

## Chapter 4 Causal Parameters

### Streptokinase and ischemic stroke

Many physicians regard ischemic stroke as the brain equivalent of a heart attack and for this reason, perhaps, it is sometimes called a “brain attack.” In a typical ischemic stroke, a thrombus clogs an artery just like it does in a heart attack, disrupting blood supply to part of the brain and causing brain cells to die. That mechanistic similarity has led many to speculate that clot-breaking drugs, known to benefit patients suffering a heart attack, would also benefit victims of ischemic stroke. In theory, quick dissolution of a clot that clogs an artery in the brain should restore blood supply to the damaged tissue and salvage brain cells that would have otherwise died.

But there was at least one important difference between ischemic stroke and heart attack—the frequency of bleeding as a result of treatment. While bleeding inside organs happens infrequently after treatment of a heart attack with a clot-breaking drug, bleeding inside the brain is not uncommon when such drugs are administered to patients suffering an ischemic stroke. Besides restoring blood supply to the brain, clot-breaking drugs may cause leakage of blood through the clogged artery, thereby transforming an ischemic stroke into a bleeding one. And the consequences of a bleeding stroke are often a more serious than the consequences of an ischemic stroke.

With hope of benefit and fear of harm, several research groups initiated randomized, placebo-controlled trials in the 1990s to estimate the effect of clot-breaking drugs on the outcome of ischemic stroke. We will use data from one of these published trials, a contrast of a drug called streptokinase with placebo, to develop a central idea in causal inquiry after which this chapter is titled: causal parameters.

A total of 270 patients participated in that randomized trial, almost evenly split between the two causal assignments. The original protocol called for a larger sample but recruitment was terminated earlier than had been planned because the estimated effect on death unfortunately substantiated pre-trial fears rather than pre-trial hopes (Table 4–1). When the trial was stopped, the odds ratio for death was 1.47 against streptokinase and the corresponding proportion ratio was 1.26. Information on person-time at risk was not provided in the article but the researchers reported a rate ratio of 1.44. (Note, incidentally, how close the odds ratio is to the rate ratio despite the fact that the event, death, was not rare at all.)

Table 4–1. Mortality of ischemic stroke patients in a randomized trial of streptokinase versus placebo

Causal Assignment	Dead within 6 months	Alive	Total
Streptokinase	61	76	137
Placebo	47	86	133

Source: Lancet, 1995: Vol 345, January 7, page 5.

All three numbers—1.47, 1.26, and 1.44—are called *point estimates* of the effect on death we attribute to the causal contrast; they are not known to be the true effect. But what do they estimate? What true quantities exist out there that we wish we had known?

The short answer is simple: *causal parameters*. The long answer for what a causal parameter is requires us, however, to commit first to a model of causation because deterministic causal parameters differ from their indeterministic counterparts. By slowly developing both types of parameters, we will gain deeper understanding of the sharp disagreement between two trails of causal inquiry: a deterministic trail and an indeterministic trail.

### **What has happened in the trial?—a deterministic view**

Inspecting Table 4–1 again, we see four cells that correspond to four possible combinations of two causal assignments (streptokinase or placebo) and two outcomes (dead or alive). Let’s consider first only one of the four cells and focus on just one patient, say, a female patient in the left upper cell—one of those 61 patients who received streptokinase and died. Why did she die?

The determinist will say that she died because a sufficient cause of death was completed during the trial. Now, either streptokinase was a component cause of that sufficient cause or it was not. If it was, then obviously her death was caused by streptokinase (along with other contributors to that sufficient cause.) If it was not a component cause, her death was neither caused by streptokinase (her causal assignment), nor by placebo (which she had not received.) In this case, her participation in the trial has made no difference: she was doomed to die from a sufficient cause that included neither streptokinase nor placebo pill. Notice that “doomed” only means, “doomed from the perspective of the causal assignments of this trial.” She would have been saved if it were possible to eliminate a component of the sufficient cause behind her death.

This unfortunate patient and all of her 60 companions who received streptokinase and died must belong to the first two rows of Table 4–2, a table that displays every possible combination of “if streptokinase were given the patient would be...” with “if placebo were given the patient would be...” (None of these 61 dead patients can belong to the third row or fourth row, which describe alive patients after streptokinase treatment.) For every recipient of streptokinase who died, the following must be deduced: either streptokinase has killed the patient (Table 4–2, first row) or the patient was doomed to die regardless of streptokinase treatment (Table 4–2, second row).

There is no way to tell, however, how many of the 61 patients belong to the first row and how many belong to the second row. To sort these patients between the categories *streptokinase causative* and *doomed*, we have to know what would have happened to each patient if he or she had received placebo. But that hasn’t happened—they were all treated with streptokinase. We see again how our inability to observe the outcome under a counter-factual causal assignment defies our attempt to know a cause. Some writers label this impassable roadblock “the problem of identifiability”: we cannot identify the deterministic classification of any person.

