

Data and Safety Monitoring Board and the *Ratio Decidendi* of the Trial

Roger Stanev*

* Banting Fellow, Ottawa Hospital Research Institute, and Part-time Professor, Department of Philosophy, University of Ottawa, Canada - email: rstanev@uottawa.ca

Abstract

Current decision-making by a Data and Safety Monitoring Board (DSMB) regarding clinical trial conduct is intricate, largely limited by cases and rules, and essentially secretive. Decision-making by court of law, by contrast, although also intricate and largely constrained by cases and rules, is essentially public. In this paper, I argue by analogy that legal decision-making, which strives for a balance between competing demands of conservatism and innovation, supplies a good basis to the logic behind DSMB decision-making. Using the doctrine of precedents in legal reasoning as my central analog will lead us to an analogy for much more systematic documentation and transparency of decisions in clinical trials. My conclusion is twofold: every DSMB decision should articulate a clear general principle (a *ratio decidendi*) that gives reason for the decision; and all such decisions should be made public. I use reported DSMB experiences of the Women's Health Initiative Clinical Trials to illustrate my analogical argument.

Introduction

Most clinical trials in the U.S. and Canada designed to assess the efficacy and safety of medical interventions require periodic assessment of evolving trial data. Such trials demand oversight by a Data and Safety Monitoring Board (DSMB). The two main mandates of the DSMB are to protect the safety of trial participants and the scientific credibility of trial results (Ellenberg et al. 2003). In order to meet these two mandates, the DSMB is guided by a trial monitoring plan (revised by the DSMB itself prior to any data collection) that includes rules such as stopping rules, which dictate when the trial might be stopped, continued, or modified given interim data. Despite consenting to the monitoring plan, the DSMB has sweeping discretion over whether or not it ought to follow its own agreed-upon rules during trial conduct. Given that the DSMB has an information monopoly during all interim analysis, also having sweeping discretion over the course of the trial precludes most meaningful oversight of its decision-making (Eckstein 2015). Decision-making discretion by the DSMB becomes particularly challenging given the added fact that most of its deliberations happen behind closed doors, routinely not reporting publicly its interim decision reasons and recommendations (Wittes 1993).¹

Although there are practical reasons for DSMBs to keep interim data analysis private (cf. Fleming et al. 2002) under the premise of confidentiality,² secret DSMB decision-making has at least one important shortcoming: the lack of publicity in decision-making prevents the public from getting a proper understanding of the reasons for the DSMB findings and final recommendation. Without a public rationale for its decisions (e.g., early stop, continue, changes to the trial), DSMB decision-making prevents others from reaching their own conclusions about the trial's ethical and scientific appropriateness. And this is an important distinction from the way decision-making by court of law happens, particularly in higher judicial decisions—e.g., setting a precedent—when a judgment is made explicitly and publicly with the inclusion of the judge's reasoning over the appropriate resolution of the legal issue.

In simple terms, decision-making in legal systems such as judge-made law strikes an optimal balance between the competing demands of conservatism (with *stare decisis*, the rule that like cases should be decided alike), and innovation (the continuous development of the legal system). Based on similar relationships in the ways DSMBs rely on rules to make decisions in clinical trials, my argument is to convey plausibility upon the need for publicity and explicitness in DSMB decision-making—contrary to current, secretive DSMB practice. If my analogy succeeds, I hope to show that a similar explanatory hypothesis in clinical trials would explain a similar consequence: DSMB decision-making striving for a balance between conservatism and innovation—avoiding dangerous medical treatments, and bringing new and effective treatments into use as rapidly as possible—promote the desiderata of publicity and explicitness.

The main point of my analogy is therefore to transfer a *pragmatic* justification for reasoning in accordance with the demands of conservatism and innovation from the legal domain to the domain of clinical trials. Seeking this analogue will lead us to an argument for much more systematic documentation of decisions in clinical trials by DSMBs; if past decisions are to be regarded as precedents, they can't be secret. The final conclusion is twofold: every DSMB decision should articulate a clear and general principle (a *ratio decidendi*) that gives reason for the decision in the trial; and all such decisions should be made public, so that we can build up a library of cases, making such decisions accountable. A crucial point of disanalogy—publicity of legal decisions vs. secrecy of current DSMB decision-making—shall be the focus of my argument with the normative upshot being that the decision process in clinical trials *ought* to be more like the legal system, i.e., decisions should be made public and based on clearly articulated general principles. I call these desiderata of publicity and explicitness.

I will focus on a particular analogical inference that I want to draw, and then assess the analogy relative to its conclusions. I use the DSMB experiences of the Women's Health Initiative Clinical Trials for illustrating that an analogy with the law is feasible. When framing and assessing the analogical argument, we need a clear logical link in the source domain (in my case, *stare decisis* and *ratio decidendi* from the law) which identifies the relevant factors that will then assist us in determining whether such link can be extended to the target domain (decision-making in clinical trials).

The idea of having DSMB interim decisions made available is not new. There are a number of guidelines that recommend reporting interim data and analyses, as well as the interim rule used to inform the DSMB decision, e.g. CONSORT (Moher et al. 2010), and DAMOCLES (2005). Korn and Freidlin go a step further than standard guidelines by proposing not just reporting what interim rule was used, “but explaining why formal monitoring was not used if it was not” (2011, 6). Very little evidence exists about how DSMBs are making interim decisions and why they are not following reporting guidelines (Eckstein 2015). My proposal differs from the guideline literature because it makes reference to a comparable and desirable functioning system of cases, rules and obligation, namely, judge-made law in common law. Having a functioning system to compare to DSMB decision-making not only has the benefit of informing it, but also gives reasons for why promoting publicity and explicitness is desirable, rather than simply assuming such ends are desirable as reporting guideline literature does. Reference to the law is made explicit through an analogical argument. Similarly to DSMB decision-making, the law also strives for a balance between competing demands of conservatism and innovation, which promotes publicity and explicitness as a standard of its decisions. With certain caveats—to be explained in §3 and §4—the law has the potential of serving as an important analog for promoting publicity and explicitness of DSMB decision-making.

My analogy with the law is further motivated by an increase in literature that focuses on making explicit otherwise implicit rules of norms by which reasons operate in institutions. Reasoning with cases, rules, and the values justifying such rules has been a common thread in much work on the literature of modeling reasoning—whether legal, scientific, or medical. Work in artificial intelligence (AI) is a good example.³ My paper, in the spirit of this literature, aims to link legal and medical reasoning by drawing on an important analogy between clinical trials and judge-made law.

The paper is organized as follows. §1 introduces DSMB decision-making and the need for an analogue such as the law. §2 presents an informal version of my argument by analogy and identifies the relevant factors in the analogy, which include legal rules, their generality, their background justification, as well as *stare decisis* and *ratio decidendi*. Here attention is given to how legal rules can be either over- or under-inclusive at times, and I explain how rules guiding clinical trials can run into similar generality issues. Additional assumptions and auxiliary hypotheses to my argument are also presented. In §3, I draw from the DSMB experiences of monitoring and reporting of the Women’s Health Initiative randomize hormone therapy trials to illustrate my argument, at which point it should be of no surprise that the fit between legal decision-making and DSMB decision-making is not without certain difficulties. Yet despite difficulties, I argue that the model of legal decision-making provides us with a plausible model for thinking about the justification of DSMB decisions and how it ought to be modified. In §4, I critically assess the analogical argument. §5 concludes with final remarks. In the Appendix, I provide a preliminary model of DSMB decision-making in light of the analogical argument.

§1. DSMB decision-making

In clinical trials designed to assess the efficacy and safety of medical interventions, evolving data are reviewed periodically by a DSMB. This board is an external and independent group—presumably the only group reviewing the data by treatment assignment. Due to its mandates, the DSMB has responsibilities to trial participants, trial sponsors, and to the scientific community at large to ensure the ethical permissibility of the study and the scientific reliability of the study results. The DSMB has a particularly important and, at times, difficult job, namely, that of deciding whether the current evidence warrants ending, continuing, or modifying the study—whether the evidence is for efficacy, harm, or futility (Ellenberg et al 2003).

A good way of characterizing what confronts DSMBs during the monitoring of clinical trials is to see them as striving for conflicting objectives of caution and expedience. This conflict between caution and expedience can be illustrated by situations when the DSMB must consider stopping the study before its intended completion, as in early stops due to efficacy and early stops due to harm.

