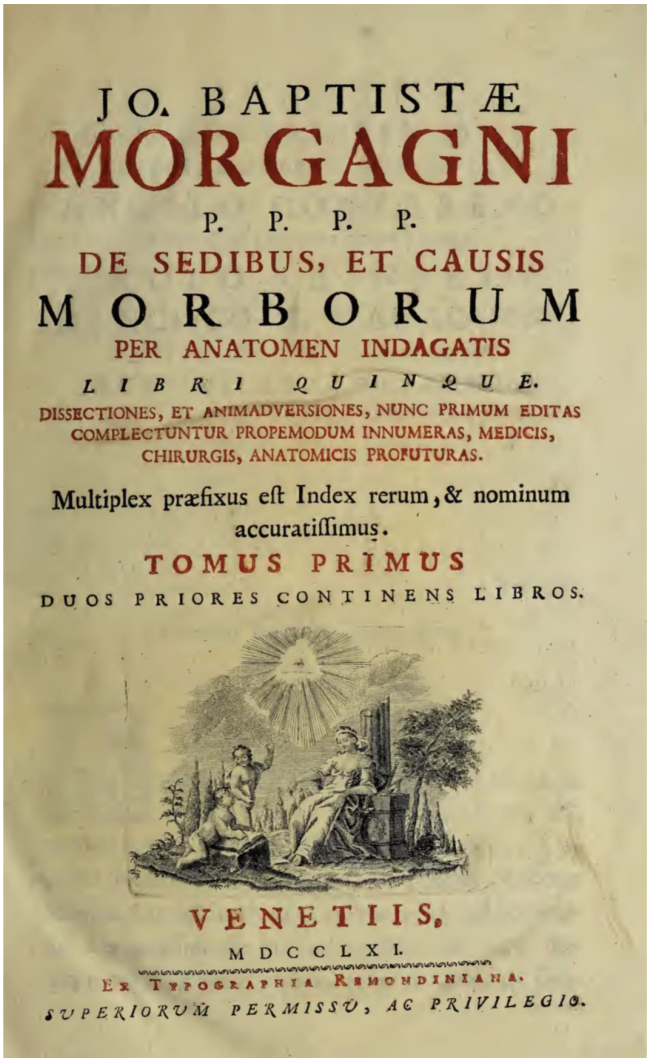

Figure 2 The title page of the book, *De sedibus et causis morborum per anatomen indagatis*, 1761

He pointedly ascribed the lesion to syphilis and described a possible pathological process. Neurological cases were also well presented and described.[6] Cases of cerebral aneurysm and rupture were directly linked to neck pain or stiffness, and subdural hematomas were described in which Morgagni speculated that trepanation might have been beneficial. Vascular dissection in the cerebellum was linked to respiratory problems and incontinence, in brain trauma the coup and contre-coup lesions were described, and head trauma was linked to epilepsy. All of these and much, much more; although it must be said that Morgagni also did some simple experiments to test his ideas.

Morgagni's circumstance is exceptional in that extreme contextual features, in this case patients so ill they died soon after consulting their doctors, are ripe for observation-based causal arguments which are more or less valid. By this I mean that the illnesses are so severe that postmortem examination will reveal clear-cut organ pathology. (In causal inference terminology, these relatively few observations provided highly probable necessary and sufficient information for causation). The causal relation between the diseased organ and illness can be reliably inferred, based on observations alone. Thus, a theory of diseases can be constructed, and tested with future observations. There should be no confounders strong enough to explain away the observed, inferred, causal relations. But in most real-world clinical situations of interest, clear cut causal inference cannot be obtained in such a straightforward manner. What should then be the appropriate theoretical framework or basis for causal inference?

A Theoretical Basis Of Causal Inference

The intuitive idea of causal inference can be illustrated with the following example. The statement, "P would have lived had the car not hit him" gives the idea behind "the car **is the cause of P's** death". That is, if there is an alternative version of events, where the car did not hit P, he would now be alive. Thus, it is the **imagined possibility of alternative events, i.e., a counterfactual,** involving the *same person or object at the same point in time*, that is the basis of causal inference.[8,9] We call this the "counterfactual" basis of causal inference. This type of causal reasoning happens, at least subconsciously, to everyone all the time. It is essentially the basis of experimentation: actively change one factor at a time, keeping everything else constant, and observe the outcome difference, assuming time-invariance.