Table 4–2. A deterministic classification of patients in a trial of streptokinase versus placebo

If streptokinase were given, the patient would be...	If placebo were given, the patient would be...	Interpretation	Shorthand description of the patient type
Dead	Alive	Streptokinase caused death	<i>Streptokinase causative</i>
Dead	Dead	Streptokinase did not cause death	<i>Doomed</i> (from this trial’s perspective)
Alive	Alive	Streptokinase did not prevent death	<i>Immune</i> (from this trial’s perspective)
Alive	Dead	Streptokinase prevented death	<i>Streptokinase preventive</i>

Let’s turn to another cell in Table 4–1, the right upper cell that contains a count of 76 patients who received streptokinase and survived. Why did a patient in that cell survive? There are two explanations again. Either streptokinase has blocked a component cause of death, thereby saving the patient’s life (Table 4–2, fourth row), or it was a coincidental bystander (Table 4–2, third row); the patient may have been immune to death as far as the causal assignments of this trial are concerned. Notice that “immune”, like “doomed”, is not an absolute assertion about longevity. The patient’s “immunity” could have been broken by some sufficient cause of death that involved neither streptokinase nor placebo.

Two other cells of Table 4–1 are left to explore, both count placebo recipients. The left lower cell contains a count of 47 recipients of placebo who died, either because they were doomed to die even if they had been given streptokinase (Table 4–2, second row) or because bad luck had prevented them from receiving streptokinase that would have saved them (Table 4–2, fourth row). The right lower cell of Table 4–1, the remaining cell, contains a count of 86 placebo recipients who survived either because they were lucky not to receive streptokinase that would have killed them (Table 4–2, first row), or because they were immune from this trial’s perspective (Table 4–2, third row). Table 4–3 reconstructs the outcome of this trial from a deterministic viewpoint.

Table 4–3. Mortality of ischemic stroke patients in a randomized trial of streptokinase versus placebo: a deterministic viewpoint

Causal Assignment	Dead within 6 months	Alive	Total
Streptokinase	<i>Streptokinase causative:</i> ?	<i>Streptokinase preventive:</i> ?	137
	+	+	
	<i>Doomed:</i> ?	<i>Immune:</i> ?	
	Total 61	Total 76	
Placebo	<i>Streptokinase preventive:</i> ?	<i>Streptokinase causative:</i> ?	133
	+	+	
	<i>Doomed:</i> ?	<i>Immune:</i> ?	
	Total 47	Total 86	

Source: Lancet, 1995: Vol 345, January 7, page 5.

Looking at Table 4–3, we may draw two interesting conclusions. First, the question marks can never be replaced with numbers, which means that no trial can tell us how many patients survived due to streptokinase, how many died due to streptokinase, and how many fates were unaffected by streptokinase.

Second and most relevant, it is not at all clear what we’ve learned from this trial, say, from computing the proportion ratio,  $(61/137)/(47/133)$ , or the proportion difference,  $(61/137)-(47/133)$ . Table 4–3 shows that behind both computations there is invisible arithmetic on proportions of *doomed*, *streptokinase causative*, and *streptokinase preventive* in two groups of patients. But why does that arithmetic on *different* patients—streptokinase recipients and placebo recipients—tell us something about the effect of that causal contrast?

### Deterministic causal parameters

To follow the logic behind the computation, we should go back to the planning stage of the trial and recall the basic deterministic question about alternative causal assignments—“what would have happened if...” When a trial is planned, that question is asked about all would-be participants: What would have happened if *all 270 would-be trial participants* were to receive streptokinase and what would have happened if all of them were to receive placebo? Had the determinist known the answer to this question, he would have been able to construct Table 4–3 for all 270 participants in the trial. What the determinist has in mind is the hypothetical trial behind Table 4–4.

Table 4–4. Hypothetical trial of streptokinase versus placebo in 270 ischemic stroke patients: a deterministic viewpoint

Causal Assignment	Dead within 6 months	Alive	Total
Streptokinase	<i>Streptokinase causative:</i> ?	<i>Streptokinase preventive:</i> ?	
	+	+	
	<i>Doomed:</i> ?	<i>Immune:</i> ?	
	Total                                      ?	Total                                      ?	
	<b>270</b>		
Placebo	<i>Streptokinase preventive:</i> ?	<i>Streptokinase causative:</i> ?	
	+	+	
	<i>Doomed:</i> ?	<i>Immune:</i> ?	
	Total                                      ?	Total                                      ?	
	<b>270</b>		

The two rows of Table 4–4, unlike the two rows of Table 4–3, describe hypothetical mortality of the same patients under two causal assignments, not actual mortality of different patients under different causal assignments. For this reason, every measure of effect from Table 4–4 will quantify the effect of streptokinase versus placebo in that group of 270 patients (ignoring for the moment the invisible arithmetic on the deterministic status of these patients.)

Any measure of effect we compute from Table 4–4 (in theory, of course) may be called a *causal parameter*—a number that exists out there but obviously cannot be known. If we decide, for example, to give streptokinase to all 270 would-be participants in the trial, we will know the proportion of deaths under streptokinase treatment, but we will not know the proportion of deaths under placebo treatment. And vice versa. We cannot simultaneously assign two causal assignments to the same patients.

How, then, does a randomized trial (Table 4–3) allow us to estimate causal parameters (Table 4–4), such as proportion ratio and proportion difference?

The task at hand is not too complicated. We have to estimate two numbers: the proportion of deaths if all patients were taking streptokinase and the proportion of deaths if all patients were taking placebo. Had we been able to estimate these two proportions, their ratio or their difference would have estimated the respective causal parameter.