In cases of early stop due to efficacy, the DSMB must balance the early evidence for efficacy against the uncertainty about long-term information concerning side-effects. This dilemma is best represented by the question: How long should the trial continue after early benefit is observed?⁴ For treatments like those for HIV/AIDS, which are presumably intended to be administered for the remainder of the patient's life, understanding the long-term effects, which include toxicity as well as whether or not long-run benefits are sustained, should be significant. The problem, however, is that keeping trial participants in the control group off treatments shown to be beneficial—at least in the interim—has critical ethical implications, given that DSMBs carry responsibilities towards securing the safety of all trial participants.

Cases of early stop due to harm illustrate another difficulty faced by DSMBs. The dilemma in harm cases, which must also grapple with the issue of how much longer to continue the study, is best represented by the question: Is the evidence for harm sufficient to rule out the possibility that the early effect is not spurious?⁵ Because the evidence demanded for demonstrating harm is less than that demanded for demonstrating efficacy, the focus of the dilemma shifts from being primarily an ethical issue (safeguarding trial participants from undue harm) to an epistemic issue, namely, whether the suggestive but inconclusive (perhaps not yet statistically significant) evidence for harm is sufficient to rule it out as due to chance. Whether we are dealing with cases of early stop due to efficacy or early stop due to harm, the DSMB continually walks a razor's edge between two opposing risks: premature stopping of dangerous or ineffective drugs, and undue delay in claiming safe, effective, and medically useful drugs available to the public.⁶

These examples serve to illustrate the conflicting objectives that DSMBs face: conservative vs. innovative objectives. Conservative because the DSMB needs to focus on protecting trial participants and future patients; and innovative because the DSMB also needs to focus on bringing new and effective treatments into use as rapidly as possible. What the examples do not illustrate, however, is my central contention with

current DSMB decision-making: the current lack of explicitness and publicity of its decisions. Because a clinical trial is a public experiment affecting human subjects, a clear rationale should be required for the decisions of the DSMB overseeing the trial. DSMB decisions are not self-justifying. They must be able to inform and meet legitimate challenges from skeptical reviewers, if such trials are to promote values such as innovation and trust in clinical trials. Both the formulation of challenges and the task of responding to challenges would be assisted if we had a public, clear and systematic way of articulating the reason for DSMB decisions—my desiderata of publicity and explicitness, and the reason for my analogy with the law.

Yet one might wonder, why not just argue directly for the two desiderata and skip any analogy with the law altogether? Couldn't one argue that rules regulating clinical trials—to which the DSMB should comply—should be made public, explicit, and justifiable to all those individuals over whom the rules purport to apply, whether researchers or participants? My concern is that even though a direct argument for the two desiderata might be possible, I suspect it would be a difficult argument to make without reference to a comparable, feasible (and desirable) available system. To clarify, clinical trials aim to promote the well-being of trial participants and to produce generalizable knowledge that shall improve medical care for future individuals, therefore demanding an optimal balance between conservatism and innovation. And without reference to a functioning system of cases, rules and obligation that also strives for a balance between competing demands of conservatism and innovation, it becomes hard to conceive of how such balance ought to work while promoting publicity and explicitness as a standard of decisions. That is why looking at the law seems promising.

Legal trials, like clinical trials, also aim at striking a balance between conservatism and innovation. An important feature of the law is that it is a system that aims at giving similar and predictable outcomes. As a system of rulings, the law aims at stability, an important form of conservatism. *Stare decisis* (i.e., standing by things decided) is what promotes stability in the law. But because the legal system also aims at innovation, allowing for the “orderly development” of its own system, *ratio decidendi* is essential. The existence of a *ratio decidendi* in judge-made decisions and its public nature enables prosecutors, defense attorneys, and higher courts to question and challenge a judge's decision. The role of *ratio decidendi* in relation to *stare decisis* explains how the law is able to promote an optimal balance between conservatism and innovation with public and explicit decisions, which I explain in §2.

My argument by analogy is an argument beginning from a set of relations in the law (my source domain). The link from the source domain is a logical association, namely, *stare decisis* and publicity promoting an optimal balance of values. I articulate the logical association in §2. My analogical argument is that a very similar logical association obtains for DSMBs. As I explain below, because there are no official models for this type of logical association in clinical trials, an informal model is called for and therefore developed in §2. Informally, my positive analogy is that *stare decisis* together with *ratio decidendi* explain legal decision-making as promoting a balance between the competing demands of conservatism and innovation. Based on a similar relationship in clinical trials, the point of my argument is to convey *plausibility* upon the need for publicity and

explicitness in DSMB decision-making (contrary to current DSMB practice). In a nutshell, the analogy is to show that a similar explanatory hypothesis in clinical trials would explain a similar consequence: DSMB decision-making that strives for a balance between conservatism and innovation promotes the desiderata of publicity and explicitness.

Forming the backdrop of my argument by analogy is Bartha's (2010) theory of analogical reasoning, which is a rich, normative theory of analogies. According to the theory, a good analogical argument shares a common logical core between source and target domain. This core is captured by two principles: prior association and potential for generalization. In contrast to classifications of analogical reasoning based simply on ('horizontal') superficial similarities between source and target domains, the theory says it is the logical ('vertical') relation that provides the key to determining the relevant similarities (and differences) to the analogy. This means that critical assessments of analogical arguments are mediated by a model of the logical relation. This model of the logical relation is what is called the prior association.⁷ This association stipulates that the source domain (here the law) must include an explicitly stated relation which the analogical argument is supposed to extend to the target domain (clinical trials). The second principle, potential for generalization, in turn stipulates the condition for *prima facie* plausibility.⁸ That is, in order for an analogical argument to be plausible (i.e., worthy of further investigation), features that play a key role in the prior association (e.g., *stare decisis*, *ratio decidendi*) should have analogs in the target domain. Together, the principles of prior association and potential for generalization define Bartha's theory, and form the backdrop of my argument by analogy.

§2. Association and preconditions

My link between the law and clinical trials is a logical association. The association is that *stare decisis* and *ratio decidendi* promote an optimal balance of competing values in the law. The two competing values are conservatism and innovation. In what follows, I briefly discuss essential elements of this prior association: the generality of legal rules, specifically how legal rules can be either over- or under-inclusive. Then, I introduce *stare decisis* (the doctrine of precedent) followed by *ratio decidendi* (a clear and general principle). The discussion will prepare the way for my argument by analogy; the argument for the *prima facie* plausibility that DSMB decision-making ought to be more like judge-made law in legal systems. The upshot is twofold: every DSMB decision should articulate a clear general principle (a *ratio decidendi*) that gives the reason for their interim monitoring decision; and all such decisions should be made public. I call these desiderata of explicitness and publicity.

Generality of rules and their background justification

Sometimes legal rules can be either over-inclusive or under-inclusive. We say a rule is over-inclusive when its reach is broader than its background justification. In common law, judges do have the power to make changes to legal rules when there are sound

policy reasons for doing so, and legislation does not indicate a contrary parliamentary intention to preserve the rule in its form. The Supreme Court of Canada (SCC) and its ruling in *R. v. Salituro* [1991] 3 SCR 654 is such an example. Here the judges found the legal rule of spousal incompetence to testify as a witness for the prosecution against the other spouse as over-inclusive and in need of change. The Court interpreted that the rule over-reached its background justification in cases where spouses were separated without any reasonable possibility of reconciliation.

The background justification of the spousal incompetence rule was the preservation of marital harmony. Although at the time of the SCC ruling the rule no longer applied to legally separated spouses (i.e., divorced couples), it still applied to spouses who were separated but not divorced, until the SCC judged it would no longer apply for aforementioned reasons.

The pertinent facts of *R. v. Salituro* included that the accused (husband) was charged with using a forged document contrary to Canadian Criminal Code. He had signed his wife's name on a check that was payable to him and then cashed it. His wife denied having given him such authority. In accepting her evidence the original trial judge convicted the husband. Although a Court of Appeal affirmed the conviction, the case was subsequently re-appealed to the Supreme Court under the question of whether the conviction violated the rule of spousal incompetence—i.e., “is there a common law exception to the rule against spousal competence for spouses who are separated”—since without the wife's testimony the appellant would not have been convicted.