The counterfactual idea is present in the randomized controlled trial (RCT), but only approximately. Because a person or biological subject cannot be used twice (in most situations) in a trial due to significant irreversible changes between interventions, or because the same initial conditions cannot be replicated for any subject at different time points, and, of course, all interventions cannot (usually) be performed at the same time on one subject (even two limbs are not identical), one practical solution is to observe *groups of subjects* randomly allocated to different interventions. This way, all groups, each sufficiently large in number, are approximately or on average the same to begin with. Then, on average, the subsequent observed differences among groups occurring within the same time frame can be attributed to (i.e., "caused by") the differences in the intervention. Thus, a group-level estimate of causal effects is used in place of the individual-based estimate. This is the essence of the RCT.

But the counterfactual idea does not necessarily have to be (approximately) realized in randomized controlled trials, i.e., within an experimental context. Indeed, frequently it cannot or is difficult to be done in practice. Instead, readily available observational data or information obtained during routine clinical work can be used for causal inference; but a suitable theoretical framework within which valid inferences can be made must be found. The first thing is to realize that in observational studies causally inferred connection between any two events are often invalidated by the presence of confounding. Without confounding, or with appropriate management of confounding under a suitable model, valid causal inferences can be achieved in observational studies. After all, humans must have been making observational causal inferences since the stone age, long before the invention of science and seem to have survived quite well.

Second, confounding effects as well as causal effects can be modeled within a graphical framework. Any model of a process where causal inferences are to be made can be written or drawn as a Bayesian causal network. Any conceivable event or "risk factor" or "confounding factor" related to an outcome event can be so modelled and the type of connection to one another will determine how the various interrelated events are to be "adjusted" to infer causality between events of interest. This graphical approach can augment standard statistical models to obtain valid causal inference.

Counterfactuals And Causal Inference In Statistics

Suppose we are looking at the effects of a binary treatment variable X on a binary outcome Y, for example, on the recurrence of cancer (recur, coded as 1; not recur, coded as 0). These treatments (values of X) are a new intervention, coded as 1 and a control, coded as 0. The objective of the study is to compare the effect of the new intervention with that of the control, in terms of cancer recurrence. In a counterfactual view of causality, for a given patient u, the causal effect of the new intervention, relative to that of the control, can be measured by the difference[8,9]

$$\Pr(Y_{x=1}(u) = 1) - \Pr(Y_{x=0}(u) = 1)$$

where $\Pr(Y_{x=1}(u) = 1)$ is the probability of recurrence ($Y_{x=1} = 1$), in patient u, given new intervention $X = 1$, and $\Pr(Y_{x=0}(u) = 1)$ is the probability of recurrence ($Y_{x=0}(u) = 1$), in the same patient u, given control $X = 0$. More commonly, the causal effect can also be measured as an Odds Ratio (OR)

$$OR = \frac{\Pr(Y_1=1)/(1-\Pr(Y_1=1))}{\Pr(Y_0=1)/(1-\Pr(Y_0=1))}$$

taking values between 0 and infinity, with an OR value of 1 interpreted as no effect, and values less than 1 as a beneficial effect (i.e., recurrence is less likely) of the new intervention $X = 1$. In this view, a Bayesian statistical approach is more appropriate since the probability is interpreted as a personal probability and not a relative frequency.

It can be seen at once that some of these probabilities are never observed in reality, and hence these individual causal effects can never be observed. This is because for a given patient u, if he or she is given the new treatment $X = 1$, then only $Y_{x=1}(u)$ is observed, not $Y_{x=0}(u)$. Conversely, if he or she is given the control treatment $X = 0$, then only $Y_{x=0}(u)$ is observed, not $Y_{x=1}(u)$. As an illustration, in table 1 we show observed outcomes and treatments in a hypothetical group of 10 patients, 5 in each actual-assigned treatment group, along with possible counterfactual, or *potential*, outcomes. For example, for patient 4 who is actually assigned to $X = 1$, the outcome in column 1, which is the actual observed outcome, is $(Y_{x=1}(4)|observed\ X = 1) = 0$; and for the same patient, in column 2, is $(Y_{x=0}(4)|observed\ X = 1) = 1$, and this latter outcome is counterfactual. In table 2, only actual observations are shown. Also shown is a confounding variable, stage of cancer Z (1 = lower, 2 = higher, stage) prior to treatment.

