

Peter Keating and Alberto Cambrosio, “Too Many Numbers: Microarrays in Clinical Cancer Research.” *Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (2012), 37-51.

Who’s Minding the Data? Data Monitoring Committees in Clinical Cancer Trials

There is irony in that, in the case of trials where we experiment on human beings, those who do so cannot be ‘trusted’ to see the data they generate while they so experiment. Surely clinical trials must be one of the few fields of scientific inquiry in an open society where researchers are so regarded. (Meinert 1998a: 542)

1. Introduction

In his highly regarded history of the rise of clinical trials in America, Harry Marks (1998) describes how their widespread adoption resulted largely from the efforts of ‘therapeutic reformers’ who sought to replace the individual expertise of clinicians with the ‘science of controlled experiment’. The transition described by Marks resembles in many respects the transition from the ‘truth-to-nature’ objectivity of individual experts to a ‘mechanical’ form of objectivity portrayed by Daston and Galison (1992, 2007). In particular, Marks details the passage from a regime of trust in expertise and experts to a regime based on the mechanical generation of data, the elimination of human judgement and the adoption of a ‘view from nowhere’ (Daston 1992) where individual and group idiosyncrasies are systematically suppressed through recourse to quantitative methods.

Close observation of the evolution of the institutions created to manage knowledge and opinion in this community suggests, however, that we are dealing with more than a simple regime change from rule by experts to rule by numbers. In our opinion, the advent of clinical trials is linked to the emergence of a new form of objectivity in biomedicine — ‘regulatory objectivity’ — that generates conventions and norms through concerted programs of action based on the use of a variety of systems for the collective production of evidence, e.g. clinical trials, that in turn call upon standard substances and practices, e.g. guidelines (Cambrosio *et al.* 2006, 2007).

Regulatory objectivity incorporates unprecedented levels of reflexivity and is less a personal attribute or a meta-epistemic virtue than a form of institutional action. It relies on the formation of an internal consensus as to how to proceed ‘objectively’ rather than on the production of abstract standards, norms and measures that express the regulatory ideal of objectivity. Regulatory objectivity is thus invested in institutions and processes rather than objects or individuals. It is consequently fair to say that objectivity in modern biomedicine is but a tool for the production of information, governance, certainty, common standards and human safety and not a stand-alone virtue. Thus, unlike mechanical objectivity, institutions of regulatory objectivity produce schemes, guidelines and models of action that seek to submit the evolution of uncertainty to commonly shared rules of action.

While this regime of regulatory objectivity saturates modern biomedicine, it is particularly prominent in clinical cancer trials that, by their very nature, deploy a number of conventional devices such as diagnostic and treatment protocols, and that require constant recourse to collective and often very heterogeneous forms of expertise. A recent and yet nonetheless central institution in this regard is the ‘Data Monitoring Committee’ (DMC), also known as ‘Data and Safety Monitoring Boards’. According to standard definitions provided by regulatory agencies such as the U.S. FDA or the EMEA in Europe, DMCs consist of a group of experts who periodically review data from an ongoing clinical trial for the purpose of monitoring safety and treatment efficacy. After review, the DMC issues one of two judgments: stop or go. While a trial’s sponsor might ignore a DMC’s recommendation, doing so would seriously jeopardize the trial’s credibility.

Of course, not all trials have or need a DMC: according to the FDA (2001a), it depends on several factors, ranging from safety and practicality concerns to issues of scientific validity. In practice, institutional factors also enter the equation. One assessment based on 1990-1995 clinical trial publications, showed that while only 48% of clinical trials had a DMC, this percentage rose to 66% in the case of NIH and NCI-sponsored multi-centred trials and dropped to 42% in the case of industry-sponsored trials (Sydes *et al.* 2004). In general, the vast majority (90%) of trials that resort to DMCs are multi-centred. More importantly, trials using DMCs are stopped prior to completion a significantly higher number of times than trials without DMCs (Kiri *et al.* 2004).

A DMC's decision to stop a trial usually follows the breach of agreed-upon rules of statistical and clinical validity although, as we will see below, how to determine when a trial no longer merits pursuit is much more than a simple actuarial calculation. Similar procedures apply when a DMC decides that a trial merits continuance: if the accumulating data fall within pre-defined limits, then the board issues a statement certifying the ongoing validity and scientific merit of the trial (EMEA 2005, FDA 2001a, Kerr *et al.* 2004). Given what we have described so far, it is easy to see how DMCs offer an unusual occasion to explore the dynamics of regulatory objectivity. In order to reach a decision, DMCs incorporate both a novel use of mechanical objectivity, in this case clinical trial statistics, and techniques for aggregating individual expert judgments. In the course of a DMC's existence, these two forms of objectivity are combined to create a novel form of contextual, collective decision-making that is neither mechanical nor individual but is nonetheless authoritative.

2. Materials and methods

In order to explore the development of DMCs and of the novel form of objectivity they enact, we draw on several sources of information. In addition to information drawn from our broader research project on the history of clinical cancer trials (e.g., Keating and Cambrosio 2007) in which we have used archives (in particular those of the US National Cancer Institute), published and unpublished documents (ranging from annual reports by major US and European cooperative oncology groups to articles in leading cancer journals) and interviews with cancer clinicians, researchers and statisticians, we have conducted specific investigations concerning the history of DMCs and their use in the field of cancer. Two sources have proven particularly useful in this respect:

- In the early 1990s the NCI's adoption of DMCs as a mandatory component of Phase III cancer trials sparked a wide-ranging discussion amongst clinical cancer trialists. Although oncology figured prominently in the discussions, the debate about the controversial role and nature of DMCs was not confined to cancer. US and European clinical trial leaders in different disease areas from academic, government and pharmaceutical industry settings met to debate the issue during a 'Workshop on practical issues in data monitoring of clinical trials' that was held in January of 1992 at the NIH. A follow-up meeting took place the next year. The proceedings

- of both meetings appeared in two special issues of *Statistics in Medicine* — one in 1993 (vol. 12, issues 5-6) and one in 1994 (vol. 12, issues 13-14). While the discussion transcripts have been ‘edited’, the editors attempted to ‘preserve the deep concern ... that was expressed by many workshop participants’ (Ellenberg *et al.* 1993).
- In 2006, the FDA issued a Guidance document concerning the use of DMCs for all clinical trials (FDA 2006, henceforth *FDA Guidance*). The Guidance was the result of a process begun in 2001 with the issuance of a provisional guidance document (FDA 2001a, henceforth *FDA Draft Guidance*). The latter followed the usual pathway for the production of a guidance document, implicating a Public Meeting (henceforth *FDA Meeting*) following the release of the document, written commentary on the document and subsequent revisions. The transcripts of the *FDA Meeting* as well as related documents (in particular, written briefs) are available online on the FDA website (FDA 2001b).

Rich data sources like the verbatim transcripts of the *FDA Meeting* provide a ‘quasi-ethnographic’ window into two turning points in the development of DMCs. Our goal, in analyzing these documents, is not to provide a comparative description of the role and activities of DMCs in different settings and countries, but to examine this novel institution as a paradigmatic instance of regulatory objectivity, understood as a socio-technical practice, rather than a meta-epistemic phenomenon.

