

HISTORIOGRAPHY TAKING ISSUE: ANALYZING AN EXPERIMENT WITH HEROIN ABUSERS*

TRUDY DEHUE

This article discusses the predicament of historians becoming part of the history they are investigating and illustrates the issue in a particular case. The case is that of the randomized controlled trial (RCT)—more specifically, its use for testing the effects of providing heroin to severe heroin abusers. I counter the established view of the RCT as a matter of timeless logic and argue that this research design was developed in the context of administrative knowledge making under twentieth-century economic liberalism of which it epitomizes some central values. I also argue that the applicability of the RCT depends on the degree to which its advocates can define the issue to be studied according to its inherent values. Next, I demonstrate how advocates of an RCT with heroin provision in the Netherlands steered the political discussion on heroin provision and how the values of economic liberalism also shaped the results of the Dutch maintenance experiment. In addition, I relate how my analysis of this experiment became part of political debates in the Netherlands. Contrary to my intentions, adversaries of heroin maintenance used my critique on the heroin RCT as an argument against heroin maintenance. Such risks are inherent to historiography and sociology of science aiming at practical relevance while challenging treasured scientific beliefs. I conclude that it still seems better to expose arguments on unjustified certainties than to suppress them for strategic reasons. © 2004 Wiley Periodicals, Inc.

Various European countries and Canada are currently conducting or planning scientific experiments with the provision of heroin to severe heroin abusers. The intention is to find out whether or not free heroin improves the physical and mental condition of despondent users and diminishes the nuisance often caused by them.

The United Kingdom, Switzerland, and particularly the Netherlands set the example. The Netherlands conducted a large-scale experiment with heroin maintenance according to the script of the randomized controlled trial (RCT).¹ RCTs, called “randomized *clinical* trials” in a medical context, compare experimental groups who get a treatment to control groups who do not. For the sake of comparability and statistical soundness, the participants are allocated to one of the groups on the basis of chance. In both medicine and social research, the RCT is known as the “ideal experiment” or “the gold standard of research.” Researchers as well as the wider public believe that it yields unambiguous results.

Yet, not everyone has that much faith in RCTs. During the years the Dutch planned and conducted the RCT with heroin provision, a range of articles and books appeared that challenged the assumption of a neutral “view from nowhere” achievable through RCTs and drew

***Editor’s Note:** This article, based on the keynote address delivered at the thirty-fifth annual meeting of the Cheiron Society at the University of New Hampshire, is the winner of the Cheiron/JHBS Award for 2003.

1. Stimson and Metrebian (2003; also on www.jrf.org.uk) give an overview of the Swiss, British, and Dutch heroin maintenance trials. For a summary of the design and results of the Dutch trial, see Van den Brink et al., 2003a, b (also on www.bmj.com). For the full report and a propaganda video on heroin maintenance, in both English and Dutch, see www.ccbh.nl.

TRUDY DEHUE is program chair of the History and Theory of Psychology Program at Groningen University. She works on the history and assumptions of research methods in the human sciences, particularly on those of the randomized controlled trial. Address for correspondence: Heymans Institute for Psychological Research, Grote Kruisstraat 2/1, 9712 TS The Netherlands. E-mail: t.dehue@ppsw.rug.nl.

attention to serious disadvantages of the design in particular settings.² Only partly aware at the time that the subject attracted the interest of a number of colleagues, I took it up as well. Focusing on the history of psychology and sociology, I argued that this type of social research did not, as textbooks suggest, develop in any natural science or in the context of clinical research but in the context of administrative knowledge making. I also argued that instead of representing an ahistorical universal logic, the RCT epitomizes some central political values of administration under twentieth-century economic liberalism (Dehue, 2001).

When the debates started in Dutch Parliament on a large-scale RCT with heroin maintenance, the example offered a challenging occasion to further test and develop my arguments. Hence, I followed the long-lasting parliamentary debates and studied the extensive experimental script. I also interviewed members of the staff conducting the experiment and representatives of users' organizations. I visited one of the stations where the heroin was provided and collected preliminary reports, letters, and newspaper clippings.

The Dutch heroin experiment was designed according to the highest practicable standards. Ingenious protocols for the production of the heroin; the design of the heroin stations; training of personnel; selection, treatment, and examination of the participants; and many other aspects of the project ensured that as little as possible was left to the discretion, preferences, or interpretations of the people conducting the experiment or those participating in it. Yet, I became increasingly skeptical about the experiment as a suitable way of establishing the effects of heroin maintenance. When the project was about to end, I clarified my misgivings in an English language article, with a subsequent translation into Dutch (Dehue, 2002a, 2002b).

Below I will elucidate my skepticism again. I will do this, however, in relation to another theme. The present article is based upon a keynote address at the thirty-fifth annual meeting of Cheiron, where the conference organizers had specifically requested that I also reflect upon the *reception* of my critique. As my previous analysis was interpreted by some journalists and politicians as a testimony against heroin maintenance rather than "just" the experiment for testing its effects, I gratefully accepted the occasion to compose my thoughts on this predicament and receive scholarly feedback at the Cheiron conference.

This article, therefore, first recounts how opponents of maintenance used my arguments on the heroin experiment. Next, I explore the general quandary of "historiography taking issue" and compare my case to that of the American historian Simon Cole, who investigated and challenged forensic fingerprint identification. Subsequently, I recap my earlier account of the historical origins of the RCT, arguing that this research design epitomizes three assumptions of twentieth-century economic liberalism. After that, I clarify my thesis that these implicit assumptions framed the Dutch discussion on heroin maintenance and shaped the results of the maintenance experiment. The final section gets back to the issue of historiography getting involved in the debates it is investigating. My tentative conclusion will be that,

2. In the 1980s, Evelleen Richards discussed experiments with cancer treatments and argued that clinical trials "inevitably embody the values or commitments of the assessors" (Richards, 1988, p. 653; Richards, 1991). Shortly thereafter, Kurt Danziger pointed at the social values hidden in educational RCTs with schoolchildren (Danziger, 1990, 2000). Harry Marks (1988, 1997) ascribed the rise of the RCTs in the medical context to an administrative culture of mistrust and described a range of far-reaching practical hazards in large-scale medical trials. Various authors published on the unavoidable ambiguity of the outcomes of drug safety tests of even the most meticulous RCTs and argued that this ambiguity generally serves the interests of the pharmaceutical industry (Abraham, 1995; Healy, 2001; Medawar, 2003). Steven Epstein studied the powerful role of AIDS activists in shaping the experimental evaluation of AIDS therapies (Epstein, 1995, 1996). Thirty years of controversy over coronary bypass grafting that could not be closed through clinical trials were surveyed by David Jones (Jones, 2000). Finally, Ted Kaptchuck, drawing attention to the progress most often shown by the control groups, pointed out that only since the canonization of the RCT have placebos become quack treatments rather than genuine cures (Kaptchuck, 1998, 2001).

regardless of the risks, it seems better to challenge untenable beliefs on straight versus non-straight facts than to stay safely hidden in academia's closet.

