ON THE IMPARTIALITY OF EARLY BRITISH CLINICAL TRIALS

David Teira

Dpto. de Lógica, Historia y Filosofía de la ciencia
UNED
Paseo de senda del rey 7
28040 Madrid
Spain

Phone: (34) 666 379 181
E-mail: dteira@fsof.uned.es

ABSTRACT

Did the impartiality of clinical trials play any role in their acceptance as regulatory standards for the safety and efficacy of drugs? According to the standard account of early British trials in the 1930s and 1940s, their impartiality was just rhetorical: the public demanded fair tests and statistical devices such as randomization created an appearance of neutrality. In fact, the design of the experiment was difficult to understand and the British authorities took advantage of it to promote their own particular interests. I claim that this account is based on a poorly defined concept of experimental fairness (derived from T. Porter's ideas). I present an alternative approach in which a test would be impartial if it incorporates warrants of non-manipulability. With this concept, I reconstruct the history of British trials showing that they were indeed fair and this fairness played a role in their acceptance as regulatory yardsticks.

KEYWORDS: clinical trials, Streptomycin, fair test, impartiality

1. THE CURRENT CONSENSUS ON BRITISH CLINICAL TRIALS

In this paper I want to re-examine the process that led to the acceptance of randomized clinical trials (RCTs) as a regulatory yardstick in Britain in the first half of the 20th century. RCTs are an experimental design aimed at assessing the comparative effects of, at least, two medical treatments on two (or more) groups of patients. The statistical analysis of the difference between the outcomes in the groups should reveal if there is any significant different showing a superior safety and efficacy of one of the treatments regarding the other(s). Throughout the second half of the 20th century and up to our days, RCTs became the testing standard for pharmaceutical agencies, so that succeeding in at least two of them is a pre-requisite for authorizing the commercial use of new compounds in Western countries.

In this sense, RCTs illustrate, once more, how statistically based decision criteria have been successfully implemented in public policy. We owe to Ted Porter's *Trust in numbers* (1995) the current paradigm for the analysis of such processes. His main thesis was summarised by Hagendijk (1999, p. 630) as follows:

> The tendency towards quantitative and standardized measures in modern society is not so much a spin-off from the development of the natural sciences, as a part of the response of weak professional groups to external social and political pressures demanding accountability.

According to Porter, in Western democracies at the turn of the 20th century, it wasn't possible for public officials to accept individual expert judgement on the sole basis of personal trust in the members of a scientific elite. The experts gradually became professional bodies whose advice was delivered complying with impersonal rules and calculations in order to exclude bias and personal preferences. The more social pressure for unified and accountable expert advice, the more likely it was that quantification succeeded among social scientists. According to Porter, this social pressure was expressed in normative terms: scientific objectivity answered to "a moral demand for impartiality and fairness" (1995, p. 8) which arose in particular social settings.

The current account of the origins of British RCTs is mostly framed in terms of Porter's paradigm. There was a public demand for impartiality in drug testing that gave the opportunity to an emerging group of trialists to impose RCTs against alternative forms of assessment (mainly, expert clinical judgment). *The thorny issue, in my view, is whether the purported impartiality of RCTs was real or just a mere rhetorical device to make their introduction acceptable*. The standard account sides mostly with this latter interpretation: e.g., the analysis of the 1948 streptomycin trial in Britain shows that both the Ministry of Health and the medical community agreed to conduct it without any substantive understanding of its statistical rationale. The trial allowed the British health authorities to manage a scarce supply of streptomycin, a drug in high demand among tuberculose patients. Private contextual interests, rather than epistemic or ethical concerns, drove the organization and conduct of the first RCT[1].

Against this consensual account, I will claim that the impartiality of RCTs (their independence regarding the interests at stake in the experiment) played an effective role in the social acceptance of controlled clinical trials, first, and then RCTs in Britain. I think Porter is right in signalling the role impartiality plays in the processes leading to the public acceptance of statistical decision-making criteria. But we need a more precise concept of impartiality in order to decide whether these criteria were actually fair. And only at that point we may assess whether this impartiality contributed to the acceptance of this criteria. In the following section, I intend to sketch my own proposal in this respect. I contend that an experiment (such as an RCT) can be deemed impartial to the extent that it incorporates debiasing procedures that block self-interested manipulation of the outcome. I will introduce a number of distinctions that will allow me to analyse two episodes in the introduction of British RCTs.

In section 3 I will discuss the antecedents of the Streptomycin trial (usually considered the first RCT worth its name), between 1920 and 1945. I am going to consider, on the one hand, the different interests at stake in the trials in order to make sense of the need of an impartial decision making process that could lead the concerned parties to an agreement. On the other hand, I will show how this necessity for

---

[1] This account is presented in a handful of excellent dissertations completed throughout the last fifteen years: namely Cox 1997, Edwards 2007, Toth 1998a, Yoshioka 1998a. This social approach has also been used in the American case: see Meldrum 1994. The reader should bear in mind that Yoshioka (1998b) has retreated somewhat from his thesis's heavy emphasis on administrative objectives, so my arguments will refer exclusively to his 1998a piece.

impartiality was understood by the trialists, without a clear consensus on its implications. In section 4, there is a discussion of the standard account of the Streptomycin trial. I will argue that randomization provided the first real warrant of the impartiality of the experiment: rather than a mere rhetorical device to comfort the British public, I contend that it was a fair enough allocation procedure for all practical purposes and, despite the self-interested motives that led to its selection, the choice of randomization deserves all praise from the standpoint of the concept of impartiality presented in section 2. Moreover, this concept does not provide an *ex post* reconstruction of the process, but rather allows us to capture (and vindicate) the actual motivations of the agents involved.