If individual causal effects (e.g., table 1) cannot all be observed (e.g., Table 2), then how is the causal effect estimated? The answer is to use the group-level estimates of the causal effect (sum over u). Instead of looking at individual, or within-row, differences in outcomes (possible only in Table 1), we look at the group, or between-column, differences (possible in both tables). In table 2, the observed group-level probability of recurrence can be estimated for columns 2 and 3 as, respectively,

$\widehat{\Pr}(Y = 1|X = 1) = \frac{2}{5} = 0.4$ (i.e., under the new treatment $X = 1$); and

$\widehat{\Pr}(Y = 1|X = 0) = \frac{1}{5} = 0.2$ (i.e., under the control treatment $X = 0$).

where we drop u and subscripts from the probabilities, and the hat symbol denotes estimation. The estimated odds ratio of new treatment vs. control is

$$\widehat{OR} = \frac{\frac{0.4}{1-0.4}}{\frac{0.2}{1-0.2}} = 2.67\ ;$$

which seems to say that the new treatment is worse (recurrence is more likely) than the control (although not statistically significant because of the small sample).

The problem with these estimates, which should be familiar to most readers, is that unless the data were obtained from well-conducted RCTs, there are likely to be effects of confounding factors lurking in the background, resulting in biased estimates. Can the reader see that tables 1 (assuming counterfactual values to be real) and 2 are likely to be the result of an observational study with stage of disease (Z) being a confounder?

A Brief Exposition Rief on Causal Inference in Observational Studies

In RCTs, the observed treatment is by random assignment, meaning that the treatment is given irrespective of the state or characteristics of the patient prior to that assignment. This also means that, for example, at both the individual and group level,

$$\Pr(Y_{x=1} = 1|X) = \Pr\ (Y_{x=1} = 1)$$

Typically, in an RCT, the assignment probability is the same for all options; thus, with 2 treatment options, the probability is 0.5 to be assigned to either treatment.

Table 1 Observed (actual) and possible (counterfactual, potential) outcomes*

Patient (u)	Recurrence (*Y*) Under *X* = 1	Recurrence (*Y*) Under *X* = 0	Observed Treatment (*X*)	Stage (*Z*)
1	0	**1**	1	2
2	1	**1**	1	2
3	**0**	0	0	1
4	0	**1**	1	1
5	**1**	0	0	1
6	**0**	1	0	1
7	1	**1**	1	2
8	**0**	0	0	2
9	**0**	0	0	1
10	0	**1**	1	2

*Counterfactual outcomes are shown in bold – these are plausible outcomes, not observable in anyway

Table 2 Observed (actual) outcomes, from table 1.

Patient (*u*)	Recurrence (*Y*) Under *X* = 1	Recurrence (*Y*) Under *X* = 0	Observed Treatment (*X*)	Stage (*Z*)
1	0		1	2
2	1		1	2
3		0	0	1
4	0		1	1
5		0	0	1
6		1	0	1
7	1		1	2
8		0	0	2
9		0	0	1
10	0		1	2

In other words, in an RCT, the probability of recurrence, if the patient were given the new treatment (X = 1), does not depend on *how* he or she was assigned (regardless of whatever value of *X* the patient was going to get, since the assignment was random). This is not true for observational studies in general where

$$\Pr(Y_{x=1} = 1|X) \neq \Pr(Y_{x=1} = 1)$$

because the assignment mechanism in observational studies in general depends on characteristics of patients which predict or are related to the outcome. In our example, if the treatment depends on cancer stage, then this assignment mechanism is directly related to or is strongly influenced by stage. Hence, the probability of recurrence if the new treatment is given, will also depend on stage, that is, on the assignment mechanism. Knowing that the patient was selected for new treatment X = 1, in our example, would mean that the probability of recurrence on new treatment, $Y_{x=1} = 1$, i.e., $\Pr(Y_{x=1} = 1|X = 1)$ could be different (i.e., worse) than that if the patient were selected for control treatment X = 0, but given the new treatment instead, i.e., $\Pr(Y_{x=1} = 1|X = 0$. That is, for observational studies in general[8],