A SKEWED BUT UNDERSTANDABLE RECEPTION

At first, I avoided entering into the public debate on the heroin experiment and even more so on heroin maintenance. Focusing on the RCT and using the Dutch heroin experiment as only an example, I expressed my views among peers in the history and sociology of science. I deliberately prepared an article for an English language historical journal rather than one in Dutch. Yet, when the main Dutch monthly of mental health proposed to publish a translation, I let myself be convinced that my views should be heard by Dutch drug professionals and policymakers. Subsequently, I consented to an interview in a national newspaper. The interview triggered attention by various other newspapers and magazines and by radio and TV programs. A prime time TV news program asked for another interview, and my critique led to questions to the Minister of Health in Dutch Parliament.

From the beginning of my studies, I did not criticize the idea of heroin maintenance (while studying the experiment, I actually became a proponent of controlled maintenance to selected users³) nor did my articles doubt the forthcoming outcomes of the experiment only in the case it would end with a *positive* conclusion. However, at a press conference in February 2002, the Minister of Health made known that the outcomes were positive. She announced that both the control group and the experimental group showed progress, but the experimental group showed about 25 percent more than the controls.⁴ In addition, the "positive responders" rapidly relapsed after the termination of maintenance. The Minister added that Parliament should soon consent to legalization and extension of controlled heroin provision, and the research team started an instantaneous promotion campaign including a pro-maintenance video on the Internet (www.ccbh.nl).

Of course, the media were particularly interested in any critique after the presentation of the outcomes. My explanation that I did not object to heroin maintenance and that the results might have been suppressed by employing the RCT as a research instrument sometimes discontinued further contacts. In other cases, journalists still wanted an interview and promised that I would not be staged as an opponent to maintenance. Yet, whereas I authorized the newspaper interview, I had no control over a misleading summary on the same issue's front page. And the interview on the TV news program was embedded into an explicit anti-maintenance item. To many media and the public, "the experiment" simply *was* heroin maintenance and anyone criticizing the experiment therefore criticized maintenance. In Parliament too, the opponents of maintenance latched on to my critique. They demanded that the Minister of Health provide an extensive written reply. Understandably from a political point of view, this ministerial rebuttal (written, of course, by the research team) took a harsh tone and, equally understandably, it mostly got me wrong (the Minister's written reply to Parliament has been reprinted together with my reply to the Minister; see Dehue, 2002c).

As of May 2004, the final decision to legalize and extend the maintenance program is still under dispute. Supporters in Parliament put forth motion after motion to speed up the process, but the present government only agreed to maintain the heroin stations set up for the

3. For arguments against and in favor of heroin maintenance, see the Heroin Maintenance Bibliography of the Drug Policy Alliance at <http://www.drugpolicy.org/library/bibliography/heroin/>.

4. To be precise, participants who inhaled heroin were rated at 22.8 percent extra progress and those who injected it at 24.3 percent. However, as I will argue below, there is hardly a reason for precision to decimals.

experiment and to give heroin to a limited number of users on a provisional basis. As I will explain in the concluding section, my critique was one of various causes leading to the delay of final decision making. Yet, the affair left me with some soul searching on the issue of history becoming part of politics. Is it a good or a bad idea to apply one's historical and theoretical work to politically hot issues? And is it a good or bad decision to talk to the press if one has a fair chance of being misunderstood, if not misused? Wasn't this at least the wrong moment to make any public comments? In order to get a better grip on the dilemma, the next section places it in the context of recent debates on the public role of historiography and sociology of science. It also discusses the comparable case of historian Simon Cole's criticism of the technique of criminal fingerprint identification.

COMBINING UNCERTAINTY AND RESPONSIBILITY

There was a time when historians of science self-evidently participated in the histories they were investigating. As has often been discussed in the pages of *JHBS*, until about the 1970s the common model in the history of science was that of the historical introduction to textbooks. Written by researchers themselves, these introductions typically scrutinized the past for notions more or less resembling the views presented in the volume's main chapters and accordingly praised or blamed the dead. In the past decades, professional historians of science began to reject such partisan stories. Ever since, a wide range of derogatory labels such as *presentist*, *finalist*, *justificationist*, *celebratory*, or *Whig* history expresses their aversion and challenges the naïve empiricism that inspired the orthodox historiography.

The new historians purposely avoided taking part in the discussions they were mapping. In the field of sociology of science, the twin position was that scientific debates should be analyzed in a "symmetrical" way, implying that students of science should remain methodologically ignorant as to which party is right and which one is mistaken. Yet, it did not take long before the community of historians and sociologists of science began to recognize a snag in this position. Claiming to be a genuinely detached outsider, arguing that, unlike presentist historians, we describe the past as it really was, is at odds with our rejection of empiricism. Many historians of science pointed out that it is simply impossible to look at the past with one's present-day eyes shut, and many sociologists raised the analogous objection that one cannot exempt oneself from the view that research cannot mirror reality. Since the 1980s, historians and sociologists of science therefore regard their own reconstructions as theoretically informed interpretations and arguments rather than straightforward representations of the world as it really was or is.

For a while, the latter position created a stable ground for further research and argumentation, but from the late 1990s on, members of the science studies community began to raise the issue that if history and sociology of science mostly is a matter of interpretation and argumentation, then we should offer our services to the morally *right* side. Once more, the symmetry principle or outsiders' position came under attack. Many students of science now also reject its moral "relativism" and call for accountable rather than detached studies of science. No one wants to return to empiricism, but many maintain that we should be explicitly partisan again. The theme of the 2002 annual conference of the European Association of Science and Technology Studies (EASST) reflected the new turn. Under the banner of "Responsibility under Uncertainty," the conference organizers argued in their electronic call for papers that the field should continue exposing unjustified certainty in science but should also "take responsibility for building a more accountable science and society."

I think I can claim that my main motive to talk to the press was to take responsibility. But I do not think the attempt was an unambiguous success. Striving for practical relevance while casting doubts on treasured beliefs makes the attempt extremely complicated. So I was intrigued to hear a talk at the conference by Michael Lynch and Simon Cole exploring this particular dilemma. The paper, presented by Lynch, discussed Cole's historical research on the technique of criminal fingerprint identification that resulted in the grand monograph *Suspect Identities: A History of Fingerprinting and Criminal Identification* (Cole, 2001). Like my work on human science experimentation, Cole's book argues that objectivity based on standardized procedures and techniques can never fully discard expert discretion and that in standardized research much deliberation, interpretation, and consensus formation is going on.

Cole particularly highlighted the interpretative aspects in establishing a match between the "latent" fingerprints found at a crime scene and the "rolled" fingerprints taken from suspects by the police. He also observed that fingerprint experts typically help the prosecution. For that reason, he decided to "no longer stay above the fray." When the defense attorney in a crime case appealed to him, Cole agreed to testify as an expert for the defense. Lynch and Cole's conference talk was an intriguing report on what happened next.⁵ It was not fingerprint identification that landed in the dock but Simon Cole's own expertise. The prosecution requested an "admissibility hearing" assessing the validity of expert knowledge in court cases. In the state of New York, where the case took place, admissibility hearings are so-called "Frye hearings," employing as a criterion of admissibility "general acceptance" in a field.