2. THE CONCEPT OF IMPARTIALITY

I am going first to sketch a concept of experimental impartiality with the sole purpose of re-appraising the current historiography of British clinical trials. This concept should stand a defence on its own, but I just want the reader to accept it for the sake of the subsequent analysis. My goal is to show that a more precise understanding of the concept allows a more optimist epistemic analysis of the processes categorized by Porter, illustrating it with the case of British trials. Suffice it to say that in my approach I am drawing on two sources: the epistemology of experimental error, of which subjective biases are just an instance; and the tradition of fair medical tests extensively documented in the James Lind Library (and articulated most prominently by Iain Chalmers)[2].

In this approach, an experiment will be impartial if it incorporates methodological devices preventing the experimenter from manipulating the results at will. A case in point that we will use as an illustration throughout this paper is randomization in the allocation of treatments. In comparative clinical trials, a random allocation of therapies prevents the experimenter to assign treatments to patients in a way that may bias the experimental outcome. If, for instance, a researcher wants to increase the chances of an experimental drug to succeed in a trial, with a view to earning a bonus if it is ever approved for commercial use, allocating the therapy to the

---

[2] The James Lind Library is an internet site "created to help people understand fair tests of treatments in health care": see http://www.jameslindlibrary.org/ (Accessed October 23th, 2012)

healthier group of patients in the trial will do it. As a debiasing method, randomization protects the experimental outcome from manipulation in the allocation of treatments.

Impartiality comes in degrees then. The more complete it is our catalogue of potential biases and debiasing methods, the more impartial our experimental procedure will be. But there is no such thing as full impartiality, because with each change in the design or implementation of the experiment a new potential source of biases emerge. Besides, notice that debiasing methods only warrant the impartiality of the experimental procedure, not the impartiality of the outcome itself. On the one hand, even if the experiment is conducted *without any manipulation whatsoever*, the results may benefit the interests of one of the concerned parties: the experimental drug may work in a trial to the great joy of the manufacturer. I contend that we should still consider the experiment impartial, since the benefits derived from the outcome did not condition in any way the experimental procedure. In other words, the external (commercial) interest driving the experiment are perfectly compatible with the internal (epistemic) interests underlying the experimental procedure[3].

On the other hand, our warrants may fail, even without intentional manipulation: what if, for instance, the younger patients get the experimental treatment by sheer chance? This is just another instance of the so-called *experimenter's regress*. Biases are just another kind of experimental error and there is no way to check whether our debiasing procedures actually work without implementing checking procedures that may themselves work or not. As I have argued elsewhere (Teira forthcoming), this is why the impartiality of the procedure counts: if we agree on a list of possible biases and the experiment implements methods to block them, this provides a good enough justification to accept the result as impartial. Were we to obtain a biased outcome nonetheless, it would not be the product of intentional manipulation.

We should distinguish then between the procedural fairness of an experiment and the self-proclaimed impartiality of scientific experts. Whereas the former is warranted by the debiasing methods implemented in the experiment, the impartiality of expert judgment as such is far more complicated to establish. Mainly, because we still

---

[3] For an extended discussion of how these internal and external approaches might be reconciled from a contractarian perspective, from which I am implicitly arguing, see Zamora 2002, 2006 for a quick presentation.

lack a good (and complete) theory of cognitive biases that would allow us to correct them as we do in an experimental setting.

Although Porter's approach remains agnostic as to the actual impartiality of either the experts or the procedures under historical analysis and focuses mostly on how their purported impartiality is used rhetorically in order to promote particular interests. However, from a philosophical standpoint at least, I think it makes a difference whether there is something more to impartiality than mere persuasion: as we are going to discuss in the following section, if the British physicians testing drugs were acting in the name of impartiality but their experiments were not in fact impartial, it seems as if the British taxpayers who funded those tests and expected their decision would have been cheated. Perhaps the historian does not want to judge the case on epistemic or moral grounds, but the philosopher should at least make clear in what sense the experiments could be impartial for the lay reader to draw her conclusions.

Finally, a word about our motivation to adopt impartial tests: why would anyone accept an experimental procedure that constrains our ability to manipulate it? We seem to have a taste for fair procedures, as documented in an extensive psychological literature about distribution processes. But we can also construct principled arguments to explain it in contractarian fashion –see Teira forthcoming. As we will see in the following two sections, every regulatory drug test should sort out a potential conflict of interests, since the concerned parties expect different things from the trial. The manufacturer wants quick access to the market, the patients want safe and effective compounds, the prescribing physicians do not want to be sued for it, etc. And they all should agree on the result of the trial instead of contesting it if it does not accord with their own preferences. As I have argued elsewhere, the impartiality of the test regarding their interests is a pre-requisite for them to accept the results. This is particularly relevant from a historical perspective, since it may explain why there was such a high demand for impartiality in drug testing and why RCTs were adopted as the regulatory standard. With this concept in sight, let me then proceed to the re-analysis of the process in which British RCTs emerged. Against the standard account, I hope to show that there was something more to their impartiality than mere rhetoric.