In a comprehensive judgment, the judges examined in detail the rationale for the spousal incompetence rule, and together concurred that the application of the rule characterized an over-reach in cases of spouses who are separated without reasonable hope of reconciliation. The court reasoned that to apply and preserve the rule to such spouses would be an unrealistic attempt to comply with the rule's rationale of preserving marital harmony.

It was an appropriate case of the court identifying rule over-reach, and the need for changing the rule. Making spouses who are irreconcilably separated competent witnesses for the prosecution against the other was the right change. And although complex changes to legal rules with uncertain consequences are typically left to legislators, this is an instance where judge-made law can and should make changes to rules, particularly when bringing them in agreement with other fundamental values such as those found in the *Canadian Charter of Rights and Freedoms*.⁹ Alignment with *Charter* values was the general principle given by the court when reaching its decision to change the rule—a principle I shall return to in more detail when explaining *ratio decidendi*.^{10,11}

There are also times when legal rules can be under-inclusive. On these occasions, the rule fails to reach instances that the direct application of the background justification behind it would hold. Consider the legal rule of limiting marriage to a union between a man and a woman, with the background justification that it optimally promotes the environment to raise children. In a unanimous decision, Iowa State Supreme Court held

that the legal rule violated the equal protection clause of its State Constitution. The decision struck down the language from Iowa code section 595.2 that limited civil marriage to a man and a woman, and further changed the rule in a manner allowing gay and lesbian individuals access to the institution of civil marriage.

In its ruling the court found that the sexual orientation classification, employed to further the goal of an optimal environment to raise children, did not pass scrutiny because it was significantly under-inclusive. The statute, the court found, is under-inclusive because it does not exclude from marriage other groups of parents such as “child abusers, sexual predators, parents neglecting to provide child support, and violent felons”, that are undeniably less than optimal parents and provide a less than optimal environment to raise children. If the marriage statute was truly focused on optimal parenting, many classifications of individuals would have had to be excluded, not merely gay and lesbian people as presumed by the rule. The rule was also deemed under-inclusive because it did not prohibit same-sex couples from raising children in Iowa.

According to the court the rule’s under-inclusiveness revealed something even more disturbing and out of step with the State Constitution which required changes to the rule, i.e., that the sexual-orientation-based classification was grounded in prejudice or “overbroad generalizations about the different talents, capacities, or preferences” of gay and lesbian people, rather than having a substantial relationship to some important objective.¹² The court concluded that a rule that limits civil marriage to opposite-sex couples is simply not substantially related to the objective of promoting the optimal environment to raise children.¹³

The issue of generality of rules applies similarly in medicine. In clinical trials, statistical stopping rules—rules that dictate when to stop the study—are also subject to issues of generality. Statistical rules can be either over- or under-inclusive. In common law, a judge may follow earlier rulings even if she has good reason not to do so. Courts may follow earlier decisions even if those earlier cases were wrongly decided according to the pre-existing legal rule—unless the court has the power to change legal rules or overrule earlier cases,¹⁴ and decides do so—as illustrated in my court of law examples. Similarly, in medicine, at times the DSMB obeys, i.e., follows the prescriptions of the statistical monitoring rule of the trial, e.g., stopping the trial at a certain statistically significance level, even if that significance level is not the same as what the DSMB thinks is the best significance given the evidence and foreseeable risks in the trial. In such cases, complying with the protocol a priori statistical stopping rule is to do something the DSMB does not think best in the trial. Obeying a stopping rule—like obeying legal precedents—is to do something that, at times, we do not think best.

This point about the over- and under-inclusiveness of legal and statistical rules has important implications for the justification of legal and DSMB decision-making. This is particularly salient when objecting to authorities. Continuing with our spousal incompetence rule example, consider a spouse who is now separated pleading to the court that even though she is still legally married to her spouse, there is no reasonable possibility of reconciliation. The thinking here is that the spouse is not claiming that the spousal incompetence rule is unclear in its application, as she might be if she could not

be compelled to testify under 'reasonable and prudent' competence rules. She is competent. Rather, the attempt by a spouse to talk her way into the trial acknowledges that the rule (according to its literal meaning) plainly applies to her—she is a spouse—yet she nevertheless claims that literal application of the rule, in this particular case, would not serve the background justification grounding the rule. She admits she is the spouse, but she certainly is not an incompetent witness to testify for the prosecution against her spouse. Or so she claims. So in clinical trials, the DSMB (or any of its dissent members) may object to the principal investigator (PI) of the study that, contrary to what the PI's claim, the DSMB was not continuing the study unsafely. The objector is not claiming that the stopping rule is unclear in the trial, as the DSMB might be if it were following a study protocol with no formal expression of what evidence is required to establish 'highly conservative p-values' or 'proof beyond reasonable doubt' in the trial.¹⁵ Rather, the DSMB decided not to follow its protocol stopping rule as it continued with the trial because it thought finding out the long term effects of the drug not only outweighed the potential benefits of early stopping, but also because the decision to continue the trial didn't put participants at unsafe risks.

In one of the earliest examples of clinical trial monitoring where it is reported that the DSMB breached the protocol stopping rule was the Coronary Drug Project (CDP) sponsored by the National Heart, Lung, and Blood Institute.¹⁶ Even though sponsors of clinical trials do not usually disclose the precise reasons of their interim monitoring policy choice and early stopping decisions, there have been numerous examples, particularly in the last 30 years, where the DSMB, retrospectively, reported not complying with the trial stopping rule when justifying interim monitoring decisions. The CDP study is one such example. This was a double-blinded, multicenter randomized controlled trial (RCT), evaluating the efficacy of lipid-influencing drugs (e.g., clofibrate vs. placebo) in the therapy of coronary heart disease.¹⁷ The CDP study involved a total of 8341 patients followed over a period of five to eight years with clinical visits every four months.

This clofibrate vs. placebo trial included a DSMB composed of four statisticians, one epidemiologist, seven cardiologists, two toxicologists, and two clinicians, responsible for reviewing interim data reports every six months. (1981, 365) During the course of the trial the DSMB observed that the z-test statistic for the clofibrate-placebo difference in proportion of deaths (i.e., the z-score as an appropriate measure of agreement or distance from the null hypothesis for mortality as the primary outcome) crossed a statistical stopping boundary "signifying a conventional p-value < 0.05" during the first 30 months of the study (1981, 372). In spite of the fact that conventional statistical significance had been reached at interim (e.g., 'reject the null'), as prescribed by the stopping rule, the DSMB decided to continue the trial thinking that future trial participants were not under unsafe risks.

The aforementioned examples illustrate an important feature about both legal and statistical rules, namely, their common generality and background justification. In contrast to specific commands—e.g., press this button to accept the software license agreement now—rules in legal and clinical trials do not speak simply to one individual trial, at one time. Rather, such rules typically address many *but not all* situations. Even

though following precedent is supposed to be applied to all courts of law (in common law) on all days under all circumstances, it may not be justified in cases where the precedent is over-inclusive nor where it is under-inclusive. Such cases occur when an appropriate court (e.g., court of appeal) concludes not to follow a precedent that prima facie applies to its case; and so with DSMBs and their statistical monitoring rules in clinical trials. They may also decide not to follow their own interim rule as illustrated by the Coronary Drug Trials earlier example. DSMBs, similarly to common law courts, can and should make incremental changes to their own guiding rules so as to make the rules serve the interests of those they bind.

Like legal rules, statistical monitoring rules are characterized by being general in just this way, but like most generalizations they might not get it right every time. Because statistical rules are general, there is always the risk that the generalization that the rule embodies will not apply in some particular clinical trial, or if it does apply to the trial it may not 'hold proper' during the entire course of the trial. Even if it is true in most instances that the DSMB should not continue a trial in violation of the trial's stopping rule, there will be instances in which the generalization that 'stop if p-value is less than alpha' will not apply (since it isn't unsafe), and when that eventuality occurs the rule is over-inclusive.¹⁸

Now that I have explained how statistical rules in clinical trials, like legal rules, can be either over- or under-inclusive in their reach, I discuss one central rule in the law, *stare decisis*: the general rule of doing a similar thing that has been done in the past just because it has been done in the past, also known as the precept that like decisions be given in like cases in the law.

Stare decisis and ratio decidendi

When examining different forms of reasoning that have been traditionally associated with law, particularly common law, one form of reasoning jumps out as central: *stare decisis*. This is the general rule that says that like cases should be decided alike.