Quotes from the proceedings of this hearing demonstrated the awkward position in which Cole ended up. The prosecuting attorney aggressively cross-examined him about general agreement on the boundaries between science and nonscience in his field. This court hearing probably was not the right place for explaining the belief that the definition of true science is unstable and may vary over times and groups. But Cole, of course, could not belie this core principle of uncertainty in his field. Thus the principle boomeranged on him. The presiding judge sided with the prosecution's arguments that Cole's expertise was to be ignored. It was not fingerprint identification but Cole's own evidence that was declared inadmissible "junk science."⁶

The course of events, Lynch and Cole argued, sheds ironic light on the conference theme. Half jokingly, they concluded that if Cole *really* had wanted to help the defendant, he simply should have appealed to Karl Popper's philosophy of science, as Popper's falsification criterion undermines the central assumption of fingerprint expertise that no two people exist with identical fingerprints. Because Cole could not do that, his attempt to make a difference for the defense had little chance of success.

So the question is whether one can contribute to delicate issues outside academia by challenging the authority of general standards on what counts as science. Can it be at all helpful in pressing matters to communicate the actual uncertainty of shared scientific beliefs? Cole's case illustrates the dilemma that sometimes bothers me too. Was I too principled or too reckless when I decided to publicly declare the Dutch heroin RCT dubious? Of course, the answer also depends on the solidity of my arguments. The next sections therefore elucidate my reasons for contesting the transcendental logic of the ideal experiment and for challenging the widespread confidence in such experiments with heroin maintenance. As readers can reach their own verdict, the final section with my own defense will be brief.

5. See also Lynch & Cole (forthcoming).

6. The transcript of the proceedings is published at http://onin.com/fp/ny_v_hyatt_frye_hearing_10oct01.pdf.

THE ORIGINS AND IMPLICIT VALUES OF RANDOMIZED CONTROLLED TRIALS

Checking the effects of an act by looking to similar instances that were left untouched most likely *is* a matter of universal logic. Examples, of course, can be easily found in the history of science and indeed of mankind. But it was not until the 1910s that researchers began to advance the crucial idea of comparing *artificial groups* and, somewhat later, of composing the groups on the basis of chance. In the eighteenth and nineteenth century, men of letters such as David Hume, Auguste Comte, and John Stuart Mill already recommended “social experiments” for acquiring knowledge (Brown, 1997; Carrithers, 1995). Yet, their notion of a social experiment differed considerably from the present-day one. It referred to disruptions of the normal social equilibrium such as a war, a commercial catastrophe, a famine, or an act of government. In the eighteenth and nineteenth century, the expression “social experiment” merely was a metaphor borrowed from natural science for indicating that careful observation of unsettling events reveals the given laws of social order. Deliberately operating on people for the sake of knowledge making was considered immoral and impractical (Dehue, in press).

The idea of active experimental policy testing arose in the last decades of the nineteenth century. The earliest proponent I have found was the British statistician, philosopher, and economist William Stanley Jevons (1835–1882). Jevons’s main example also concerned an attempt to control drug abuse through drug maintenance. In the 1880 article “Experimental Legislation and the Drink Traffic,” he discussed the legalization of the free trade of beer. The intent of the “Beer Act,” allowing the sale of beer in shops, was to drive back the consumption of gin. Jevons, however, regarded the act as a “salient example of bad legislation.” Rather than being based on the “unanimous wisdom of Parliament,” he argued, crucial interventions should be scientifically tested before being brought into operation. Had this been done with the Beer Act, the government would have known beforehand that it would only enhance “the beastly state of drunkenness among the working classes” (Jevons, 1880, p. 183).

Jevons seems to have been the first author to argue that deliberate experiments with human beings can and should be done for the sake of administrative knowledge making. He also added the condition of comparison, arguing that it was possible to “experiment upon an object of varying conditions, provided we can find two objects which vary similarly; we then operate upon the one, and observe how it subsequently differs from the other” (Jevons, 1880, p. 185). Most importantly, his proposal formed part of his views on social change and the extension of government. Apart from being an advocate of experimental policy evaluation, Jevons also was an early exponent of the turn-of-the-century upper-middle-class movement striving for some limits on economic liberalism. In his 1882 volume *The State in Relation to Labour*, he agreed that people’s success and misfortune is largely of their own doing. But he also pleaded for some help for the poor and some state intervention into the allegedly independent laws of economy. Such appeals for combining the basic principles of economic liberalism with reluctant redistribution of income provided the context in which the notion emerged of policy experimentation as comparing randomly composed artificial groups. Three interrelated maxims of twentieth-century economic liberalism (or “welfare liberalism”) were crucial to the gradual emergence of the ideal experiment.

Individual responsibility remained the first and foremost maxim. That social success or lack thereof was basically an individual affair implied, of course, that the degree of social care should be limited. It also implied that ameliorative attempts were to be directed first and foremost at problematic individuals rather than structural social change. Helping people largely meant treating, educating, punishing, or rewarding individuals in order to turn them into self-supporting citizens. The second maxim was that of *efficiency*. The more hesitant a society is

about state charity, the larger the fear of squandering public funds and the stronger the urge to look for singular causal factors of misery and backwardness. Ameliorative actions financed with public money had to produce instant results with simple economical means. And the third maxim was that of *impersonal procedures*. The stronger the fear of abuse of social services, the less the belief in people's own stories of needs and the more pressing the call for impersonal techniques establishing the truth "behind" that story. In addition, it was not only the self-assessment of the interested recipients of help that was to be distrusted, but also that of the politicians and administrators providing help. Measurement also had to control administrators' claims of efficiency. When Jevons ridiculed the "unanimous wisdom of Parliament" in the 1880s and argued that it should make way for scientific proof, he heralded an era in which government officials increasingly had to base their authority on impersonal knowledge produced according to pre-established procedures.

From the beginning of the century, academic experts on psychological, sociological, political, and economical matters recognized the social niche created by the rise of welfare liberalism. They adapted their questions and approaches to the new demands and began to substitute their traditional organicism for modern elementaristic thinking. They too focused on social progress by improving individuals, on means to enhance the efficiency of such attempts, and impersonal procedures to evaluate their success.

It was no coincidence that these developments took place firstly and most strongly in Britain and the United States. Even today, these countries remain the prototypes of the "liberal variety" of twentieth-century welfare societies, leaving comparatively much to people's own responsibility and focusing more strongly on procedures than "social Democrat" and "corporatist" or "conservative" welfare regimes (Esping-Andersen, 1990). Following the administrative turn of their discipline, American social scientists began to devise techniques for measuring individual abilities, intelligence, motives, and attitudes, and social phenomena such as delinquency, illiteracy, poverty, and drug abuse. Soon some of them conceived the idea that these measurement instruments might as well be used for assessing any differences before and after ameliorative actions. Not surprisingly at times of strong emphasis on impersonality, they also reckoned that studying only one group would leave open too much room for wishful thinking as to the real causes of effects. The need to find comparable groups became a much debated issue.

For further historical details on this issue, see my earlier article (Dehue, 2001). Here it suffices to recount that some authoritative authors rejected the idea of group comparison, arguing that no two equal groups exist (and that it is immoral to subject people to an experimental regime for a considerable period of time). Psychologists working in the context of educational administration were the ones to introduce the decisive solution of comparing *artificially composed* groups rather than trying to find equal natural ones. In the early 1900s, educational psychologists advanced the idea of creating comparable groups. The earliest technique of matching the groups on particular features, however, was elaborate and hence expensive, and, worse, still needed personal imagination as to the aspects the groups should be equalized on. Psychologists at Columbia University in New York, renowned at the time for their rigorous turn to standardized educational research, were the first to suggest fully impersonal *random* allocation of the schoolchildren to the groups. Random assignment, they argued, would be an economical way of guaranteeing full comparability (McCall, 1923).

