Studies in History and Philosophy of Biological and Biomedical Sciences 40 (2009) 204–209

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

Semmelweis's methodology from the modern stand-point: intervention studies and causal ontology

Johannes Persson

Department of Philosophy, Lund University, Kunghuset, Lundagård, SE-222 22 Lund, Sweden

ARTICLE INFO

Article history:
Received 24 November 2008
Received in revised form 7 March 2009

Keywords:
External validity
Internal validity
Randomized controlled trial
Ontology
Cause
Semmelweis

ABSTRACT

Semmelweis's work predates the discovery of the power of randomization in medicine by almost a century. Although Semmelweis would not have consciously used a randomized controlled trial (RCT), some features of his material—the allocation of patients to the first and second clinics—did involve what was in fact a randomization, though this was not realised at the time. This article begins by explaining why Semmelweis's methodology, nevertheless, did not amount to the use of a RCT. It then shows why it is descriptively and normatively interesting to compare what he did with the modern approach using RCTs. The argumentation centres on causal inferences and the contrast between Semmelweis's causal concept and that deployed by many advocates of RCTs. It is argued that Semmelweis's approach has implications for matters of explanation and medical practice.

When citing this paper, please use the full journal title *Studies in History and Philosophy of Biological and Biomedical Sciences*

1. Introduction

Ignaz Semmelweis (1818–1865) is famous for his enquiries into the causes of childbed fever. His contribution to the field of birthing sciences has been said to be among 'the most moving, persuasive, and revolutionary works in the history of science' (Codell Carter, 1983, p. ix) and is used as example in leading textbooks in philosophy of science. It will be shown in this article that while Semmelweis performed several clinical trials, neither randomization nor control groups were involved in these. In other words, he did not perform randomized controlled trials (RCTs). This is interesting since today RCT is the 'gold standard' for judging whether a treatment does more good than harm (Sackett et al., 1996, p. 72). In particular, few challenge the merits of RCT when it comes to warranting inference to causes.[1] Should the traditional assessment of Semmelweis's contribution be revised downwards?

As we shall see, Semmelweis's conception of causation probably differed from that deployed by many of his contemporaries. It differs from the concept utilized by students of healthcare today—especially those advocating randomized trials—as well. His necessitarian causal ontology makes inference to causes demanding in a way that ensures that such inferences are not dramatically facilitated by RCT. Supporters of RCT, on the other hand, sometimes help themselves to a causal concept which makes inference to *local* causes, that is, the internal validity (Campbell, 1957) of causal inferences, a rather trivial matter as soon as randomization can be implemented while at the same time rendering causal *generalisation*, that is, the external validity of causal inferences,[2] problematic (cf. Kristiansen and Mooney, 2004, p. 8).

E-mail address: johannes.persson@fil.lu.se

[1] However, James Le Fanu (1999), p. 406, has claimed that 'this statistically derived knowledge . . . has consistently been shown to be unreliable, promoting the patently absurd as proven fact'.

[2] Two comments are needed at this early point. First, the statistical relevance required in order for something to come out as causal in the RCT may be difficult to meet in certain plausibly causal contexts. Many examples from epidemiology testify to this point. It is explicitly remarked in Sackett et al. (1996) that 'we should try to avoid the non-experimental approaches, since these routinely lead to false-positive conclusions about efficacy'. Nevertheless, *inference* to causes in favourable circumstances is uniquely unproblematic on this view. It is only the latter feature of RCT this article criticizes. Second, note that Cronbach (1982) defines internal and external validity in a slightly different way than Campbell. According to Cronbach, statistical generalisations—whether causal or not—may be instances of internally valid inferences. But even on his definition, causal generalisations to different units, treatments, or observations are matters of external validity. While being explicitly directed at Campbell's conception most conclusions in this article should apply to Cronbach's conception of internal validity as well.

1369-8486
doi:10.1016/j.shpsc.2009.06.003

The contrast, drawn in this article, between the 'gold standard' and Semmelweis's research illustrates the way in which ontology typically influences epistemology and vice versa. It also shows that just what an RCT uniquely adds depends on logically independent assumptions concerning the nature of causation, and on whether internal or external validity is at issue.

2. Historical reasons

One obvious reason why Semmelweis's research does not live up to the current gold standard is that randomization was not incorporated in any comparable standard operating in his own time. Randomized trials clearly did not have the status in the 1840s they have now. Exactly when randomization becomes evidentially important we do not know. There are requirements of randomization in psychic research in the 1890s (Hacking, 1988).