It is easy to estimate the proportion of deaths in any group from a random sample of that group. But how would we take a random sample of 270 recipients of streptokinase (or of 270 recipients of placebo) without first giving streptokinase (or placebo) to every patient? Here comes to play the procedure of randomization—random allocation to one of two causal assignments. The central idea is this: *If we decide at random whether a patient will receive streptokinase or placebo, then each treatment group will be, in effect, a random sample of 270 patients who would have received that treatment.* Therefore, each row total of

Table 4–3 from the actual randomized trial (137 patients and 133 patients) may be viewed as a random sample of the corresponding row total in Table 4–4, the hypothetical trial!

I had disturbing thoughts about the last paragraph, mainly because it shows how randomization substitutes for random sampling while these distinct concepts get confused in many minds. So I should make one point clear. Although randomization substitutes for random sampling in the logic of a randomized trial, random sampling (the procedure of actual sampling from some population) does not substitute for randomization and bears limited relevance to causal inquiry.

Back to the streptokinase trial. After randomization, 137 patients received streptokinase and 133 received placebo, two random samples of what would have happened to all 270 patients under the respective treatment. When the trial was terminated, 44% of streptokinase recipients were dead (61/137) as were 35% of placebo recipients (47/133). We have estimated two proportions of deaths in 270 trial participants and therefore, we can estimate the proportion ratio and proportion difference in Table 4–4, two causal parameters. How close the estimates are to the true values is a different matter.

The principles of our discussion equally apply to every causal parameter: odds-based, rate-based, and even to the arithmetic mean difference and the geometric mean ratio. In every randomized trial the causal parameter, whatever form or shape it takes, refers to hypothetical outcome of the entire study group under two causal assignments.

### Deterministic causal parameters: probing deeper

Since every participant in the streptokinase trial must fit a deterministic classification (Table 4–2), the trial population must be composed of four kinds of patients.

$$A_{\text{streptokinase causative}} + B_{\text{streptokinase preventive}} + C_{\text{doomed}} + D_{\text{immune}} = 270$$

where each letter denotes an unknown number of patients.

After dividing both sides by 270, we obtain the following equation:

$$P_{\text{streptokinase causative}} + P_{\text{streptokinase preventive}} + P_{\text{doomed}} + P_{\text{immune}} = 1 \quad (\text{Equation 4-1})$$

where each subscripted P denotes the proportion of that class of patients.

Another look at Table 4–4, the source of causal parameters, should convince us that it is possible to express the proportion of (real and hypothetical) deaths under streptokinase treatment as the sum of two subscripted proportions. If all 270 patients were treated with streptokinase, then two types of patients would have died: those who were *streptokinase causative* and those who were *doomed* (Table 4-4, left upper cell). Therefore the proportion of deaths under this causal assignment should be equal to  $P_{\text{streptokinase causative}} + P_{\text{doomed}}$ .

It is similarly possible to express the proportion of deaths under placebo treatment as the sum of two subscripted proportions (Table 4–4, left lower cell.) If all 270 patients were treated with placebo, the *doomed* would still have died but now those who were *streptokinase preventive* would also have died. (They would not have died if they were

treated with streptokinase). Therefore the proportion of deaths under this causal assignment should be equal to  $P_{streptokinase\ preventive} + P_{doomed}$ .

We are just one step away from deriving formulas for the causal parameters using the hidden arithmetic of deterministic classification. We will derive the formulas for the proportion ratio and the proportion difference, adding the subscript “causal” to highlight the parametric quality of these quantities.

$$\begin{aligned} \text{Proportion Ratio}_{CAUSAL} &= \frac{\text{Proportion of deaths, if all 270 patients were getting streptokinase}}{\text{Proportion of deaths, if all 270 patients were getting placebo}} \\ &= \frac{P_{streptokinase\ causative} + P_{doomed}}{P_{streptokinase\ preventive} + P_{doomed}} \end{aligned} \quad (\text{Equation 4-2})$$

Proportion Difference  $_{CAUSAL} =$

Proportion of deaths, if all 270 patients were getting streptokinase  
*minus*  
 Proportion of deaths, if all 270 patients were getting placebo =

$$\begin{aligned} & (P_{streptokinase\ causative} + P_{doomed}) - (P_{streptokinase\ preventive} + P_{doomed}) = \\ & P_{streptokinase\ causative} - P_{streptokinase\ preventive} \end{aligned} \quad (\text{Equation 4-3})$$

In the next section we’ll discuss striking implications of these two formulas, but first let’s see another possible view of the causal parameters from a randomized trial. So far I presented these parameters as quantitative properties of all 270 patients, estimated by randomization as a substitute for random sampling from a hypothetical trial. All of this is true, but not the whole truth. Having understood the idea of deterministic classification, you will see that the causal parameters we have just computed also apply to *each treatment group* and that a different light can be shed on the purpose of randomization.