Experimenting with randomly composed groups perfectly epitomized the maxims of twentieth-century liberalism that interventions should be directed at turning individuals into self-supporting citizens, that instantaneous effectiveness of treatments should be demonstrated, and that this was to be done via standardized procedures. After Ronald A. Fisher's

handbook *The Design of Experiments* (1935) appeared, the authority of statistics provided extra ammunition to the advocates of randomization as a means of controlling interested administrators and researchers.

THE DISSEMINATION OF THE EXPERIMENTAL MODEL

Whereas nineteenth century *laissez-faire* liberalism prohibited experiments with human beings, twentieth-century welfare liberalism amply allowed for it. Gradually, experiments were conducted with adults as well, albeit adults with little more social power than the school-children for whom the RCT was first developed.

From the 1930s, random allocation to groups was proposed (if, for practical reasons, not always applied) in studies with soldiers, prisoners, and patients, and from the 1960s on, social experimentation truly flourished. Particularly after President Johnson's "war on poverty," large-scale experiments were conducted for studying publicity campaigns, income-maintenance programs, employment programs, social housing, marriage courses, or safe-sex campaigns. Ever since, welfare recipients, slum-dwellers, spouse-beaters, incompetent parents, wild teenagers, drug abusers, and various other people in need of advancement have been randomly assigned to experimental and control groups for studying attempts at improving their behavior and performance.

In medical research, experimentation and random allocation also developed in the context of administrative concerns. Comparing groups was first introduced to assess the claims of pharmaceutical industries, and from there, medical administrators disseminated it to clinical practice. It seems medical reformers first learned about random allocation from Fisher's statistics. Nevertheless, just like experimental psychologists and social researchers, the advocates of experimentation in medicine mainly advanced it as an extra means to control discretion and not, as Fisher himself did, mainly for reasons of sound statistics. With random allocation, no charges could be made afterward that any hopes for a favorable or unfavorable outcome steered the composition of the groups (Chalmers, 2001; Marks, 1997). Practicing physicians objected to the transfer of this administrative procedure into a clinical context. They argued that hopes often are a substantial part of the treatment itself and maintained that subjecting patients to standardized tests rather than examining them in a truly individual way would harm rather than enhance the effectiveness of diagnoses and treatments. They also raised moral objections against assigning patients to treatments on the basis of chance and keeping it a secret which treatment a patient actually gets. Yet the advocates of stringent control gradually won the day. Moral objections against RCTs never fully died down, but they were countered with "remedying" regulations like the requirement of informed consent.⁷ Appeals to seasoned clinical judgment or beneficial doctor-patient relations did not disappear either, but they acquired a suspect sanctimonious reputation. Today, protesting doctors easily call the suspicion on themselves of being unwilling to give up an outdated elitist position for the one and only scientific attitude.

In the meantime, the realm of medicine rapidly extended. Increasingly, all kinds of deviations from the norm such as unwanted behavior or "personality disorders" became defined as medical problems. Many critical commentators have argued that medicalization is the preeminent strategy of recasting social problems in terms of individual deficiencies (for an overview,

7. The requirement of "informed consent" implies informing the patients that they might be assigned to a control group. Often, however, people feel dependent on the doctors asking for their consent. In addition, informed consent counteracts blindness. Participants start guessing about the group they are in and their guesses influence their response (Fisher & Greenberg, 1993; Kaptchuck, 2001).

see Nye, 2003). It looks for problems inside the individual rather than at social systems while it simultaneously transfers power from the people to quantitative science. It forcefully communicates the obligation to make good use of any help available in a relentless effort to meet established standards of good health and independency (Nelkin & Tancredi, 1994; Stone, 1993).

From the 1960s on, the American definition of scientific human research with the RCT as its gold standard was also exported to other countries. Imperceptibly, the international human science community imported the political values of individual responsibility, efficiency, and impersonality with it. To be sure, depending on the problems at hand, these values can be appropriate and useful. Moreover, the value-ladenness of the RCT does not straightforwardly turn it into an invalid research instrument. Yet, it is important to note that its validity is not a transcendental or obvious one. First, only if we accept that the problem to be studied indeed is a problem of unconnected individuals rather than the result of extra-individual social or cultural processes does it make sense to take people out of their regular group and allocate them randomly to artificial experimental and control groups lacking natural social cohesion. Second, only if we believe that social and historical patterns indeed are insignificant is it justified to focus on a single isolated "independent variable." And third, only if the judgments of responsible politicians, the subjects themselves, their physicians, psychologists, families, and friends really are irrelevant can it be rational to rely mostly on preestablished standardized procedures.

The applicability of the RCT therefore depends on the degree to which its advocates can convincingly define the problem at hand according to the inherent values of the RCT. The initiators of the Dutch heroin maintenance experiment are experienced researchers, extensively trained, moreover, to work at the latter condition as well. They strongly advocated the supremacy of the experimental stance together with a representation of the heroin problem according to its values. I did not intend to cast doubt on their integrity. Yet, I argued that it actually took quite some squeezing to fit the heroin maintenance issue into the experimental template as well as the blindfold of research tradition to overlook the traces of the effort in the final results.

IMPOSING THE EXPERIMENTAL STANCE

In 1995, a newly established government proposed the idea of a heroin maintenance experiment. At that time there were an estimated 24,000 users in the Netherlands. About 50 percent were in methadone programs, which had little beneficial effect on about two-thirds of them. The Minister of Health proposed an experiment with these "therapy-resistant" users. Randomly composed experimental and control groups should be compared, the first one receiving heroin plus methadone and the second only methadone. Heroin stations would be established in six Dutch cities, and the participants would be closely studied during the experiment and afterward.

The Minister of Health was a member of the social-liberal party that strongly promotes scientific approaches to social problems and was in favor of the heroin maintenance experiment. The Labor Party and the Green Left Party also endorsed the proposal, but the Christian-Democrat Party (and smaller Christian parties) and the Socialist Party maintained that it was unethical to give up on allegedly hopeless cases and pleaded for detoxification programs. The powerful Dutch right-wing party of Liberal Democrats opposed the plan too. This party mainly worried about the project's high *material* costs (of, indeed, nearly 23 million Euros). In this case, more fearful than ever of misuse of public services, the latter party also foresaw true assaults on the heroin distribution stations and severe trouble in their environment.⁸

8. On an American scale, this Dutch right-wing party with its distinct "liberalism" in a moral sense most likely would count as a Democratic rather than a Republican Party.