3. Histories of DMCs

Those implicated in the development of DMCs have provided two different accounts of the origins of this institution, both focusing on *outcomes*. After examining both versions, we offer an alternative, yet complementary account that foregrounds the socio-technical dimension of DMCs and, accordingly, focuses on *process*.

3.1 Participant history 1: ‘DMCs emerged in the late 1960s because of an administrative innovation at the National Heart Institute’

According to many participant histories including the ‘official’ history provided in the *FDA Guidance* document, DMCs originated in the late 1960s at the National Heart Institute of the NIH. The founding document in

this genealogy is the ‘Greenberg Report’, an internal NIH report produced by a committee headed by Bernard Greenberg, then a professor of biostatistics at the University of North Carolina. The Report developed a model for the administration of clinical trials, referred to retrospectively as the ‘NIH Model’. In 1988, following the decision by major cooperative oncology groups such as SWOG (Southwest Oncology Group) and ECOG (Eastern Cooperative Oncology Group) to make DMCs mandatory for all Phase II/III clinical trials, the journal *Controlled Clinical Trials* reproduced the ‘Greenberg Report’ in order ‘to preserve it as a citable document’ (Greenberg Report 1988). The Report has been cited ever since as the first official document to recognize ‘the need for a mechanism for terminating a trial early if it became evident that it could not meet its objectives or new information rendered it superfluous’ (Ellenberg *et al.* 2002: 5).

The idea that the Greenberg Report foresaw the rise of DMCs is more than implied in the flow chart used by Fisher *et al.* (2001) to summarize the NIH model and that was presumably taken directly from the Report (*Figure 1*). As can be seen, the ‘Data Monitoring Committee’ is part of the ‘Policy Board’. *Figure 1*, however, does not correspond to the flow chart published in Greenberg’s original report (*Figure 2*). There is, of course, significant overlap between the two renderings, but ‘Data Monitoring Committee’ is clearly a late addition to the function of the Greenberg’s originally unadorned ‘Policy or Advisory Board’.

Clear differences between the present and the past emerge when we turn to the definition of the Policy or Advisory Board. Its members, according to Greenberg, are ‘senior scientists, experts in the field of study but not data-contributing participants in it’, and they perform three functions: ‘Such a group can review the overall plan, make recommendations on any possible changes (including changes in protocol and operating procedures), adjudicate controversies that may develop, and advise the National Heart Institute on such matters as the addition of new participants or the dropping of nonproductive units [i.e., hospitals that fail to enrol a significant number of patients] (Greenberg Report 1988: 143). No mention here of interim review. The periodic review of the data is recommended, but it falls within the purview of the ‘Chairman’ of the study — ‘the most important position in a cooperative project’ — and his Executive Committee. According to the Greenberg report (1988: 142), their mandate includes ‘the responsibility ... to review the data analysis coming from the Coordinating Center and to

prepare *frequent reports* to the participants, as well as annual reports to the National Heart Institute’.

The idea, in other words, of interim reviews in 1967 is hardly to restrict access to data but to put as much data into circulation as possible. While the question of the early termination of a trial is indeed discussed, the issue is viewed separately from the continual review of data and is seen as something that might occur only under ‘unusual circumstances’. Those circumstances are: ‘if the accumulated data answer the original question sooner than anticipated, if it is apparent that the study will not or cannot achieve its stated aims, or if scientific advances since initiation render continuation superfluous’. As to who or how this decision should be made, the Greenberg report (1988: 146) gave no answer. It merely suggested that ‘a mechanism must be developed for early termination’. It is doubtful that Greenberg had present-day DMCs in mind as a possible mechanism.

3.2 Participant history 2: ‘DMCs emerged as the result of innovations in the field of statistics’

Statisticians are quick to observe that DMC decisions often implicate the use of ‘stopping rules’. Statisticians also observe that the latter were developed to solve a problem inherent to clinical statistics, awareness of which emerged at about the same time as the Greenberg Report in the late 1960s. There is thus a second, technical/theoretical as opposed to administrative account of the origins of DMCs. It runs as follows. In a 1969 article in the *Journal of the Royal Statistical Society* (Armitage *et al.* 1969) and in a 1974 article in the *New England Journal of Medicine* (McPherson 1974), a group of pre-eminent statisticians in the field of clinical trials forcefully argued that conducting repeated significance tests of clinical trial data as they emerged could lead to the production of false positive or false negative results simply through the laws of chance. A table in the 1974 *NEJM* article (*Figure 3*) showed that, for instance, at a nominal significance level of 5% (the value above which p-values are held to be significant), 25 interim readings would result in a 26.6% overall probability of achieving significant results when in fact there was no difference between two treatments.

Armitage and McPherson’s proposed remedy was to vary the nominal significance levels for each interim reading of the data according to the number of interim readings or analyses that the investigators planned to

make: the more significance tests one conducted, the higher the initial p-values would have to be in order to reduce the chance of calculating misleading results. There are now many ways to calculate such a sliding scale of significance levels and these different techniques are called ‘stopping rules’. The problem raised by Armitage did not go unnoticed by statisticians who, since the late 1960s, had learned to deal with the issue informally without the aid of DMCs. David Harrington (interview, 22 April 2005, Boston) former chief statistician for ECOG and member of many DMCs recalled:

In the early days of ECOG — and it’s still true — we would look at the data formally every six months ... but we knew even as the theory was just developing, every statistician, every trained statistician knows that if you look too often you might stop a trial for which there is no difference; you stop it by chance alone. ... And those were rules of thumb, quite useful rules and there was a very early paper by Haybittle [1971] that made that a bit more formal, that said ... any decent rule of thumb in sequential design says: ‘look as often as you want’, well, not exactly but: ‘look fairly often, and only stop the trial or think about stopping the trial if you see a difference that’s more than 3 standard deviations above the null, or below the null’.

Statistical techniques associated with interim analysis evolved in parallel with the set of problems they were designed to address. The data analyzed in the course of a trial are incomplete and subject to modification as further events are recorded. This problem was initially tackled by the *sequential methods* that emerged during WWII. As initially formulated, however, sequential trials required the analysis of patient data as each new pair of patients was treated. While this might have been appropriate for small trials undertaken at a single institution, as clinical trials became multi-centred and as the time-lag between the start of treatment and the measurement of responses (e.g., disease-free survival) became increasing large, sequential methods fell into disuse. In the mid 1970s, and following Peter Armitage’s cue, Stuart Pocock (1977), an ECOG statistician, generated the first *group sequential designs* (not to be confused with sequential methods) for clinical trials. These techniques allowed trialists to overcome two problems: they could analyze groups of patients at predefined intervals, say every six months to correspond with cooperative group meetings, and they could set pre-calculated significance levels or ‘stopping rules’ that would compensate for the fact that repeated testing would make the result of a positive outcome more likely (Canner 1977: 603). Once group sequential methods became available, the cooperative groups quickly adopted them. In this view, then,

the DMCs were simply an adjunct to a practice of reviewing the data at regular intervals in a way that avoided looking at the data too often.

