Pre-print draft of October 2016 – please see published version of *The Bloomsbury Companion to Contemporary Philosophy of Medicine* (James Marcum, ed.) for final text

12

Causation in Medicine

Brendan Clark and Federica Russo

Understanding medicine

Our aim in this chapter is to discuss causation in medicine. One key claim for us here is that there are different ways of understanding causes, and these different ways of understanding causes are more or less useful depending on the medical context. One difficulty at the outset is to try and describe the range of contexts that we are interested in. We locate our interest in medicine within the philosophy of science in practice tradition. This means that our foremost aim is to engage with the details of scientific practice. However, what kind(s) of scientific practice does this chapter therefore choose to engage with? We can think of no generally recognized piece of terminology that picks out our field(s) of interest. Rather than invent a new word, for the purposes of this chapter, we will instead define "medicine" in the broadest possible terms to include all clinical, scientific, and political forms of engagement with health and disease. We recognize, however, that our *medicine* is a disunion, with the consequence that our definition might well include forms of practice that not all would recognize as *medicine* in some more restrictive way. Specifically, here, we are thinking of fields such as:

- Clinical practice, including primary care and hospital medicine
- Preventive medicine and public health (more policy-oriented)
- Epidemiology

Just as these fields are many, so are their theories. Again, while our broad definition might include a few marginal accounts of how health and disease should be understood, we aim to capture the several more major theories found in contemporary medicine. In this chapter, we intend to involve and encompass several of these, including, for instance:

- evidence-based practice and associated practice like guideline development and evidence reviews
- narrative medicine
- personalized medicine
- gender medicine
- alternative, or complementary medicine

It is worth being inclusive here because, although some of the accounts mentioned earlier may be perceived as too heterodox to be included in "medicine," in their attempt to understanding causes and effects of health and disease, they may eventually feed more mainstream medical approaches.

For instance, according to well-accredited evidence-based medicine (EBM) approaches to evidence, narrative medicine fares rather poorly. However, we want to leave the door open in case we develop methods for evaluating evidence of mechanisms that makes patients' or doctors' narratives useful. Likewise, we want to leave the door open for alternative or complementary medicine, in case they manage to provide sound (scientific) evidence for the efficacy of treatments that are currently nonstandard. It is worth emphasizing that this does not allow an "anything goes" kind of argument. We agree that for some of the aforementioned accounts we currently lack explanations of health and disease that are also compatible with our best scientific theories. However, as history of science (and of medicine) teaches us, knowledge is not a simple, progressive accumulations of facts. We had to go through wrong theories and concepts to make progress in our understanding of the natural and social world, and of the functioning of the body. Conversely, many of the concepts and theories that today are considered well established were not received positively by the scientific communities of the time. Examples abound. So while we do want to be cautious

about what counts as a *scientific* approach to studying causes and effects of health and disease, we do *not* want to adopt a narrow-minded attitude that a priori excludes an approach, just because *today* it does not fit with our best science. In this sense, history of science (and of medicine, for the matter) is an integral part of the philosophy of science—as well as of the scientific practices we engage with (Chang, 1999).

In sum, in this chapter we would like to include any scientific approach that engages with health, disease, and well-being of individuals and populations. These two groups—one of kinds of practice, the other of theories of practice—are each large and disparate. However, for our purposes, they have one goal in common: in whole, or in part, they involve reasoning about the *causes and effects* of health and disease. Here, we are trying not to presuppose that a particular way of understanding how this reasoning about cause and effect should be privileged. Commonly used umbrella terms, such as "biomedicine," emphasize one part of medicine in a way that is unhelpful to us. Biomedicine, as the name might suggest, emphasizes the biological aspects of medicine, while "medical practice" emphasizes the clinical aspects of medicine at the expense of policy or of fundamental research. We think that very different ways of reasoning about cause and effect are found in the many practices that each of these umbrella terms picks out.

For example, the rise of evidence-based approaches over the past 20 years has prioritized causal strategies capable of demonstrating the efficacy of a treatment (cause) with respect to an outcome measure (effect). If we treated evidence-based practice as if it were all of medicine (in our broad sense), our approach would be dominated by approaches that emphasize efficacy over other goals of causal analysis, such as the search for causes of effects. But this, as Miriam Solomon (2011) has argued, would be to mistake a mode of reasoning that is common with one that is universal. This worry is our motivation for our inclusive approach, and possibly explanatory of our earlier terminological quibbling.

On our part, this claim about the universality of causation remains controversial, in that causal language or concepts may not be explicit in many areas of medical practice. The best

example that we know of is the move toward describing the *risk factors* of a disease, rather than their causes. This is a change that began in academic epidemiology during the 1970s, but has become very widely adopted in this medicine that we describe here. While some medical doctors and epidemiologists have been explicit in defending and discussing a causalist approach in medicine (see, e.g., Vineis, 2003; Vandenbroucke, 2009), there also have been much more skeptical attitudes. Russo and Williamson (2007) review some of these skeptical approaches and instead argue in favor of including causal notions explicitly in medical theory and practice.

Our aim in what follows is to try and understand causation in medicine by looking at some of its manifestations in practice. These, which we have dubbed "episodes of medical causation," aim to show what causation looks like all over the (deliberately) broad field of medicine as we understand it. We will then try to come to some more general understanding of causation. However, as we have already written, it would be misleading to claim that this process of moving from accounts of causes in medicine to an account of causation in medicine more generally will be simple. Medical practices differ from each other, and the causal problems of one field may be very unlike those found elsewhere. Anything like a simple accumulation of lessons learned about causes won't do here. Instead, we introduce a central metaphor—that of the *causal mosaic*—in order to try and assist us draw some general lessons from these particular cases. Our intention here is to use this central metaphor as a way of understanding the ways that causal thinking might work in different contexts.

Philosophical theorizing about medical causation

As just mentioned, this chapter makes use of "episodes of medical causation" to come up, at the end of the chapter, with a general approach to causation in medicine, one that hinges upon the metaphor of building a mosaic. An important source of inspiration in this is the work of Jonsen and Toulmin (1988). In this book, they describe an ambitious project to resuscitate one of the dirtiest words in moral philosophy: "casuistry." This practice of basing moral decisions largely on the details and

contingencies of individual cases has been viciously attacked as a philosophical stance that is itself "an invitation to excuse the inexcusable" (p. 11).

In their 1988 book, Jonsen and Toulmin instead argue that thinking about cases is a necessary component of effective moral reasoning. Inflexible rules cannot account for all moral decisions; some flexibility is needed; the bounds of this flexibility are formed by the details of the case in question. This is the role for casuistry outlined here: as shock-absorber between inflexible moral rules and flexible nature:

> [A]nyone who has occasion to consider moral issues in actual detail knows that morally significant differences between cases can be as vital as their likenesses. We need to respect not only the general principles that require us to treat similar cases alike but also those crucial distinctions that justify treating dissimilar cases differently. (p. 14)

Here, Jonsen and Toulmin are interested in the way that the details of individual cases interact with general moral principles. We aim to achieve a similar project for causation by thinking carefully about the roles, methods, and practices of causal reasoning in medicine. We select these from different contexts—research and clinical, preventive and curative, commonplace and oddball. They also come from different geographical locations and different times (although, admittedly, with a modest bias toward the contemporary) and have different implications for involved actors. What unites them, though, is that each of these episodes are defined by a central problem: reasoning about causes and effects of health and disease, of one kind or another.

Each of these episodes points to important and controversial issues that emerge in medical practice (but not exclusively there) and that are debated in the philosophy of causality.