3. EARLY BRITISH TRIALS

Between 1900 and 1950 expert clinical judgment was the main approach in the assessment of the properties of pharmaceutical compounds in Britain. An experienced clinician would administer the drug to a series of patients he considered more apt to benefit from it. His conclusions would be presented as a case report, with the details of each patient's reaction to the treatment. The alternatives were first laboratory experiments and then controlled clinical trials (from which RCTs would later emerge). The former would proceed either *in vitro* or *in vivo* (on animals and patients): considered superior by clinicians with a scientific background, its scope was usually restricted to safety considerations. It soon gave way to comparative trials, in which two treatments were alternated on the same patient or administered in two groups of patients (simultaneously or not). The arrangements to secure the comparability of the two treatments were the *controls* and they adopted different forms. The following items counted as controls in these trials: the patients' eligibility criteria, the way treatments were allocated (alternation and randomization), uniformity in administration of treatments and patients' blinding. They were not used necessarily all at once. Statistical reports from controlled trials conveyed their results with different degrees of sophistication. Significance testing features only occasionally before 1950.

The assessment of medical therapies became an issue of general concern with the creation of the first public insurance schemes in the 1910s. With the implementation of the National Health Insurance Act, passed in Parliament in 1911, the health budget started to grow significantly. The British administration became a massive purchaser of drugs, increasing yearly its pharmaceutical expenses. In 1919 the Ministry of Health was created with the mission, among others, of controlling these expenses. A better management of this budget required discerning the most efficient drug on sale in the market.

During its first decade, the Ministry resorted to various forms of clinical expertise to decide about the treatments to subsidize (for anti-tuberculosis treatments, for instance). Gradually, the Medical Research Council began to establish itself as the leading public authority on drugs[4]. Originally created to conduct research on tuberculosis with funds provided by the NHI Act, its scope became wider after the First

---

[4] The standard sources on the MRC are the two volumes of Landsborough Thomson 1973 and Austoker & Bryder 1989

World War. The MRC contributed then to prevent the distribution of defective salvarsan, a treatment for syphilis, among British troops, providing laboratory tests to check its purity[5]. The reputation thereby earned made the Ministry defer to the MRC the standardisation of "biologicals" of pharmaceutical use. This would be regulated in 1925 with the Therapeutic Substances Act.

Before 1925, the British pharmaceutical market was almost unregulated and, with just a small duty payment, everybody could sell as pharmaceutical products compounds of undisclosed nature, the so-called *secret remedies*. In case of nil or adverse effects it was difficult to adjudicate in court the responsibility of the producer without some sort of assessment of the composition and properties of the remedy. This led to all sorts of clinical tests, usually controversial in their results[6]. Among these controversies, the MRC gradually became the sole arbiter in drug assessment in Britain. According to Desirée Cox (1997), this was due, in a great deal, to the vision and determination of Walter Fletcher, its first secretary.

Fletcher's moral convictions about the impartial role of the civil servant informed the organization of the MRC during this period[7]. Fletcher believed that the MRC should contribute to the British common good through the disinterested investigations of its "noble scientists". Unlike ordinary physicians, they should have a scientific education that allowed them to provide an objective assessment of the nature and efficacy of medical treatments. For the MRC scientists it was a moral duty to be objective: they were safeguarding the justice of the transactions in the British pharmaceutical market. This is, at least, what Fletcher wrote to the Lord President of the Privy Council, Arthur Balfour, in defence of the Therapeutic Substances Act: if the MRC verified that each compound was standardised in accordance to its purported therapeutic properties, the consumers (be they patients or doctors) could be certain they

---

[5] On the salvarsan study see Toth 1998, pp. 125-138. The indications provided by the MRC regarding the administration of the drug were based on case reports: see MRC 1920, pp. 27-33

[6] For an illustration of these early clinical trials see Cox 1997, pp. 26-93. For the legal framework, see Abraham 1995, pp. 53-56. The Select Committee of the House of Commons had relied namely on laboratory tests in order to identify the composition of secret remedies for their 1914 report (Select Committee on Patent Medicines 1914, pp. 11-12).

[7] The following account is based on Cox 1997, pp. 94-188.

would get what they paid for. Standardisation would thus improve the safety of the consumers and the international sales prospects of the British pharmaceutical industry[8].

In order to guarantee the objectivity of the assessment, Fletcher organized the MRC trials in such a way as to prevent any conflict of either intellectual or economic interests. In exchange for the material support they would receive for each trial, these *noble scientists* should commit themselves to the research agenda set by the MRC and disclose whatever results they reached. In case there was a dispute about the priority in any finding, Fletcher would mediate. Making financial profits from trials was discouraged and Fletcher saw to it, excluding very qualified clinicians from the trials due to their potential stakes in their result. However, the MRC ideology of impartiality was not implemented in any methodological principles aimed at excluding personal biases. According to Cox, physicians "used these trials in the name of biological standardisation as opportunities to express their clinical prowess and demonstrate their individuality" (1997, p. 175).