With the advent of DMCs, a number of different ways of implementing stopping rules appeared. The graph in *Figure 4*, often reproduced in statistics and DMC textbooks, shows how the results of the three most popular techniques vary during the course of a clinical trial. Given this variation, failure to stipulate which stopping rules were to be used in a particular trial was occasionally the source of some mischief. As one trialist reported in 1993: ‘I was involved in a trial in which there were no boundaries set in advance and the co-ordinating centre said that if we had used Pocock [stopping rules] we would have just barely crossed, if we used O’Brien-Fleming we wouldn’t have, if we had used stochastic curtailing... and it became completely meaningless’ (Proschan quoted in Anonymous 1993: 586). There were thus clear advantages to stipulating the stopping rules beforehand.

Recent contributions claim that these seemingly incommensurable measures are ‘in fact mere transformations of each other’ and that ‘the definition of any boundary on one of the scales uniquely determines a boundary on any of the other scales’ (Ellenberg *et al.* 2002: 133). This structural equivalence is further reinforced by the existence of software that makes such test statistics immediately available. The underlying mathematical harmony does not, however, lighten the load for the DMC. Even if they are presented with all the stopping boundaries at once, they still have to choose amongst the various indicators and decide which is the most appropriate for the decision at hand. If, for example, the values are expected to fluctuate wildly at the beginning of a trial — and this expectation would be based on a clinical judgment — then statisticians generally recommend the selection of a rule that generates a more conservative boundary.

Despite this temporary consensus, controversies remain. Consider, for example, the somewhat fluid notion of a ‘convincing’ trial. Although all trialists agree that a trial should produce convincing evidence, the underlying question is: convincing to whom? Some investigators believe that a clinical trial should produce evidence that is convincing enough to change medical practice or, in other words, to convince average medical practitioners. Others may believe, sometimes on ethical grounds, that the numbers of patients required and the consequent p-values are simply too high a price to pay to convince the medical everyman; it is sufficient, in their eyes, to persuade

clinical researchers in the area in question. Such a level of persuasion would require fewer patients and lower p-values.

3.3 What do the official histories leave out?

These two participant histories obscure the fact that DMCs did not become widely adopted in the field of cancer until the end of the 1980s, almost twenty years after some of the original institutional and statistical innovations that made them possible. What took so long? What is missing from these accounts?

Between, on the one hand, the initial call for some device to regulate the early stopping of a clinical trial and the creation of formal techniques to calculate the stopping moment in the late 1960s and early 1970s, and, on the other hand, the widespread institution of DMCs in the 1980s, the field of clinical cancer research underwent a fundamental change. Clinical cancer trials stopped being simply a means to test drugs or to test a clinical hypothesis and became the full-time occupation of a number of researchers. Briefly and simply, the cancer trial metamorphosed from a tool to a style of practice (Keating and Cambrosio 2007). By the 1970s, cooperative oncology groups ran hundreds of Phase II and Phase III multi-institutional trials and, in the eyes of NCI administrators, group members and the permanent staff of the Statistical Centres had also emerged as interested parties in the conduct of clinical cancer trials.

The racing heart of the clinical cancer trial system was the aforementioned group statistical centre, the first of which was created in 1970 by Marvin Zelen for ECOG. Staffed with statisticians and ‘data managers’, and charged with the design and analysis of clinical trial protocols, these institutions are key to the performance and regulation of clinical trials. In the field of cancer, regulators see DMCs as a counterweight to the all-powerful statistical centre. DMCs, in other words, have emerged as part of a network of institutions whose collective output is a new form of objectivity that assures both the evidential and ethical nature of the enterprise known as clinical research. Accordingly, the original impetus to set up the DMCs in the cooperative groups was both to protect the data during long-term trials and to maintain the integrity of the clinical trial process.

With regards to the statistical techniques used in DMCs, given that their widespread adoption had to wait almost a decade, it is clear that their mere availability did not occasion their immediate use. Rather, in the late 1980s, within the cooperative oncology groups, there was a growing realization that many trials failed to reach completion precisely because of the publicity generated by favourable and unfavourable interim results. If clinicians came to know that drug A was succeeding or failing, they would have little impetus to put their patients on trial.

Until the mid 1980s, SWOG, for example, conducted all trials without monitoring committees. Aware of the problem of widespread knowledge of early results, SWOG compared a series of 14 SWOG trials with those sponsored by the North Central Cancer Treatment Group (NCCTG), founded in 1977 by members of the Mayo Clinic. The NCCTG had developed an early form of a DMC composed solely of the study investigators and statisticians from the Group's data centre. The purpose of the DMC was to restrict circulation of interim data rather than sharing with all members of the Group, as was the usual practice (Ellenberg *et al.* 2002: 8). As suspected, unlike the NCCTG, many of the SWOG trials had suffered declining accrual rates that had followed what appeared to be convincing positive or negative results. Thus, to save future data from similar corrupting influences, SWOG adopted a policy of restricted circulation of interim results, confining them to members of the DMC (Green *et al.* 1987).

Only after the groups had begun to use DMCs to limit the interim circulation of findings did the NCI make DMCs mandatory for all Cooperative groups in 1992. In its initial formulation, the NCI guideline for DMCs required the cooperative groups to eliminate possible conflicts of interest and establish DMCs that were 'independent of trial investigators', reasoning that 'because clinical trials are under increasing public scrutiny, we must use procedures that *protect our research* against the *appearance* of impropriety' (Simon and Ungerleider, quoted in Green *et al.* 1997: 90). The rules stipulated that all Phase III clinical cancer trials set up a DMC consisting of at least the Group Chair and the Group Statistician. NCI allowed the Groups to continue to appoint other members to the committees as they saw fit.

By the mid-1990s, clinical trial administrators began to question whether the DMC should be an outgrowth of the organizational machinery of the trial or whether it should consist of independent, external experts (Meinert 1998b, Ellenberg and George 2004). An 'internal' DMC, composed of researchers

directly involved in the trial, might be willing to pursue a trial in spite of serious problems, but an ‘external’ independent committee might lack both ‘clinical sensitivity’ (Simon 1993: 525) and an intimate knowledge of the trial’s intricate design and machinery.

In short: it was not at all clear how one should establish a committee that would further the interests of objectivity. In the early 1990s, some trialists believed that data were best monitored by those involved in the study with outside representatives from the NCI and the Group itself. Expertise, in other words, narrowed the number of possible participants. Yet another school of thought believed that the wider the spectrum of interests represented on the committee, the greater the objectivity of the outcomes of the committee’s deliberations. Thus, SWOG initially included representatives of the pharmaceutical industry on its monitoring committees even though the former openly admitted that one of the priorities of pharmaceutical companies was indeed ‘to make a profit and to pay dividends to shareholders’ (Rochold and Enas 1993). The experience proved ‘unsatisfactory’ and drug company representatives subsequently disappeared from SWOG’s DMC rosters (Green *et al.* 1997: 91).