The popularity of randomization develops from R. A. Fisher's later methodology for experiments in agriculture. Success did not come immediately, as Hacking reminds us:

> In 1932, when Fisher had a research student write a dissertation on randomized experimental design (at Rothamstead, but for a University of London degree), no one was willing to examine it, even though at the time Britain was still the leading center of pure and applied statistical theory. (Ibid., p. 429)

RCT is said to have found its way into medicine and healthcare via the work of B. A. Hill (1937) and studies of the efficacy of drugs (Pedersen, 2004). In light of this historical fact we should not expect to find randomization in Semmelweis's work (which, to clarify the timeline, predates the discovery of the power of randomization in medicine by almost a century).

3. An institution of birth and death

A great deal of information about the problem Semmelweis worked on and the context in which it arose is available. The General Hospital of Vienna housed an enormous maternity wing, catering for about 8,000 patients a year, in the mid-nineteenth century when Semmelweis began his career. No other hospital in the world had such a high reputation for the teaching of obstetrics (Loudon, 1992, p. 65). Maternity care was provided in two clinics from 1833 onwards: the First Maternity Division and the Second Maternity Division. After 1840 only the First Clinic (as it is normally called) was used for the instruction of male medical students; the Second Clinic was reserved for the instruction of midwives.

Childbed fever haunted the First Clinic. It was called 'puerperal fever' because it often occurred during the *puerperium* (approximately six weeks after childbirth) when the womb returns to its normal shape. Between 1833 and 1840 death rates in the two clinics were comparable, but in the period 1841–1846 the death rate was 9.92% in the First Clinic and 3.88% in the second (Gillies, 2005, p. 161). In fact the difference was even more pronounced than these numbers suggest, since in severe cases of puerperal fever patients were sometimes removed from the First Clinic and placed in the general hospital, where they normally died—thereby failing to be registered in the First Clinic's mortality statistics (Semmelweis, 1983, pp. 64–65). There is thus a sense in which the Vienna Maternity Hospital was indeed, as a student of Semmelweis's once remarked, 'truly an institution of death'.[3]

4. Non-interventionist refutations

Following Hempel (1966), introductions to the philosophy of science often refer to Semmelweis in connection with his convincing enquiry into the causes of the higher death rate from childbed fever in the First Clinic. Although Semmelweis's work on childbed fever had forerunners in the investigations conducted by Alexander Gordon (1752–1799) and Oliver Wendell Holmes (1809–1894), his method, involving hypothesis-testing in clinical trials, has a special significance. He had reason to look for causes inside the hospital, since maternity hospital closures, though drastic, were known to be an efficient ways of curtailing outbreaks of childbed fever:

> Hospitals are closed not to force maternity patients to die elsewhere, but because of the belief that if patients deliver in the hospital they are subject to epidemic influences, whereas if they deliver elsewhere they will remain healthy. (Semmelweis, 1983, pp. 66–67)

Semmelweis examined, and swiftly eliminated, some rather obvious but erroneous causal hypotheses relating to hospital management. A few examples: the incidence of childbed fever is raised by the clinic's practice of admitting only single women in desperate circumstances; childbed fever is caused by the poor ventilation; it spreads through the laundry process (where a clinic's laundry was mixed with that of the general hospital); and it results from dietary mistakes.

Semmelweis's elimination of these hypotheses fits well with the hypothetico-deductive method associated with Hempel. What Hempel does not mention is that randomization was at least unintentionally in play at this stage in Semmelweis's enquiry. Women were admitted to the two clinics on alternate days (Loudon, 1992, p. 65). Exploiting this mechanism so as to control for relevant differences among women in the two clinics strengthens the assumption that the cause of the fever was to be found in hospital management. It also increases the evidential value of Semmelweis's observations that the clinics were ventilated in the same way, that the laundry contractor mixed both the laundry of the first and the Second Clinic with that of the general hospital, and that the food provided was the same in both clinics. The suggested hypotheses above are incompatible with these facts.

Does this imply that, implicitly, Semmelweis was conducting RCTs after all? No: there may have been a control group, and randomization may have operated, in the early phase of his research, but the most essential component is lacking: the *intervention*. The early phase is an intellectual one relying mostly on information that has been collected before the testing of the hypotheses.