We use different episodes, rather than one single "toy" example, because causal questions, while often formulated as "Does *C* cause *E*?" do not, in fact, reduce to this one. There are instead a plethora of issues involved, all causal, one way or another. Illari and Russo (2014) distinguish five scientific challenges that causal theory is confronted with: inference, explanation, prediction, control, and reasoning. Episodes of medical causation are also confronted with different challenges, but not all at *one* time. Explaining disease is not the same as formulating prognosis; reasoning about

causes and effects of health and disease can take different forms, depending whether at stake is individual diagnoses, epidemiological studies, or other.

This is also reflected in the various methods used in medicine. In the previous section, we highlighted that medicine constitutes an umbrella term including areas as diverse as clinical medicine, epidemiology, EBM, and so on. It is therefore not surprising that, in the selection of the episodes discussed later, we have been inclusive also with respect to their methods. Some of them employ statistical methods, whether observational or experimental (e.g., randomized controlled trials (RCTs)), others rely on experiments done in labs, some others employ case reports, and so on. Each method has its own virtues and vices, but we will not enter this debate here. The form of causal pluralism that we defend toward the end of the chapter goes hand in hand with a form of *methodological* pluralism, that we also embrace, and for which arguments have been offered elsewhere (for a discussion, see, e.g., Illari and Russo, 2014; Russo, 2009, 2014). In a nutshell, such form of methodological pluralism holds that there is not a fixed and immutable ranking of methods from best to worst. Differently put, there is not a gold standard that any other method is supposed to reach, as any method has to be assessed taking into account the specific scientific challenge it is designed to address.

Thus, for instance, RCTs are not intrinsically better than observational studies. A well-conducted observational study is better than a poorly planned RCT, even if, arguably, observational studies cannot grant us direct epistemic access to causal relations, which is instead possible to have with (carefully planned) experimental methods. This, however, seems to presuppose that scientists are free to choose whatever method they want, but that is not always the case. It is universally recognized today that forcing people to smoke, or to expose them to excessive amount of tar, is not just unfeasible, but deeply unethical. Thus, if we cannot make experiments in such cases, an RCT in this case is simply *not* an option. Instead, in this case, *observational* studies are the only option and a sensible question concerns what type of observational study would deliver better results. A corollary of this position is that it makes little if no sense at all to downgrade an observational

study, just because it did not implement the methodology of RCTs, administering a treatment to a test group and not to a control group.

The argument can be pushed even further as to include qualitative studies. The information provided by a case report on *one* patient (or a very small group of patients) is clearly not the same as the information gathered in a large scale cohort study. Yet, this is no *principled* reason to rank it lower; it is instead an empirical matter that rests on the specification of the questions and problems behind the chosen methods. For instance, if the goal is to reconstruct the reasoning that solved a diagnostic problem in a difficult case, a well and carefully written case report may be useful. But if we are trying to establish the population-level impact of a public health intervention, case reports most probably will be of a lesser help.

Thus, using different episodes serves the purpose of illustrating different scientific challenges and, consequently, different philosophical accounts, concepts, or notions to address them. Different episodes do so in a way that any toy example is unable to do: capturing the complexities and subtleties of concepts, forms of reasoning, and of methods in medicine. One virtue of toy example is that they can "distill" the essential properties of a complex problem, and so they enhance philosophical theorizing and debate to the extent that they are used for a specific purpose.

But what we aim to do in this chapter is not *just* assessing the pros and cons of *one* philosophical account with respect to the episodes. Instead, we aim at building a sophisticated philosophical view about medical causation, one that will allow us to analyze specific philosophical or methodological problems, when we encounter them in different episodes. This sophisticated view is presented later in the chapter, when the metaphor of the "causal mosaic" is discussed. Briefly put, the general approach to causation in medicine that we advocate is a *pluralistic* one. The rich range of accounts available in the philosophical literature cannot address all scientific challenges simultaneously—in medicine nor in any other scientific context. Hence, as we argue, a qualified version of causal pluralism is in order. The discussion of each episode of medical causation contains a brief discussion of some key philosophical issues that arise therein. Most of these issues have

been discussed, in the philosophy of causality, as candidates for *The-One* theory or account of causality (Reiss (2015, p. 20) refers to these as *straightjacket* theories). While the coming sections provide a descriptive overview of the diversity of causal concepts and accounts involved in medicine, toward the end of the chapter we provide more theoretical arguments in favor of a pluralistic approach for causation in medicine.

Episodes in medical causation

Hypertension treatments: efficacy and effectiveness, and the reference class problem

In this episode, we discuss a pair of closely related problems that arise when using current EBM methods in clinical practice. EBM makes a nonempirical claim about applied epistemology in medicine. In essence, this claim is that the best way to aid decision-making about the care of an individual patient is by reference to "current best evidence," which almost always takes the form of a clinical study of some kind. The canonical presentation of this method concerns a clinical question about a patient with epilepsy (Guyatt et al., 1992). The question is as follows: after an isolated, idiopathic seizure, what is the chance of recurrence? This question is answered by referring to a methodologically valid clinical study that studied a group of similar subjects, and reported the overall probability of recurrence at one year, which was found to be between 43 percent and 51 percent. This group probability is then transferred to the individual, meaning that their chance of suffering a subsequent seizure during the first year is also somewhere in the range of 43–51 percent. The intuition is that a methodologically valid clinical study that deals with the disease process or intervention that we are interested in straightforwardly provides evidence that can guide the care of an individual.

In the following section, we discuss two reasons to be somewhat suspicious of this straightforward transition from knowledge of groups to decision-making about an individual that are to be found in the case study of treatment for hypertension. While this is a paradigmatic example of EBM methodology in action in some respects, notably the emphasis given to using large

randomized-control clinical studies to gather evidence, we think that it also demonstrates many of the difficulties that might affect inferences from groups to individuals using that evidence.

Efficacy versus effectiveness

As we suggested earlier, the relationship between individual interventions (such as drug prescription) and knowledge about population outcomes for a drug may be complex. We think that this complexity may arise in part from biological variability in response to medical interventions. We won't defend this claim in a general way, but instead we set out a more specific argument that is (we think) part of this general thesis about variability. We claim that treatment populations, and the individuals in those populations, tend to be unlike trial populations.

It is worth beginning this argument by noting a terminological distinction that suggests that this difference is already well known to clinical trial researchers. Different words are used to refer to the effects of an intervention in a trial, versus the effects of an intervention in the real world. "Efficacy" describes effects in a trial, while "effectiveness" describes effects in populations (Gartlehner et al., 2006). To adopt this medical terminology, we think that part of the difference between efficacy and effectiveness lies in differences between trial and treatment populations.

According to the 2010 Health Survey for England, over 30 percent of adults have high blood pressure (see http://www.hscic.gov.uk/pubs/hse10trends). As roughly half of these adults receive some medical treatment, the resulting number of prescriptions for antihypertensive drugs is very large indeed. Moreover, prescription decisions for high blood pressure are complex. Many different drugs, belonging to different pharmacological families, are available and a very large number of clinical trials, and reviews of those trials, have been performed on these agents. While it is not easy to characterize the 13,603 clinical trials for antihypertensive drugs on human subjects, systematic reviews, and meta-analyses from the past 10 years currently listed by PubMed (August 2015), one source of difference between many of these trials, and many of the clinical decisions that they are intended to guide, concerns comorbidities. As Fortin et al. (2006) argue, most patients in practice have multiple comorbidities, but most subjects in clinical trials do not. In more detail, 54.5 percent

of patients in their sample of 424 individuals encountered in practice had hyperlipidemia, and a very large minority had existing heart disease (40.1 percent). While the trials studied by Fortin et al. do not straightforwardly report the number of trial subjects with comorbidities, we note that most trial subjects do not appear to have similar comorbidities (p. 106). This is important, because the presence or absence of other diseases is likely to change the effectiveness of an intervention. For example, hypertension in those with hyperlipidemia carries a much greater risk of adverse cardiovascular outcomes than without. Given that the aim of prescribing antihypertensive treatments is not to reduce blood pressure per se, but to reduce the incidence of these adverse cardiovascular events, it is at least an open question whether treatment with antihypertensives in patients with existing comorbidities has a comparable effect on cardiovascular outcomes to those reported by clinical trials. To illustrate, responses to a particular treatment may be very different in individuals with isolated hypertension compared with individuals with hypertension secondary to renal disease, largely because the mechanisms by which these two kinds of hypertension come about are different.