When the treatment status (streptokinase or placebo) is assigned at random “enough times”, the two groups should have a similar distribution of every possible attribute including the (unknown) attribute called deterministic class. For example, both 137 streptokinase recipients and 133 placebo recipients may be composed of 20% *streptokinase causative*, 24% *doomed*, 11% *streptokinase preventive*, and 45% *immune*. (Notice that my numbers happen to agree with the proportions of death in the trial: 20%+24%=44%; 11%+24%=35%.) Since my hypothetical distribution inevitably holds in the combined group of 270 patients, all three groups—streptokinase recipients, placebo recipients, and all trial participants—must share identical causal parameters, such as the proportion ratio! Proportion Ratio  $_{CAUSAL}$  of 270 patients = Proportion Ratio  $_{CAUSAL}$  of 137 streptokinase recipients = Proportion Ratio  $_{CAUSAL}$  of 134 placebo recipients. If so, we can estimate these parameters for whichever group we choose and the estimate will apply to all three.

For example: To estimate the proportion difference or the proportion ratio for 137 streptokinase recipients alone, we have to estimate their hypothetical outcome under placebo treatment, which is the sum of  $P_{\text{streptokinase preventive}}$  and  $P_{\text{doomed in that group of patients}}$ . Recalling that successful randomization should have yielded identical proportions of every deterministic class in the two treatment groups, we may estimate this sum from the placebo group. And that’s easy: it is  $47/133=35\%$  (Table 4–3). Similarly, to estimate either causal parameter for 133 placebo recipients, we have to estimate their hypothetical outcome under streptokinase treatment, which is the sum of  $P_{\text{streptokinase causative}}$  and  $P_{\text{doomed in that group of patients}}$ . Again, thanks to randomization these proportions are hopefully identical in streptokinase recipients and the estimate of their sum is known:  $61/137=44\%$ . To sum up: 1) Randomization allows us to “peek” at what would have happened to each treatment group had it received the counter-factual treatment—the treatment it had not received. 2) Any estimate of a causal parameter from a randomized trial applies not only to the entire trial population but also to each of the randomized groups.

When would the above reasoning fail? When randomization has not succeeded to match the distribution of the four classes in the two treatment groups, either because the sample size was not large enough or because we were unlucky. After all, neither the earlier idea of randomization-as-random-sampling, nor the present idea of randomization-to-match-a-distribution, guarantees a good estimate of the proportion difference or the proportion ratio.

### Deterministic causal parameters: derivations

This was a long trail to full exposition of just two deterministic causal parameters, but worth following for several interesting findings at the end. First, we may now understand why the determinist prefers a difference measure of effect. When the proportion difference is computed (equation 4–3), the *doomed* group cancels out—a desired property of a deterministic causal parameter because this group bears no relevance to the contrast between streptokinase and placebo. Both  $P_{\text{doomed}}$  and  $P_{\text{immune}}$  are what we call “nuisance parameters.”

In contrast to the proportion difference, the proportion ratio contains  $P_{\text{doomed}}$  in both the numerator and the denominator (equation 4–2) and it therefore depends on the proportion of *doomed* in the study group, which is both unknown and unpredictable. If, for example, that group of 270 patients happened to include 90 percent of *doomed* patients, we will be estimating a Proportion Ratio  $_{\text{CAUSAL}}$  that is close to 1 regardless of how different  $P_{\text{streptokinase causative}}$  and  $P_{\text{streptokinase preventive}}$  are. A simple matter of math:

$$\frac{0.9}{1} \leq \frac{P_{\text{streptokinase causative}} + 0.9}{P_{\text{streptokinase preventive}} + 0.9} \leq \frac{1}{0.9}$$

Having convinced ourselves that the proportion difference is the preferred causal parameter, we will turn to a second interesting observation. The proportion difference,  $P_{\text{streptokinase causative}} - P_{\text{streptokinase preventive}}$ , describes the net effect of harm and benefit in the trial population. If Nature has decided that



$P_{\text{streptokinase causative}} = P_{\text{streptokinase preventive}} = 0.1$  in that group of 270 patients, the trial will be estimating a proportion difference of zero—that is, a causal parameter of zero. We will have killed 10% of streptokinase recipients and saved the lives of another 10%, yet say that streptokinase is no different from placebo. Unfortunately (from one point of view), or maybe fortunately (from another), no future trial will be able reveal what we have done.

Next, we may finally realize why the determinist has objected earlier to substituting the word probability for proportion and why he does not consider the odds a probabilistic measure. Neither 44% nor 35% —the proportions of deaths in the trial—describe chance-related events. And certainly, the proportions they estimate have no probabilistic content. The quantities “ $P_{\text{streptokinase causative}} + P_{\text{doomed}}$ ” and “ $P_{\text{streptokinase preventive}} + P_{\text{doomed}}$ ” in the trial population were predetermined by Nature when she assigned a deterministic status to each patient; they did not result from any process that may be called probabilistic realization.

Last, and perhaps most important, the deterministic causal parameter always depends on  $P_{\text{streptokinase causative}}$  and  $P_{\text{streptokinase preventive}}$  in the study (and on  $P_{\text{doomed}}$  when a ratio is computed.) As the formulas show, it is fully determined by these proportions in the group under study, which means that each trial of streptokinase and placebo may be estimating a different causal parameter. In one trial, we may estimate a proportion difference of 0.2 in favor of streptokinase; in another trial, a proportion difference of 0.2 *against* streptokinase; and in a third trial, a proportion difference of zero. Yet from the viewpoint of determinism, all three numbers are coherent results, each of which might have hit on the trial’s causal parameter. Diverging results of studies are the natural expectation from a deterministic model of causation—not the exception—since there is no reason to expect similar proportions of *causative*, *preventive*, (and *doomed*) in different studies. In short, there is no obvious reason to call conflicting scientific results “conflicting results.” A startling derivation, I think.