Before she started a political career, the Minister had been a full professor of the evaluation of medical treatment. Combining the skills of a clinical researcher and a politician, she dealt with the issue in an outstandingly apt way. Step by step, she educated Parliament about the benefits and requirements of impersonal judgment in this complicated political issue. Her first major step was to convince Parliament that further decision making should await a detailed research design worked out by scientific experts. In 1997, the Central Committee on the Treatment of Heroin Addicts (CCBH) presented an elaborate scientific protocol, including a full experimental design with accompanying statistical techniques. Whereas the independent variable in heroin maintenance experiments is not the drug itself but the way users *acquire* it, the report consistently spoke of “a medical experiment.” And whereas heroin is not considered a medicine, the authors called the experimental intervention “a medical treatment.” The entire report was written in clinical language, indicating the envisioned participants as “patients” who are “therapy-resistant” if (and only if) they did not benefit from methadone, and to whom heroin would be “prescribed” rather than simply provided.⁹

The authority of the medical voice did not completely stifle opponents in Parliament. Heroin maintenance would be considered a success in case of at least 20 percent positive responders, with positive response defined as at least 20 percent progress in medical or psychosocial condition, plus no more than 10 percent deterioration in either physical or psychosocial condition and no more than 20 percent increase in drug consumption.¹⁰ On the basis of such statistical parameters, the CCBH arrived at a minimum number of 656 participants and, taking a particular dropout rate into account, the committee argued that 750 heroin recipients should be enrolled, 50 percent heroin injectors and 50 percent heroin inhalers. Members of Parliament mainly lodged protests against the much higher number of participants than the Minister had spoken of before and also to the fact that *each* participant, including the ones assigned to the control groups, would have to be given heroin for a particular period of time.

Methodological experts discussed the experimental design long before the CCBH report was written. In earlier publications, they had raised the problem of how to entice the participants who were assigned to the control group to profound examinations of their mental and physical state during a couple of months. As just methadone and a participant fee would not suffice, the only option seemed to be promising heroin to controls during a *next* period of

9. At first sight, another rationale for medical language in relation to the heroin problem is that heroin is harmful to people's health and that addiction is an illness. Both assumptions, however, are questionable. The medical definition of the heroin problem actually is a clear example of a complex social problem redefined in terms of individual deficiencies. Historians such as Michael de Ridder (2000) and David Courtwright (2001) related that heroin was originally introduced at the beginning of the twentieth century when pharmaceutical companies marketed it as a harmless cough repressant. Initially, limited and oral heroin use hardly had undesirable side effects. But heroin also was a forceful mood enhancer, and in the first decades of the century a subsection of users, usually people living in grim circumstances, began to smoke or inject it in large quantities and became addicted. In the 1930s, American regulatory authorities responded by reclassifying heroin from an over-the-counter drug to an “on-prescription drug” to be given only to already addicted people. American doctors, however, were not too happy with the task of policing heroin users and successfully lobbied for a reclassification of heroin into an entirely illegal substance. This course of events created the persona of the miserable heroin addict. The prototypical junkie is not simply the victim of his or her own internal makeup, as biological theories of addiction have it, but is just as much a product of a pharmaceutical accident, drug legislation driving users into delinquency, and criminal dealers messing with the quality of the heroin and driving up its price. There are well-to-do heroin users who have regular jobs, have their use under control, and are healthy (Hanson, Beschner, Walters, & Bovelie, 1985; Zinberg, 1984). It is not obvious that reallocating the heroin problem to the medical sector is the best solution. Most doctors in the United Kingdom who allow heroin on medical prescription refrain from getting a license for “treating” these “patients” (Stimson & Metrebian, 2003).

10. The end report states that later on “the original requirement for response of at least 20% change on any of the outcome domains was adapted to at least 40% change” (Central Committee on the Treatment of Heroin Addicts, 2002, chapter 4).

time. If we do that, the architects of the experiment also reasoned, we might as well continue heroin prescription to the experimental group during the period of reward to the controls. This procedure will enable comparison of a long-term heroin group to a short-term heroin group. That is to say, this can be done if a *third* group of heroin users acts as controls for the full period of time. And, of course, this third group will have to be motivated by the promise of heroin provision after the entire period of collaboration.

The 1997 CCBH report argued that, for the sake of a truly scientific experiment, the controls would have to be given heroin too and that it would be necessary to compare a six-month heroin group to a twelve-month heroin group. Opponents in Parliament tried to challenge the research design, but proponents argued that politicians, not being scientists themselves, were not entitled to tinker with methodological imperatives and could only accept or reject the entire package. So the discussion went on, mainly about the danger of too much enthusiasm on the users' part and of trouble in the neighborhoods where the heroin would be provided. At this point, the Minister once more proposed a deal. Explaining that scientific experiments are always preceded by a preliminary pilot study, she promised that if any serious trouble would occur during a pilot with only 50 heroin recipients in Amsterdam and Rotterdam, the entire project would be called off.

In 1998, an independent committee reported that no trouble occurred at all during the pilot. This implied that nothing was in the way of a "go" to the full project. The Minister urged Parliament to complete its decision making as soon as possible. With the danger of assaults on the heroin stations being taken away, the Liberal Democrats were reassured. By the end of 1999, the Minister gained the majority in Parliament. The full experiment began, and in February 2002, the end-report appeared with the positive results and the "excellent adherence by the participants." As about 25 percent extra progress in the experimental group was statistically significant, the researchers and the Minister self-confidently concluded that supervised prescription of heroin together with methadone is more effective than methadone alone. They strongly pleaded for legalization and extension of controlled heroin provision.

The Minister urged Parliament to vote for extended heroin maintenance programs before the forthcoming elections of May 2002. She added that, in expectation of final decision making, the existing heroin stations and their staff should not be wasted. New elections in the meantime brought an end to the Minister's term, but other proponents of maintenance continued to lobby for heroin provision on a permanent basis at a larger scale. The CCBH's successor, the Committee on the Treatment of Heroin Addiction (CIBH) pleads for 15 heroin units that will treat some 1,600 users with "heroin therapy," though not more than 1,000 simultaneously.¹¹

QUESTIONING THE GOLD STANDARD

As mentioned earlier, I did not oppose heroin maintenance but challenged the disproportionate authority of experimental human science. The Minister no doubt followed her clever strategy with the best intentions, but its result was that decision making was relegated to a large extent from politics to seemingly neutral science. From the moment her opponents agreed to await the scientific protocol by the CCBH, the experimental approach, so to say, became a trap. Step by step, science began to determine the agenda and outcome of parliamentary debates.

Stated differently, I expressed concerns about experimental science taking the power of defining moral standards and decision making out of the hands of responsible politicians and

11. The 2003 CIBH report is available in Dutch at <http://www.ccbh.nl>.

the people they represent. Since the early twentieth century, this tactic has been justified by the assumption that experimental research offers neutral ground in knotty social issues. As I argued, this assumption was questionable from the very start. Even the ideal experiment comparing randomized experimental and control groups can only be neutral to those who either overlook its inherent political assumptions or regard them as universal truths. RCTs represent the values of twentieth-century economic liberalism, which in the Netherlands of the 1990s were advocated by the Minister's social liberal party, the Liberal Democrats, and increasingly by Labor, too. Under the guise of neutrality, the experiment actually strengthened the power of the three parties in the governing coalition while diminishing that of their opponents. Maybe it is appropriate to mention here that I am a member of the Labor Party myself. Nevertheless, my studies of the heroin maintenance problem confirmed my view that too much emphasis on individual responsibility, efficiency, and impersonal procedures is detrimental to democracy and to the solution of social and scientific problems alike. Moreover, it is particularly inappropriate in the context of the heroin maintenance issue. Troubles during the Dutch heroin experiment can illustrate the point.