A central feature of *stare decisis* is that decision-making by courts respect precedent. In jurisprudence, precedents have 'binding effects.' In common law, precedent is 'binding' when the court of the present case is obligated to ('bound to') decide the case in the same way as the previous similar case, even when the current case appears to dictate a different outcome, or even if the court has good reason not to do so. Most legal cases do not create precedents, however. Precedents are those legal cases that require courts (e.g. Courts of Appeal, Supreme Courts) to resolve a dispute over the law.¹⁹ Lower courts are said to be strictly bound because they have no power to overrule higher court's decisions.

Even though the binding force of *stare decisis* is ubiquitous in common law it is decidedly non-absolute (Duxbury 2008; Shauer 2009). That is, even though a subsequent judge is 'bound' to follow a precedent, this obligation is not an absolute obligation to follow. Because common law is revisable in a manner that enacted legislated laws are not, precedents do not bind as statutes do. In judge-made law,

judges may interpret legal rules and statutes by scrutinizing their background justification so as to overrule precedents. This binding difference between legislated laws and judge-made laws has puzzled jurisprudential scholars and philosophers for years. Reflecting on the difference between the 'gravitational' force of precedents and the binding force of enacted laws, Dworkin makes this point of distinction, i.e., "the gravitational force of precedents cannot be captured by any theory that takes its full force to be its enactment force as a piece of legislation" (1978, 112-13). When judges resort to a precedent in their decision-making, they are reasoning for the purposes of a particular ruling, and deciding in a manner consistent with earlier decisions. The precedent constrains the decision-making. For Dworkin, this constraint means deciding on the basis of consistency considerations, or by an underlying principle of fairness, i.e., treating like cases alike. As he puts it:

The gravitational force of a precedent might be explained by appeal, not to the wisdom of enforcing enactments, but to the fairness of treating like cases alike ... [The judge] must limit the gravitational force of earlier decisions to the extension of the arguments of principle necessary to justify those decisions. If an earlier decision were taken to be entirely justified by some argument of policy, it would have no gravitational force. (1978, 318)

This tells us that for Dworkin, to decide in a manner that is consistent with earlier decisions means to take account of the *ratio decidendi*, i.e., the principle necessary to justify an earlier decision, which in turn provides a coherent justification for *stare decisis* and all common law precedents.

In practice, special circumstances may allow the overturning of precedents. Special circumstance might be justified when innovation trumps conservatism. Overturning precedents can and should occur when the highest court, e.g. the Canadian Supreme Court, overrides the obligation to follow *stare decisis* on the basis of having to promote values in accordance with the *Charter*. Either way, overturning a precedent requires the identification of its *ratio decidendi*. The *ratio decidendi* means either 'reason for the decision' or 'reason for deciding'.²⁰ It is the public authoritative portion of a precedent—the portion of a precedent that justifies its binding effect. It is the general principle that *gives reason for the decision*. Therefore, to say that a judge or court is 'bound' to follow a precedent is to say that the court is constrained in its decision-making by uncovering the ratio of the precedent for the purpose of its ruling.

In agreement with this interpretation of *ratio decidendi* is the distinction between the *ratio decidendi* and *obiter dicta*. In contrast with the *ratio decidendi*, *obiter dicta* is the portion of the judicial reasoning that is not necessary for the ruling. It represents other views or opinions in the judgment that are not binding on lower courts. *Obiter*, typically, is information related to the ruling but not directly in dispute. It may also be digressions. Given that a good portion of a lawyer's job is preoccupied with determining how to characterize the court's ruling—e.g., what portion of the precedent has authority—this characterization of *ratio decidendi* has its advantages but is not without challenges. One difficulty is proposing a "bullet-proof test" that distinguishes *ratio decidendi* information from *obiter dicta* (Duxbury 2008), since the two can at times blur into one another.²¹

Despite such difficulty, characterizing the *ratio decidendi* this way and in terms of a principle is common and appealing. Under this characterization, when a judge interprets a legislated rule in the process of reaching its decision, it means the ratio is what the judge thinks to be the best interpretation of the legislated rule, namely, the judge's ruling rather than the legislated rule per se (cf. Duxbury 2008). To the extent that legislation is interpreted by courts, then the process in deciding cases is governed by *stare decisis* and by the requirement that the rationale be published and clear.

I should add, however, that different characterizations of *ratio decidendi* have been proposed in the law (see Lamond 2014). The two most-cited characterizations are the views that *ratio decidendi* is: (1) the precedent's justification in the form of a rule or principle, and (2) simply the material facts of a case, i.e., all the relevant facts that were necessary for the result of the case (see Goodhart 1959, for such characterization). Despite different characterizations, the first one seems the most common among jurisprudence writers, and the one I adopt in my analogy.

With this characterization, a related feature of the law is that it is a system that aims at giving similar and predictable outcomes. Here, we say the law aims at stability, an important form of conservatism. *Stare decisis* is, in this way, *the* mechanism of attaining conservatism. But because the legal system also aims at innovation, allowing for the "orderly development" of its own system, *ratio decidendi* is key. It is key because through its identification, one can call *stare decisis* into question. Prosecutors, defense attorneys, and higher courts are given the means to challenge a judge's past decision, typically on the grounds that the judge's decision ought *not* to be extended to justify similar relationships in new cases. For this reason, *ratio decidendi*, understood in its relation to *stare decisis*, best explains how the law is able to promote an optimal balance between conservatism and innovation. Now that I have identified the relevant elements of the logical association in the law, I seek a model of DSMB decision-making with similar logical association.

***Ratio decidendi* in clinical trials**

Whitehead describes the nature of DSMB decision-making the following way:

A consequence of [sequential analysis of clinical trials] is that if no formal study design is used, then no valid analysis is possible. If the stopping rule is violated, either by premature stopping for some other cause or by continuation beyond the indicated stopping time, this too may invalidate the analysis. It is usual for major studies of life-threatening diseases to have a data and safety monitoring board. The DSMB must reserve the right to stop the trial at any time for any reason if this is felt to be ethically proper. The committee should also review the decision to stop at the time indicated by the stopping rule. They should have the power to override the stopping rule. (...) An inescapable conflict between the welfare of study subjects and the scientific goals of the investigators exists in all clinical trials. The scientific validity of the analysis is pure only if the stopping rule is obeyed to the letter. The interests of study subjects are protected only when the DSMB is empowered to break the rule. This conflict can be minimized if the stopping rule is designed to call off the trial

automatically in the most probable and foreseeable situations of patient disadvantage. (Whitehead 1997, 12)

The *stare decisis* analog in clinical trials is captured by the scientific requirements of clinical trials referred to by Whitehead. The scientific requirements include both ethical and epistemic requirements of the trial. By including both sets of requirements, a clinical trial aims at giving permissible, reproducible, and predictable outcomes, without which “no valid analysis is possible.” Statistical considerations in the form of controlling significance levels—i.e., the control of type I errors—and the statistical power of the trial are typical expressions of such requirements.²² Similar trials are expected to be designed, monitored, and decided in a like manner general principle.²³

Decision-making by the DSMB respects the trial’s scientific requirements. For instance, conventional statistical monitoring rules that underlie the design of the trial dictate DSMB having a protocol stopping rule for monitoring. This rule works similarly to precedents in the law by bringing a ‘binding effect’ to DSMB decision-making. The rule constrains DSMB decision-making by presenting a set of stopping conditions with corresponding courses of action (a.k.a. stopping boundaries) which the DSMB is ‘bound to’ follow, even when the current interim evidence appears to dictate a different course of action for the trial, i.e., the DSMB has reason not to comply with what the stopping rule demands. However, most interim analyses do not raise difficulties to DSMBs. In most instances the DSMB simply follows what the stopping rule says.

Even though stopping rules seem ubiquitous in clinical trials, as mentioned in §1, it is certainly a non-absolute DSMB obligation to follow them, since the DSMB has an information monopoly over interim analysis and discretion over the course of the trial. Because the reasons for early stopping (i.e., early stopping principle) is revisable during the course of the trial in a manner that the protocol stopping rule is typically not, the stopping rule does not bind as the early stopping principle. In hard ethical cases, the DSMB may find itself having to revisit its stopping rule and interpret the rationale underlying the rule so as to overrule what the rule dictates. That is why making a distinction between on the one hand the early stopping principle (analogous to *ratio decidendi* in legal trials) and on the other the stopping rule per se is helpful in understanding what the DSMB can and should do.