At most, we may claim to extend the results from our study to some larger, target population from which we have sampled, or pretend to have sampled, the study group, but that larger population must still be *finite*. The so-called target population may include 1,548 patients (or mice) or 10 million patients (or mice), but it must contain a fixed number of members if it were to deliver any causal parameter. And when the last member of the target population expires, we should shred the results from all nested studies: they apply to no other patients (or to no other mice). Now go tell this conclusion to scientists who are estimating effects and are searching for universal causal laws in whatever field they are working in—medicine, epidemiology, biology, sociology—and ask them whether they still hold to a deterministic model of causation.

## Indeterministic causal parameters

For the indeterminist, a recipient of streptokinase who died did not die because she belonged to the class we called *streptokinase causative* nor because she belonged to the class we called *doomed*. There are no such classes in indeterminism. Her death was a matter of chance—probabilistic realization of a causal propensity to die. And if the treatments in the trial indeed make up a causal variable, that propensity also depended on whether she was treated with streptokinase or placebo.

Instead of carrying along an unknown fate, each patient who entered the trial was harboring a propensity to die that was generated by an unknown number of contributors: the size of the stroke, its location, a co-existing disease, the patient’s age and sex—to name hypothetical few. When streptokinase was added to the list, or when placebo was added, that propensity—that causal force—might have gone up, might have gone down, or might have not changed. But whichever had happened, the outcome of the patient has remained undetermined. From an indeterministic viewpoint, causal assignments may be blamed for altering propensities to bring about death, but they should not be indicted as *the* causes of any particular death. To talk about the causes of a particular event is to speak the language of determinism, a slip of the tongue that spares no one—not even a committed indeterminist.

To define the indeterministic causal parameter, we’ll first imagine a possible set of Nature-determined causal assignments that generates some propensity to die after falling victim to an ischemic stroke: perhaps suffering a large stroke, age of 74, male sex, having a lung disease, and carrying an abnormal gene. After adding streptokinase treatment or placebo treatment to this background set, a new pair of propensities to die is possibly generated. Let’s call these propensities, whether indeed new or not,  $P_{streptokinase}$  and  $P_{placebo}$ , keeping in mind that a set of other causes contributes to both. In symbols:

$$P_{streptokinase} \equiv P_{[streptokinase, large\ stroke, age\ 74, male, lung\ disease, abnormal\ gene]}$$

$$P_{placebo} \equiv P_{[placebo, large\ stroke, age\ 74, male, lung\ disease, abnormal\ gene]}$$

If this model is true, the ratio of  $P_{streptokinase}$  to  $P_{placebo}$  compares two causal forces and quantifies the effect of the causal contrast between streptokinase treatment and placebo treatment. A ratio greater than one implies that streptokinase causes death relative to placebo in the context of that background set, whereas a ratio smaller than one implies that streptokinase prevents death relative to placebo. A ratio of one, precisely one, implies that neither streptokinase nor placebo changes the background propensity to die or that both change it by the same amount. Whatever its value may be, the ratio of  $P_{streptokinase}$  to  $P_{placebo}$  is a quantitative law of Nature, a numerical descriptor of universal reality that cannot be known but may be estimated. In other words, it is a causal parameter.

## How many values a causal parameter may take?

In a deterministic world this question is not particularly interesting and the answer is simple: almost as many values as the number of groups one could assemble. For example, the Proportion Difference  $P_{CAUSAL}$  can take as many values as the expression “ $P_{streptokinase\ causative} - P_{streptokinase\ preventive}$ ” may take. And as we’ve already seen, the proportions in that expression are expected to vary from one study to another, from one group of patients to another. There are no quantitative laws of causation in a deterministic world.

In an indeterministic world the short answer is, “We’ll never know.” The long answer will be revealed shortly but it may begin with a short statement: That’s what causal inquiry is all about!

Although Nature has not shared with us her causal design, logic dictates a choice from three possible indeterministic structures, say, for the triad of death, streptokinase, and placebo.

In the first structure, the ratio of  $P_{streptokinase}$  to  $P_{placebo}$  takes a single value regardless of the background set of causal assignments to which streptokinase or placebo are added. For instance, the propensity to die after streptokinase treatment may be twice as strong as the propensity to die after placebo treatment whether the background propensity is generated by the hypothetical set [*large stroke, age of 74, male sex, co-existing lung disease, an abnormal gene*] or by any other conceivable set. The same value of 2 holds, for example, for the background set [*small stroke, age of 33, female sex, no lung disease, a normal gene*]. In symbols:

$$P_{[streptokinase, background set]} / P_{[placebo, background set]} = 2$$

In the next possible structure, Nature has allowed  $P_{streptokinase} / P_{placebo}$  to take any number of values, linking each value to some background propensity, to some background causal set. The number of values may be as small as two or as large as it may be, but it must be finite. At the lower end of variation, Nature might have allowed the causal parameter to take only two values, perhaps one for men and another for women. For example,

$$P_{[streptokinase, male]} / P_{[placebo, male]} = 2.5$$

$$P_{[streptokinase, female]} / P_{[placebo, female]} = 1.5$$

At the higher end, Nature may have expanded these two rows to any number of rows and any lengthy, but finite, list of contributors to the background set. An example of a modestly expanded structure is shown below:

$$P_{[streptokinase, male, large stroke]} / P_{[placebo, male, large stroke]} = 3.0$$

$$P_{[streptokinase, male, medium stroke]} / P_{[placebo, male, medium stroke]} = 1.9$$

$$P_{[streptokinase, male, small stroke]} / P_{[placebo, male, small stroke]} = 0.5$$

$$P_{[streptokinase, female, large stroke]} / P_{[placebo, female, large stroke]} = 2.0$$

$$P_{[streptokinase, female, medium stroke]} / P_{[placebo, female, medium stroke]} = 1.0$$

$$P_{[streptokinase, female, small stroke]} / P_{[placebo, female, small stroke]} = 0.9$$