The pilot study in Amsterdam and Rotterdam had also taught something quite unexpected to both the proponents and the opponents of the experiment. The Amsterdam and Rotterdam heroin users did not refrain from besieging the heroin units but, quite the contrary, proved hard to enroll. The pilot was conducted with 141 participants instead of the planned 150, of whom 45 instead of 50 were assigned to the experimental group. The government heroin was chemically purer than street heroin, but newspapers reported that participants sometimes quit the experiment at an early stage, commenting that "this dope does not taste good" or even that it was "nasty rotgut."

In Parliament, worries that not enough participants could be attracted replaced the former fears that heroin fiends would storm the maintenance units. The Minister again soothed the opposition, arguing that statistical soundness would still be guaranteed with 625 instead of 656 participants. Changing the quantity of caffeine added in order to ease resorption would solve the problem of the heroin's taste. With the danger of assaults on the stations taken away, the Liberal Democrats were reassured, and the Minister gained the majority in the house.

During the main experiment, however, heroin users kept on complaining about the quality of the heroin, saying, for instance, that "the flash" of the government heroin is not as good. In addition, recruitment of participants remained quite difficult. Soon after the start of the experiment, local authorities of the participating cities collectively alerted the Minister to the need of extra money because of the slow enrollment of participants and the consequent longer duration of the project. In a 2001 interim report, the Minister mentioned the number of 600 participants, and the final reports presented results for 549 participants. After the disclosure of the long-awaited results, no one recalled the original number of 656 to 750 dictated by the rules of statistics, nor any later minimum number mentioned.

I highlighted this troublesome aspect of the project, not, however, to challenge the statistical soundness of the results but because the heroin users' reluctance should ring another bell. In 1998, after the pilot study, the experimenters ascribed the enrollment problem to the diminishing popularity of injecting heroin, but that explanation is not convincing. To begin with, it was at odds with another report by the Department of Health stating that the number of heroin injectors has been stable for years (Ministerie van Volksgezondheid, 2001). Moreover, the experimenters' expectations on willingness to participate had to be corrected downwardly *over and again* and this had to be done in the rare event of an experiment with a treatment to which the envisioned participants were addicted and even the control group was

promised the treatment. These were all good reasons to seriously wonder why the expected rush on the free heroin did not materialize.¹²

Acknowledging the recruitment problem, I argued, offers a view on the hidden, and in this case unwarranted, assumption of experimentation that the problem at hand is a purely individual affair. Let wine lovers ask themselves whether they would fancy glasses of the finest quality vintage “on medical prescription” in exchange for participation in an experiment. In order to acquire their regular quantity, they would have to attend a wine station three times a day and seven days a week. They are not allowed to linger near the station and a camera at the front door is recording them. Behind the door, they pass a metal detector and a porter behind a glass screen. They collect their drink through a box office, and while they consume it, nurses behind a screen constantly keep an eye on them in order to keep them from smuggling some out. If they drink slowly or perhaps close their eyes while enjoying the drink, they are urged to speed up. Once a week they can be randomly selected for a urine test that controls for additional use of drugs, women are tested for pregnancy every month, and everyone undergoes a physical examination every other month requiring them, for instance, to open their mouths for a check of their teeth. They also have to answer all kinds of questions inspired by the standards of teetotalers. They are asked about their contacts outside the wine fiend community, the way they acquired their income, and the places where they urinated.

Even if wine would be far beyond their financial means, not many wine buffs would be eager to take part, and those who did most likely would hardly enjoy their drinks. The point, of course, is that one consumes heroin, like wine, in the right situation with the right people, the right equipment, and with all kinds of other rituals. That is to say, the expectation that the sheer offer of heroin is attractive enough to enroll many eager users, and particularly the fear that too many users would show up, is based on a model of addiction as something purely physical. Many drug experts reject this individualized view of heroin addiction. They do not deny the physical effects of heroin but argue that users learn to interpret these effects in the setting where they started using. Using drugs also is a way of life. It is part of a historically constituted subculture, or rather subcultures differing among themselves, which are at least as important as the drug itself. Participation in the experiment demanded that users leave their normal life and adapt to a stern regimen established by regular culture.

Taking the cultural aspect of heroin use into account explains why heroin users proved hard to enroll. But it also brings serious problems of the experimental approach to light that would have existed even if there had been a rush on the free heroin. First, it indicates that those who did participate most likely are not representative of the intended population of the most miserable users. It is doubtful that users living on the outskirts of the cities would have been attracted by the maintenance experiment, as these people are beyond reach even of regular social outreach, which, literally or metaphorically, does not speak their language. A spokesperson for the Rotterdam Junkie League commented that this experiment is for “yuppie junkies” only, indicating that successful participation supposed good understanding of regular culture and the ability to follow the demanding script of the experiment (N. Storm, personal communication, August 23, 2000).

Second, it is well known that there are subcultures of heroin use. Dutch-Surinam, Dutch-Turkish, Dutch-Chinese, and ethnic Dutch users all differ in some respects and differ among one another. Experiments cannot possibly study the varying effects of maintenance in *natural* groups, because the rules of experimentation prescribe that artificial groups are created in

12. In a Swiss heroin trial, it also was hard to attract participants and there also were complaints about the quality of the heroin (Bourgeois, 2002).

which systematic differences are canceled out. Also, experiments must employ standardized diagnostic instruments, applying the same criteria to each participant. Diagnosis of progress and decline cannot occur in ways tailored to cultural differences, let alone to personal ones.

Third, the experiment produced results of very limited value for yet another reason. It does not seem far-fetched to look upon this test as a trial in the sense of a hurdle race in which much was at stake. Participants in the experimental group had to make sure to pass the bi-monthly tests. Conversely, participants in the control group had to take care not to progress too much. This hurdle-like aspect is a substantial problem in any social and medical experiment testing a treatment wanted by the participants. If possible, the problem is countered by keeping the participants ignorant about the group they are in. In heroin experiments, however (as in most policy experiments and many medical trials), it is impossible to make the experimental treatment indistinguishable from the control treatment. And even if it is possible, participants' responses will be steered by their (often correct) guesses on the condition they are in (Fisher & Greenberg, 1993; Kaptchuck, 2001).

In the Dutch heroin experiments, this problem was magnified to the extreme. A person could at one time be a participant in the control group and at the other in the experimental group, which meant that he or she at one time had to react negatively, at the other time positively. Even worse, not only did the participants not know whether they were receiving heroin, and have high hopes of positive results, but also, for ethical reasons, they had to be told that in case of considerable *regress* while receiving heroin, they would be excluded from the project. And for the same reasons, they had to be told that in case of considerable *progress* in the control condition, they would lose the prospect of heroin maintenance at a later stage. They also knew that the combination of progress during maintenance and quick relapse soon thereafter would give them a fair chance of continued maintenance after the experiment (CCBH, 1997). It is therefore puzzling that, according to the experimenters, "no sanctions" were "connected to the content of the answers" (Van den Brink et al., 2003a, p. 312).