Similarly to justices in the law, if the DSMB finds that its stopping rule is out of step with one of its moral mandates—e.g., safeguarding trial participants—the DSMB may scrutinize its rule closely. For instance, while the stopping rule may monitor primary efficacy endpoint(s) and perhaps also one or two endpoint(s) for serious adverse effects known to be associated with the intervention being assessed, the DSMB may think it needs to evaluate the net benefit of a therapy going beyond its stopping rule. Such examples abound in the clinical trial literature, particularly for cases of early stop due to efficacy or safety (for such examples see DeMets et al. 2006). DSMB decision-making may put greater emphasis on safety concerns, not only for known or suspected adverse effects that are captured by the stopping rule, but also for the possibility of unexpected toxicities.

Yet despite these logical analogs in the application of rules and their epistemic criteria, in contrast with legal decision-making, the rationale for either stopping or continuing a clinical trial at any given interim point is private, not public. If there is indeed an early stopping principle that is used by the DSMB, this principle is left private, nearly never spelled out in writing and made public. Despite federal regulatory guidelines on the operation of DSMBs, and guidelines from statements such as CONSORT (Moher et al. 2001) which demand explicit articulation of interim decision-making,²⁴ mandates have been mostly ignored in practice. As a matter of fact, until recently, a clinical trial could be conducted almost completely in secret. Claiming proprietary rights, trial sponsors could keep all information about the conduct of the trial, including its very existence, private. This meant that, if a drug had important adverse effects, such information might never reach the public. (Drazen et al. 2007) If adverse effects had been observed at interim, neither the public nor regulatory agencies had the proper means to question DSMB decisions.

Since 2007, however, important changes have been made in the operation of clinical trials with the FDA Revitalization Act. For instance, today, “once a clinical trial is mounted, the sponsor has an ethical obligation to publicly acknowledge the contribution of the participants and the risk they have taken by ensuring that information about the conduct of the trial and its principal results are in the public domain” (2007, 1756). Although the Act takes steps in the right direction, ‘public acknowledgment of trial participation’ is still a far cry from making DSMB decisions explicit and public. And this is an important difference from the way legal courts articulate their decisions. That is because in legal courts, such as those in common law, the law recognizes a duty to provide reasons for their decisions under appeal; ‘the rationale for the decision’ is a duty to the public at large. This duty to articulate explicitly and publicly the rationale for the decision or ‘adequate reasoned judgment’ implies telling the public why the decision was made.

In a court of law, judges are not only responsible for evaluating the evidence presented in the case but also deciding on the relevant facts of the case. For instance, in higher courts (Appeals or Supreme), judges may review legal and factual arguments before deliberation. The court may hear oral arguments from the attorneys before deciding the case and issuing its conclusions. The justices meet privately behind closed doors to deliberate and vote on how the case should be resolved. The justices weigh their options prior to member voting, and then make the court’s final decision in accordance with rules of law and jurisdiction. The decision is often made through a consensus or simple majority depending on the type of legal case. By parallel reasoning, in clinical trials the DSMB is responsible for evaluating the evidence presented during the trial and deciding on the relevant facts of the case. Similar to higher courts of law procedures, during closed door deliberations the DSMB also weighs its options prior to member voting, followed then by a final committee decision (either to stop the trial or continue) in accordance with the study protocol rules and its own DSMB charter.²⁵ The DSMB decision also aims at a consensus or the application of a majority vote rule.

But unlike the legal setting where the reason for the decision is spelled out publicly in writing by the court as a matter of duty to the public, the reason for an interim decision is

not spelled out publicly—not by the DSMB, nor the study principal investigator, nor by anyone else involved in the trial. This lack of publicity and explicitness in clinical trial decisions leave the public at a loss to understand the reason for the trial decision. Was it because the DSMB thought the long term effects of AZT couldn't outweigh its potential benefits of early stopping? Can the DSMB decision to stop the trial be rationalized as an instance of the maximization of expected utilities? Was the early stopping of the 'estrogen alone' trial based on a minimax regret principle, i.e., DSMB chose the action whose maximum regret is minimal? Or was it based on the maximization of the minimum foreseeable benefits of the intervention, i.e., that stopping the trial halfway safeguarded trial participants optimally? Without a *ratio decidendi* such questions cannot be answered.

Given the ways clinical trials are currently reported there is simply not enough information in public reports (e.g., journal publications) to allow the public a proper understanding of the trial findings (cf. Montori et al. 2005; Mills et al. 2006). And there is evidence to suggest that regulatory agencies such as the U.S. FDA, which is in charge of overseeing clinical research in the interest of the public at large and has the legal authority to obtain DSMBs meeting reports (both open and closed meeting minutes), is arguably in no better position than the general public when it comes to assessing *ratio decidendi* of clinical trials.²⁶

In sum, the similarities between legal and clinical trial decision-making seem to lend some prima facie plausibility to the conclusion that DSMB decision-making *ought* to be more like legal decision-making. If I am right in my analogy, the link of striking a balance between conservatism and innovation 'entails' a *ratio decidendi*. To appreciate the analogy further before assessing it in §4, I examine the DSMB experiences of the Women's Health Initiative randomize hormone therapy trials (WHI). This example, which is based on the experiences and thinking of DSMB members who have actually done data and safety monitoring of the WHI trials, provides one case that, in practice, the analogy with the law might work. The WHI trials also serve to document empirically where the discussion of DSMBs (and their decision-making) in the medical literature is currently deficient, and to show how that analysis is likely to be enriched by the analogy with the law along the lines of what I argue here.

§3. Women's Health Initiative Clinical Trials

The Women's Health Initiative Clinical Trials (WHI) were a group of large double-blind randomized controlled trials sponsored by the National Heart Lung and Blood Institute (NHLBI) studying chronic disease prevention in postmenopausal women. WHI tested three strategies for disease prevention: hormone therapy, a low-fat diet, and calcium plus vitamin D supplements. These interventions were assessed in terms of their "risks and benefits" for reducing the incidence of coronary heart disease (primary outcome measure), breast and colorectal cancer, as well as hip fracture, stroke, and other secondary outcomes (Wittes et al. 2007).

The hormone therapy component of WHI involved testing two separate interventions in different subgroups of women: estrogen plus progestin (E+P) in women with an intact uterus, and estrogen alone (E-alone) in women with prior hysterectomy (Anderson et al. 2007). The E+P enrolled over 16,000 postmenopausal women, whereas the E-alone enrolled about 10,000 participants. In April of 2002, against much prevailing wisdom regarding the benefits of hormone therapy, and at odds with the evidence from a large set of observational studies, the DSMB decided to stop the E+P trial—three years earlier than its intended original completion—given what it saw at interim analysis. The decision to stop the trial was based on interim assessments that “the health risks exceeded benefits for disease prevention in postmenopausal women” (2007). This decision followed the trial’s protocol monitoring rule and stopping criteria for adverse effect—to be explained below. The upshot: the NHLBI accepted the DSMB decision to stop E+P given the DSMB’s reason for the decision.

As acknowledged by Goodman (2007), the WHI trials marked a touchstone in the history of clinical trials. This landmark can be explained by the scale and complexity of the trials, as well as the considerable efforts made by the DSMB “to develop monitoring procedures with the right degree of structure and flexibility” (ibid, 205) that promoted the publicity and explicitness of decisions. As explained in §2, understanding the nature of statistical monitoring rules as to what they are, i.e., general rules, and foreseeing that the adoption of such rules may fall short of the rules’ intent, the DSMB in its wisdom devised an explicit monitoring plan that could strike a proper balance between competing demands of conservatism and innovation while rejecting stopping criteria deemed ad hoc.

The plan was devised in order to provide the foundation of every interim analyses and their proper public reporting. The plan made explicit a number of things, including criteria for early stopping, namely, that early stopping considerations for adverse effects would be triggered, if and only if, “any disease outcome crossed its associated lower statistical boundary *and* the global index was suggestive of overall harm” (ibid, 209) [emphasis is mine]. Using explicit stopping criteria, the DSMB also made explicit to the NHLBI the presentation of interim results, interim decision based on those results, and the reason for such decision.