In 1994 the NCI established a policy requiring DMCs to become independent of study investigators. The statisticians for the seven major cooperative oncology groups wrote a joint paper arguing against such independence and the ‘zero tolerance’ attitude that it implied (Harrington *et al.* 1994). The crux of their brief questioned the possibility and the utility of separating questions of risk from the scientific questions involved in the innovation process. Arguing that ‘[t]he research program of a multi-center clinical trials organization is an elaborate chess game and expert moves require knowledge of current theories of cancer etiology and treatment, experience in clinical practice, an awareness of current research in other organizations, and no illusions about the strength and weaknesses of that organization’, they suggested (1413) that:

Monitoring a clinical trial is a complex task, best conducted by a small group integrating all the issues that must be addressed. This group should freely and confidentially discuss patient safety, the accuracy of accrual and outcome predictions made at the design stage, protocol compliance, unexpected side effects, recent results from other, similar trials, the place of the trial in the scientific program of the organization conducting the study, and, when appropriate, outcome data. These issues are not separable and should not be discussed in isolation.

Moreover, trials often generate unexpected information that jeopardizes a mechanical handling of their results. A clinician-member of a DMC, reported that his role was ‘to recognize unexpected problems that surface, trends that appear in subgroups that have not been previously defined or even known about, so that I provide a clinical alertness quite different from just arbitrarily accepting whether a line has been crossed’ (Bristow, quoted in Anonymous 1993: 584). As one of the inventors of the numerous ‘stopping rules’ used in interim analysis, Peter Armitage (2001: 2545), has suggested, in particularly complex trials ‘it is better to play the matter by ear’.

Since the ‘statisticians’ revolt’, the NCI policy has undergone several revisions. The 1997 version of the NCI guidelines sought a compromise between the two extremes that had prompted the 1994 debate by mandating that the DMCs be composed of a majority of external members with a few insiders, as well as a patient representative. Moreover, the specific stopping rules to be used by the DMC to monitor the trial would henceforth be developed *in advance* by the study committee and not applied on an *ad hoc* basis by the DMC. In addition, some of the original opponents of independent committees have come to see them as a virtue. According to David Harrington (interview, 25 April 2005, Boston), former chief statistician at ECOG and one of the original opponents of independent DMCs, the change in viewpoint is largely due to the fact that the external monitors have learned to do what the internal monitors had done previously: acquire the expertise needed to understand the trial and not mechanically apply stopping rules.

4. DMCs as an institution of objectivity

Participants’ history 1 and 2, with their restricted focus on administrative and statistical developments, took for granted the emergence of clinical trials as institutional devices: our complementary account corrected that view. DMCs are a late addition to the ‘clinical trial machinery’, a further component of the regulatory arrangements ensuring that the trial system generates ‘factual information’ about new therapies and, more recently, the biological mechanisms underlying them.

DMC decisions should in principle be guided by an ‘objective’ assessment of interim data, but the search for objectivity, far from resting solely on

methodological principles, resorts to a set of reflexive practices that focus on the qualifications of DMC members, their attributes and prerogatives, and the extent to which a DMC should be able to redefine the parameters of a trial and to restrict the circulation of data. Reflexive forms of criticism such as those embodied in DMCs are not new. Indeed, Lynch (1982: 501) has convincingly argued that ‘endogenous critical inquiry’ is a constitutive feature of scientific activities, insofar as it ‘practically “solves” the issue of what counts as the relevant contingent relationship between “objective” and “socio-historical” states of affairs’. In Lynch’s investigations, however, this form of criticism was embedded in ordinary scientific actions, whereas in the case of DMCs designated actors conduct the criticism within a distinctive organizational frame: in short, endogenous critical inquiry has been institutionalized and framed by explicit socio-technical arrangements defining the structure and role of a DMC.

The reflexive engagement of DMC members in the governance of the institution can be easily retrieved from the discussions sparked by the 2001 *FDA Draft Guidance*. A focus on the issues raised during the *FDA Meeting* provides us with an ideal opportunity to observe, so to speak, the institution of an institution. These discussions, in other words, allow us to examine regulatory objectivity in action.

4.1 The role and purposes of a DMC

When DMCs first emerged as a going concern for the NCI, administrators believed that a statistician and a clinician would suffice to undertake the interim analysis. This was based on the belief that there were essentially two kinds of facts to be judged: statistical and clinical. And one could go further: if we exclude trials stopped on the basis of excessive toxicity or untoward clinical events, one might easily surmise that there is really only one kind of fact, namely statistical. In brief, it is possible to reduce the work of a DMC to a simple calculation: if, during the interim analysis, the trial crosses a pre-set boundary, then it must be stopped.

As experienced members of DMCs frequently point out, however, crossing the statistical line is never the end of the matter. As already mentioned, trials generate data on a number of variables and the context within which a trial is conducted constantly evolves. Parallel or recently terminated trials, for example, can generate data that will directly affect the interpretation of

interim results, just as research within the wider field of cancer can produce surprising results that change expectations and thus modify interpretations. In sum: '[t]he process of making such judgments is rarely straightforward; the observation of a low p-value or a sequential boundary crossed represents only the beginning of this process, not the end' (Ellenberg *et al.* 2002: 45).

One of the more contentious issues relating to the overall remit of a DMC concerned its power to change a clinical trial protocol and the conditions under which it could or should do so. On initial reading, the *FDA Draft Guidance* seemed quite generous in this regard. According to Gregory Campbell (*FDA Meeting*), Director of the Division of Biostatistics at the FDA Center for Devices and Radiological Health, who presented this section of the document at the Public Meeting, DMCs should be allowed considerable leeway to modify the original trial protocol. This expansive power followed on from the nature of clinical trials:

A clinical trial, after all, is not a fixed quantity. It's almost like a living thing. It evolves; it changes; it can change. One of the obvious ways in which a clinical trial might need to be modified has to do with the sample size. When the sample size is calculated, different things are assumed about the rate in the control arm, the rate in the treatment arm. Those assumptions may or may not be valid. And it may turn out that the trial is underpowered and the sample size needs to be adjusted. A data monitoring committee, although it's not easy, can grapple with this. If it's left only to a sponsor, it creates difficulties. There are questions about scientific validity in those cases.

Campbell's presentation prompted immediate disagreement from some members of the assembled expert panel. Janet Wittes (*FDA Meeting*), a prominent statistical consultant, felt that far from resolving the issue, Campbell had merely kicked the problem of validity further on down the road. Just as it would be invalid for the sponsor of a clinical trial to change protocol in the course of the trial as such changes could easily lead to charges that this had been done to facilitate a favourable outcome, it was equally as invalid for the DMC to do so, for a DMC 'is seeing data on efficacy and for it to have the ability and the right to change end points and to change crucial aspects of design I think can sacrifice the integrity of the design'. In other words, in the interest of the maintenance of integrity, the duties of a DMC with regards to protocols should be restricted to issuing orders to stop or go. As the only persons privy to the evolving data, however, the members of a DMC were ultimately the only ones who could recommend protocol changes based on knowledge of the data. According to

one of the authors of the Guidance document, Susan Ellenberg (*FDA Meeting*), the situation was paradoxical, for ‘if the only people who are in a position to make the change [in sample size] are people who have seen the interim data, you have no way out of this sort of conundrum’.