5. Two interventionist studies

To make a stronger case for the possibility that important parts of Semmelweis's research were in practice conducted in accordance with the guidelines of evidence-based medicine, so that evidence from his enquiries would be not only acceptable but of highest rank, we need to examine later phases of the enquiry where Semmelweis put more promising hypotheses to the test. The studies here are clearly interventionist and, in a broad sense of the word, 'experimental'. We shall use two of these intervention studies for illustrative purposes.

To begin with, then, one of the first hypotheses Semmelweis tested through intervention was based on the following conclusion of a hospital commission:

[3] A comment by one of Semmelweis's students, quoted from Loudon (1992), p. 68.

THE MALE FOREIGNER INTERVENTION

> Toward the end of 1846 an opinion prevailed in one commission that the disease originated from damage to the birth canal inflicted during the examinations that were part of the instructional process. However, since similar examinations were part of the instructions of midwives, the increased incidence of disease in the clinic for physicians was made intelligible by assuming that male students, particularly foreigners, were too rough in their examinations. As a result of this opinion the number of students was reduced from forty-two to twenty. Foreigners were almost entirely excluded, and examinations were reduced to a minimum. The mortality rate did decline significantly in December 1846, and in January, February, and March of 1847. But in spite of these measures, fifty-seven patients died in April and thirty-six more in May. This demonstrated to everyone that the view was groundless. (Semmelweis, 1983, p. 84)

The hypothesis in the second example involves Semmelweis's own invention. Semmelweis developed the hypothesis after learning that one of his colleagues, Dr Kolletschka, died from a disease 'identical to that from which so many hundred maternity patients had also died' (ibid., 1983, p. 88) and that Kolletschka had received a puncture wound 'with the same knife that was being used in the autopsy' (ibid., p. 87):

THE CHLORINA LIQUIDA INTERVENTION

> Suppose cadaverous particles adhering to hands cause the same disease among maternity patients that cadaverous particles adhering to the knife caused in Kolletschka. Then if those particles are destroyed chemically, so that in examinations patients are touched by fingers but not by cadaverous particles, the disease must be reduced. This seemed all the more likely, since I knew that when decomposing organic material is brought into contact with living organisms it may bring on decomposition. To destroy cadaverous matter adhering to hands I used *chlorina liquida*. This practice began in the middle of May 1847; I no longer remember the specific day. Both the students and I were required to wash before examinations. After a time I ceased to use *chlorina liquida* because of its high price, and I adopted the less expensive chlorinated lime. In May 1847, during the second half of which chlorine washings were first introduced, 36 patients died—this was 12.24 percent of 294 deliveries. In the remaining seven months of 1847, the mortality rate was below that of the patients in the second clinic. In these seven months, of the 1841 maternity patients cared for, 56 died ... Since the chlorine washings were instituted with such dramatic success, not even the smallest additional changes in the procedures of the first clinic were adopted to which the decline in mortality could be even partially attributed. (Ibid., pp. 91–92)

Are these intervention studies perhaps cases of RCT? John Matthews (2006, p. 3) has claimed that if they were RCT studies, they would involve (1) a population of eligible patients; (2) a group of patients recruited from this population; (3) at least two treatment groups; and (4) randomized allocations of treatment. It would also be necessary (5) for outcome measures in the treatment groups to be compared at the end of the trial.

It should be conceded at once that there is reference to the Second Clinic in the *male foreigner* and the *chlorina liquida* intervention studies. In fact, Lipton (2004, pp. 75–79) defends the view that Semmelweis attached great importance to comparisons between First Clinic and Second Clinic. Hence component (3) seems to be in place in both cases. Equally, the same kind of random allocation mechanism as before is in place. Loudon (1992, p. 65) claims that there was a 'system of random allocation' in play in Semmelweis's studies. Hence conditions (1), (2) and (4) appear to be fulfilled. Finally, there are some post-test comparisons between the first and second clinics, so an element of (5) is present as well.

It is certainly tempting to conclude that the tension between RCT studies in evidence-based medicine and Semmelweis's research, set up in this article, is exaggerated; but this would be a mistake. The reconciliatory attempt above is driven by superficial features. It may be true that (1)–(5) can be used to describe Semmelweis's two interventions studies, but the description is only partly accurate as a reconstruction of what was going on.