Anderson et al. (2007) and Wittes et al. (2007) are unusual examples of DSMB reflecting retrospectively on the process of coming up with a monitoring plan that meets explicitness and publicity of decisions. These postmortem reflections supplement the WHI original publications in important ways. They make clearer why and how early stopping criteria were adopted in the trials, and how the DSMB decision ensued from such criteria. As in a court of law, Wittes et al. (2007) make public DSMB summaries of the arguments for and against stopping the trial (see Exhibit 1, pp.232-33). Of particular relevance were split opinions regarding what kind of data would convince clinicians. Members against stopping E+P early thought interim results wouldn’t persuade skeptical clinicians of the E+P harms; those in favor of stopping the trial early thought results could be compelling enough to change clinical practice provided that the reason for stopping were also spelled out clearly and publicly. And despite differences of

opinions among DSMB members, once the interim results had reached the DSMB explicit criteria for stopping, the decision to stop E+P early was unanimous.

As to how the WHI stopping criteria were derived so to avoid unduly ad hoc decisions, Anderson et al. (2007) report:

The plan was informed by *a priori* expectations of benefits and risks of hormone therapy and developed around key design assumptions, the protocol-defined analysis plan, and standard statistical methods for multiple outcomes and interim analyses. It was tailored to support the clinical and ethical judgment of the DSMB by an iterative process of assessing the DSMB's responses to hypothetical scenarios of interim results. This development process was instrumental in bringing forward various perspectives of board members in their assessment of the clinical environment in which these data would be viewed. The statistical framework for the stopping boundaries was derived to present the data in alignment with their judgment. The concept of a global index of risks and benefits was one particularly useful aspect of the plan that arose from these discussions. The primary role of the global index in this plan was to promote caution in early termination if there was emerging evidence in both directions (2007, 215).

The doctrine of precedent was applied during the WHI trials, where “the principles that guide the reporting of the E+P trial were employed in reporting the parallel WHI trial of E-Along” (Anderson et al., 2007, 208). Similar to the law, there was a deliberative attempt from the DSMB to decide and report the E-alone trial according to principles used in the precedent E+P case, so that similar issues or facts may yield similar predictable DSMB decision.

This time, however, the arguments for stopping the E-alone trial early focused on the strong evidence of stroke (i.e., an unbalanced global index) with the data showing no evidence either of harm or benefit for coronary heart disease (Wittes et al. 2007, 228). The argument for stopping the trial surmised that if the interim data became available to the public, no one would take estrogen to prevent coronary heart disease. But different than the E+P trial, where the DSMB voted to stop the trial unanimously, this time the DSMB was hung after extended deliberation. The DSMB voted 5 to 4 to continue the trial with each member also reporting “a strength of conviction in their decision” on a scale of 0 (definitely stop) to 100 (definitely continue). This procedure resulted in a DSMB summary of confidence ranging from 45 to 55, thus a ‘hung trial’ (ibid).

Given the split decision, the DSMB led the NHLBI to convene a new panel to review the interim data and to make an independent recommendation as to whether to stop or continue E-alone trial. After new panel deliberations, and meetings with representatives from other NIH institutes, the argument for stopping the E-alone trial early won the day. And the final rationale for the DSMB decision was that the elevated risk of stroke, and lack of coronary heart disease benefit, would harm postmenopausal women more than benefit them, even though no stopping boundaries for harm had been crossed; unlike the E+P precedent trial where the data did cross its stopping boundary for harm, resulting in a unanimous decision.

§4. Assessing the argument

We are now in a position to critically assess the analogical argument. I do so in the following way: I assess whether the attempt to transfer the positive analogy to the target domain is blocked in some way (no-critical difference condition).

Blocking the positive analogy transfer

One disanalogy is that there really is no higher-level body in clinical trials analogous to higher courts (Appeals Courts, Supreme Court) that could overrule any doctrine of precedent. Judges obviously do not like to have their decisions overturned and have strong motivation to apply precedents with great care. This is not a decisive disanalogy, however. A key point of the analogy is that the proposal (explicitness, publicity, use of past cases) offers a rational way to promote the core values espoused by any DSMB. Moreover, there might be separate reasons not to take the extra step of introducing something like a regulatory body analogous to a court of appeal in clinical trials due to its cost. Premised in terms of cost-saving practices meeting market demands, a secondary objective of clinical research is to strike a balance between regulation and cost, so long as safety and scientific validity are not compromised in important ways.

A second disanalogy is the independent importance and validation of statistical methodology in DSMB decisions vs. ordinary legal decision-making. Simply put, there are scientific imperatives that dictate policy in the DSMB case, while there is no such thing in the case of legal reasoning—e.g., empirical adequacy or truth. But this difference, although deep, does not defeat the analogy. On the one hand, so long as both conservative and innovative values remain central to both domains, the fact that certain scientific imperatives differ from the judicial (e.g., justice) is somewhat irrelevant to the argument, because it does not figure directly in the justification for the desiderata of explicitness and publicity of decisions. The point of the analogy is to transfer a pragmatic justification for reasoning in accordance with the demands of conservatism and innovation from the law to clinical trials. Even though the disanalogy may suggest that the pragmatic justification is unlikely to be transferred successfully to clinical trials, in fact, in evaluating the argument, what matters is whether the differences are critical relative to the logical association. The difference just noted is ultimately irrelevant, because the positive analogy abstracts from it.

Given the legal analogy, what to make of DSMB legal liabilities? That is, given the analogy, a core legal question is whether legal liability should attach to DSMBs, institutions, or individual DSMB members. Here I suspect there might be good arguments on both sides. An argument for DSMB liability would have that the DSMB should take its moral mandates very seriously—perhaps more seriously than it currently does—and produce a careful report that justifies its actions. An argument against it would have that making the DSMB legally liable would have undesirable effects on clinical research, since very few individuals would agree to serve on DSMBs. If this is correct, then making DSMB legally liable might hold up drugs for years and years, making drugs more expensive and less expedient to patients.

A third line of thinking is that there might be some intermediate measure, e.g., limited liability, or liability of some sponsoring institution, given that legal liability can come with different degrees of acuity, and therefore avoid opposites such as "complete immunity" or "strict liability". There could be limited liability in gross criminal behavior (e.g., taking bribes to produce particular decisions, acting with malice towards the interests of trial participants), liability for willful, reckless endangerment, liability for negligence or failures of due diligence, or liability for failure to exercise competent judgment.

Moreover, if there is a public good at stake, and there is the potential for legal accountability, then it seems reasonable to require DSMBs to keep somewhat detailed records of decision-making justifications, accessible by courts of law with warrants in case there is a matter of legal fault involved. Perhaps there is still a reason to keep these proceedings confidential by default, yet that in itself is not incompatible with insisting that records be made and kept, and that a public rationale for interim decision be made available along the lines I have argued in this paper.

Surely research institutions should be held liable for harms that are foreseeable and avoidable, and neither negated by the valid consent of participants to risk such harms nor outweighed by considerations of social benefit expected from such research. If institutions are held liable, then there are grounds to hold DSMBs and their members to a weaker standard of legal liability, since they have an agency relationship to their larger institutions that bear full legal liability. If institutions have fairly strict liability for harms, then they have incentives to operate effective DSMBs to avoid being held accountable for harms.

If the differing purposes and mandates imposed on DSMBs might be in conflict, it is either necessary to have explicit rules and principles for resolving the conflicting purposes, or else have the functions of the boards dis-aggregated, so that a board that is charged with the ethical protection of research subjects and scientific validity of the study is not also charged with looking out for the broader economic interests of the institution.

§5. Final remarks

In this article, I have argued for certain policies in DSMB decision-making, based on the similarity of DSMB decisions and legal decisions. I set out an argument by analogy for DSMB decision-making striving for a balance between competing demands of conservatism and innovation promoting publicity and explicitness of decisions. My argument is subject to the assumption that *stare decisis* and *ratio decidendi* promotes the explicitness and publicity of legal decisions. My way of dealing with the lack of explicitness and publicity of DSMB decision-making supplements reporting guidelines literature and current approaches to the ethics of clinical research. It does so because it rests on an ethics of public reasoning that is close to the law. DSMB decisions are not self-justifying. They must be able to inform the public and meet legitimate assessments and challenges from reviewers. The WHI trials and their DSMB experiences are a living testament that promoting publicity and explicitness of decisions in clinical trials might be

possible. My argument by analogy does not rest on any deep metaphysical assumptions about the law nor scientific methodology. It is not subject to the objection of existing disanalogies between the law and clinical trials in the ways I discussed earlier. Its soundness is compatible with the need for revising the current ethics of DSMB decision-making.