In some minds, the debate over a DMC’s power to change protocols turned on a prior distinction between internal and external information. As argued by Thomas Fleming (*FDA Meeting*), a statistician from the University of Washington and a principal contributor to the thinking behind the Guidance document, the original objection to changing clinical trial protocols confounded this distinction:

Not all studies are confirmatory but those studies that are confirmatory, I’d like to be able to interpret them in that manner. ... I’d like to have a prespecified hypothesis that I then confirm. At the same time, there can well be, during the course of a long trial, external information that could enlighten us as to what the hypothesis really ought to be. I actually don’t have a problem if I’m certain that it’s external data that leads to refinement and this is the essence of where this independence and separation enables or empowers the sponsor to have that flexibility.

Even with this distinction, the problem nonetheless remained. The DMC was privy to information that could lead to protocol change, but such change would potentially bias the entire trial. In this regard, it is not a coincidence that the FDA has more recently undertaken a campaign to encourage companies to replace sequential methods and stopping rules with ‘adaptive designs’ (Vastag 2006). Rather than attempt to foresee all contingencies and to specify under what conditions a trial should be stopped, adaptive designs offer DMCs the opportunity to redesign the trial based on interim results, although ‘the types of adaptation need not be pre-specified in the protocol’. In short, in contrast to ‘traditional committees [that] simply collate data against pre-determined stopping points ... the new committees [would] have much more power’ (Vastag 2006: 1043). At that point, however, they would no longer be monitoring the trial: they would be conducting the trial, by continuously modifying its content.

To sum up: the question of whether the DMCs have the power to change trial protocols raised the issue of whether clinical trials should be conceived of as self-contained devices or, to the contrary, as open to modifications prompted by ‘external’ information. Initially established to ensure the integrity of data, DMCs could become a threat to the machinery that

produced those data by exposing it to ‘external’ influences and by re-evaluating the role of one of its key procedural components: blinding. In short, endogenous critical inquiry addressed the critical nexus between the institution being regulated and the institution in charge of its regulation.

4.2 The structure and membership of DMCs

Unlike the supposedly context-free nature of statistical algorithms, the size and structure of DMCs are known to vary from disease to disease and country to country. While in the UK, for example, DMCs organized by the MRC tend to have 3 members — two clinicians and a statistician — (Grant *et al.* 2005: 96) in the United States, DMCs can balloon to as many as 15 members although the NCI counsels no more than 10 (Smith *et al.* 1997). According to the *FDA guidance*, in addition to statisticians and clinicians DMC organizers should also consider adding toxicologists, pathologists, clinical epidemiologists, laboratory scientists, pharmacologists, ethicists, and patient representatives or, in other words, consider rendering the composition of the DMC representative of the participants of the clinical trial process. Thus, resonating with the mandatory ‘inclusion’ of women and ethnic or racial minorities in trial populations (Epstein 2007), the *FDA Guidance* (9) suggested that ‘[a]ppropriate representation of gender and ethnic groups may be of particular importance for some trials’, and that ‘DMCs for international trials will usually include representatives from at least a subset of participating countries or regions’. Grafted on top of the statistical machinery of trials that was designed to eliminate the subjectivity of expertise, stipulations such as these for DMCs threatened to reintroduce the vagaries and biases of the external world.

In any event, subsidiary debates focused on two main topics:

- The kind(s) of expertise that DMCs need in order to function properly, i.e., to reach ‘objective’ decisions, and the conditions under which such and expertise could/should function.
- Potential conflicts of interests that, in a reflexive turn, became the object of topical investigations as to what actually amounts to a conflict given the present configuration of biomedical research. Since potential sources of bias appear to be all-pervasive — they include not only financial interests but also program objectives of public agencies or the pursuit of prestige and reputation — the issue was less one of drawing a boundary between the presence and absence of conflict of

interest, but of defining what would amount to an ‘acceptable level’ of conflict of interest. Participants made full use of their ‘critical capacity’ (Boltanski and Thévenot 1999) in performing this task. Due to space constraints, our discussion below will consider only the first topic.

The kinds of facts that enter into the determination of an interim analysis — statistics, clinical findings, regulatory events, etc. — require multiple experts and opinions to make sense of the evidence. As previously mentioned, clinicians and statisticians provide the information on which a DMC must ultimately rest its decisions, but the issue is not solely one of specific skills and expertise; familiarity with procedures and a healthy dose of pragmatism is needed. For instance, clinicians enable the committee to separate adverse reactions from the disease process itself. They also enable the DMC to stay abreast of ‘external information’ such as clinical research in the field and thus determine whether or not results relevant to the decision-making process have emerged from other studies. But too rigid a clinical stance could become a hindrance (Ferris, *FDA Meeting*):

One of the problems that I’ve seen over the years with clinicians on data monitoring committee is by nature we’re interested in individuals and what happens to this individual and at times some of the clinicians have asked literally for every case report. Bring in the wheel barrows because they want to see every last piece of data.

Similar considerations apply to other possible members of a DMCs. Robert J. Levine, Director of the Law Policy and Ethics Core at the Yale University Medical School complained (*FDA Meeting*) that: ‘There are a lot of tyroethicists who can be really very disruptive, thinking they’re going to apply their principles in the field of clinical trials’. Given the competing values of competing forms of expertise, a key issue, in this respect, is to prevent DMC deliberations from descending to the level of advocacy. Minimally, according to Fleming (*FDA Meeting*), a DMC’s decisions should be made by consensus emerging from lengthy and detailed discussions, rather than by votes.

While explicit rules, procedures and techniques for generating consensus are an essential lubricant of DMCs activities, there is something like a DMC know-how that has to be acquired through by experience and transmitted to potential DMC members. One of the expert panellists at the public meeting recalled how when she first joined the National Heart and Lung Biostatistics

Branch in the early 1980s she quickly learned that DMCs' 'decisions and the discussions are very complicated, they're very nuanced, and they reflect a certain sociology of a committee that varies from committee to committee', which is why senior colleagues advised her to 'go to every DMC ... that you can go to because you can learn a lot, it's the only way you're going to understand it and it's really fun', an opportunity that no longer exists, she lamented, given 'the many more rules about who can attend and who cannot attend' (Wittes, *FDA Meeting*). As a possible solution to this problem, Fleming (*FDA Meeting*) proposed a sort of 'apprentice' system whereby inexperienced statisticians would be paired with experienced statisticians on DMCs. Although this suggestion did not make it into the final document, it shows how DMCs have become an institution that creates and requires a new form of expertise just as it tries to minimize individual expertise as a source of authority.

Consensus within DMCs is usually built around statistical information. In turn, statisticians carry out the data analysis and deal with statistical issues as they arise. Maintaining objectivity among DMC statisticians, however, raised many issues. One of the more contentious concerned the notion of an 'independent statistician'. The *FDA Draft Guidance* (21) proposed that trials with DMCs be provided with three statisticians, for the 'integrity of the trial is best protected when the statistician preparing unblinded data for the DMC is external to the sponsor'. The statistician member of the trial's protocol team or steering committee would remain blinded to the data, the one reporting to the DMC would have unblinded access to the data but be independent of the trial sponsor, and the third statistician would be the statutory member of the DMC. This proliferation of statisticians did not go unnoticed: 'Some wag said that this document is a full employment act for statisticians' (DeMets, *FDA Meeting*).