We should also mention the similar philosophical conversation about external validity—the idea that "the 'same treatment' has the 'same result' in a specific target as it did in the study" (Cartwright and Hardie, 2012, p. 46). While the terminology is different, the underlying issue is the same. Trials are often short on external validity, and we think this occurs because trial populations tend to be unlike treatment populations.

The reference class problem

The reference class problem has a lengthy pedigree in philosophy of science, as it traces back to classical and influential works such as Reichenbach (1949). Simply put, the problem describes a difficulty that affects drawing inferences about individuals from knowledge of groups. Imagine a statistical inference of this kind—say, calculating the probability of my house burning down in the next year. As my house is in Europe, we might look up the overall probability of house fires in Europe, and use this group probability as an estimate of the individual probability of my house

burning down. In this case, we have assigned our house to the reference class of houses in Europe. However, this reference class is not the only one to which we might assign my house. We could also have assigned it to the reference class of houses in the United Kingdom, or houses in England, or houses in London, or houses in South London. The difficulty is that each of these reference classes may have a different incidence of house fires. As we can assert that they are all appropriate reference classes, because my house is a member of each, this means that we will end up with different estimates of the individual probability of my house burning down depending on the way that we assign reference class membership.[1]

Given this, how should we select the most appropriate reference class? The bad news is that there appears to be no simple answer to this question—there is no objectively correct reference class to which my house should be assigned. Most attempts to find a solution to the reference class have provided ways of picking out extremely narrow reference classes on the grounds that sufficiently narrow reference classes should contain only identical members. We might term reference classes like this "fully homogeneous." So Wesley Salmon (1989, p. 171), for example, argued that it must be possible to find some reference classes of this kind in nature:

> I maintain, for example, that the class of carbon-14 atoms is objectively homogeneous with respect to the attribute of spontaneous radioactive decay within 5730 years. If our concept does not fit cases of that sort, our explication must be at fault.

Not finding homogeneous reference classes must be, for Salmon, a philosopher's failure, rather than any necessary consequence of ontology. Yet it is very hard to see how such fully homogeneous reference classes might be sought in medicine—even monozygotic twins have different environmental exposures, and so on. The hidden assumption here concerns the pervasiveness of natural kinds as objects of scientific inquiry, and it is one that we are suspicious of (largely because of Dupré, 1993).

Salmon also suggested that, in the absence of fully homogeneous reference classes, we should prefer the most narrow reference class for which reliable statistics are available. Salmon calls this practice the epistemic version of the "reference class rule" (Salmon et al., 1971, p. 43).

While this avoids the ontological pitfalls of homogeneous reference classes, it leaves the selection of a particular reference class as a matter of subjective choice. Given the convincing arguments supplied by Hájek (2007), we think that questions regarding the selection of reference classes are most suitably addressed in this epistemic manner. Hájek argues that the reference class problem arises from a fundamental mistake at the roots of contemporary probability theory. While an exhaustive discussion of this argument is beyond the scope of this chapter, Hájek argues that Kolmogorov's definition of conditional probability as "a ratio of unconditional probabilities" (p. 580) is the source of the reference class problem. He argues that the relationship between conditional and unconditional probabilities is the wrong way round. Instead of understanding conditional probabilities as derivative of unconditional ones, we should instead take conditional probabilities as primitive. This would free us from the misleading idea that we can ever generate a fully objective, unconditional, probability for a single event—something that Hájek suggests is at the root of the reference class problem. To illustrate using the example earlier, rather than trying to produce an unconditional probability estimate of the form p(my house burning down next year), we should instead look for an individual conditional probability estimate, such as p(my house burning down | house in the United Kingdom). This explicit statement of what we are conditioning our probability over would then avoid the reference class problem.

However, we still have to select a reference class to which we can assign our event of interest. While, unfortunately, we cannot supply a simple answer as to which that should be, we can outline some of the principles that might be used to guide reference class selection. One example might be to think about the breadth of reference class. Both broad and narrow reference classes have characteristic advantages and disadvantages, and we can describe these in a general way.

Broad reference classes (like that of houses in the United Kingdom in the example) tend to contain lots of members, and we are more likely to be able to find statistical data that characterizes them well. On the other hand, it might be that our individual is a highly atypical member of a broad class—I might collect fireworks in my house, or something—meaning that the probability of our

individual event might reasonably be expected to be very different from that of the broad group. Narrow reference classes (like that of houses in my street) tend to contain fewer members, meaning that our statistical characterization might be less precise. However, these weaker statistics might fit our individual much more closely because an individual is more likely to conform to the typical properties of a narrow reference class as compared to a broad one.

We can find evidence of such a trade-off between different kinds of reference class when we look at contemporary drug policy. The current UK NICE guidance on the treatment of high blood pressure picks out some specific reference classes, and members of these reference classes receive different drugs from nonmembers. Specifically, different prescription practices are advised for hypertension sufferers of different ages, from different ethnic backgrounds, and suffering different comorbidities. The usual first-line treatment for high blood pressure in people under 55 would be an ACE inhibitor (ACEI) or angiotensin receptor blocker (ARB):

> 1.6.6 Offer people aged under 55 years step 1 antihypertensive treatment with an angiotensin-converting enzyme (ACE) inhibitor or a low-cost angiotensin-II receptor blocker (ARB). (NICE, 2011, p. 17)

However, prescription practices should vary because of age, ethnicity, and other factors:

> 1.6.8 Offer step 1 antihypertensive treatment with a calcium-channel blocker (CCB) to people aged over 55 years and to black people of African or Caribbean family origin of any age. If a CCB is not suitable, for example because of oedema or intolerance, or if there is evidence of heart failure or a high risk of heart failure, offer a thiazide-like diuretic. (p. 17)

The underlying logic is (we think) closely related to Hájek's principle of prioritizing conditional probabilities as a way of drawing the teeth of the reference class problem. Here, several different reference classes are picked out—those under 55 years of age, those over 55 years of age, those of African or Caribbean family origin, and those with other illnesses. This makes explicit the kind of causal reasoning involved. Not "what is the probability a given drug relieving hypertension," but "what is the probability of this drug relieving hypertension given *x*, *y* and *z*."

Deficiency diseases: causation by omission, evidence, and inference

Causation by omission is a worry for most philosophical theories of causation. Why? Largely because absences are (ontologically speaking) just that, absences. Yet causal theories usually feature some commitment to existent objects or events. While the precise commitments vary, most regularity theories are committed to causation as a relation between events (Beebee, 2004). Process theories and mechanisms, on the other hand, understand causation as involving some kind of physical connection between cause and effect (Dowe, 2004; Machamer et al., 2000). Worries about the difficulty of treating nonexisting things as either relations or as physical connections has produced a variety of accounts that treat causation by absence as something other than causation proper—see the faintly pejorative terminology for absence causation—"fake causation" used by Persson (2002) or "causation*" used by Dowe (2000).