$$\Pr(Y_{x=1} = 1|X = 1) \neq \Pr(Y_{x=1} = 1|X = 0)$$

But if, in our observational study, cancer stage was the *only* important factor in treatment assignment, then given the same cancer stage, e.g., Z = 1 (or Z = 2), the probability of recurrence if the new treatment were given, should be the same irrespective of whether X = 1 or X = 0 was ultimately selected, i.e.,

$$\Pr(Y_{x=1} = 1|X = 1, Z) = \Pr(Y_{x=1} = 1|X = 0, Z)$$

That is,

$$\Pr(Y_{x=1} = 1|X, Z) = \Pr(Y_{x=1} = 1|Z)$$

Thus, in our example, *given Z*, the probability of recurrence (on new treatment) should be same regardless of the actual treatment assignment. This last condition, i.e., the independence of outcome from actual assignment mechanism given some predictive characteristics, *if assumed to be true for all observational studies*, is called the *"ignorability" assumption*[8]. The ignorability assumption is a means of treating observational studies as if they were RCTs, that is, to *ignore the treatment assignment mechanism*. If two people were to have identical Z values, then the *observed* assignment of one to X = 1 and the other to X = 0 can be considered essentially random. Then, the *unconditional* causal effect mentioned at the beginning, e.g., the comparison (at the group level)

$$\Pr(Y_{x=1} = 1) - \Pr(Y_{x=0} = 1)$$

can be estimated even for observational studies, via the conditional effects

$$\Pr(Y_{x=1} = 1|Z) - \Pr(Y_{x=0} = 1|Z)$$

but at a price of having to identify and condition on an appropriate set of variables Z, which are commonly confounders.

Note that there is still a counterfactual element in the above comparison. To eliminate this element, and to use actual observational data for the estimates of interest, some *consistency assumptions* must be made. That is, there is consistency between counterfactuals and actual observations: for a patient or a group of patients with the same Z, the probability of *potential* outcome when

given a counterfactual assignment X = 1 should be the same as that of the outcome when actually assigned X = 1; that is

$$\Pr(Y_{x=1} = 1|Z) = \Pr(Y = 1|X = 1, Z)$$

where the left-hand expression refers to counterfactuals while the right refers to actual probabilities based on observational data. Subscripts are dropped whenever real observed data are referred to.

To estimate the unconditional causal effect, we use a basic theorem in probability theory along with the consistency assumption (here, $\Sigma_Z f(Z)$ denotes summation of f over all discrete values of Z):

$$\Pr(Y_{X=1} = 1) - \Pr(Y_{X=0} = 1) = \sum_z \{\Pr(Y = 1|X = 1, Z) - \Pr(Y = 1|X = 0, Z)\} \Pr(Z)$$
$$= \sum_z \frac{\Pr(Y = 1, X = 1, Z)}{\Pr(X = 1|Z)} - \sum_z \frac{\Pr(Y = 1, X = 0, Z)}{\Pr(X = 0|Z)}$$

Since the joint probabilities $\Pr(Y = 1, X, Z)$ can be approximated directly from the data, this is a weighted sum of observed differences, inversely weighted by $\Pr(X|Z)$ (which cannot be 0 or 1); the weights are also known as the *propensity scores*. The propensity score could be obtained from some statistical model based on the data. This last expression is the well-known (inverse) propensity score-weighted causal estimates.[8]

Some readers might observe that this causal approach is not so different from any multivariable regression approach commonly seen in "non-causal" statistics. This is partially true but misses the point. The characteristics Z are related directly to assignment, not necessarily to the outcome. The relations between Y, X, and Z are the important features here, which are often not addressed appropriately in regression analysis, and so the latter may lead to inappropriate or misleading statistical models. The special relations are causal in nature, which must be postulated from some background biological knowledge.[9] These relations are not statistical, and are best described through the use of causal diagrams.

Causal Diagrams And Causal Inference[9-11]

For simplicity, consider three variables, Y, X, Z as in our earlier example. Some important insights into the use of causal ideas can be gleaned even from such simple cases. A typical relationship where Z is a "conventional" confounder is shown in Figure 3. This diagram is an example of an acyclical, directed graph with 3 nodes and 3 edges. It is directed because there are definite directions of causal effect (arrows), and is acylical because the arrows do not all point in the same clockwise or counterclockwise direction in a closed cycle. There are obviously 3 nodes, one for each variable, and 3 directed edges connecting them. In causal language, treatment X is a cause of recurrence Y, while cancer stage Z is a cause of both.