Clinical trials became more important in the 1930s when the Association of British Chemical Manufacturers (ABCM) approached the MRC demanding an external assessment of their products. At this point their sales were decreasing due to the effects of the Great Depression and rogue competition from small distributors who cut prices producing low quality drugs (Abraham 1995, pp. 53-54). According to the meeting's record, the ABCM argued that British pharmaceutical companies could not get their remedies tested and approved by the medical community as was already happening in Germany: "doctors were afraid of publishing clinical trials of new drugs in this country lest they should be suspected of having a pecuniary interest in doing so" (Cox 1997, p. 197). This request certainly suited the ideology of fairness that presided over the MRC: the Therapeutic Trials Committee (TTC) was formed in 1931 and a public announcement was made.

The members of this committee were again "noble scientists", who perceived themselves as "the disinterested juror, managing the interest of the pharmaceutical industry, the public, and the profession" (Cox 1997, p. 199). The TTC established a set of rules aimed at giving equal opportunities to all manufacturers who wished to test their products, provided their applications complied with certain conditions (e.g.,

---

[8] For an international perspective on the British pharmaceutical industry, see Liebenau 1990.

disclosure of the composition). Once accepted, the TTC would appoint external experts to conduct a trial. The manufacturer could not contact the expert during the test and should not arrange a separate trial until the TTC reached a conclusion. The TTC would decide on the publication of the results and only in this case could the manufacturer make public use of the trial (without naming the expert who conducted it).

We may now wonder in what sense the trials supported this impartial judgment. Guidelines were established at a 1931 conference in which the MRC attempted to reorganize the multi-centre trial of the antipneumococcal serum against pneumonia (Edwards 2007, pp. 106-112). The approved "Standard scheme of inquiry" for this trial specified that patients should be allocated by alternation to a control group. According to Cox (1997, p. 206), this was the "uncontested position" of the TTC. Nonetheless, as Martin Edwards (2007, pp. 22-23) observes, the TTC never imposed a unified methodology for its trials. They were often case reports, and only a quarter of them used some sort of comparative control. Significance testing was used only in one of these studies (Snodgrass et al. 1937).

In other words, expert clinical judgment was still the main criterion applied in the TTC trials. According to Toth, comparative trials were only advocated by the TTC "when a British drug was being compared to a foreign produced variant" (1998, p. 156). Martin Edwards (2007, p. 112) claims that the scheme was adopted in order to control the very idiosyncratic practices of the clinicians involved in the trial. In fact, the MRC paid no attention to the methodological flaws uncovered by Bradford Hill in the allocation process of the lobar pneumonia trial (MRC 1934) and never corrected them in any other test conducted during the 1930s[9]. On this basis, the standard account of the TTC trials concludes that there wasn't any real understanding of the epistemic properties of a fair test in the MRC: when such tests were implemented it was more as a means to achieve non-epistemic goals. The same applies to the ideology of fairness promoted by Fletcher: as Cox (1997) shows in detail, the MRC allowed exceptions to its own rules if more mundane considerations advised it. In Cox's account, impartiality, be it methodological or moral, was mainly a rhetorical tool for the MRC to establish itself as the sole voice for medical research during this period.

---

[9] The alternation of treatments have produced unbalanced groups in respect of age, which turned out to be a variable correlated with the effect of the serum. See Chalmers 2002 and Edwards 2007, pp. 93-119 for comments on this trial.

In my view, the evidence available allows for a more charitable epistemic interpretation. We need to separate the impartiality of the expert from the impartiality of the trial. As to the former, it was clearly admitted by all parties that there was a source of subjective bias: the commercial interests of the pharmaceutical manufacturers. Fletcher tried to warrant the impartiality of his experts cutting all their financial ties to the industry[10]. However, there seem to have been diverging views within the MRC team about selection biases from non-economic sources. To put it bluntly, even if a minority desired to avoid them (through the "Standard scheme of inquiry"), a majority of the MRC sponsored clinicians considered themselves perfectly capable of conducting a trial and interpreting its results impartially. In other words, they considered themselves impervious to every other type of bias –as most of us still do, according to current psychological research –see, e.g., Dana 2009. They might have been wrong, but nobody questioned the principle that biases should be prevented. Hence, the MRC clinicians did not see the necessity of constraining their studies with debiasing procedures.

We should also separate the impartiality of the MRC trials from the partiality of the goals they served. For instance, establishing the MRC as the British authority on drugs. As Daniel Carpenter has argued regarding the American Food and Drug Administration "there is nothing inherently 'selfish' or 'inefficient' about reputation protection" (Carpenter 2004, p. 54). If the MRC tried to establish itself as the sole arbiter of British trials it is just because this reputation would allow them to "more easily attract desired personnel; fend off budget cuts; and lobby for the programs, funds, and other things they desire" (*ibid.*)[11]. This is not incompatible *per se* with the impartiality of the MRC trials. Only a manipulation of these tests with a view to promoting the goals of the sponsoring body would have breached their impartiality.

4. THE STREPTOMYCIN TRIAL

By the end of the 1930s there seems to have been widespread agreement within the British medical community about the necessity of impartiality in trials, although

---

[10] Even if the MRC made exceptions to its own rules regarding the independence of trials from the industry, this does not imply that the principle of impartiality was questioned. Social norms do not need to be systematically obeyed to remain valid: as empirical studies show, compliance is never complete – see Bicchieri 2006 for an overview. If the exception had become the rule, the reputation of the TTC would have collapsed at some point, which apparently never happened.