There are both descriptive and normative reasons for this. To begin with the descriptive reasons: Semmelweis is more concerned with comparisons between pre- and post-tests within the First Clinic than he is with comparison of the first and second clinics. There might be an element of (3) and (5) in what Semmelweis does, but it does not seem very important to him. Moreover, there is no evidence that Semmelweis was content with the random allocation and the way in which the alleged control group was set up. He does not seem to be in control of the situation. If this is true, it immediately tells against ascribing the RCT design to him. Experimental designs in general and RCTs in particular are often characterised in terms of the investigator's control, or lack of control, of the allocation of participants to intervention groups (cf. Deeks et al., 2003, p. 2). On the basis of the admittedly weak evidence we have of Semmelweis's intentions and administrative authority, it would be hard to come to the decisive conclusion that any of these interventions were cases of RCT—at least, as we know it today.

More strikingly, there are normative reasons to doubt that Semmelweis utilised the Second Clinic in order to perform intervention studies with a control group. On the assumption that Semmelweis is right about his hypotheses there are major differences between the intervention group and the control group before treatment. In the male foreigner intervention, to begin with, the supposed control group consists exclusively of midwives who are supposed to be gentler than staff at the First Clinic. In particular, it is implicit in the discussion that there were no foreign men among these midwives. In the *chlorina liquida* intervention (to move on to the next case) the cadaveric matter hypothesis postulates a major difference between the groups. According to that hypothesis, the intervention group inhabits an environment that supposedly contains much more in the way of causally active cadaverous particles than the control group does. It is evident that the two interventions render the two clinics more similar than they were before the treatment. This makes the two intervention studies bad examples of RCT. It is hard to believe that Semmelweis—a hero in the annals of hypothetico-deductive research—would make such a mistake at all, let alone twice.

It comes as no surprise, then, that in the two intervention studies performed the effects of the manipulations are measured against how things were before the intervention took place in the First Clinic rather than how things were in the Second Clinic. Not only is randomization absent in Semmelweis's well known intervention studies, but the idea of a simultaneous control group seems to have been of little interest to him as well.

6. The ontology and epistemology of causation

Sometimes we mistakenly infer causal relations from real enough, but misleading correlations. Smoking and lung cancer are probably related as cause to effect, but there is a small possibility that both have a common cause. As long as we have suspicions as to what this common cause might be we can devise comparative studies to investigate the matter; but often we do not know, have no idea, what the common cause might be. The popularity of RCT is

due in part to its ability to handle exactly this type of situation. Randomization is a handy tool that helps us to control for undetected common causes—and, in that way, improve on matters of internal validity.

To be in a position to make valid causal inferences from RCT outcomes we seem to require extra assumptions. Yet more than one advocate of randomization adopts a view on which RCT alone underwrites a positive causal inference. Consider David Papineau for example:

> You take a sample of people with the disease. You divide them into two groups at random. You give one group the treatment, withhold it from the other ... and judge on this basis whether the probability of recovery in the former group is higher. If it is, then T [treatment] must now cause R [recovery], for the randomization will have eliminated the danger of any confounding factors which might be responsible for a spurious correlation. (Papineau, 1994, p. 439)

This is excessively optimistic for reasons having to do with the possible artefacts of randomization (cf. Shadish et al., 2002, Ch. 2), but that is, not the present point. Let us assume that randomization is successful in the desired respect. Papineau's modified position seems to rely on a concept of causation given which *in the relevant cases* causation is entailed by (i.e. is unproblematically inferable from) the fact that the relative frequency of R in the intervention group is higher than it is in the control. Thus, for instance, the concept of cause employed is not that causes are sufficient in the circumstances, nor that they are necessary. This is plainly not so since neither kind of causation is entailed by the experimental fact. Let us reserve 'P-causation' for the type of causal ontology compatible with Papineau's general approach.[4]

Now consider Semmelweis's intervention studies. They reveal almost as much about his concept of cause, 'S-causation', as the above quotation reveals about P-causation. First, Semmelweis's argument from the mix of laundry from the First Clinic and Second Clinic, together with the differences in mortality between the two clinics, to the conclusion that mixing the laundry with laundry from the general hospital does not cause childbed fever suggests that he is looking for a cause potent enough to actually result in effects in the relevant circumstances. In this respect S-causation resembles P-causation.