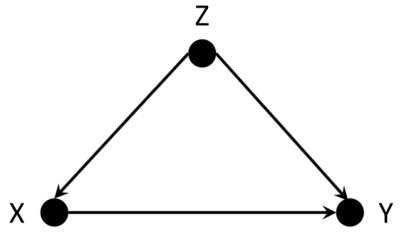

Figure 3 Causal diagram with Z confounding the effect of X on Y

Suppose that X and Z are independent causes of Y. Then the causal diagram would be as in Figure 4. This is a straightforward case where the causal effect of X on Y does not depend on Z and no conditional probabilities or adjustments for Z are required for unbiased estimation. The argument for the truth of Figure 3 as opposed to Figure 4 depends more on background knowledge of cancer biology than on observed statistical relations. But if Figure 4 is true, then adjusting for Z in a multivariable analysis should be unnecessary (and misleading) if the main aim is to estimate the causal effect of X on Y.

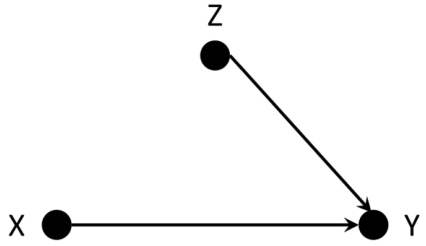

Figure 4 Causal diagram with independent causes X, Z on Y

Can Figure 3 become figure 4, such that the causal effect of X on Y can be estimated in a direct and unbiased way? Yes – one way is through doing an RCT. RCTs are designed so that causal effects of confounders Z on X are eliminated, and the causal effect of X on Y can be estimated directly without bias (on average). But appropriate adjusted analysis in observational studies can do the same. If the analysis is done separately, in Figure 3, for each value of Z, then the confounding effect of Z is eliminated. By fixing the value of a particular confounding variable, Figure 3 becomes Figure 4.

This is called a stratified analysis: stratified by the levels of *Z*. The estimated causal effect of *X* on *Y* is then the weighted average of the direct causal effect of *X* on *Y* at each *Z* level. We can say that the stratified analysis as described is a prototype of "simulating" RCTs estimates through using observational data.

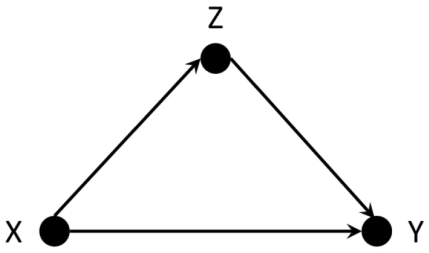

Figure 5 A cause X with direct and indirect (through Z) effects on Y

What if the causal relations are as in Figure 5, where now *Z* is *post-treatment* cancer stage? Treatment is a cause of both post-treatment cancer stage and recurrence, and post-treatment stage causes recurrence. If we assume no causal effects of *pre-treatment* cancer stage on treatment, and on post-treatment cancer stage, then the causal effect of *X* on *Y* is not confounded by *Z*. However, according to some current epidemiological definitions of confounding, *Z* should be a confounder in this case, since it is still related to both *X* and *Y*. This shows how causal thinking can clarify confusions regarding what a confounder is. *Z* is not a confounder simply because it does not cause *X*, in contrast to *Z* of Figure 3. Hence, the causal effect of *X* on *Y* can be estimated without bias while ignoring *Z*. This is called the total causal effect, because some of *X*'s causal effect on *Y* is indirectly through *Z*. If we did an adjusted analysis by adjusting for *Z*, we would get only a partial, but direct, causal effect of *X* on *Y*.

The diagram in Figure 6 is biologically impossible but has interesting implications. It says that treatment is not a cause of recurrence, but both are causes of cancer stage. Without adjustment, statistical analysis should not show any significant relation between *X* and *Y*. If this were true, adjustment for Z would create an illusion of causation between treatment and recurrence. Why is this? If *X* and *Y* both cause *Z*, for example, if both the new treatment and recurrence cause higher cancer stage, then if *Z* were fixed at some value, for example at the higher stage ($Z = 2$), statistical analysis would show that the new treatment is related to lower recurrence risk, or the control to higher recurrence risk. This would be a misleading result. It is the worst outcome of inappropriate statistical adjustment.