[11] For an extensive discussion of the role of reputation in the analysis of regulatory agencies, see Carpenter 2010, pp. 33-70.

with no consensus as to its practical requirements. This consensus began to emerge after the war on one particular debiasing procedure, randomization, after it was implemented in the famous British Streptomycin trial. Again, the standard interpretation of the trial suggests that randomization provided just an appearance of impartiality. Applying the distinctions presented in section 2, I will defend that randomization in the Streptomycing trial was as impartial as we may expect. But let me first put this trial in context.

In 1943 Alfred Schatz and Selman Waksman, at Rutgers University, established that an antibiotic substance produced by the actinomycetes *Streptomyces griseus* halted the growth of various bacteria *in vitro*. Soon it was tested on *M tuberculosis* and by 1945 the first study on the use of streptomycin in human tuberculosis was published. Merck started to manufacture it and in 1946 a large multi-centre trial was conducted by the Veterans Administration in collaboration with the US Public Health Service, the National Research Council and the National Tuberculosis Association. Due to the short amount of streptomycin available, this first streptomycin trial was not randomized in order to maximize the number of patients receiving the drug. The majority of them improved from their lesions, though there were adverse effects and signs of the bacteria developing resistance to the drug. In 1947 these preliminary results were published and guidelines set on the therapeutic use of streptomycin. Further trials were called for[12].

Though the production (and commercial success) of streptomycin multiplied tenfold in a couple of years, these figures could not be matched in Britain until 1948, despite the licensing agreements. However, early in 1946 the MRC was expressing interest in the drug only to find the opposition of the Ministry of Health that feared public demand for the drug at a time it could not be met. However, expressions of public interest in the drug were already arriving at the Ministry. The National Health System had just been created and one of its founding principles, stated in the Beveridge Report, established that every British citizen was entitled to receive whatever available treatments were needed[13]. Did this principle apply to streptomycin? Again, for the

---

[12] See Yoshioka 1998a, pp. 18-35 and Toth 1998, pp. 222-248.

[13] "A comprehensive national health service will ensure that for every citizen there is available whatever medical treatment he requires, in whatever form he requires it, domiciliary or institutional, general, specialist or consultant, and will ensure also the provision of dental, ophthalmic and surgical appliances, nursing and midwifery and rehabilitation after accidents" (Beveridge, 1942, p. 158). See Toth 1998, pp. 232-234 and Yoshioka 1998a, pp. 40-42 for comments.

Ministry of Health the issue was whether this new drug was effective enough to deserve the investment and the MRC was commissioned to carry out a clinical trial.

In July 1946, the main decisions about the streptomycin trial were adopted in a series of meetings organized by the MRC: the types of tuberculosis and patients to be treated, the participating hospitals and laboratories and the methodology of the trial[14]. For the main group of patients in the trial (which finally would be those suffering from "acute, rapidly progressive, bilateral pulmonary tuberculosis of recent development, unsuitable for collapse therapy, bacteriologically proven, age limits 15-30") it was agreed that a double-blind trial with untreated control cases was necessary[15]. All the other trials were planned without this methodological proviso. Randomization is mentioned just once in the archival record (Yoshioka 1998a, pp. 177-178). The minutes of the meetings do not provide therefore many insights about the methodological discussion within the MRC, had there been any.

In 1947, once the doses were secured, the Streptomycin in Tuberculosis Trials Committee conducted the different tests. The patients received treatment between four and six months: admission started in January and ended in September. The number of patients in the pulmonary tuberculosis trial considered in the final report was 107. In September 1948, a draft paper on the trial written by M. Daniels and P. D'Arcy Hart was thoroughly discussed by the Committee and sent for publication in the October 30th issue of the *British Medical Journal*.

In the paper we find a quick defence of the need for controls, given the variability in the natural course of the disease. The most cited passage is probably the presentation of the randomizing procedure ("The Control Scheme"):

> Determination of whether a patient would be treated by streptomycin and bed-rest (S case) or by bed-rest alone (C case) was made by reference to a statistical series based on random sampling numbers drawn up for each sex at each centre by Professor Bradford Hill; the details of the series were unknown to any of the investigators or to the co-ordinator and were contained in a set of sealed envelopes, each bearing on the outside only the name of the hospital and a number. After acceptance of a patient by the panel, and before admission to the streptomycin centre, the appropriate numbered envelope was opened at the central office the card inside told if the patient was to be an S or a C case, and this information was then given to the medical officer of the centre. Patients were not told before admission that they were to get special treatment. C patients did

---

[14] See Yoshioka 1998a, pp. 165-186.