But second, and more importantly, the argument from the fact that hospital closures elsewhere have prevented further outbreaks of childbed fever to the conclusion that we need to look for causes in this hospital environment strongly suggests that Semmelweis conceives of causes as *necessary* for effects. This suspicion is confirmed by close examination of Semmelweis's refutation in the male foreigner intervention. The withdrawal of foreign males is accompanied by lower mortality rates for several months. Then, and still within the temporal boundaries of the intervention, mortality increases. Semmelweis's summary, recall, was straightforward: 'This demonstrated to everyone that the view was groundless'. But unless S-causation is tied to unique causes, Semmelweis's claim is too strong. Perhaps, we might think, another cause of childbed fever, compensating for the non-occurrence of X, emerged towards the end of the intervention at $T*$.

The most interesting contrast here is with P-causation. Imagine that two *successfully* randomized, controlled trials are run at T and $T*$ respectively. In each case, the intervention consists in random withdrawal of male foreigners. Let us focus on the first trial, where we assume that there is an increase in relative frequency. In this case—given what Papineau says—withdrawal is a P-cause of the effect. This is clearly compatible with there being no increase in relative frequency in the other trial. This result should, for example, be expected if 'the other cause' of childbed fever was present in the intervention and control group in the second trial but not the first.

Looking at matters more generally, we find direct indications that Semmelweis's concept of causation is linked to that of necessity in his many causal claims. For example: 'In order for childbed fever to occur, it is a *condition sine qua non* that decaying matter is introduced into the genitals' (Semmelweis, 1983, p. 149). And, of course, others have ascribed a necessitarian notion of causation to Semmelweis:

> Semmelweis seems to have been among the first to conceive of puerperal fever in a way such that it would have a necessary cause—all of his contemporaries seem to have been thinking in altogether different ways. (Codell Carter, 1983, pp. xlix–l)

> In the second half of the nineteenth century ... it became increasingly standard to try to characterise a disease in terms of a single necessary cause ... Semmelweis was in fact one of the first to adopt this new approach to causality ... but the very novelty of the approach must have made it hard for his contemporaries to understand and accept. (Gillies, 2005, pp. 175–176)

The reception of Semmelweis's conclusions, as reported in Codell Carter's introduction to Semmelweis's *The etiology, concept, and prophylaxis of childbed fever* (1983), displays a preoccupation with the nature of causation. On the one hand, there are people like Semmelweis's influential friend Josef Skoda. Skoda announces Semmelweis's results, but overlooks, or at least downplays, any notion that he claimed to have found the necessary cause of childbed fever. Skoda even denies this ambition in official formulations such as: '[Semmelweis was] not seeking to explain all the causes of puerperal fever, but only to find and to circumvent the causes of the excessive mortality in the first clinic' (quoted from Codell Carter, p. xxx n. 156). On the other hand, there are those who correctly understand Semmelweis's causal claim but adjudge the reasons to accept it wanting. For instance, Hermann Lebert, professor of clinical medicine at Breslau, might have accepted that Semmelweis discovered *one* cause of childbed fever, but he was not at all convinced of its uniqueness, that is, necessity. A number of victims of childbed fever, Breslau conjectured, had never been exposed to cadaveric matter:

> It is questionable whether those who have died of this disease can have been directly inoculated by poison from corpses. Semmelweis has elevated this possibility into a system. In any case this would be only one of many possibilities of conveyance. (Quoted from ibid., p. xxxiii)

Those who accepted and those who rejected Semmelweis's conclusions about *what* it was that caused childbed fever shared a sceptical view of his ontological views on causation.

The implications of this clash between causal ontologies will be discussed further in the next section. It should be noted that some of the evidence can reasonably be said only to support the weaker interpretation that Semmelweis conceived of *this* specific cause as

[4] We should be careful not to take it for granted that what has been quoted is Papineau's actual position. In a few passages he adds the further and complicating constraints that what is detected is an *objective probability* or even an *single-case objective probability*. Every such addition—especially of the former kind—decreases the internal validity, that is, which causal claims can actually be demonstrated in the ideal RCT, by making P-causation more distant from what is manifested in ideal RCTs. It should be noted that several other philosophers of causation express views that are in line with P-causation. For instance, Cartwright (2009), p. 129, states that 'it is possible to show that in an "ideal" RCT a positive result deductively implies the conclusion under test: If there is a higher probability of O in the treatment group in an "ideal" RCT than in the control group, it follows deductively that T causes O in the experimental population under the experimental conditions'.

necessary. This means that the above observations are even more important in settling the ontological question. His refutations of alternative hypotheses, more than his presentation of the cadaveric matter hypothesis, show that Semmelweis was only looking for a necessary cause. It was no coincidence that the cause he found was of this kind.