In Figure 6, suppose *X* is ovarian cancer (yes = 1; no = 0), *Y* is breast cancer (yes = 1; no = 0) and *Z* is hospitalization (hospitalized = 1; not hospitalized = 0). Now it is plausible that both *X* and *Y* cause *Z*. Both cancers, if sufficiently advanced, will require hospitalization, at least for surgical treatment. In this case, if Z were fixed at hospitalization (Z = 1), then for hospitalized patients there will seem to be an inverse relation between breast cancer and ovarian cancer, even though both are independent events. That is, if the hospitalized patient has breast cancer, she is less likely to have ovarian cancer and vice versa. This spurious (negative in this case) association between different cancers in hospitalized patients is a common finding in observational research. The next time the reader sees research on hospitalized patients, be wary that any claimed association between diseases might just be an illusion.

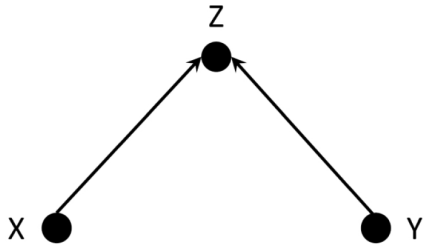

Figure 6 *Collider bias*: both *X* and *Y* independently cause *Z* but adjusting for *Z* induces correlation between *X* and *Y*

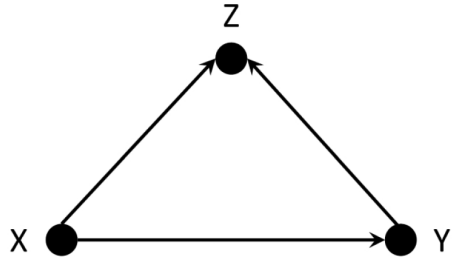

Figure 7 Only *X* causes *Y*, and *Z* is *not* a confounder: adjusting for *Z* results in collider bias

In the final diagram, in Figure 7, there is a causal effect between *X* and *Y*, but both *X* and *Y* cause *Z*. In many epidemiology textbooks, because it is related to both *X* and *Y*, Z would be considered a confounder. But it is not. *Z* neither causes *X* nor *Y*, and is therefore irrelevant for estimating the causal effect of *X* on *Y*. There is no need to adjust for *Z*. If one adjusts for *Z*, collider bias would appear, as described previously in Figure 6, and might distort or eliminate the real causal effect of interest.

Causal diagrams of real-life clinical phenomena can be created for purposes of causal inference in observational studies. These real-life diagrams will incorporate some or all of the "elementary" 3-way diagrams in Figures 3 to 7, as well as others not shown here[8], and will inform the researcher or statistician as to what clinical variables to include or not include in the analysis, and which variables should be considered confounders or otherwise. The selection of variables in some multivariable analysis will thus be knowledge-based, i.e., based mainly on known biological and causal principles, and relying less on *ad hoc* statistical procedures such as the various stepwise selection or penalty-based criterion including the Akaike information criterion (AIC) and its variants.[11] Misleading or even wrong analysis by the researcher, for example adjusting for collider variables (Figures 6 and 7) when he or she should not, could be avoided. Analysis of interventions and their effects on outcomes can be done and estimated via observational studies with similar results as for RCTs, given appropriate causal diagrams and data. Further, under suitable models, reasonably precise quantitative predictions may be possible.

Causal Inference In Physics And The Future Of Causal Inference In Medicine

In physics, causal inference is the rule. All major equations of physics are causal in nature. There are explicit quantitative relations between variables and not just relations in causal diagrams. Causal inference in clinical medicine may someday be captured as a collection of similar equations, all based on some set of accepted biological principles. But currently it is mainly in the physical sciences where we find causal relationships encoded as precise equations.