[15] Given the length of the trial, the idea of using a placebo for the control group (injections of an inert substance) "was felt to be impractical" by the planners (Yoshioka 1998a, p. 176).

not know throughout their stay in hospital that they were control patients in a special study; they were in fact treated as they would have been in the past, the sole difference being that they had been admitted to the centre more rapidly than was normal. Usually they were not in the same wards as S patients, but the same regime was maintained. (MRC 1948, p. 770)[16]

The other statistical feature of the trial, the use of significance tests, is not so often discussed. The trial had been planned, claim the authors, to answer the question "Is streptomycin of value in the treatment of pulmonary tuberculosis?" (MRC 1948, p. 780). The main criterion to decide about the efficacy of the drug was a comparison between chest radiographs of the patients at the beginning and at the end of the trial. Interestingly enough, there is no explanation of the statistical rationale of either randomization or significance testing –this latter is only clarified with occasional allusions to the probability of a difference between two percentages occurring by chance[17]. However, the streptomycin trial is usually considered a landmark in the history of statistics in medicine: after centuries of medical resistance to statistical thinking, the epistemic weight of Fisher's case for randomized experiments would have persuaded at least a minority of enlightened clinicians to implement the first solid RCT. Thereafter, it would become a paradigm for trialists all over the world[18].

Against this interpretation we find again an alternative approach focused on the worldly interests that pervaded the streptomycin trial, namely defended by Alan Yoshioka and Benjamin Toth. According to Yoshioka (1998a, p.185), the scientists involved in the trial "treated research as a means to achieve administrative ends" and, in retrospect, we may conclude they did it efficiently (pp. 223-224). As for Toth (1998, p. 252), the streptomycin trial methodology "does not so much reflect the interests of the groups involved as help to shape the relationships between the groups". In other words, it was not the epistemic weight of any statistical argument that made this trial a landmark RCT: the methodology was chosen because it was instrumental for the achievement of more mundane ends. Had it not been conducive to those ends, a

---

[16] Notice, though, that the trial was not blinded.

[17] In the 1948 volume of the British Medical Journal we find 18 mentions of the phrase "statistically significant" (including the MRC streptomycin paper). Since 1840, the phrase had appeared in 153 occasions, the first one in 1904 and the rest of them after 1927. The calculation is rarely provided.

[18] See the collection of acknowledgments compiled in Yoshioka 1998a, p. 182, n. 123 and p. 183, n. 126. The success of the trial owed more to its methodology than to the actual effects of the treatment: further studies showed that the early improvement detected in the 1948 trial did not necessarily indicate a better long-term prognosis: Florey 1961, pp. 132-133. I owe this reference to Donald Gillies.

different methodology might have been chosen. Let us briefly discuss this case examining the interests of each party involved in the trial.

As the archival record shows, from September 1945 through December 1946 the public was expressing more and more interest about streptomycin, requesting information and access to the new drug from the British authorities (Yoshioka 1998a, pp. 45-89). The only way to meet this demand was importing doses from the United States. From the point of view of the Ministry of Health, the goals were, on the one hand, to avoid wasting the scarce supplies available treating the wrong patients and, on the other hand, to avoid rationing, still darkly associated to the War years. According to Yoshioka (1998a, p.90), the Ministry decided to manage the situation allocating all the available doses to a clinical trial, in which economic or political influence gave no priority access to the drug. In order to justify this test, the Ministry fabricated a "toxicity scare" about streptomycin (Yoshioka 1998a, p. 92), stating time and again that its therapeutic value was still uncertain. Still according to Yoshioka (1998a), this was a manipulation of public opinion, since the certainties accumulated about the drug in America were enough to justify its mass production and this was acknowledged in private by the Ministry officers.

As for the MRC, to quote Yoshioka, it made a careful use of "its scientific expertise and credibility in order to achieve its institutional goals in the scramble for resources" (Yoshioka 1998a, p. 95). In other words, establishing itself as the British sole arbiter on drug assessment. The methodology of the trial allowed the MRC to exert a centralized control on the administration of the drug, preventing British clinicians from choosing the patients who would receive it –this was their methodology of choice, as we saw in the previous section. However, in this case, it might have been well received by the medical profession, since it relieved the clinician from the responsibility of selecting patients: despite the Ministry, the drug was thought a life-saving treatment and it was certainly hard to refuse it to a patient (Yoshioka 1998a, 178-179)[19]. Finally, regarding the pharmaceutical industry, no manufacturer in Britain was in a position to benefit from streptomycin production at the time of the trial, and their production plans were mostly decided by the MRC at its own convenience (Yoshioka 1997, pp. 94-127).

---

[19] This is an argument advanced in the editorial note that accompanied the publication of the trial in the British Medical Journal.

All these mundane interests would have driven the organization of the streptomycin trial: randomization was only used because it suited all parties involved, be it for the purposes of rationing, control or relief of responsibility, but not for any appreciation of its statistical value. The lack of epistemic weight of statistics in this trial can be inferred from the archival record, since there is no mention of randomization in the minutes of the meetings in which the trial was planned and just occasional allusions in a few related documents. Moreover, Bradford Hill only joined the actual discussions once the main methodological issues had been decided, so he could not possibly rehearse any Fisherian argument to inspire the trial (and very few would have been able to understand him). The use of statistical controls would have been mostly rhetoric, conclude in different ways Toth (1998) and Edwards (2007)

I disagree with this conclusion. Again we should distinguish between two separate epistemic dimensions in this trial: the control for biases, on the one hand, and the statistical interpretation of the result, on the other hand. Iain Chalmers (2005) has argued indeed that Bradford Hill was not following Fisher's theory about the statistical design of experiments when he prepared his "control scheme" for the streptomycin trial. He was just concerned with concealing the way treatments were allocated among the clinicians in charge of recruiting patients. This way, no conscious or unconscious selection of patients would bias the test. Hill's concern was shared by other members of the committee (such as P. D'Arcy Hart and G. Scadding) who had used alternation for the same purpose in the 1944 patulin trial (Chalmers & Clarke 2004). In sum, the epistemic weight of the control scheme lay mostly in the warrant provided of the fairness of the trial, not in the statistical weight of the conclusions. Bradford Hill was certainly a pioneer in medical statistics, but we may suspect that had he appealed to significance tests alone, he would not have found much support for the streptomycin trial.