Whatever the reason, causation by absence is troublesome, and usually regarded as an odd or exceptional species of causation. Yet in medicine, absences often participate in causal reasoning. This section will discuss one example of medical causation by absence, which is the case of vitamin D deficiency.

Vitamin D participates in bone mineralization. Its biosynthesis depends on sunlight exposure to the skin. In adults, an absence of sun exposure can lead to the development of osteomalacia, which is a disease of inadequate bone mineralization. In children, vitamin D deficiency causes rickets, which has different clinical manifestations. However, the philosophical analysis in both cases is much the same: *absence* of sunlight causes a disease. Schematically, the mechanism of this disease is as follows. First, persistent lack of sunlight causes reduced cutaneous UV-B exposure. This in turn causes a decrease in the rate of cholecalciferol synthesis, which (over time) can cause osteomalacia.

Here, then, are at least three mechanistic links where an absence causes something to happen. First, an absence of sunlight reduces cutaneous UV-B exposure. Next, reduced UV-B exposure causes a reduction in cholecalciferol synthesis. Finally, the absence of cholecalciferol, over time, causes osteomalacia to develop.

Some philosophers worry about the metaphysics of causation by omission for reasons explained earlier. Yet, others do accept those kinds of causal links, but explain them, for instance, saying that omissions and absences are largely a matter of *language*. For any absence there is something that instead is active or activates, so we could recast the links above using "active" language. This is the strategy adopted by Machamer (2004). In any case, McGrath (2005, p. 125) expresses the general concern elegantly:

> DILEMMA: Either there is no causation by omission, or there is far more than common sense says there is.

Diseases due to deficiency are widespread in medicine. Thus, we take side with McGrath's second way: medicine suggests that causation by omission is very common. The challenge for a philosophical theorizing about causation is to make sense of the causal power of omissions and absences, beyond a common sense understanding.

One reason why philosophers find causation by omission tricky is because they want to find worldly causes that make causal claims true. Thus, if we say that vitamin D *deficiency* causes rickets in children, it will be hard to find, in the world, such a thing as a *deficiency*. However, philosophical theorizing about causation is very rich and other accounts allow us to make sense of causal claims about absences and omissions, but we have to change the conceptual framework. Notably, instead of looking for causes, mechanisms, or processes *in the world* we can work at the epistemological level and understand what *evidence* is required to establish causal claims like the one just mentioned and what *inferences* such evidence licenses.

Emphasis on evidence, and specifically about its multifarious character, has been given in the work of Russo and Williamson (2007), Illari (2011), and Clarke et al. (2014). Simply put, this body of work aims to spell out the evidential basis of causal claims (in medicine, but also elsewhere). The thesis, also known as "Russo–Williamson Thesis" (RWT), states that to establish a causal claim we need evidence *that* a cause makes a difference to the occurrence of the effect *and* evidence about *how* a cause produces the effect. This pluralistic view about evidence does not entail that there exist *difference types* of evidence. Instead, RWT offers an understanding of what aspects

are important to establish causal knowledge. Moreover, RWT does not imply that difference-making and production are exclusive categories. Quite the contrary is the case. On the one hand, they help each other with establishing causal claims. On the other hand, difference-making considerations are often involved in studying the mechanisms of disease causation and, conversely, considerations about causal production often guide the study of correlation and dependencies. This is explained by Clarke et al. (2014) using the analogy of "reinforced concrete," which is much stronger than concrete and steel *alone*, because these materials are resistant under different types of stress, and *together* they resist better.

Thus, in this conceptual framework, it makes perfect sense to say that vitamin D *deficiency* makes a difference to rickets. The difference it makes may be tracked in population-level correlations, for instance. Then, such claim may be supported by biochemical explanations of the causal pathways that are blocked (or activated) in presence (or absence) of a causal factor such as vitamin D.

Evidential pluralism can also be explained in terms of Susan Haack's view on evidence, which she explains using the metaphor of crosswords. According to Haack (2009), crosswords provide a useful analogy to explain the *structure* of evidence. The way in which we cross words is analogous to the way in which scientists try to find mutual support for their beliefs about the phenomenon under study. The pieces of evidence we put in the "crossword" are not just those pieces of information coming out of lab experiments or randomized studies, but also include a person's experiential evidence. Haack's account clearly has a strong pragmatist flavor, as the way in which we cross the various pieces of evidence does not *just* depend on what the world is like, but also on the interaction of the scientist with the world, on (scientific) language, and on a number of contextual factors.

A pragmatist flavor, although with a slight different emphasis, is also recognizable in Julian Reiss (2015). Simply put, Reiss is interested in what inferences we are (not) allowed to make, depending on the evidence base at our disposal and also on the target of these inferences. An

interesting corollary of this view is that the inferences about disease causation licensed by the results of an RCT may not be the same if what is at stake is instead the evaluation of a public health intervention, or the diagnosis of an individual patient.

Interestingly, Reiss argues in favor of this inferentialist view, as it helps regain the broad understanding of (testing) causal relations advocated by Sir Bradford Hill (1965). Hill famously discussed nine *viewpoints*, or aspects, to be considered when analyzing a putative causal relation. Frumkin (2000) summaries them thus:

1. Strength of Association. The stronger the relationship between the independent variable and the dependent variable, the less likely it is that the relationship is due to an extraneous variable.
2. Temporality. It is logically necessary for a cause to precede an effect in time.
3. Consistency. Multiple observations, of an association, with different people under different circumstances and with different measurement instruments increase the credibility of a finding.
4. Theoretical Plausibility. It is easier to accept an association as causal when there is a rational and theoretical basis for such a conclusion.
5. Coherence. A cause-and-effect interpretation for an association is clearest when it does not conflict with what is known about the variables under study and when there are no plausible competing theories or rival hypotheses. In other words, the association must be coherent with other knowledge.
6. Specificity in the causes. In the ideal situation, the effect has only one cause. In other words, showing that an outcome is best predicted by one primary factor adds credibility to a causal claim.
7. Dose–Response Relationship. There should be a direct relationship between the risk factor (i.e., the independent variable) and people's status on the disease variable (i.e., the dependent variable).

8. Experimental Evidence. Any related research that is based on experiments will make a causal inference more plausible.
9. Analogy. Sometimes a commonly accepted phenomenon in one area can be applied to another area.

Reiss correctly points out that, in Hill's account, these viewpoints do not correspond to necessary and sufficient conditions for causality. These are all fallible (yet useful) indicators, rather than strict criteria for causality. This is important because the inferentialist rejects any gold standard for evidence and instead pleads for putting together a "convincing case" for or against a causal claim, on the basis of how different pieces of evidence fit together. According to Reiss, Hill's viewpoints provide useful pragmatic "criteria" that help the inferentialist to come to a judgment about a given causal claim. This is strikingly close to RWT-like arguments, which are nonetheless a bit more specific about the *content* of the (pragmatic) indications given by the viewpoints: viewpoint 1, 3, 7, 8 concern evidence of difference-making or correlation, while viewpoints 2, 4, 5, 8, 9 concern evidence of production or mechanisms.

Ottawa ankle rules and backward causal reasoning

The tight relation between causal reasoning, evidence, and inference is recognizable also in other forms of medical reasoning. In the previous section, we highlighted how inferentialist positions à la Reiss are compatible with RWT, with other versions of evidential pluralism such as Haack's, and with Bradford Hill's viewpoints, which remain widely accepted by the medical community. In this section, we highlight another aspect of causal reasoning, that is, its *direction*. While we often reason from causes to effects, in other cases we may also reason from *effects* to causes.