Keep in mind that every contributor to the background set must be a causal assignment in a causal variable. It has to affect the propensity to die.

In the third and last possible structure, Nature has linked the values of the causal parameter to an ever expanding list of background sets, a list as long as the list of patients on earth: past, present, and future patients. In this causal structure each victim of ischemic stroke carries along his or her own value of  $P_{streptokinase} / P_{placebo}$  and it may be that no two victims share the same value, not even twins. The list may look like the following:

Patient #1:  $P_{[streptokinase]} / P_{[placebo]} = 3.2$

Patient #2:  $P_{[streptokinase]} / P_{[placebo]} = 1.2$

Patient #3:  $P_{[streptokinase]} / P_{[placebo]} = 0.4$

Patient #4:  $P_{[streptokinase]} / P_{[placebo]} = 2.3$

.  
. .  
.

In symbolic language, we'll add the subscript  $i$  to denote the  $i$ th patient, the patient's identity:

$$P_{[streptokinase]} / P_{[placebo]} = C_i \quad (C_i > 0)$$

Notice that in the last two structures, the value of the causal parameter applies to momentary propensities produced by a fixed background set. If the background set changes over time, as may well happen, one's value of the causal parameter will change as well. At one moment a patient's value may be 3.0 against streptokinase whereas at the next moment it may be 0.5 *for* streptokinase. In between the stroke may have shrunk, for example.

### What have we learned from the trial?—an indeterministic view

Karl Popper, a twentieth century philosopher of science and an articulate proponent of indeterminism, has contributed several ideas to human thought that should help us understand the indeterministic trail of causal inquiry. The first idea says that there is no such thing as a theory-free observation: we don't simply record what we see or hear or smell or feel. Every observation we make in this world, be it the rising sun or the results of a trial, is contrasted in our conscious or subconscious mind with a theory we were holding just before the observation was made. If the observation corroborates our theory, no important impression is made on our mind. If the observation conflicts with our theory, we have a sense of surprise and we search for a better theory to explain what we have observed. For example, I am not surprised to see that my outside thermometer is showing  $-10^{\circ}\text{F}$  right now because it's winter in Minnesota and I had expected to see a low temperature. But if the thermometer were showing  $70^{\circ}\text{F}$ , I would have been surprised—the number would have clashed with my theory of what the number should be—and I would have come up with a new explanatory theory, say, the thermometer was broken. A subsequent observation of a broken thermometer would have been contrasted, in turn, with my explanatory theory but would not have surprised me, of course.

If we accept Popper's idea, it should be legitimate to ask what theory was held in the human mind before any trial has studied the effect of streptokinase versus placebo in patients with ischemic stroke. According to Popper's viewpoint, whatever we observed in the first trial must corroborate or refute prior expectation, a prior theory. We don't just read numbers off a printout and "discover" effects.

But what is an indeterministic causal theory about the effect of streptokinase versus placebo if not a statement about the value of  $P_{[streptokinase]} / P_{[placebo]}$ ? And what statement

would the human mind make before any trial was conducted if not “The ratio  $P_{streptokinase} / P_{placebo}$  may take any (positive) number”? In the absence of an earlier trial, any observed estimate of the causal parameter would not surprise us. Any proposed number is as good as any other number. The ratio of these two causal propensities may be 1, may be close to 0, or may be extremely large. Who knows? (At best, one may be willing to commit to a range, extrapolating perhaps from experience with trials of other drugs.)

This vague causal theory marks the starting point of conjectural knowledge about the effect of streptokinase on death as compared with placebo. It is indeed primitive and uninformative but nonetheless a theory, a human-made hypothesis about causal reality. As soon as the first trial is completed, however, the “any number” theory faces an empirical challenge by the result we compute; we now have a better theory. Not necessarily a true theory as you’ll shortly see, but a better theory that commits to one value of the causal parameter instead of committing to any value, which is not a commitment at all. On the rate scale for propensities, the new theory states that  $P_{streptokinase} / P_{placebo} = 1.44$ .

Our new theory still needs to be tied with those three indeterministic structures we considered in the previous section. To which of these structures does the number 1.44 apply? Obviously, it must apply to the first structure because we produced only one estimate of the causal parameter. Our post-trial theory states that the ratio of  $P_{streptokinase}$  to  $P_{placebo}$  is 1.44 for any background set of causal propensities. In symbols:

$$P_{[streptokinase, background\ set]} / P_{[placebo, background\ set]} = 1.44$$

Is this theory true? Maybe not. If Nature has varied the value of the causal parameter according to the background set (the second structure) or according to the patient’s identity (the third structure), the theory is false. But even if this were the case, which theory would you choose to contrast a future observation about streptokinase, placebo, and death: the “any number theory” we held before the trial or the number 1.44 we hold after the trial? I hope we’ve all given the same answer, unless fraud was suspected. Will the true causal structure ever be known? No, Truth is never known in scientific inquiry; only truth seeking is possible. Is there any way to continue to explore the possible falsehood of the new theory? Yes, there is.