The most immediate and obvious objection to the idea of an independent statistician was that those with the most intimate knowledge of the data would be excluded from DMC deliberations. According to Joe Constantino (*FDA Meeting*), the Associate Director of the Data Coordinating Center at the University of Pittsburgh, separating trial statisticians from the DMC would create an unwanted and unhelpful barrier between the DMC and knowledge of the actual conduct of the trial. Indeed, the idea of an independent statistician amounted to pitting objectivity against patient safety, the very thing that objectivity was intended to preserve. European statisticians, such as Dr. Richard Sylvester (2002) of the Brussels-based

EORTC Data Centre, expressed similar objections that put ‘independence’ in competition with ‘expertise’ in the common pursuit of objectivity.

In an attempt to defuse the dilemma, Thomas Fleming (*FDA Meeting*) pointed out that it was not a question of either/or but a problem of maximization. Even if cooperative group statisticians, for example, were to be prevented from serving as statisticians to the DMC, this did not necessarily entail that the DMC would be deprived of their knowledge. The structure of DMC meetings had been set up to enable alternative channels of communication. During ‘open sessions’, members of the protocol committee might be advised by a DMC on a number of issues (e.g., slow patient enrolment rates) that do not break the confidentiality of the safety and efficacy data, and, in turn, ‘[o]pen sessions allow for further sharing of insights by those individuals [i.e. cooperative group statisticians] who have unique insights who aren’t members of the [DMC]’ (Fleming, *FDA Meeting*). Once again, procedural arrangements solved an epistemic conundrum.

5. Conclusion

We began this paper with a quote from a well-known and outspoken statistician from Johns Hopkins University’s Center for Clinical Trials, Curtis Meinert, in which he expressed amazement at the convoluted arrangements embedded in DMCs that have been established to ensure the objectivity and trustworthiness of clinical trial results. As complicated as they may have been in 1998 when Meinert made his comments, DMCs have continued to evolve and now, a decade later, constitute a still controversial, but increasingly common component of large-scale cancer clinical trials and an essential part of their methodology and practice. The relevant sociological question thus becomes: how should we account for the emergence and development of this peculiar socio-technical device?

In order to answer this question, we need to link two separate bodies of literature: one focusing on the contingent character of scientific objectivity as embodied, for instance, in processes of measurement, representation and consensus formation; and a second dealing with biomedical regulation that has traditionally investigated professional, specialty and drug regulation. Only occasionally have scholars fully recognized the interweaving of processes of objectification aimed at eliminating human subjectivity and of

regulatory activities aimed at standardizing, for instance, the activities leading to the acceptance of therapeutic substances. Yet, as shown in this paper, the two practices are mutually constitutive, which is why we have introduced the notion of ‘regulatory objectivity’.

DMCs are a paradigmatic instance of regulatory objectivity. The kind of objectivity they enact is not reducible to an absence of subjective bias: the ‘integrity’ of the clinical trial process is also a sought-after good. As in most other institutions of regulatory objectivity in biomedicine, the debates and controversies surrounding the development of DMCs originated with the trialists, clinicians and statisticians: in other words, participants in the field of cancer clinical trials, far from ignoring the arguments and processes analyzed in this paper, have discussed and participated in them as part of the endogenous form of criticism that is constitutive of their activities. We can distinguish, in this respect, two different arenas of action. On the one hand, DMCs have become part of the design of large-scale cancer clinical trials within which they operate as an institutionalized form of endogenous inquiry. On the other hand, the rules that guide the functioning of DMCs are also subject to ongoing debate and criticism, occasionally prompted by proposals issued by public agencies such as the FDA and the NCI but often also derived from the tensions produced by the practical engagement with the operating procedures of the DMCs. It is precisely this reflexive layering and interlocking of different kinds of regulatory activities that qualifies DMCs as an emblematic example of regulatory objectivity.

To sum up: this paper has made two related but distinct contributions. The first refers to the analysis of the emergence and development of a new form of objectivity *and* regulation that increasingly characterizes clinical activities in domains such as oncology that are at the forefront of contemporary medical research. A second, less obvious contribution refers to the fact that while it has become commonplace for sociologists to emphasize the situated and contingent nature of clinical activities, and while this remains an important insight, it is also of limited value. As we saw amply above, clinicians and researchers, in spite (or maybe because) of the present hype over evidence-based medicine, are quite aware of the contingent nature of their activities. The interesting issue then is no longer *that* their activities are contingent but *how* practitioners *reflexively* try to work around these contingencies.

Confronted with the classical Wittgensteinian problem according to which (methodological) rules such as the stopping rules described in this paper provide no solution to the contingency issue since every rule requires an interpretation (a process leading to an infinite regress), medical practitioners cannot adopt the 'sociological solution' available to anyone who has been properly socialized in a given community, that consists simply in stating: 'this is how we do it'. Maintaining the integrity of the process of clinical trials demands that when challenged about a given rule, clinicians and researchers say and do something more than refer to shared, tacit practices and values. Accordingly, the present paper has described an activity devoted to the creation of mutually reinforcing administrative and statistical rules and their interpretation that provide a pragmatic solution to the infinite regress conundrum: such a regress does not occur in such an open-ended activity since, as we have seen, the rules are constantly revised in light of alternative and contradictory interpretations.

Acknowledgements: Research for this paper was made possible by grants from the Canadian Institutes for Health Research (CIHR MOP-64372), the Fonds Québécois de la Recherche sur la Société et la Culture (FQRSC ER-95786), and the Social Sciences and Humanities Research Council of Canada (SSHRC 410-2002-1453).

References

Anonymous. (1993) Panel discussion: scientific issues in data monitoring, *Statistics in Medicine*, 12, 583-600.

Armitage, P. (2001) Theory and practice in medical statistics, *Medical Statistics*, 20, 2537-2548.

Armitage, P., McPherson, C.K. and Rowe, B.C. (1969) Repeated analysis of clinical trial data, *Journal of the Royal Statistical Society, Series A*, 132, 235-244.

Boltanski, L. and Thévenot L. (1999) The sociology of critical capacity, *European Journal of Social Theory*, 2, 359-377.

Cambrosio, A., Keating, P., Schlich, T. and Weisz, G. (2006) Regulatory objectivity and the generation and management of evidence in medicine, *Social Science & Medicine*, 63, 189-199.

Cambrosio, A., Keating, P. and Bourret, P. (2007) Objectivité régulatoire et systèmes de preuves en médecine : Le cas de la cancérologie. In Tournay, V. (ed.) *La gouvernance des innovations biomédicales*. Paris: Presses Universitaires de France. pp. 155-175.

Canner, P.L. (1977) Monitoring treatment differences in long-term clinical trials, *Biometrics*, 33, 603-615.

Daston, L. (1992) Objectivity and the escape from perspective, *Social Studies of Science*, 22, 597-618.

Daston, L. and Galison, P. (1992) The image of objectivity, *Representations*, 40, 81-128.

Daston, L. and Galison, P. (2007) *Objectivity*. New York: Zone Books.