Ankle injuries are common, and can be broadly classified into two groups. The first group are injuries that involve a fracture of one of the bones of the ankle. These injuries are treated by plaster immobilization (and sometimes surgical fixation). The second large group of ankle injuries involves ligaments alone. These are treated more conservatively by rest and appropriate anti-inflammatory drugs. While only about 15 percent of ankle injuries are fractures, the serious

consequences of failing to treat them adequately by immobilization have traditionally led emergency practitioners to adopt a precautionary approach. This has led to the widespread practice of X-raying most ankle injuries. Yet this is wasteful and (population-wide) harmful. How then might clinicians (a) do fewer X-rays while (b) ensuring that ankle fractures are identified and treated appropriately?

The Ottawa ankle rules are a set of rules by which clinicians can make informed decisions about which painful ankles require X-raying, and which do not. They involve reasoning, in a practical way, from effects (in this case symptoms) to causes (in this case, either fracture or ligamentous injury of the ankle). The basis for this reasoning is that the effects of either fractures or ligamentous injuries differ. Specifically, ankle fractures produce a characteristic pattern of tenderness (pain evoked by pressing) over the bony parts of the ankle. Additionally, those with ankle fractures experience more pain (and for longer) than those with ligament injuries when standing—in the jargon, they have a protracted inability to weight-bear. This leads to a very simple procedure that can be carried out in the emergency department: the patient is asked to walk four steps, and the ankle is palpated for tenderness over "the bony portions of the lateral and medial malleoli, the navicular, and the proximal fifth metatarsal" (Mayer, 2009, p. b2901).

Despite their simplicity, the Ottawa ankle rules have surprisingly high predictive power. Using the base-rate of about 15 percent fractures given earlier, the probability of fracture is "less than 1 percent" for a negative result, and "about 22 percent" if positive. In conclusion, "a negative result on applying this rule means an x ray of the ankle is not necessary" (Mayer, 2009, p. b2901). Put another way, this is an excellent example of strong causal reasoning in practice that demonstrates the benefits of adopting a pluralist approach: if all reasoning about cause and effect were forced into the "Does *C* cause *E*?" straightjacket (as discussed earlier), the utility of the Ottawa ankle rules would be an anomaly.

Gender medicine: confounding and generic versus single-case

Yentl is the name of the female protagonist of a play by Isaac Bashevis Singer. She truly is a heroine, as she disguises herself like a man in order to keep studying the Talmud. In the early 1990s, Bernadine Healy (1991) used the term "Yentl syndrome" to indicate a huge gap between genders concerning treatment and diagnosis. She bases her claims about gender inequality on data from hospital admission. Healy showed that there exist significant differences in the diagnosis of heart attack for men and women. This is due to the fact that symptoms for the same disease are very different in men and women, due to slightly different physiology. Nearly 30 years later, such differences still exist (Merz and Bairey, 2011). The studies of Healy have been followed up for many other health conditions, and the subfield of "gender medicine" has gradually emerged as a recognized and autonomous area of investigation within medicine (Signani, 2013; Baggio et al., 2013).

The focus on women is perhaps peculiar in that feminist demands for equal treatment in society are now transforming into recognizing important (biological) differences between men and women, these differences being at the base of unequal treatment in health. But the phenomenon is in fact much broader than that. Recent studies show widespread bias in research because it mainly focuses on "WEIRD" people, namely, on individuals coming from countries that are *W*estern, *E*ducated, *I*ndustrialized, *R*ich, and *D*emocratic (Henrich et al., 2010a, b).

These two episodes help us identify further conceptual complexities in medical causation. One such complexity is *heterogeneity* of individuals, discussed in the hypertension section. Gender medicine and studies on WEIRD people point in fact to the problem of correctly specifying the population of reference and, within this population, to specify the most appropriate reference classes. The issue is well known in the social sciences and in the philosophy of causality.

Another is problem is confounding, to be understood more broadly than just in statistical terms. In statistics, confounding variables (*C*) are variables associated both with the putative outcome (*Y*) (for instance, a disease) and with its putative cause (X). Because of the association between *C* and *X*, the causal effect of *X* on *Y* cannot be properly disentangled. The solution is to

control for confounding variables. But in the case of gender medicine, or of studies on "WEIRD" people, variables like gender, education, socioeconomic status are more than confounders in the statistical sense. They confound the *mechanisms* of disease causation. So in order to understand variations in health and disease, it will not be enough to control for gender, or for different levels of education or socioeconomic status. What looms large is a question about biological uniformity across individuals on the one hand, and about complex biosocial mechanisms on the other hand. This is the idea that humans are sufficiently similar in the relevant respects, thus justifying the generalization of causal claims about disease mechanisms as well as efficacy of treatments. This hypothesis is also at the basis of RCTs, but is sometimes questioned (Victora et al., 2004).

Another problem still is the "generic versus single-case." One might argue in fact that wrong diagnoses of heart attacks in women (partly) depend on using wrong medical knowledge. The medical sciences aim at building *generic* knowledge about health and disease, which is knowledge that is valid for the whole population. In turn, such knowledge will help in making inferences in the single case, for instance, in diagnosis and prognosis. Gender medicine is pointing to the fact that health and disease (at least *some* aspects of them) may be very different for men and women. This means that we cannot treat *all* individuals alike, but that we must study them separately, and in their specificities.

These specificities may be biological or social. For instance, gender medicine is raising the question of how drugs are tested: typically, on male patients in a certain age group, with some average body-mass index, and so on. This, however, ignores the possible effects of drugs on women, who typically have a lower body-mass index and that experience hormonal changes every month, and during the course of their lives. Just as specificities of women have been neglected, so have those of men. For instance, although much rarer in men than in women, diagnosis, prognosis, and treatment of breast cancer does occur in men too. Yet, much of what is known and done for men is extrapolated from studies and treatments for women.

So defining the "correct" population of reference is important not only for building medical knowledge that is valid at the generic level, but also for making inferences concerning individual patients. Relevant causal factors may vary greatly across ethnic groups, men and women, social classes, and so on. Perhaps this variation could be understood as an example of the reference class problem (see reference class problem section).

Asbestos-related deaths: social factors of disease and multifactorialism

According to today's knowledge, exposure to asbestos is undoubtedly related to fatal diseases such as asbestosis and lung cancer. Current biomedicine has sound insights about the biochemical mechanisms explaining such diseases—see, for instance, IARC monograph (IARC Working Group, 2012). These issues have been studied extensively in fields such as occupational medicine—the relations among work, work hazards, and health have a long history (Gochfeld, 2005). A paradigmatic example in this respect is asbestos-related deaths close to asbestos factories. Examples abound across different geographical locations. We might mention, for instance, Barking in the United Kingdom (Greenberg, 2003) and Eternit in Italy, for which a memorable sentence has been issued in 2009 after a long and difficult trial (Mossano, 2011; Allen and Kazan-Allen, 2012). Many aspects related to latency are, however, still objects of dispute (see, e.g., Terracini et al., 2014; La Vecchia and Boffetta, 2014).

Studies on effects of asbestos exposure help us highlight another dimension of medicine, namely, its interests in understanding health and disease studying factors *other than* biological. Arguably, in fact, the mechanisms leading to asbestos-related deaths include life style, socioeconomic status, and other social determinants. The argument is not new, and yet it remains controversial (Freese and Lutfey, 2011). For one thing, the inclusion of social factors in the explanation of disease and in public health policy was the very basis of the vision of pioneer healthcare officers during the nineteenth century. Also, social epidemiology has had ups and downs in the past decades, sometimes having to justify why social components of health should be considered at all (for an overview, see Kelly et al., 2014). In recent times, as the literature on the

social determinants of health has grown, the idea that social determinants of health act *directly*, rather than via other risk factors, has gained much credibility (see Marmot and Wilkinson, 2005, for example). Yet these social determinants still lack an identified *mechanism*. In fact, while epidemiology has sufficiently proven *that* social factors make a difference to health conditions, a convincing explanation of *how* that is the case is still lacking. The literature, in the "in and out" flutter of fashion, pretty much converged toward the idea that social determinants are *distal*, while biological causes are *proximal*. This gives social determinants a role in establishing difference-making relations, but not mechanisms of disease. These discussions are currently unsettled and it is to be hoped for a genuine integration of sociopsychological approaches and biomedicine to provide broader explanation of health and disease.