## **What may be learned in the future?—an indeterministic view**

Two complementary routes may help us to revise knowledge about the effect of streptokinase on death as compared with placebo: one is to search for heterogeneous effects within the trial; the other is to conduct more trials and look for inconsistent effects. In the first route we may hypothesize, for example, that the number 1.44 does not describe causal reality for every background propensity, hiding instead two propensity ratios: perhaps one for male patients and another for female patients. To pursue this hypothesis, we will compute sex-specific rate ratios, one for men and another for women (ignoring for a moment the question of how good these estimates would be.) In statistical-epidemiological jargon, we’ll say that we have searched for heterogeneity of the causal parameter by sex (does its value depend on the patient’s sex?) by stratifying on sex (by splitting the sample into two groups: men and women.) If the sex-specific rate ratios turned out to be identical, our hypothesis of heterogeneous effects by sex was false. If

they differed, say, one was 1.2 and the other was 1.9, we would call sex a *modifier* of the effect of streptokinase on death (as compared with placebo) and revise our conjectural knowledge as follows:

$$P_{[\text{streptokinase, male}]} / P_{[\text{placebo, male}]} = 1.2$$

$$P_{[\text{streptokinase, female}]} / P_{[\text{placebo, female}]} = 1.9$$

Life isn't that simple, though. I can assure you and you can assure me that each time we will stratify the study group on some variable, whether on sex or on eye color, the stratum-specific rate ratios will not be identical. And if we stratify on several variables at once, we'll compute many different rate ratio estimates. Does this mean we should a priori reject the value of 1.44, endorse every series of stratum-specific estimates, and conclude that the first causal structure is always false? The answer is, no. Since estimation is the name of the game, no estimate deserves unlimited faith and some estimates carry more credibility than others. We will address the credibility issue time and again throughout the book but be forewarned that no recipe is waiting around the corner. If there were a recipe to reach the Truth, it would not have been called scientific inquiry.

The second route to challenge the theory that the number 1.44 applies to any background propensity should not be news to anyone: it is replication—the cornerstone of all science. If  $P_{\text{streptokinase}} / P_{\text{placebo}}$  indeed takes a single value, regardless of the background set of causal propensities, and if this value is well estimated by the number 1.44, then other trials, likely embedding various background propensities, should come up with a similar number. And if conflicting results are produced, we should look for new explanatory theories. Perhaps  $P_{\text{streptokinase}} / P_{\text{placebo}}$  does not take the same value for every background propensity, or perhaps something went wrong in some of the trials, or perhaps some estimates are not as credible as others.

It may be an appropriate moment to bring determinism back to the picture.

What will we learn from these exploratory routes if we hold a deterministic model of causation? What will we learn from computing a different rate ratio for men and women within the first streptokinase trial, or from observing five vastly different rate ratios in five repeated trials? The answer is “not much.” These observations corroborate a deterministic view of causation—not clash with it. Recall that each group of patients, whether nested in one trial or recruited in another trial, is expected to carry its own value of the deterministic causal parameter. Heterogeneity rules within a trial; inconsistency rules across trials. And I am still waiting for someone from the camp of determinism to explain why estimates from repeated trials turn out similar so often. Is it a miracle or is determinism a false model?

### A third structure

As we have seen earlier, in the third possible structure  $P_{[\text{streptokinase}]} / P_{[\text{placebo}]} = C_i$ , where  $i$  stands for the patient's identity. This structure is sometimes called stochastic causation—the law of causation for streptokinase, placebo and death was uniquely written for each patient, one line at a time. Where does this structure take us?

If  $P_{streptokinase} / P_{placebo}$  is indeed linked to the patient's identity, our search for these values is futile. To estimate these values we'll have to stratify on the patients' identities, creating strata that contain one patient each, but no estimate can be computed from a one-person stratum because the outcome under one causal assignment is unknown. If this is indeed the causal structure for the triad of streptokinase, placebo, and death, we'll continue forever to search for deeper heterogeneity of the causal parameter, by an ever-growing list of background propensities, and the search will lead nowhere. There is no chance of hitting on the truth.

Four possible responses may be given. I like three of them, which are linked, and dislike the fourth. First, we do not know whether the third structure governs all causal relations, some of them, or none of them. In the absence of knowledge or means to know we have nothing better than to search for structures that may be within our reach: homogeneity of the causal parameter (one value across all background propensities) or heterogeneity by any number of background propensities, as large as that number may be. Second, who says that a path of scientific inquiry should lead anywhere? Who says we are guaranteed to learn anything new about hidden reality? Scientific inquiry is a risky business: we may be pursuing false routes, wasting some of our time or all of our time. Third, falsehood should not be equated with futility. Anyone who thinks that there is neither merit in false scientific theories nor practical benefit to be gained is invited to read the progress of science in any field she may choose: physics, chemistry, biology, epidemiology. What is scientific progress if not successive replacement of old theories (presumably false) with new theories (presumably true)? It is better to know a rate ratio of 1.44 that is false—say, because the world is stochastic—than to know “any number will do”.