Again, note that these two concepts of causation, P- and S-causation, are automatically built into neither the single group design implemented by Semmelweis nor the RCT design advocated by Papineau. For instance, if the claim of this article—that there is in fact no role for randomization in Semmelweis's intervention studies—is mistaken, he may well be conducting RCTs *with* a necessitarian concept of cause. And this would still be helpful in partly handling ignorance. The reason why Semmelweis chooses S-causation instead of P-causation is probably connected with the divergent consequences of P- and S-causation, and especially with their difference with regard to external validity. Certainly, when we examine Semmelweis's reason for rejecting the causal claim of the commission we find that this difference is both visible and significant. But before giving this illustration let us first briefly mention a difference that seems more directly induced both by the adoption of an RCT-perspective and Semmelweis's hypothesis-driven approach.

In Semmelweis's approach the intervention, as well as the rest of the experimental situation, typically belongs to the realm of empirical consequences rather than reflecting the causal hypothesis itself. Hence, a theoretical, or ontological, distance between a relevant cause and what is intervened on and tested in experimental studies is inevitably assumed in the latter case. This is less inevitable with the P-causation approach. On this approach it becomes natural to identify the alleged cause with the withdrawal of male foreigners, to be sure. But when describing Semmelweis's case it is clearly preferable to identify the alleged cause with that indicated in Semmelweis's original text when he says that 'the damage to the birth canal inflicted during the examinations that were part of the instructional process'.

Back to the reason for refusal: Semmelweis has an ontological reason to reject the hypothesis and to say that it is groundless that an advocate of P-causation would not have. A truly necessary cause cannot be absent while the effect is present. It does not help that when it is absent at T the effect is absent as well. It must be the case that when it is absent at $T*$ the effect is absent again. S-causes come with entailments about other situations in a way that P-causes do not.[5] Actually, given what Papineau has told us about P-causes, the only thing we can do with them is to plead ignorant—in neither case has Semmelweis provided us with the kind of experimental design that allows us to determine inferentially whether P-causes are present or not.

7. The importance of practice and explanation for ontology

The primary context in which randomization is discussed is epistemological (Papineau, 1994, is typical in this respect). This contextual reality emphasizes that the need for randomization is positively correlated with ignorance about the local case under scrutiny. It is often in order to control for unknown bias that we randomize. It drives us towards a concern with matters of internal validity.

This article has pointed to a different, but complementary, correlation between randomization and ontology. Sometimes the belief in the power of experimentation spawns constructive ideas about causal ontology, like P-causation. Given P-causation, the perfectly randomized controlled trial validates inferences to causes. The presumption that a successful RCT proves causation, and that therefore causation is P-causation, is widespread even among critics of evidence-based medicine (see e.g. Daly, 2005, p. 6). Randomization has an ontological, as well as purely epistemological, interface.

On the assumption that the combination of RCT and P-causation resolves the perennial problem of causal inference, and assuming also that it is a good thing to find any treatment that increases the probability of curing the patient or preventing a disease, should we not let the 'gold standard' of evidence-based medicine guide our ontology to P-causation?

Not so quickly. The choice of ontology has consequences that are less straightforwardly epistemological than the above focus on internal validity suggests. Two of these concern practice and explanation. The comparison of S- and P-conceptions of childbed fever in these respects helps us to understand Semmelweis's struggle, as I shall try to explain in the next two subsections.

7.1. Implications for practice

The ontological conflict between Semmelweis and his colleagues was at least partly grounded in disagreements over the way in which diseases should be characterized and defined. It had been the case that morphological features were decisive. With such a concept of disease the possibility that each disease has various unrelated causes is clearly not ruled out. Semmelweis's work was partly inspired by another, causal, conception of disease, and on this conception diseases must have necessary causes—it is impossible for them not to do so. Semmelweis was not afraid of the re-categorization that might follow from this. For instance, the concept of childbed fever might have dissolved completely: 'Kolletschka also had this disease. Thus childbed fever is not a species of disease; rather it is a type of pyemia' (Semmelweis, 1983, p. 117). In other words, the ontological difference between P- and S-causation and the differing views on how to define diseases have obvious corollaries in medical practice.