For our purposes, these debates are most interesting because they are related to a topic now central in the philosophy of causality: the different characterization, role, and use of "difference-making" and of "mechanisms" in disease causation. These have already been introduced earlier in the deficiency diseases section, but here we can add a further layer of sophistication to the discussion. In fact, a large part of the debate on evidence of mechanisms focused on *biological* mechanism. The question of *how* causes produce their effect is by and large understood in terms of what processes are triggered in our bodies. These processes, in turn, can be described going down to the molecular level.

Yet, episodes like this should at least trigger (or rather, revive) interest for mechanisms that see individuals embedded in the network of social relations. The environment, in fact, is *also* made of our family and peers, the workplace we routinely attend, the infrastructures we have access to, and so on. Thus, an explanation of asbestos-related death should not be limited to the identification of the biochemical processes triggered in our bodies, once exposed to the hazard. Such an explanation should also endeavor to spell out the social, behavioral, or psychological mechanisms involved. In this specific case, for instance, the workplace (Barking, Casale Monferrato) *is* part of the mechanism of exposure, and not just for the workers of asbestos factories. For instance, the

wives of workers of the Barking or Eternit factory were also exposed to asbestos fibers because they washed their husbands' coveralls. Or, living nearby the factory, and even downwind, may also increase exposure. Understanding how these elements enter the mechanisms of disease causation helps identify (precautionary) interventions at a more appropriate level.

Finally, episodes like this point also to problem that is rather ubiquitous in medicine: diseases often have more than one cause. In medicine, and especially in epidemiology, this phenomenon has been labeled as "multifactorialism" (as opposed to *mono*factorialism). Medical theory has long recognized that many diseases may have multiple causes (while acknowledging that others, instead, still admit only one). The most famous epidemiological model developed to account for that is Rothman's (1976; Rothman et al., 2008) "causal pies." Causal pies display all known factors that play a role in the occurrence of a particular disease. Pie charts, in a sense, visualize contingency tables, but with an important difference: in a pie chart etiological fractions—that is, the components—do not have to sum up to one. This means that we are not looking for the *sum* of the components that make the effect necessary. It is a useful heuristic way of thinking about the multifactorial character of most diseases.

Rothman (1976) made the point that, in many cases, what we call "causes" should instead be understood as *components* of sufficient causes, but that are not themselves sufficient. According to this model, a sufficient cause of a disease is generally not one *single* causal factor, but a complete "causal mechanism." Rothman, to be precise, takes a causal mechanism to be a minimal set of conditions and events that are sufficient for the disease to occur. In this perspective, no specific event condition or characteristic is sufficient, by itself, to produce the disease. So the definition of "cause" does not describe a complete causal mechanism, but only a component of it. It is worth noting that Rothman's use of the term "mechanism" differs from the one discussed earlier, for instance, in relation to RWT. Consider the episode of this section: exposure to asbestos dust is a component of a "bigger" cause that may also include lifestyle (and smoking, dietary habits, etc.), altogether leading to developing lung cancer. To give another example, measles virus is said to be

the cause of measles; however, the "complete sufficient cause" of measles also includes lack of immunity to the virus and exposure to the virus.

In this line of reasoning philosophers will readily recognize the "INUS causes" of Mackie (1974). INUS stands for an *I*nsufficient, but *N*onredundant part of an *U*nnecessary but *S*ufficient condition (p. 62). Concepts like INUS or "causal pies" are useful because they allow us to characterize much more precisely the nature of many of the causes that do actually interest us. This can sharpen our reasoning, and thereby have an impact on our practices of inference, prediction, and control. In particular, in adopting a multifactorial view, we are more likely to identify those factors that, though not part of the biological mechanism of disease causation, have an active causal role.

The causal mosaic

In previous sections, we showed that medicine is a highly heterogeneous field, where reasoning about causes and effects of health and disease has not a univocal, fixed, or predetermined meaning. Most importantly, causation in medicine amounts to asking a plethora of questions, rather than just one. A straightforward consequence of this analysis of medical practices is that there is not one single causal theory that fits the bill. While pluralism becomes the most plausible candidate, *any* form of pluralism won't do either. In the following, we sketch the main features of a particular form of pluralism. This pluralism draws on the metaphor of building a (causal) "mosaic," where causal theories are not simply and ad hoc juxtaposed, but chosen and placed according to how they fare with different types of philosophical questions and with different types of scientific challenges. While a thorough presentation can be found in Illari and Russo (2014), we highlight aspects that are relevant for building a *medical* causal mosaic.

Thinking about causality has a long tradition in philosophy and in science. Some of the discussions carried out in Greek or Medieval thought, or in the development of the scientific method in modern times, may look very different from our current way of thinking about causality. Admittedly, causal thinking had several ups and downs, including vigorous attacks in the early twentieth century due to Ernst Mach in physics, Karl Pearson in statistics, and Bertrand Russell in

philosophy. Yet, philosophical thinking did not fade away. Instead, two main strands in philosophical theorizing can be identified since the 1970s and 1980s. One is the counterfactual approach developed mainly by David Lewis (1973, 1983), and the other is the process-tracing, or causal–mechanical, account mainly developed by Wesley Salmon (1984). The past two decades witnessed a steady increase of interest in causal questions, also thanks to the approach based on graphical models as developed by Judea Pearl (2000), Spirtes et al. (1993), and their collaborators. These new, more science-oriented approaches led—more or less directly—to important changes in addressing causal questions (for a discussion, see also Illari et al., 2011; Illari and Russo, 2014).

As far as medicine is concerned, causal thinking has always been central, and this well before Aristotle's doctrine of the four causes or Galenic physiology (for a discussion, see Rabins, 2013), and after (in particular, Rabin discusses medieval conceptions of medical causation, largely neglected in current debates). Arguably, the experimental method of Claude Bernard (1856) started a tradition in medicine that looked for biological causes of health and disease, and nowadays is giving rise to investigations of disease mechanisms at the molecular level. Yet, at the same time, medicine also developed *other* ways of studying health and disease, most prominently epidemiology—which is based on statistical analyses of data—but also narrative medicine—that clearly distances itself from views that "biologize" health and disease.

For one thing, philosophers have being paying increasing attention to the special sciences, for example, economics and other social sciences, biology, or epidemiology. Dedicated discussions about epidemiology and medicine (such as Broadbent, 2013; Howick, 2011) are also part of this heightened specialization of philosophy of science, and notably of philosophy of causality. Consequently, considerable effort was put in trying to formulate questions that are tailored to specific fields, and particularly to the methodological challenges of these fields. This is indeed the rationale behind laying down "episodes" of medical causation discussed earlier, and from these to discuss relevant aspects of causal theories.