Ellenberg, S.S. and George, S.L. (2004) Should statisticians reporting to data monitoring committees be independent of the trial sponsor and leadership?, *Statistics in Medicine*, 23, 1503-1505.

Ellenberg, S., Geller, N., Simon, R. and Yusuf, S. (1993) Preface, *Statistics in Medicine*, 12, 451.

Ellenberg, S., Fleming, T. and DeMets, D.L. (2002) *Data Monitoring Committees in Clinical Trials. A Practical Perspective*. Hoboken, N.J.: John Wiley.

EMA. (2005) European Medicine Agency (EMA), Committee for Medicinal Products for Human Use (CHMP), *Guideline on Data Monitoring Committees* (London, 27 July 2005)
<http://www.emea.eu.int/pdfs/human/ewp/587203en.pdf>

Epstein, S. (2007) *Inclusion: The Politics of Difference in Medical Research*. Chicago: University of Chicago Press.

FDA (2001a) Food and Drug Administration, *Guidance for Clinical Trial Sponsors. On the Establishment and Operation of Clinical Trial Data Monitoring Committees. Draft Guidance*.
<http://www.fda.gov/Cber/gdlns/clindatmon.pdf>

FDA (2001b) FDA Guidance on Clinical Trial Data Monitoring Committees (DMCs). *Open Public Meeting, FDA Guidance on Clinical Trial Data Monitoring Committees (DMCs)*. Tuesday, November 27, 2001. Hyatt Regency Bethesda, Rockville (MD)
 part 1: <http://www.fda.gov/cber/minutes/clinmon1101p1.htm>
 part 2: <http://www.fda.gov/cber/minutes/clinmon1101p2.htm>

FDA (2006) Food and Drug Administration, *Guidance for Clinical Trial Sponsors. Establishment and Operation of Clinical Trial Data Monitoring Committees*
<http://www.fda.gov/CBER/gdlns/clintrialdmc.pdf>

Fisher, M., Roeker, E.B. and DeMets, D.L. (2001) The role of an independent statistical analysis center in the industry-modified National Institutes of Health model, *Drug Information Journal*, 35, 115-129.

Grant, A.M., Altman, D.G., Babiker, A.B., Campbell, M.K., Clemens, F.J., Darbyshire, J.H., *et al.* (2005) Issues in data monitoring and interim analysis of trials, *Health Technology Assessment*, 9, 7, xiv + 238 pp.

Green, S.J., Fleming, T.R. and O'Fallon J.R. (1987) Policies for study monitoring and interim reporting of results, *Journal of Clinical Oncology*, 5, 1477-1484.

Green, S., Benedetti, J. and Crowley, J. (1997) *Clinical Trials in Oncology*. New York: Chapman and Hall.

Greenberg Report. (1988) Organization, review, and administration of cooperative studies: a report from the Heart Special Project Committee to the National Advisory Heart Council, May 1967, *Controlled Clinical Trials*, 9, 137-148.

Harrington, D., Crowley, J., George, S.L., Pajak, T., Redmond, C. and Wieand S. (1994) The case against independent monitoring committees, *Statistics in Medicine*, 13, 1411-1414.

Haybittle, J.L. (1971) Repeated assessment of results in clinical trials of cancer treatment, *British Journal of Radiology*, 44, 793-797.

Keating, P. and Cambrosio, A. (2007) Cancer clinical trials: the emergence and development of a new style of practice, *Bulletin of the History of Medicine*, 81, 197-223.

Kerr, D., McBride, R. and Shemanski, L. (2004) Data monitoring committees in practice. Tips on using DMCs to improve trial efficiency and safety, *Applied Clinical Trials*
<http://www.actmagazine.com/appliedclinicaltrials/article/articleDetail.jsp?id=87247>.

Kiri, A., Tonascia, S. and Meinert, C.L. (2004) Treatment effects monitoring committees and early stopping in large clinical trials, *Clinical Trials*, 1, 40-47.

Lynch, M.E. (1982) Technical work and critical inquiry: investigations in a scientific laboratory, *Social Studies of Science*, 12, 499-533.

Marks, H. (1998) *The Progress of Experiment: Science and Therapeutic Reform in the United States, 1900-1990*. Cambridge: Cambridge University Press.

McPherson, K. (1974) Statistics: the problem of examining accumulating data more than once, *New England Journal of Medicine*, 290, 501-502.

Meinert, C.L. (1998a) Rejoinder, *Controlled Clinical Trials*, 19, 539-543.

Meinert, C.L. (1998b) Masked monitoring in clinical trials. Blind stupidity?, *New England Journal of Medicine*, 338, 1381-1382.

Pocock S.J. (1977) Group sequential methods in the design and analysis of clinical trials, *Biometrika*, 64, 191-199.

Rochold, F.W. and Enas, G.G. (1993) Data monitoring and interim analysis in the pharmaceutical industry: ethical and logistical considerations, *Statistics in Medicine*, 12, 471-479.

Simon, R. (1993) Remarks quoted in 'Discussion', *Statistics in Medicine*, 12, 521-525.

Slutsky, A.S. and Lavery, J.V. (2004) Data safety and monitoring boards, *New England Journal of Medicine*, 350, 1143.

Smith, M.A., Ungerleider, R.S., Korn, E.L., Rubenstein, L. and Simon, R. (1997) Role of independent data-monitoring committees in randomized clinical trials sponsored by the National Cancer Institute, *Journal of Clinical Oncology*, 14, 2736-2743.

Sydes, M.R., Altman, D.G., Babiker, A.B., Parmar, M.K.B., Spiegelhalter, D.J. and the DAMOCLES Group. (2004) Reported use of data monitoring committees in the main published reports of randomized controlled trials: a cross-sectional study, *Clinical Trials*, 1, 48-59.

Sylvester, R. (2002) Comment Number EC -7 on the Draft Guidance: Clinical Trial Sponsors on the Establishment & Operation of Clinical Trial DMCs (Docket: 01D-0489)

<http://www.fda.gov/ohrms/dockets/dailys/02/Feb02/022702/01D-0489-EC-7.html>

Vastag, B. (2006) New clinical trials policy at FDA, *Nature Biotechnology*, 24, 1043.