Even if the control scheme was instrumental in respect of different practical goals, we should notice that achieving these goals depended on the impartiality of the control scheme. For all parties involved the more serious problem seems to have been how to allocate a scarce good (streptomycin). The standard solution in many different social settings is to organize a lottery, whose fairness depends on the chances you have of getting an equal share of the good (Stone 2007). For the Ministry of Health and the British patients randomization might have seem just one such lottery: as long as it was

unbiased (i.e., chances are evenly distributed among the participants) it is acceptable[20]. By the same token, we may well understand that the clinician was relieved of any personal responsibility in the choice of treatment: all the patients in the trial were equally entitled to receive streptomycin and they all had the same chance of getting it. If the practical goal to be achieved was the distribution of a scarce good without controversy, not every allocation scheme could have been used, but just fair mechanisms. In RCTs the social fairness of the allocation procedure goes hand in hand with its epistemic weight: the more impartial it is, the better it will prevent biases.

5. CONCLUDING REMARKS

In this paper, I have tried to show that a number of British landmark trials between 1930 and 1950 were more impartial than the standard historical account suggests. In my view, this account depends on an objectionable concept of impartiality, quite popular in social studies of statistical methods after Porter (1996). I think that Porter's main claim is correct: the intervention of the British health authorities in the pharmaceutical market was acceptable to the different stakeholders (pharmaceutical manufacturers, patients, physicians…) to the extent it appeared impartial. However, we should be more precise in the definition of this concept, if we want to explain why the results of MRC trials came to define a public standard in drug testing.

We should distinguish in the first place the impartiality of the expert and the impartiality of the trial. Whereas there is no clear consensus, still today, about how to warrant the former, early in the 1930s emerged a view about the necessity of experimental fairness, as defined, for instance, in the "Standard scheme of inquiry". A test would only be impartial if it incorporated debiasing procedures that prevented self-interested manipulations of the outcome. The Streptomycin trial showed how randomization provided one such warrant of impartiality, enough to create a social consensus on the experiment. The difficulty of finding similar warrants for simple expert judgment may probably explain why the mechanical objectivity of trials came to prevail in regulatory decisions.

However, trials can be put to a strategic use, as the British tests show: be it the promotion of national compounds, discipline clinicians through the research protocol or

---

[20] In *The Times*, October 29th, 1948, p. 2. we read for instance: "To remove every possibility of bias each patient's group was decided by a statistician using an ingenious system of sealed envelopes".

delay the introduction of a drug in the market. My claim is that such partial use of trials will only damage their impartiality if the research protocol is manipulated to meet those goals. I do not find evidence in any of the accounts discussed about such manipulations. Hence, I conclude that an alternative interpretation is at least defensible: the adoption of randomized clinical trials as a regulatory yardstick was propelled by the warrants of impartiality they gradually incorporated (e.g., blinding, pre-established rules of interpretation of their results [significance testing], etc.). The implementation of these warrants does not imply that the outcome will be unbiased. But, as far as we can discern today, they fared better than any alternative testing procedure in use till then.

The current design of RCTs was shaped in a process in which the public demand for impartiality in drug tests was growing. As the late Harry Marks (2000) argued, by the 1940s the pharmaceutical industry was deeply mistrusted by the public: commercial and scientific interests were deemed contradictory and RCTs were supposed to reconcile them. If we accept the concept of impartiality suggested in this paper, I think epistemic reasons will contribute as much as particular interests to explain the history of this process[21].

6. REFERENCES

Abraham, J. (1995). *Science, politics and the pharmaceutical industry*. N. York: St. Martin's Press.

Austoker, J., & Bryder, L. (1989). *Historical perspectives on the role of the MRC : essays in the history of the Medical Research Council of the United Kingdom and its predecessor, the Medical Research Committee, 1913-1953*. Oxford ; New York: Oxford University Press.

Beveridge, W. H. B. (1942). *The Beveridge Report in brief*. London: H.M.S.O.

Bicchieri, C. (2006). *The grammar of society : the nature and dynamics of social norms*. New York: Cambridge University Press.

Carpenter, D. (2004). The Political Economy of FDA Drug Review: Processing, Politics, and Lessons for Policy. *Health Affairs, 23*, 52-63.

Carpenter, D. P. (2010). *Reputation and power : organizational image and pharmaceutical regulation at the FDA*. Princeton: Princeton University Press.

---

[21] As a reviewer reminds me though, the methodology of the British Streptomycin trial might have been impartial, but we should not conclude that the warrants of fairness incorporated into the design of the experiment (randomization, significance testing) suffice today to grant the impartiality of pharmaceutically sponsored trial (see Goldacre 2012 and Dumit 2012 for an updated review). For a discussion see Teira & Reiss 2013.