In recent years, another question gained ground: how to *use* causal knowledge, namely, what to do once we know (or do not know) causal relations. This happened thanks to work of Donald Gillies (2005), who formulated an "action-oriented theory of causality" and of Nancy Cartwright (2007), who drew philosophers' (as well scientists' and policymakers') attention to very delicate causal issues lingering in policy-making and to the scope of the results of trials. Questions about use are particularly pressing in medicine and public health. This is because we want to use medical knowledge to cure people and save lives. However, *any* medical knowledge won't do here. What is required to diagnose a particular patient carefully differs from what is required to make sound claims about the efficacy of a drug, which is again different from what is required to make sound claims about the efficacy of a public health intervention. Thus, it seems that questions about use cannot be asked in "absolute terms," but need to be specified with respect to the context in which they arise.

A bird's eye view on the vast literature produced in the past decades reveals a sprawling of accounts, concepts, and notions. Causation has been examined in terms of, or in relation to, counterfactuals, mechanisms, probabilities, inference, necessary and sufficient conditions, and many others. Illari and Russo (2014) present these debates and explain how these different accounts can be of help in addressing typical scientific problems about causality, namely, inference, explanation, prediction, control. Illari and Russo conclude, from such diversity and variety of accounts, that it is reasonable to give up on *The One Theory of Causality*. We certainly do not have it as of today; this does not exclude, as a matter of principle, that we will find one theory able to encompass all cases, episodes, and types of causation in the future. While their argument spans philosophy of causality *across* the sciences, this chapter further supported it, making the case *specifically* for medicine.

Even if we agree, prima facie, that there exists a plurality of accounts, concepts, and approaches, we still face the difficulty of making sense of such diversity and variety. How can these accounts coexist together? It is here that the metaphor of a "causal mosaic" proves useful. A mosaic

is an image formed of several tiles, possibly different in shape and color. When the tiles are appropriately located one next to the other, an image appears. Each tile participates in the formation of the image, when correctly positioned. Individually, tiles may have beautiful colors or shapes, but they do not have a clearly recognizable meaning. That is to say, the tiles acquire meaning in relation to other tiles, and in the context of the mosaic being built. Philosophical theorizing about causality works just like that. Available accounts of causality all have a role in our understanding of causality. But to see the whole causal picture, we need to place the tiles—the causal accounts—according to how they respond to different philosophical questions (metaphysical, epistemological, semantic, methodological), and according to how they help us with specific scientific challenges (inference, explanation, prediction, control).

Causation in medicine is *one* of the mosaics we might want to build. If we want to understand "medical causation," we have to start with more specific questions. What kind of practice is at stake? What is the specific causal question being asked? What is the underlying philosophical problem or scientific challenge? And so on. Philosophical accounts will help answer such questions by providing an appropriate conceptual framework. Thus, for instance, a theory of evidence based on the mechanism—difference-making distinction—might prove useful to answer questions about drug or treatment efficacy. An account of the levels of causation and of Simpson's paradox might help address the problem of reference classes and of external validity. An approach to causal reasoning might lend support to guidelines for prescribing diagnostic exams such as X-rays. And so on and so forth.

The metaphor is quite telling about philosophical methodology and philosophical practice too. In fact, we clearly do not know in advance what the final causal mosaic may look like and philosophers need to work together, and with scientists, in order to figure out how best to place the tiles. Also, the final picture is clearly dynamic, as the scientific challenges evolves and so does the philosophical theorizing that accompanies it. Last, but not least, causal mosaics may be very idiosyncratic, which should be a reason to foster dialogue and collaboration among philosophers,

among scientists, and among those two. Again, causation in medicine is no exception in this respect. The philosophical community interested in medicine grew significantly in the recent past. We moved from a philosophy of medicine that was by and large populated by questions about ethics, to a much larger group of scholars equally interested in epistemological and methodological questions about medicine. We need to build a *medical* causal mosaic also thinking prospectively in terms of collaborations with ethicists or political philosophers. In fact, the "phil sci" question of how to use medical knowledge easily turns into a moral or political question about whether (and how) we should use such knowledge, or whether we are legitimized in using it, and so on.

We developed the chapter precisely along these ideas, and so ours has been an exercise in building *a* causal mosaic for medical causation, or causation in medicine. What we suggest is *a* reconstruction of how causal questions arise in medicine (broadly construed) and of how to make (philosophical) sense of them. The picture that we lay down reflects some specific philosophical presuppositions and methodology. We espouse the goals and methods of the "Causality in the Sciences" (CitS) approach, which proposes looking at scientific practice in order to select philosophical questions and to "test" them. While scientific practice help philosophical theorizing, philosophical accounts in turn aid scientific practice. CitS advocates in favor of this iterative relation of mutual aid between philosophy and the sciences. It is worth noting that this is very much in line with other approaches, notably those of the developing "Philosophy of Science in Practice" and "History and Philosophy of Science" movements. All look to the practice of science—contemporary or historical—in search of fruitful interactions between the fields of philosophy and of science. CitS also shares much with the philosophy of information, which advocates developing a "timely philosophy," namely, a philosophy that helps address the challenges of science and society *today*.

In sum, we hope that this chapter successfully conveyed the message that causation in medicine is a fertile and growing area of investigation. What we presented here does not even get close to building a full mosaic. But this was not our intention. We wanted instead to propose an

approach to *how* one can contribute to building a mosaic of medical causation. This, we maintain, is a task that only the scientific and philosophical communities, together, can achieve.

Acknowledgments

The authors gratefully acknowledge financial support from the AHRC (UK) for the research done under grant n. AH/M005917/1 ("Evaluating Evidence in Medicine"). An early version of this chapter was presented at the EBM+ project workshop at the University of Kent, Canterbury, in January 2015. We would like to extend our thanks to the workshop participants, especially Jon Williamson, Michael Wilde, Phyllis Illari, and Veli-Pekka Parkkinen, for their thoughtful comments and invaluable input.

References

Allen, D., and Kazan-Allen, L. (2012). *Eternit and the Great Asbestos Trial*. London: The International Ban Asbestos Secretariat. http://www.ibasecretariat.org/eternit-great-asbestos-trial-toc.htm.

Baggio, G., Corsini, A., Floreani, A., Giannini, S., and Zagonel, V. (2013). Gender medicine: A task for the third millennium. *Clinical Chemistry and Laboratory Medicine* 51(4): 713–17. doi: 10.1515/cclm-2012-0849.

Beebee, H. (2004). Causing and nothingness. In *Causation and Counterfactuals*, eds. John Collins, Ned Hall, and L. A. Paul. Cambridge, MA: MIT Press, pp. 291–308.

Bernard, C. (1856). *An Introduction to the Study of Experimental Medicine*, 1927th ed. New York: Macmillan.

Bradford Hill, A. (1965). The environment of disease: Association or causation? *Proceedings of the Royal Society of Medicine* 58(5): 295–300.

Broadbent, A. (2013). *Philosophy of Epidemiology*. Palgrave Macmillan.

Cartwright, N. (2007). *Hunting Causes and Using Them*. Cambridge: Cambridge University Press.

Cartwright, N., and Hardie, J. (2012). *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press.

Chang, H. (1999). History and philosophy of science as a continuation of science by other means. *Science & Education* 8(4): 413–25. doi:10.1023/A:1008650325798.

Clarke, B., Gillies, D., Illari, P., Russo, F., and Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi* 33(2): 339–60.

Dowe, P. (2000). *Physical Causation*. Cambridge: Cambridge University Press.

Dowe, P. (2004). Causes are physically connected to their effects: Why preventers and omissions are not causes. In *Contemporary Debates in Philosophy of Science*, eds. Christopher Hitchcock. Oxford: Blackwell, pp. 189–96.

Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.

Fortin, M., Dionne, J., Pinho, G., Gignac, J., Almirall, J., and Lapointe, L. (2006). Randomized controlled trials: Do they have external validity for patients with multiple comorbidities? *Annals of Family Medicine* 4(2): 104–108.