The research team also praised the excellent adherence by the participants. It should not come as a surprise, though, that those users who did participate did their utmost best to run the full race. Opponents of heroin maintenance interpreted this aspect of my critique as an allegation that the results were manipulated in a positive sense. Most likely, however, the hurdle aspect of the experiment *suppressed* the results. Think of yourself participating in an RCT with something you definitely need to feel well (say food or clothes) but can only get illegally and with great effort. People with a fairly different lifestyle and in many respects different criteria of a decent life (heroin users?) composed the tests for establishing the quality of your response. They want you to show progress at times when they give you what you need and they want you not to progress at times when you have to get it yourself. Representatives of junkie organizations reported that the participants were extremely nervous about their bimonthly examinations. To the participants, the experiment indeed was a true trial. More so than the effects of heroin provision, the level seems to have been measured to which a selection of "yuppie junkies" managed to deal with the extraordinary situation.

There is much reason to expect large benefits from taking users out of the atmosphere of criminality and ruling out criminal dealers who drive up the price and adulterate the quality of heroin. My original article ended with the suggestion that if, apart from sound political thinking informed by seasoned expert judgment, further empirical research of heroin provision is needed, then high-quality ethnographic research seems more adequate than experiments. The results would have been more informative if progress had been studied in less biased ways, if the groups were natural rather than artificial ones, if the views and interpretations of the participants were taken into account, if criteria of progress had been adapted to the subcultures of heroin use, and

if the conclusions were embedded in well-informed historical and cultural studies of the problem. To be sure, this kind of research is not airtight, but neither is experimental research that purports to be so. Ethnography is meant to offer well-informed interpretations and to instigate debate.

STAYING OUT OR MINGLING IN?

Can it be helpful in sensitive issues to challenge hard-won agreements on the scientific approach? The Minister's 2002 rebuttal pointed out that this is not the right context to argue for "alternative definitions of science." There might be some truth in that. My position was comparable to that of Simon Cole, who in the fingerprint court case may have picked the wrong place and time to question mainstream views. Wanting to help the defendant, Cole could have done so by pragmatically declaring Popper's falsificationism as the standard of true science. Similarly, I, not wanting to be an obstacle to the introduction of heroin maintenance, could have withheld my doubts or postponed my critique. Nevertheless, I still think I made the right decision to mingle in at the moment I did.

Another comparison may help to make my case. In a number of respects, my comments paralleled those of the sociologist Warwick Anderson, who studied an RCT with syringe maintenance to heroin users (Anderson, 1991). Anderson discussed this trial, conducted in New York in the 1980s, as an illusory attempt to relegate a politically hot issue to seemingly neutral science. Quite ironically in the present context, he added a compliment to countries such as the Netherlands where politicians still take their own responsibility and do not need seemingly neutral science. Anderson was spared a dilemma, because the New York syringe trial ended with negative conclusions. In a sense, however, the Dutch heroin maintenance trial ended negatively too. The proponents of maintenance may maintain that the results were positive, but their opponents were not impressed.

After an initial attempt to challenge the outcomes, the Christian Democrats changed their tack. They now used the results as an argument *against* heroin maintenance. Whereas the CCBH emphasized that, according to regular standards of *medical* effectiveness the success of heroin maintenance was quite high, the Christian Democrats maintained that the results nevertheless were too poor to legalize and extend the heroin maintenance program. Furthermore, they repeated their former moral objections.

In addition, the May 2002 elections completely turned the political tables. In spring 2002, the extreme right-wing politician Pim Fortuyn, founder of the political party "The list Pim Fortuyn," quickly gained massive popularity. Ten days before the elections, Fortuyn was killed by an environmental activist, and the shocking murder further fueled the general discontentment already stirred up by Fortuyn. Even many of his political adversaries agreed that it was time for radical transformations . . . and voted for the Christian Democrats. When the issue of heroin maintenance was on the agenda again, the main Christian party had fully recovered from longtime decline.¹³

13. Spring 2002 was a turbulent period for yet another reason. In the first week of April, the historian/professor Hans Blom and his team presented a long-awaited report on the devastating events in Srebrenica in 1995 where Slobodan Milosevic's general, Ratko Mladic, overruled Dutch militaries and brutally murdered 7,500 Muslim men. The issue became an even more harrowing example of historiography becoming part of politics, as the government drew the conclusion that it should immediately resign without awaiting the forthcoming elections. So, one month before the elections, the government stepped down. This happened much to Blom's dismay, as he commented that the government resigned before anyone could have read the long Srebrenica report and that his words were interpreted wrongly. Other leading historians, however, repeated what they said before, namely that Blom and his team should never have accepted the order of doing this study because it would unavoidably force him into the role of a politician, if not a judge (Blom, 2002).

To the proponents of maintenance there is no way back. Having endorsed the heroin experiment for many years, they cannot rejoin the now powerful Christian Democrats by arguing that the outcomes should not be taken that seriously.¹⁴ I, however, can. I agree with the opponents of maintenance that, practically spoken, the outcome of maximal 25 percent extra progress is not that impressive. The CCBH even came to percentages of 45 to 78 participants who “did not respond to the co prescribed heroin” (Van den Brink et al., 2003b). Yet, as I have argued, there actually is very little reason to make much of these figures.

Huge amounts of money and years of time and energy—of politicians, researchers, and heroin users alike—have been spent on an experiment that hardly provides reliable information. With a fraction of the material and immaterial costs, members of Parliament could have been profoundly informed by national and foreign experts on the historical, legal, moral, social, and psychological aspects of heroin maintenance, as well as by studies of the cultures of heroin use. Also with a fraction of the costs, political and moral debates would have been possible that were largely cut short by years of focusing on experimental details. If truly political discussions on maintenance had taken place, the opponents of maintenance would have had a chance to elucidate their ethical views and the proponents would have had a chance to give ethical counterarguments.

Of course, at the time I decided to talk to the press, I could not foresee this governmental change of power. Yet, in democratic societies, fundamental political change may always take place. And even if no revolt would have occurred, I still think it was better to speak up than remain silent. The best defense of my coming out is that in the entire chorus of voices it was my part to question assumptions of scientific research. After all, there are many experts on drug issues but not that many whose job it is to track down hidden assumptions of scientific beliefs. Cases such as the Dutch heroin experiment offer a rare opportunity to draw public attention to arguments that normally get a hearing only in small communities of historians of science. For that reason, I think it is important to strike while the iron is hot. After all, more harm has been done in history by unjustified certainty than by justified doubt.

ACKNOWLEDGMENTS

This article is based on a keynote lecture to the thirty-fifth annual meeting of Cheiron. The author thanks Ben Harris and Homer Stavelly for their kind invitation and the discussants at the conference for their helpful comments. I also thank Alf Berendse, Loes Linssen, and Freek Polak for their comments and for illuminating me on the intricate issue of heroin maintenance. Finally, I thank Michael Lynch and Simon Cole for mailing the EASST conference paper on the case of criminal fingerprint identification.

