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.

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/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
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.