Freese, J., and Lutfey, K. (2011). Fundamental causality: Challenges of an animating concept for medical sociology. In *Handbook of the Sociology of Health, Illness, and Healing: A Blueprint for the 21st Century*, eds. Bernice A. Pescosolido, Jack K. Martin, Jane D. McLeod, and Anne Rogers. New York, NY: Springer, pp. 67–81.

Frumkin, H. (2000). *Causation in Medicine*. http://www.aoec.org/CEEM/methods/emory2.html.

Gartlehner, G., Hansen, R. A., Nissman, D., Lohr, K. N., and Carey, T. S. (2006). Criteria for distinguishing effectiveness from efficacy trials in systematic reviews. 12. Agency for Healthcare Research and Quality (US). http://www.ncbi.nlm.nih.gov/books/NBK44029/.

Gillies, D. (2005). An action-related theory of causality. *British Journal for the Philosophy of Science* 56(4): 823–42.

Gochfeld, M. (2005). Chronologic history of occupational medicine. *Journal of Occupational and Environmental Medicine* 47(2): 96–114.

Greenberg, M. (2003). Cape asbestos, barking, health and environment: 1928–1946. *American Journal of Industrial Medicine* 43(2): 109–19.

Guyatt, G., Cairns, J., Churchill, D., Cook, D., Haynes, B., Hirsh, J., Irvine, J., et al. (1992). Evidence-based medicine: A new approach to teaching the practice of medicine. *Journal of the American Medical Association* 268(17): 2420.

Haack, S. (2009). *Evidence and Inquiry: A Pragmatist Reconstruction of Epistemology*. London: Prometheus Books.

Hájek, A. (2007). The reference class problem is your problem too. *Synthese* 156(3): 563–85.

Healy, B. (1991). The Yentl Syndrome. *New England Journal of Medicine* 325(4): 274–76.

Henrich, J., Heine, S., and Norenzayan, A. (2010a). Most people are not WEIRD. *Nature* 466(7302): 29. doi:doi:10.1038/466029a.

Henrich, J., Heine, S., and Norenzayan, A. (2010b). The weirdest people in the world? 139. Rates für Sozial- und Wirtschaftsdaten. http://www.econstor.eu/bitstream/10419/43616/1/626014360.pdf.

Howick, J. H. (2011). *The Philosophy of Evidence-Based Medicine*. John Wiley & Sons.

IARC Working Group. (2012). *Arsenic, Metals, Fibres and Dusts*. IARC. http://monographs.iarc.fr/ENG/Monographs/vol100C/index.php.

Illari, P. (2011). Mechanistic evidence: Disambiguating the Russo–Williamson thesis. *International Studies in the Philosophy of Science* 25(2): 139–57.

Illari, P., and Russo, F. (2014). *Causality: Philosophical Theory Meets Scientific Practice*. Oxford: Oxford University Press.

Illari, P., Russo, F., and Williamson, J. (2011). Why look at causality in the sciences? In *Causality in the Sciences*, eds. Phyllis Illari, Federica Russo, and Jon Williamson. Oxford: Oxford University Press, pp. 3–22.

Jonsen, A. R., and Toulmin, S. (1988). *The Abuse of Casuistry: A History of Moral Reasoning*. Berkeley, CA: University of California Press.

Kelly, M., Kelly, R., and Russo, F. (2014). The integration of social, behavioural, and biological mechanisms in models of pathogenesis. *Perspectives in Biology and Medicine* 57(3): 308–28.

Lewis, D. K. (1973). *Counterfactuals*. Oxford: Blackwell.

Lewis, D. K. (1983). *Philosophical Papers*. Vol. 1. Oxford: Oxford University Press.

Machamer, P. (2004). Activities and causation: The metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science* 18(1): 27–39.

Machamer, P., Darden, L., and Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science* 67(1): 1–25.

Mackie, J. L. (1974). *The Cement of the Universe: A Study on Causation*. Oxford: Oxford University Press.

Marmot, M., and Wilkinson, R. (2005). *Social Determinants of Health*. Oxford: Oxford University Press.

Mayer, D. (2009). The injured ankle and foot. *BMJ* 339: b2901.

McGrath, S. (2005). Causation by omission: A dilemma. *Philosophical Studies* 123(1): 125–48.

Merz, C., and Bairey, N. (2011). The Yentl Syndrome is alive and well. *European Heart Journal* 32(11): 1313–15. doi: 10.1093/eurheartj/ehr083.

Mossano, S. (2011). The eternit trial: The verdict is close. *Epidemiologia E Prevensione* 35(3–4): 175–77.

NICE. (2011). Hypertension: Quick reference guide. The National Institute for Health and Care Excellence. www.nice.org.uk.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Persson, J. (2002). Cause, effect, and fake causation. *Synthese* 131(1): 129–43.

Rabins, P. V. (2013). *The Why of Things: Causality in Science, Medicine, and Life*. New York, NY: Columbia University Press.

Reichenbach, H. (1949). *The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*. University.

Reiss, J. (2015). *Causation, Evidence, and Inference*. London: Routledge.

Rothman, K. J. (1976). Causes. *American Journal of Epidemiology* 104(6): 587–92.

Rothman, K. J., Greenland, S., and Lash, T. L. (2008). *Modern Epidemiology*. Philadelphia, PA: Wolters Kluwer/Lippincott Williams and Wilkins.

Russo, F. (2009). *Causality and Causal Modelling in the Social Sciences: Measuring Variations*. Methodos Series. New York: Springer.

Russo, F. (2014). What Invariance is and how to test for it. *International Studies in the Philosophy of Science* 28(2): 157–83.

Russo, F., and Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in Philosophy of Science* 21(2): 157–70.

Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Salmon, W. C. (1989). *Four Decades of Scientific Explanation*. Minneapolis, MN: University of Minnesota Press.

Salmon, W. C., Jeffrey, R. C., and Greeno, J. G. (1971). *Statistical Explanation and Statistical Relevance.* Pittsburgh, PA: University of Pittsburgh Press.

Signani, F. (2013). *La Salute Su Misura. Medicina Di Genere Non È Medicina Delle Donne*. Este Edition.

Solomon, M. (2011). Just a paradigm: Evidence-based medicine in epistemological context. *European Journal for Philosophy of Science* 1: 451–66.

Spirtes, P., Glymour, C., and Scheines, R. (1993). *Causation, Prediction, and Search*. Second (2000). Cambridge, MA: MIT Press.

Terracini, B., Mirabelli, D., Magnani, C., Ferrante, D., Barone-Adesi, F., and Bertolotti, M. (2014). A critique to a review on the relationship between asbestos exposure and the risk of

mesothelioma. *European Journal of Cancer Prevention* 23(5): 492–94. doi:10.1097/CEJ.0000000000000057.

Vandenbroucke, J. (2009). Commentary: "Smoking and lung cancer"—the embryogenesis of modern epidemiology. *International Journal of Epidemiology* 38(5): 1193–96.

Vecchia, C. La, and Boffetta, P. (2014). A critique of a review on the relationship between asbestos exposure and the risk of mesothelioma: Reply. *European Journal of Cancer Prevention* 23(5): 494–96. doi:10.1097/CEJ.0000000000000051.

Victora, C. G., Habicht, J.-P., and Bryce, J. (2004). Evidence-based public health: Moving beyond randomized trials. *American Journal of Public Health* 94(3): 400–405.

Vineis, P. (2003). Causality in epidemiology. *Social and Preventive Medicine* 48(2): 80–87.

Note

1. We are indebted to Connor Cummings, who developed this example in his BSc dissertation at the University College London on reference classes.