REFERENCES

- Abraham, J. (1995). *Science, politics and the pharmaceutical industry: Controversy and bias in drug regulation*. London: UCL Press.
- Anderson, W. (1991). The New York needle trial: The politics of public health in the age of AIDS. *American Journal of Public Health*, 81, 1506–1517.
- Blom, H. (2002). Politiek versus historie. *Nederlands Instituut voor Oorlogsdocumentatie maakt onderscheid tussen moraal en wetenschap*. *Academische Boekengids*, 3336, 4–7.
- Bourgeois, P. (2002). Anthropology and epidemiology on drugs: The challenges of cross-methodological and theoretical dialogue. *International Journal of Drug Policy*, 13, 259–269.
- Brown, R. (1997). The delayed birth of social experiments. *History of the Human Sciences*, 10, 1–23.

14. Representatives of heroin user organizations with whom I had long e-mail correspondences largely agreed with my critique on the experiment. Understandably, however, in the political debate on heroin maintenance, they too put their bets on the experiment and underlined that the outcomes were positive according to previously established criteria.

- Carrithers, D. (1995). The enlightenment science of society. In C. Fox, R. Porter, & R. Wokler (Eds.), *Inventing human science. Eighteenth-century domains* (pp. 232–270). Berkeley, CA: University of California Press.
- Central Committee on the Treatment of Heroin Addicts (CCBH). (1997). *Onderzoek naar Heroïne op Medisch Voorschrift. I Het Onderzoek. D. Deelnemersinformatie*. Unpublished report, Utrecht, the Netherlands.
- Central Committee on the Treatment of Heroin Addicts (CCBH). (2002). *Medical co-prescription of heroin. Two randomized trials*. Retrieved March 31, 2004, from <http://www.ccbh.nl>
- Chalmers, I. (2001). Comparing like with like: Some historical milestones in the evolution of methods to create unbiased comparison of groups in therapeutic experiments. *International Journal of Epidemiology*, 30, 1156–1164.
- Cole, S. (2001). *Suspect identities: A history of fingerprinting and criminal identification*. Cambridge, MA: Harvard University Press.
- Courtwright, D. (2001). *Dark paradise: A history of opiate addiction in America*. Cambridge, MA: Harvard University Press.
- Danziger, K. (1990). *Constructing the subject*. New York: Cambridge University Press.
- Danziger, K. (2000). Making social psychology experimental: A conceptual history, 1920–1970. *Journal of the History of the Behavioral Sciences*, 26, 329–349.
- Dehue, T. (2001). Establishing the experimenting society. *American Journal of Psychology*, 114, 283–302.
- Dehue, T. (2002a). A Dutch treat. Randomized controlled experimentation and the case of heroin maintenance in the Netherlands. *History of the Human Sciences*, 15, 75–98.
- Dehue, T. (2002b). Over de assumpties van RCT's, geïllustreerd aan het Nederlandse experiment met verstrekking van heroïne. *Maandblad voor Geestelijke Volksgezondheid*, 3, 230–249.
- Dehue, T. (2002c). Antwoord op de brief van Minister Borst. *Maandblad voor Geestelijke Volksgezondheid*, 9, 806–811.
- Dehue, T. (in press). The history of social experiments. From disturbing events to social science research designs. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement*. New York: Academic Press.
- de Ridder, M. (2000). *Heroin. Vom Arzneimittel zur Droge*. Frankfurt am Main, Campus.
- Epstein, S. (1995). The construction of lay expertise: AIDS, activism and the forging of credibility in the reform of clinical trials. *Science, Technology & Human Values*, 20, 408–437.
- Epstein, S. (1996). *Impure science: AIDS, activism and the politics of knowledge*. San Diego, CA: University of California Press.
- Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Cambridge: Polity Press.
- Fisher, R. (1935). *The design of experiments*. Edinburgh: Oliver & Boyd.
- Fisher, S., & Greenberg, R. P. (Eds.). (1993). *The curse of the placebo: Fanciful pursuit of a pure biological therapy*. New York: Wiley.
- Hanson, B., Beschner, G., Walters, J. M., & Bovelle, E. (1985). *Life with heroin: Voices from the inner city*. Lexington, MA: Lexington Books.
- Healy, D. (2001). The dilemmas posed by new and fashionable treatments. *Advances in Psychiatric Treatment*, 7, 322–327.
- Jevons, W. S. (1880). Experimental legislation and the drink traffic. *Contemporary Review*, 7, 177–192.
- Jevons, W. S. (1882). *The state in relation to labour*. London: MacMillan.
- Jones, D. S. (2000). Visions of a cure: Visualisation, clinical trials, and controversies in cardiac therapeutics, 1968–1998. *Isis*, 91, 504–541.
- Kaptchuck, T. J. (1998). Intentional ignorance: A history of blind assessment and placebo controls in medicine. *Bulletin of the History of Medicine*, 72, 389–433.
- Kaptchuck, T. J. (2001). The double-blind randomized, placebo-controlled trial: Gold standard or golden calf? *Journal of Clinical Epidemiology*, 54, 541–549.
- Lynch, M., & Cole, S. (forthcoming). STS on trial. Dilemmas of expertise.
- Marks, H. M. (1988). Notes from the underground: The social organization of therapeutic research. In R. C. Maulitz & D. E. Long (Eds.), *Grand rounds. One hundred years of internal medicine* (pp. 297–336). Philadelphia: University of Pennsylvania Press.
- Marks, H. M. (1997). *The progress of experiment: Science and therapeutic reform in the United States, 1900–1990*. New York: Cambridge University Press.
- McCall, W. A. (1923). *How to experiment in education*. New York: Macmillan.
- Medawar, C. (2003). The antidepressant web: Marketing depression and making medicines work. Retrieved March 31, 2004, from www.socialaudit.org.uk.
- Ministerie van Volksgezondheid. (2001). *Voortgangsrapportage drugbeleid, September 1999–September 2001*.
- Nelkin, D., & Tancredi, L. (1994). *Dangerous diagnostics: The social power of biological information*. Chicago, IL: The University of Chicago Press.
- Nye, R. A. (2003). The evolutions of the concept of medicalization in the late twentieth century. *Journal of the History of the Behavioral Sciences*, 39, 115–131.
- Richards, E. (1988). The politics of therapeutic evaluation: Vitamin C and cancer. *Social Studies of Science*, 18, 653–701.
- Richards, E. (1991). *Vitamin C and cancer: Medicine or politics?* London: Macmillan.
- Stimson, G. V., & Metrebian, N. (2003). *Prescribing heroin: What is the evidence?* York: Joseph Rowntree Foundation.

- Stone, D. A. (1993). Clinical authority in the construction of citizenship. In H. M. Ingram & S. R. Smith (Eds.), *Public policy for democracy* (pp. 45–68). Washington, DC: The Brookings Institution.
- Van den Brink, W., Hendriks, V. M., Blanken, P., Koeter, M. W. J., Van Zwieten, B. J., & Van Ree, J. M. (2003a). Medical prescription of heroin to treatment resistant heroin addicts: Two randomised controlled trials. *British Medical Journal*, 327, 310–312.
- Van den Brink, W., Hendriks, V. M., Blanken, P., Koeter, M. W. J., Van Zwieten, B. J., & Van Ree, J. M. (2003b). Correction for Van den Brink et al, *BMJ*, 327, 310–312. *British Medical Journal*, 327, 724.
- Zinberg, N. E. (1984). *Drug, set and setting: The basis for controlled intoxicant use*. New Haven, CT: Yale University Press.