Chalmers, I. (2002). MRC Therapeutic Trials Committee's report on serum treatment of lobar pneumonia. *JLL Bulletin: Commentaries on the history of treatment evaluation (www.jameslindlibrary.org).*

Chalmers, I. (2005). Statistical Theory Was Not the Reason That Randomization Was Used in the British Medical Research Council's Clinical Trial of Streptomycin for Pulmonary Tuberculosis. In G. Jorland, G. Weisz & A. Opinel (Eds.), *Body counts: medical quantification in historical and sociological perspective* (pp. 309-334). Montreal: McGill-Queen's Press.

Chalmers, I., & Clarke, M. (2004). The 1944 Patulin Trial: the first properly controlled multicentre trial conducted under the aegis of the British Medical Research Council. *International Journal of Epidemiology, 32.* 253-260.

Committee, S. (1947). The effect of streptomycin upon pulmonary tuberculosis. *American Review of Tuberculosis, 56*, 485–507.

Council, M. R. (1948). Streptomycin treatment of pulmonary tuberculosis. *British Medical Journal, 2*, 769-782.

Cox, D. (1997). *The Making of the Clinical Trial in Britain, 1910-1945: Expertise, the State and the Public.* Ph.D. Thesis, Cambridge University, Cambridge.

Dana, J. (2009). How Psychological Research Can Inform Policies for Dealing with Conflicts of Interest in Medicine. In Institute of. Medicine (ed.), *Conflict of Interest in Medical Research, Education, and Practice.* Washington, DC: The National Academies Press.

Dumit, J. (2012). *Drugs for life: how pharmaceutical companies define our health.* Durham, NC: Duke University Press.

Edwards, M. N. (2007). *Control and the therapeutic trial: rhetoric and experimentation in Britain, 1918-48.* Amsterdam: Editions Rodopi.

Florey, M. E. (1961). *The Clinical Applications of Antibiotics. Vol. II: Streptomycin and other Antibiotics Active Against Tuberculosis.* London: Oxford University Press.

Goldacre, B. (2012). *Bad Pharma.* London: Fourth State.

Hagendijk, R. (1999). An Agenda for STS: Porter on Trust and Quantification in Science, Politics and Society. *Social Studies of Science, 29*, 629-637.

Landsborough Thomson, A. (1973). *The Medical Research Council : half a century of medical research.* London: H.M.S.O.

Liebenau, J. (1990). *The twentieth-century British pharmaceutical industry in international context.* Madison, Wis. American Institute of the History of Pharmacy, 1990.

Marks, H. (2000). Trust and Mistrust in the Marketplace: Statistics and Clinical Research, 1945-1960. *History of Science*, 38, 343-355.

Medical Research Council (Great Britain), & Great Britain. Privy Council. (1920). *Report of the Medical Research Council* (Vol. 1919/20-1963/65.). London: H.M.S.O.

Medical Research Council (Great Britain), & Great Britain. Privy Council. (1934). *Report of the Medical Research Council* (Vol. 1919/20-1963/65.). London: H.M.S.O.

Meldrum, M. L., & State University of New York. (1994). *Departures from the design : the randomized clinical trial in historical context, 1946-1970.*

Porter, T. M. (1995). *Trust in numbers : the pursuit of objectivity in science and public life.* Princeton, N.J.: Princeton University Press.

Select Committee on Patent Medicines. (1914). *Report from the Select Committee on Patent Medicines: together with the proceedings of the Committee, minutes of evidence, and appendices.* Londond: Wyman.

Snodgrass, W. R., & Anderson, T. (1937). Sulphanilamide in the Treatment of Erysipelas. *Br Med J, 2*, 1156-1159.

Stone, P. (2007). Why lotteries are just? *The Journal of Political Philosophy, 15*, 276-295.

Teira, D. (forthcoming). A contractarian solution to the experimenter's regress. *Philosophy of science (PSA 2012 Proceedings).*

Teira, D., & Reiss, J. (2013). Causality, Impartiality and Evidence-Based Policy. In H.-K. Chao, S.-T. Chen & R. Millstein (Eds.), *Mechanism and causality in biology and economics.* London: Springer.

Therapeutic Trials Committee (1934). The Serum Treatment of Lobar Pneumonia: A Report of the Therapeutic Trials Committee of the Medical Research Council. *Br Med J, 1*, 241-245.

Toth, B. (1998). *Clinical Trials in British Medicine 1858-1948, with Special Reference to the Development of the Randomised Controlled Trial.* Ph.D. Thesis. University of Bristol, Bristol.

Yoshioka, A. (1998a). *Streptomycin, 1946: British Central Administration of Supplies of a New Drug of American Origin with Special Reference to Clinical Trials in Tuberculosis.* Ph.D. Thesis. Imperial College, London.

Yoshioka, A. (1998b). Use of randomisation in the Medical Research Council's clinical trial of streptomycin in pulmonary tuberculosis in the 1940s. *BMJ*, 317, 1220-1223.

Zamora, J. (2002). Scientific Inference and the Pursuit of Fame: A Contractarian Approach. *Philosophy of Science, 69*, 300-323.

Zamora, J. (2006). Science Studies and the Theory of Games. *Perspectives on Science, 14*, 525-559.