The fourth reply throws the towel in. It says that every measure of effect we may compute—whether a rate ratio or a mean difference—is just estimating an average of unknown causal parameters. If the second structure holds, the number 1.44 estimates the average of those parameters that were represented in the trial: perhaps 2, perhaps 20, perhaps more. And if the third structure holds, it estimates the average of 270 causal parameters. The fourth reply tells us to lower our scientific aspirations, not claim to be searching for the Truth, and be satisfied with the average of Truths. It is, in my view, an average reply indeed.

\*

Without the (fallible) scientific hypothesis of homogeneity of the causal parameter—at some finite level other than personal identity—we cannot produce quantitative (albeit conjectural) knowledge about causal effects. And we cannot propose homogeneity of the causal parameter (at any level) neither under a stochastic model of causation nor under a deterministic model of causation. If one of these models is true, all that we end up estimating is average effects over finite populations. In that case, all of our estimates are doomed to become historical information at some point, and, therefore, they should not carry the title "scientific knowledge" of causal effects. Conjectural scientific knowledge, such as a rate ratio of 1.44 for the triad of streptokinase, placebo, and death, makes a (fallible) claim about the unobserved—about future causal effects, which are unconstrained by identity and time.

Those who categorically deny the homogeneity of the causal parameter at any level of covariates' stratification (invoking the "implausibility argument" or a deterministic model of causation) must face two challenges: First, they should explain why we should ever

worry about inconsistent results in repeated studies of the same causal hypothesis (Perhaps we should not?) Second, they should explain why the results of any study ever apply to a person who did not participate in the study (Perhaps they never do?).

## **What will be learned in the future?—a social account**

Charting the road of causal inquiry may be intellectually rewarding. You can see the logical obstacles, the slopes, the side trails, and still hope that a road is there to follow. *Reality* of causal inquiry is a different story with other players on stage: convictions, norms, social arguments, hopes, and even the courts. There are fashions of the day, rules of conduct, and non-removable constraints.

Consider a rapidly fatal cancer for which no treatment has been known, and imagine a first randomized trial of a newly synthesized drug reporting a rate ratio for death of 0.5 for the drug as compared with placebo. Will that trial ever be replicated? If the trial was reasonably large, I doubt it, as you do as well. The ethicist will say that another trial would be unethical because the drug proved to save lives, and nobody will ask her what “proved” means in science or what model of causation she holds in her mind. And the patients will demand that drug, never willing to participate in another trial even if certified ethical. All of the conundrums, uncertainties, and difficulties we have charted so far are forgotten and gone. Philosophy asides. The drug works!

Will possible heterogeneity of the causal parameter by sex or ethnicity be pursued within the trial? Maybe. But I am not sure what reaction would be generated if the rate ratio were found, for example, to be 0.35 among women and 0.95 among men. An army of statisticians may line up to argue that subgroup analysis is risky, or bogus, or unworthy, and that chance may account for any difference between men and women or between blacks and whites. And nobody will ask them what model of causation underlies the computation, what sort of chance they have in mind, and how they know that Nature has chosen the first causal structure for the cause-and-effect in question. (Recall what “The drug works!” means: the value 0.5 holds for every background propensity.)

Now imagine a different trial, a first trial to report that one of two anti-hypertension drugs lowers systolic blood pressure 10-millimeter mercury more than the other. Will replication be deemed unethical? Probably not. Will heterogeneity by sex or ethnicity be deemed possible? Possibly. The social pendulum has swung 180 degrees and what many regarded as solid truth for one cause-and-effect relation is now regarded as only possible truth for another. What have we done to change our mind about causal reality? Not much more than changing the context of the story and the names of the variables: cancer became hypertension, death was replaced by blood pressure, and the causal contrast of a drug with placebo was called a causal contrast between two drugs. Now, a question: who is calling the shots about causal reality—our swinging mind or an objective Nature?

Even when ethical and practical constraints permit searching for deeper layers of heterogeneity, such avenues of research are not guaranteed. To produce credible estimates from stratified analysis, we have to ensure large enough strata, which means that a price must be paid: many more people have to be studied. But in a world with limited resources, social forces must decide whether to spend the money on deep-level estimation of the effect of one cause, or on first-level estimation of multiple causes. For a long time the social climate largely favored the latter, but the pendulum started swinging around the 1980s when Western societies became increasingly concerned about social justice by



sex and race. Since then it has been deemed worthwhile to explore heterogeneity of biological relations by sex or by race (one level deep) or by sex *and* race (two levels deep.) More recently, this trend has expanded when heterogeneity by genotype was added to the list of fashionable research.

Do I sound a little cynical? Perhaps. But what bother me most in these trends are not the social norms or the fashion-of-the-day, which is unavoidable human reality. What bothers me most is the attempt to disguise societal preference (to search for heterogeneity by sex, for example) or technology-driven preference (heterogeneity by genotype, for example) as preferences that were guided by some deep scientific reasoning. They were not. As far as heterogeneity of effects is concerned, the preference list is a secret of Nature's. And she does not play by societal rules, I think.