Acknowledgements

An earlier version of this paper was presented at Models and Decisions conference, Munich Center for Mathematical Philosophy, April 2013. I would like to thank all those who have discussed and commented on the paper, both during its presentation and submission, and, informally, at the Ottawa Hospital Research Institute and the Department of Philosophy at the University of Ottawa. Special thanks go to Paul Bartha, Scott Anderson, Karine Millaire, David Hyder, Paul Rusnock, Vincent Bergeron, Scott Findlay, Jan Sprenger, Frederik Herzberg, Dean Fergusson, Bill Cameron, Michael Strevens, Jason Borenstein, Stephan Hartmann, Jeremy Simon, Mohammed Ansari, and Nick Barrowman. I would also like to thank the generous support from the Canadian Institutes of Health Research via its Banting Fellowship award.

References

- Anderson G.L., Kooperberg C., Geller N., Rossouw J.E., Pettinger M., Prentice R.L. 2007 "Monitoring and reporting of the Women's Health Initiative randomized hormone therapy trials." *Clinical Trials* 4:207-217.
- Ashley, K. 1991. *Modeling Legal Argument*. MIT Press.
- Bartha, P. 2010. *By Parallel Reasoning*. Oxford University Press.
- Branting, K. 2000. *Reasoning with Rules and Precedents*. Kluwer, Dordrecht.
- Bench-Capon, T., and Sartor, G. 2003. "A model of legal reasoning with cases incorporating theories and values." *Artificial Intelligence* 150:97-143.
- Berner, E.S. 2007. *Clinical Decision Support Systems: Theory and Practice*. Springer.
- DAMOCLES Study Group. 2005. "A proposed charter for clinical trial 2005 data monitoring committees: helping them do their job well." *Lancet* 365:711-22.
- DeMets, D., Furberg, C.D., and Friedman, L.M. 2006. *Data Monitoring in Clinical Trials: A Case Studies Approach*. Springer.
- Duxbury, N. 2008. *The Nature and Authority of Precedent*. Cambridge University Press.
- Dworkin, R.M. 1978. *Taking Rights Seriously*. Harvard University Press.

- Eckstein, L. 2015. "Building a More Connected DSMB: Better Integrating Ethics Review and Safety Monitoring." *Accountability in Research*. 22:81-105.
- Ellenberg, S., et al. 2003. *Data Monitoring Committees in Clinical Trials*. John Wiley & Sons, LTD.
- Goodhart, A.L. 1959. "The Ratio Decidendi of a Case." *Modern Law Review* 22: 117–124.
- Hart, H.L.A. 1994. *The Concept of Law*. 2nd ed. Oxford University Press.
- Korn, E. L. and Freidlin, B. 2011. "Inefficacy interim monitoring procedures in randomized clinical trials: The need to report." *The American Journal of Bioethics* 11(3):2–10.
- Lamond, G. 2014. "Precedent and Analogy in Legal Reasoning." *The Stanford Encyclopedia of Philosophy* (Spring 2014), Edward N. Zalta (ed.).
<http://plato.stanford.edu/archives/spr2014/entries/legal-reas-prec/>
- Laudan, L. 2006. *Truth, Error and Criminal Law: An Essay in Legal Epistemology*. Cambridge University Press.
- McCormick, C.T. 1984. *McCormick on Evidence (Practitioner Treatise Series)*. West Group.
- Mills et al. 2006. "Randomized Trials Stopped Early for Harm in HIV/AIDS: A systematic Survey." *HIV Clinical Trials* vol.7 1:24-33.
- Moher et al. 2001. The CONSORT statement. *BMC Medical Research Methodology* 1:2 DOI: 10.1186/1471-2288-1-2.
- Montori, V.M., et al. 2005. "Randomized Trials Stopped Early for Benefit: A Systematic Review." *Journal of the American Medical Association* 294, 2203–2209.
- R. v. Salituro*, [1991] 3 S.C.R. 654, <https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/820/index.do>
- Salbu, S. 1999. "FDA and Public Access to New Drugs." *Boston University Law Review* 79:93-152.
- Schauer, F. 2009. *Thinking Like a Lawyer*. Oxford University Press.
- Stanev, R. 2011. "Statistical decisions and the interim analyses of clinical trials." *Theoretical Medicine and Bioethics* vol. 32 1:61-74.
- Stanev, R. 2012a "Modelling and simulating early stopping of RCTs: a case study of early stop due to harm." *Journal of Experimental & Theoretical Artificial Intelligence* vol. 24 4:513-526.
- Stanev, R. 2012b. "The Epistemology and Ethics of Early Stopping Decisions in Randomized Controlled Trials." Ph.D. dissertation, The University of British Columbia.

The Coronary Drug Project Research Group. 1981. "Practical aspects of decision making in clinical trials: the coronary drug project as a case study." *Controlled Clinical Trials* 1:363-376.

Whitehead, J. 1997. *The Design and Analysis of Sequential Clinical Trials*. John Wiley & Sons.

Wittes, J. 1993. "Behind Closed Doors: The Data Monitoring Board in Randomized Clinical Trials." *Statistics in Medicine* vol. 12, 419-424.

Wittes J., Barrett-Connor E., Braunwald E., Chesney M., Cohen H.J., Demets D., Dunn L., Dwyer J., Heaney R.P., Vogel V., Walters L., Yusuf S. 2007. "Monitoring the randomized trials of the Women's Health Initiative: the experience of the Data and Safety Monitoring Board." *Clinical Trials* 4:218-234.

Woodward, J. 2005. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.

Appendix: Sketch model of DSMB decision-making²⁷

According to the sketched model, the dual risks of DSMB decision-making should refer to typical factors appealed to by DSMBs. These factors are contextual. **Table 1** below displays the link between the type of early stopping and general DSMB decision, as well as a list of typical factors corresponding to each general decision.

Factors are contextualized to the circumstances of the case. This contextualization permits a link to what might be deemed proper decision criteria. For instance, the degree of conservatism of the decision criterion should correlate positively with the availability of affordable, safe and effective alternatives to the new treatment under evaluation. The urgency of the situation should justify the stringency of the decision criterion. The more effective and safe the available treatments, the weaker the patients' need for a new drug, the less urgent the public interest in alternatives, and consequently, the more conservative the decision criterion.

By way of quick illustration: in the mid-1980's when HIV/AIDS was fatal and well-tested, safe, and effective treatments were not available, less conservative criteria were justified. This contrasts with the situation of the second half of the mid 1990's when drug cocktails containing protease inhibitors were relatively effective and available. The situations before and after the availability of effective drugs for treating HIV/AIDS warrant different decision criteria—with different levels of conservatism.

<i>Type of early stop decision</i>	<i>Interim</i>	<i>Dilemma</i>	<i>Contextual factors</i>	<i>Desideratum</i>	<i>Standard of proof as conditions for early stopping</i>
For efficacy	Statistically significant	<i>How much longer should the trial continue given that early benefit is observed?</i>	Patient health status {healthy, unhealthy}	Weighing early favorable effect vs. uncertainty of late unfavorable effects	If unhealthy, stat. significance is (often) sufficient but not necessary. If healthy, stat. significance is (often) necessary but not sufficient.
			Treatment alternative {available, unavailable}	Conservatism of decision criterion (greater availability, greater stringency)	If unavailable, stat. significance is (often) sufficient but not necessary. If available, stat. significance is (often) necessary but not sufficient.
For harm	Not statistically significant	<i>How much longer should the trial continue, so that evidence for harm sufficiently rules out the possibility of a spurious effect?</i>	Type of trial {placebo-control, active-control}	Minimal clinical importance (<i>mci</i>)	If placebo-control, then if <i>mci</i> is unlikely, stop trial. If active-control, then if <i>mci</i> is unlikely, then consider treatment current usage.
			Treatment current usage {widespread, not in use}	Secondary outcomes	If widespread use, then if no secondary benefits, stop trial.