Medical practice benefits enormously from the discovery of necessary causes that can be blocked, or annulled, before they become causally efficacious. But characterizing causes so that each disease can have various unrelated causes makes it difficult to generate effective techniques for controlling the disease in some cases. Physicians have identified both excessive and inadequate diet as causes of the same disorder. In such cases it is even possible for the measures adopted to correct one of these factors to cause the disorder by bringing about the other factor (Codell Carter, 1983, p. xxvii). Regardless of how the world really is, the policy of searching only for necessary causes—adopting a causally necessitarian worldview—to a large extent circumvents this problem in medical practice.

However, this focus on the virtues of necessity reflects directly only one reason that the choice between P- and S-causation is important for practice. Of even greater practical importance is the difference arising for matters of external validity, or causal generalization. To recap: the few things we have so far assumed about P-causation permit us to conclude that proving P-causation has few implications for external validity. From within the P-causation framework, questions of internal validity (proving causation) seem firmly separated from questions of external validity (justifying causal generalizations). Indeed, in order to limit misunderstandings

[5] Cronbach (1982), pp. 137–138, makes a similar point, but, arguably, he misidentifies the reason the conclusion holds. He argues that internal validity focuses on local situations in the *past*. My argument is that some of those pursuing the internal validity issue adopt a P-causation conception that fails to support these entailments. P-causation formulated as in the quotation from Papineau clearly has this problem, and so has the position reported in Cartwright (2009), and it still exists if what is claimed to be detected is a single-case probability (cf. n. 4).

about the implications of 'internal validity', D. T. Campbell has proposed a new label: 'local molar causal validity'. The word 'local' emphasizes the idea that the validity being claimed is limited to 'the context of particular treatments, outcomes, times, settings, and persons studied' (Shadish et al., 2002, p. 54). That is, randomization in the sense we are discussing here is not supposed to have bearing on external validity but is all important so far as *local* matters of P-causation are concerned. In this respect, as we have already remarked, S- and P-causation clearly diverge.

This difference is not dependent on P-causation being non-necessitarian as such: it is due to it not being as universal. Even if (in accordance with endnote 2) we grant that, under circumstances where local P-causes have been sampled carefully, P-causation claims statistically generalize to the corresponding population, nothing whatsoever is entailed about slightly different circumstances. Plainly this has implications for practice as well. Lee J. Cronbach has made this point:

> I consider it pointless to speak of causes when all that can be validly meant by reference to a cause in a particular instance is that, on one trial of a partially specified manipulation *t* under conditions A, B, and C, along with other conditions not named, phenomenon P was observed. To introduce the word *cause* seems pointless. Campbell's writings make internal validity a property of trivial, past-tense, and local statements. (Cronbach, 1982, p. 137)

At this stage S-causation (but not P-causation) seems to be best used for medical decision-making. At least, with P-causation there is a need for extra assumptions grounding the external validity of relevant P-causation claims.[6]

7.2. Implications for explanation

One cannot consider Semmelweis's research without reflecting on the question why he failed to convince others. The story of Semmelweis is a tragic one at this personal level. Donald Gillies (2005) tells a plausible Kuhnian story about dominant paradigms and the importance of a new hypothesis emerging in revolutionary times in order to stand the chance of being accepted. According to him, Semmelweis's theory clashed with the dominant paradigm. As we have seen, an obvious problem with S-causation was that it did not harmonize with entrenched ontological beliefs in medical science.

Disregarding these entrenched beliefs but keeping Gillies's idea about the importance of a theory or paradigm in mind, we might conjecture that S-causation has explanatory potential that P-causation lacks. With P-causation it is difficult to generate relevant explanations for the observed facts.[7] The reason is straightforward: what we nearly always succeed in intervening on in RCT studies are complex entities rather than any specific property or variable. Similarly, the outcome derived from an intervention—or indeed any causal process—is a complex entity. What we manage to do is multifaceted, and this creates an inability to distinguish causal from causally irrelevant relations displayed in the study. The more coarse-grained and indirect the intervention is, the more difficult it will be to disentangle the causally relevant and causally irrelevant factors that are components both of the intervention and the observed outcome. But disentanglement is needed if we are to explain why something happened (cf. Persson & Sahlin, 2009). As has been shown in the literature on scientific explanation, irrelevant information efficiently destroys explanatory power: that water dissolves salt may be a satisfactory, although shallow, explanation of the fact that a piece of salt 'disappeared' when put in water. However, add the information that the water was holy and the explanation vanishes into thin air—similarly, of course, with other interventions containing a blend of relevant and irrelevant explanatory features. In sum, there are a number of interesting reasons not to take the modern stand-point with regard to intervention studies and causal ontology for granted.