Imagine asking about counterfactuals in thermodynamics. Look at the equation of state of an ideal gas:

$$PV = nRT$$

If we assume no measurement error or negligible error and negligible chance variation, something that distinguishes simple systems (e.g., gas) from complex systems (e.g., humans), then any counterfactual value of the gas pressure *P*, say, can be calculated if the volume (*V*), temperature (*T*), and number of gas molecules in moles (*n*) are known (*R* is the gas constant). Without having measured or observing anything, we can imagine any scenario to be observed, in the future or otherwise, that the states of the gas can go through. We can hope that, looking ahead, various specialties of medicine may have similar "equations of state" for various diseases, where we can modify certain state variables using other state variables to achieve clinical cure. This can all be planned through solving the equations of state. Similar to the equations of state in physics, in biomedical causal inference we have *structural equations*. These can be used to make quantitative predictions (e.g., predicted mean values) subjected to random error. We will not consider them in this article, but the reader can easily find relevant introductory books for further information.[9,10]

The incomplete knowledge of human biology and pathology, the still unknown effects of various treatments of diseases, and the complexity of living organism are all working against any dream of a comprehensive system of such medical "equations of state" existing anytime in the foreseeable future. The simplistic view that there should exist a small set of predictive equations in medicine is probably unrealistic. Quantitative predictions in medicine in the near future is likely to come from structural causal models created by artificial intelligence (AI) systems under human guidance, with real-time data input and flexible self-learning prediction algorithms. Indeed, such learning algorithms with causal reasoning abilities are currently being developed.[10] It is exciting to see what the future may bring to the science of clinical prediction and prognostication, or of pattern recognition in general.

Why Experiment?

While observational studies can provide evidence for causal inference, such inference hinges on the validity of the available causal model. What if there is no such model? Then we must resort to creating some background knowledge, perhaps through experimentation, to obtain sufficient data to build a causal model.

Hypothetical causal models can also be constructed based on incomplete knowledge, and tested using observational data. But in certain situations, there is no replacement for experimental studies. Causal claims based on observational data in many cases will require experimental confirmation, especially if these claims are weak.

Conclusion

Science, including that of medicine, is about causal reasoning and finding causal relationships. Quantitative approaches in medicine are dominated by the conventional, statistical way of thinking, which is empirically oriented and skeptical of observation-based causal claims. This is not necessarily a bad thing, but it must be realized that causal thinking and causal claims can be made through appropriate observational framework, not just through experimental studies. History teaches us that some valid causal inferences were made based on accurate clinicopathological observations. The causal inference framework presented in this article is a more recent attempt to build a systematic, valid approach to making causal claims from observational data, something that clinicians including surgeons can readily participate in. If clinicians see it as a mandate of the profession to participate in research, under public acceptance, then they are clinician-scientists. We hope that this article can stimulate some of our colleagues to become clinician-scientists, to look at the value of observational data from a higher perspective, to see that making accurate, relevant clinical observations are good enough for high quality research, and is particularly cost-effective when compared with RCTs. Statisticians working in medicine are encouraged to empower themselves with another powerful set of tools still under active development, with a potential to arrive at conclusions of a causal nature with more confidence than ever before.

References

1. Miettinen O. Toward scientific medicine. New York: Springer; 2014.
2. Childers CP, Maggard-Gibbons M. Same data, opposite results? A call to improve surgical database. JAMA 2021;156:219-20.
3. Meyer MN, Heck PR, Holtzman GS, et al. Objecting to experiments that compare two unobjectionable policies or treatments. PNAS 2019; doi:10.1073/pnas.1820701116.
4. Singer C, Underwood EA. A short history of medicine. 2nd ed. Oxford: Clarendon Press; 1962.
5. Zampieri F, Zanatta A, Basso C, Thiene G. Caediovascular medicine in Morgagni's De Sedibus: dawn of cardiovascular pathology. Cardiovasc Pathol 2015;25:443-52.
6. Giovanni Battista Morgagni (1682-1771): his anatomic majesty's contributions to the neurosciences. Childs Nerv Syst 2012;28:1099-102.
7. Morgagni GB. De sedibus et causis morborum per anatomen indagatis libri quinque. Venice, 1761. English translation: Alexander B (transl). The seats and causes of diseases investigated by anatomy in five books. London: Miller & Cadell, and Johnson & Payne; 1769.
8. Rubin DB. Matched sampling for causal effects. Cambridge: Cambridge University Press; 2006.
9. Pearl J, Glymour M, Jewell NP. Causal inference in statistics: a primer. Chichester: John Wiley & Sons; 2016.
10. Pearl J, Mackenzie D. The book of why: the new science of cause and effect. London: Penguin Books; 2018.
11. Pearl J. Causality: models, reasoning, and inference. 2nd ed (corrected). Cambridge: Cambridge University Press; 2013.