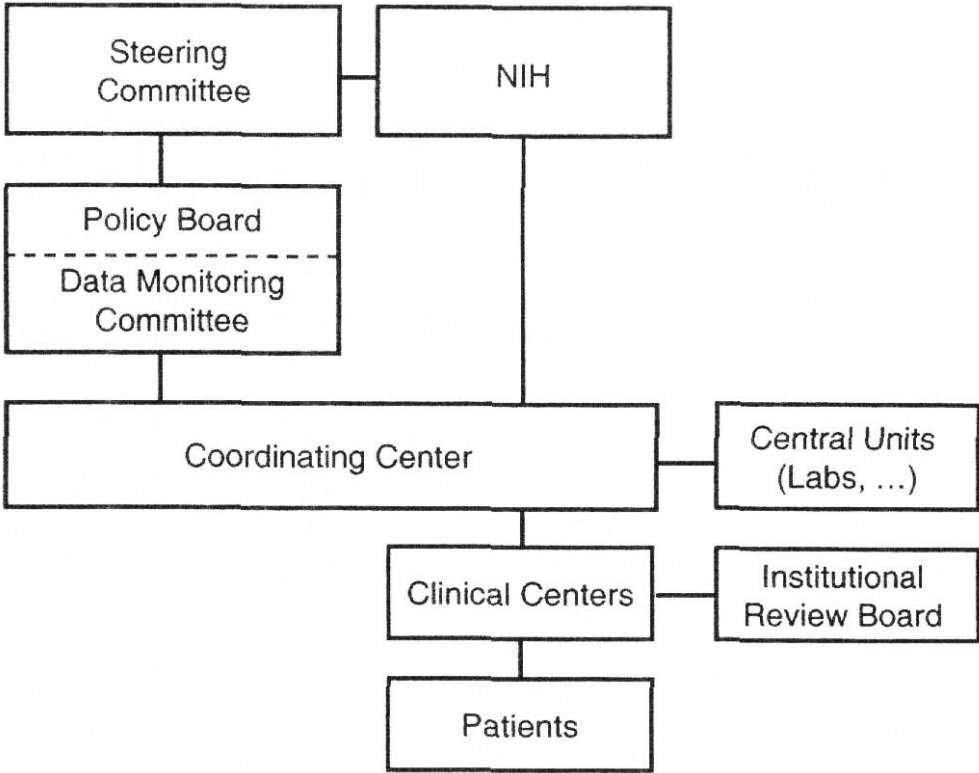


Figure 1: The 'NIH model' as represented in 2001. Reprinted from Fisher et al. (2001: 116), Copyright © 2001, with permission from the Drug Information Association.
366x287mm (72 x 72 DPI)

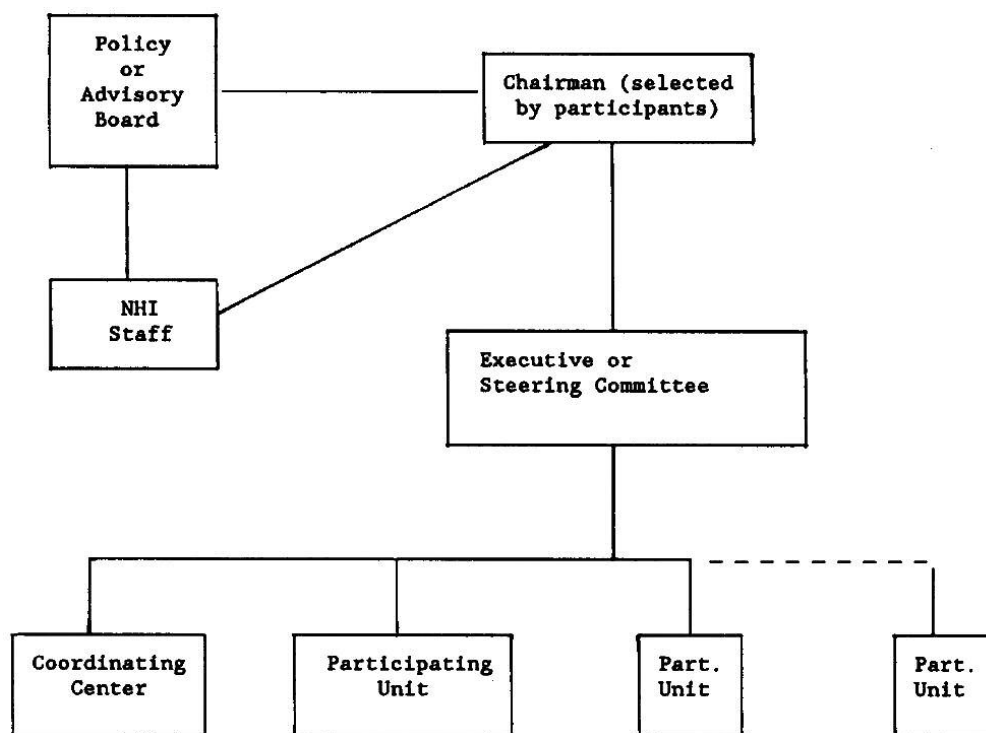


Figure 2: The 1967 representation of the 'NIH model'. Reprinted from Greenberg Report (1988: 138), Copyright © 1988, with permission from Elsevier.
347x255mm (72 x 72 DPI)

Table 1. Overall Probability of Achieving a Result with a Given Nominal Significance after n Repeated Tests when There Is No Difference in the Effects of Two Treatments (Expressed as a Percentage).

NOMINAL SIGNIFICANCE LEVEL (%)	NO. OF REPEATED SIGNIFICANCE TESTS (N)								
	1	2	3	4	5	10	25	50	200
1	1	1.8	2.4	2.9	3.3	4.7	7.0	8.8	12.6
5	5	8.3	10.7	12.6	14.2	19.3	26.6	32.0	42.4
10	10	16.0	20.2	23.4	26.0	34.2	44.9	52.4	65.2

Figure 3: McPherson's 1974 table illustrating the 'problem of examining accumulating data more than once'. Reprinted from McPherson (1974: 502), Copyright © 1974, with permission from the Massachusetts Medical Society. All rights reserved.
448x205mm (72 x 72 DPI)

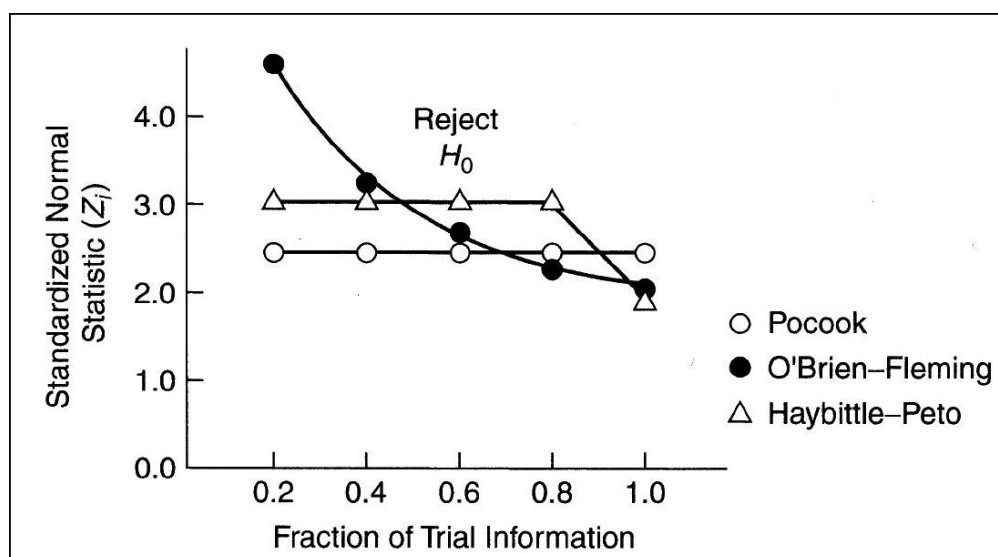


Figure 4: Variation of the thresholds set by three different stopping rules during a clinical trial. Reprinted from Ellenberg et al. (2002: 124), Copyright © 2002, with permission from John Wiley and Sons.
93x51mm (300 x 300 DPI)