Acknowledgements

The author would like to thank Ingalill Rahm Hallberg, Nils-Eric Sahlin, Lena Wahlberg, Annika Wallin, and an anonymous referee for a number of valuable comments. Funding for this work was provided by The Vårdal Foundation, The Swedish Research Council, and Lund University.

References

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin, 54*, 297–312.
Cartwright, N. (2009). Evidence-based policy: What's to be done about relevance? *Philosophical Studies, 143*, 127–136.
Codell Carter, K. (1983). Translator's introduction. In I. Semmelweis. *The etiology, concept, and prophylaxis of childbed fever* (K. Codell Carter, Trans.) (pp. iii–lviii). Madison, WI: The University of Wisconsin Press.
Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Fransisco: Jossey-Bass.
Daly, J. (2005). *Evidence-based medicine and the search for a science of clinical care*. Berkeley: University of California Press.
Deeks, J. J., Dinnes, J., D'Amico, R., Sowden, A. J., Sakarovitch, C., Song, F., Petticrow, M., & Altman, D. G. (2003). Evaluating non-randomised intervention studies. *Health Technology Assessment*, *7*(27). http://www.hta.ac.uk/fullmono/mon727.pdf.
Feinstein, A. R., & Horwitz, R. I. (1997). Problems in the 'evidence' of 'evidence-based medicine'. *The American Journal of Medicine, 103*(December), 529–535.
Gillies, D. (2005). Hempelian and Kuhnian approaches in the philosophy of medicine: The Semmelweis case. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 159–181.
Hacking, I. (1988). Telepathy: Origins of randomization in experimental design. *Isis, 79*(3), 427–451.
Hempel, C. (1966). *Philosophy of natural science*. Englewood Cliffs, NJ: Prentice Hall.
Hill, B. A. (1937). *Principles of medical statistics*. London: Oxford University Press.
Kristiansen, I. S., & Mooney, G. (Eds.). (2004). *Evidence-based medicine in its place*. London: Routledge.
Le Fanu, J. (1999). *The rise and fall of modern medicine*. London: Little, Brown & Company.
Lipton, P. (2004). *Inference to the best explanation*. London: Routledge.
Loudon, I. (1992). *Death in childbirth*. Oxford: Clarendon Press.
Matthews, J. N. S. (2006). *Introduction to randomized controlled clinical trials* (2nd ed.). Boca Raton: Chapman & Hall.
Papineau, D. (1994). The virtues of randomization. *British Journal for the Philosophy of Science, 45*(2), 437–450.
Pedersen, K. M. (2004). Randomised controlled trials in drug policies: Can the best be the enemy of the good? In I. S. Kristiansen, & G. Mooney (Eds.), *Evidence-based medicine in its place* (pp. 124–140). London: Routledge.
Persson, J., & Sahlin, N.-E. (2009). A philosophical account of interventions and causal representation in nursing research: a discussion paper. *International Journal of Nursing Studies, 46*(4), 547–556.
Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence-based medicine: What it is and what it isn't. *British Medical Journal, 312*(7023), 71–72.
Semmelweis, I. (1983). *The etiology, concept, and prophylaxis of childbed fever (K. Codell Carter, Trans.)*. Madison, WI: University of Wisconsin Press.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.

[6] That this need is not easily met can be seen when we contemplate another common criticism of RCT as the best kind of evidence that evidence-based medicine can produce. For instance, it has been argued by Feinstein & Horwitz (1997), p. 529, that such generalizations will disclose at best the 'comparative efficacy of treatment for an "average" randomized patient'.

[7] Codell Carter (1983), p. 27, makes almost the same claim, but he mistakenly attributes this problem to morphological concepts of disease. However, diseases defined that way can still be combined with a concept of causation of Semmelweis's kind, and then the difficulty disappears. A combination of causally defined diseases and a conception of P-causation is also possible—although, arguably, it yields similar problems for explanation.