Table 1: Summary of the main relevant factors, desiderata, standard of proof, and respective early stopping conditions for the two types of DSMB decisions

¹ As expressed by Wittes (1993):

Sometimes a few written words pass from the DSMB to the outside world; a sanitized version of the minutes may become public or a report of a specific problem may find its way to the literature, but the animated discussions, the wrangling, the agonizing over ambiguities in the ongoing data are lost. Hours of argument about the propriety of continuing a trial become a sentence in the minutes, “The Data Safety Monitoring Board unanimously recommended continuation of the trial.” (1993, 420)

² Paternalistic considerations are the prevalent ones. The main idea is that by having the DSMB to deliberate secretly, it guards against erroneous conclusions about trial results and potentially premature adoption of study findings. A premise that begs the question in DSMB confidentiality arguments is the assumption that medical practitioners, trial participants, and other laypersons may over-react to the interim data if the data is made publicly available, unless they have been properly trained in the science of distinguishing the statistically probable from the merely possible.

³ Whether modelers are working on the burgeoning field of clinical decision support systems (cf. Berner 2007) or the more recent domain of evidence-based medicine, central to much of the work are views regarding the relationship between past medical cases, regulatory rules, and duties. A similar predicament is found in the modeling of legal reasoning. A retrospective look at 25 years of international conferences on AI and law gives an idea of the variety of models of binding precedents, as well as statutory and regulatory laws. Like the modeling of medical reasoning, the common assumption is that legal reasoning is far from arbitrary, working instead to promote a mix of values in the system, e.g. stability and innovation of the law. Prominent examples of such work are Ashley (1991), Branting (2000), and Bench-Capon & Sartor (2003).

⁴ Stanev (2011) is an example of modeling early stop due to efficacy decisions.

⁵ Stanev (2012a) is an example of modeling early stop due to harm decisions.

⁶ Salbu (1999) presents a similar characterization of the “fundamental dilemma” that the FDA faces. When looking at the conflicting objectives of caution and expedience that confront the Agency during each new drug application, Salbu explains how while fulfilling its mission to monitor and control the safety and efficacy of drugs, the federal agency is concerned with “the unavoidable tradeoffs between under-conservatism and over-conservatism in the drug approval process.” (97) The predicament is characterized as “the FDA’s fundamental dilemma.” With some qualifications, much of the FDA’s dilemma is also applicable to the central dilemma that DSMBs have to face when monitoring clinical trials and have to decide on whether to stop, modify, or continue them.

⁷ Examples of such models are abductive, predictive, probabilistic, and decision-theoretic associations.

⁸ Bartha, following Salmon, identifies two distinct conceptions of plausibility: probabilistic (with degrees of plausibility) and modal (*prima facie* plausibility). In this paper, the modal conception suffices. It suffices because to say “it is plausible that Q,” means, roughly speaking, “there are sufficient grounds for taking Q seriously”, which suffices for my purposes.

⁹ In his reasons for the ruling judge Blair J.A. citing pertinent judicial decisions dismisses the appeal and cite this passage from *McCormick on Evidence* (1984): “Family harmony is nearly always past saving when the spouse is willing to aid the prosecution. The privilege is an archaic survival of mystical religious dogma and a way of thinking about the marital relation that is today outmoded.” (p. 354)

¹⁰ “Where the principles underlying a common law rule are out of step with Charter values, the courts should scrutinize the rule closely. If it is possible to change the rule so as to make it consistent with such values, without upsetting the proper balance between judicial and legislative action, the rule ought to be changed” (*R. v. Salituro* [1991] 3 SCR 654).

In scrutinizing the background justification for the spousal incompetence rule, SCC judges reasoned that in its origins the rule “followed naturally” from the legal position of the wife at the time, reflecting a view of the role of women which is now archaic and no longer compatible with the gender equality value enshrined in the Charter.

¹¹ Another example that serves to illustrate the intuition about how legal rules can be over-inclusive is this simplified example borrowed from Schauer (2009), i.e., speed limit of 55 miles per hour. Suppose that the law says that a driver should not drive at speed greater than 55 miles per hour. Why? Because driving at

a speed greater than 55 miles per hour is unsafe. But even if it is true in most instances—driving at a speed greater than 55 mph is unsafe—consider someone driving safely at 70 miles per hour in ideal traffic conditions. In this case, the background justification does not apply because the driving is not unsafe. In situations like this, the reach of the ‘no driving at speeds more than 55 miles per hour’ rule is broader than the reach of the rule’s background justification, and so we say that the rule is over-inclusive in this instance. Important caveats are in order with this example. Speed limit laws are not judge-made laws and thus are not governed by the process of *stare decisis*—i.e., not requiring a public rationale. The process of deciding whether a 55 mile per hour speed limit becomes the law is made through the legislative process rather than judge-made law. Moreover, in terms of enforcing the speed limit rule, unlike judicial cases that are fact-sensitive, the rule 55 mile per hour speed limit is unambiguous. The reason for the driver to exceed that limit is generally not relevant. Granted, police officers are generally permitted some discretion in enforcement, but that is quite unlike the requirements of case law where the rationale is given and published—and must comport with prior opinions. This way, the 55 mile per hour speed limit rule only serves to illustrate the *intuition* that applying the rule in this instance over-reaches its justification that it is unsafe to violate it. It does not serve as an analogy of judge-made law changes. I am thankful for one of reviewers for pointing out this important difference in the analogy.

¹² See the official summary of the Iowa Supreme Court ruling in *Varnum v. Brien*, 763 NW 2d 862 (2009).

¹³ Another simplified example that serves to illustrate the intuition is no-vehicles-in-the-park rule. If a ‘no vehicles in the park’ rule is meant to prevent noise, it will be under-inclusive with respect to things other than ‘vehicles’ in the park that are noisy, such as loud portable radios—or an ambulance which might fall within the literal scope of the ‘no vehicles in the park’ rule, yet the vehicle is justified in the park.

¹⁴ In common law there is a practice whereby certain courts are given a limited but an important power to deny earlier decision of their binding status on the basis that the earlier decision was wrongly decided. Appeal courts exemplify having such power. Supreme Court of Canada is another example. It is not bound to follow decisions of appeal courts and therefore free to overrule such decisions if it takes a different view of how the case should have been decided.

¹⁵ See Fisherian examples in clinical trials, e.g. section 5.6 of U.K.’s Clinical Trial and Services Unit in <http://www.ctsu.ox.ac.uk/reports/ebctcg-1990/section5> (accessed September 9, 2015).

¹⁶ The Coronary Drug Project Research Group (1981) “Practical aspects of decision making in clinical trials: the coronary drug project as a case study.” *Controlled Clinical Trials* 1:363-376.

¹⁷ Coronary Drug Project Research Group: “Clofibrate and niacin in coronary heart disease.” *JAMA* 231:360-381, 1975.

¹⁸ Conversely, instances in which safety demands extreme significance values (e.g. p-value<0.001) the generalization that ‘stopping when p-value is less than 0.05’ is safe, can be under-inclusive.

¹⁹ In the Concept of Law, Hart makes this point that because the law at times lacks specificity, judges though courts not only can but should have the power to revise and improve the law.

²⁰ From *Halsbury’s Laws of England*:

The enunciation of the reason or principle upon which a question before a court has been decided is alone binding as a precedent. This underlying principle is called the *ratio decidendi* (1979, p. 292).

²¹ One proposal for distinguishing the *ratio decidendi* from *obiter dicta* of a case is to use a counter-factual test. That is, determining the ratio is to identify that set of proposition in the judgment which, were its meaning to be inverted, would have altered the judge’s decision. Counter-factual explanations are common views among philosophers, particularly philosophers of science. See Woodward 2008 for such an account.

²² In scientific terms, type I error or “false positive” is the incorrect rejection of a true (often the null) hypothesis, whereas type II error or “false negative” is the failure to reject a false hypothesis. Power here is a statistical measure, meaning 1 – (probability of type II error). The control of such errors is an essential requirement for the scientific acceptability of any trial.

²³ Laudan (2006) points out a related logical similarity between legal decision-making and the scientific requirements of clinical trials that is somewhat pertinent here. I say somewhat because Laudan is referring to criminal trials not higher court trials. That is, law and clinical trials share a logical relation in their *standards of proof*, with the following addition: it is also difficult to reject (disprove) the default

assumption in a trial. In clinical trials, statistical monitoring rules are implemented with the null-hypothesis as the default assumption, namely, that the new intervention does no better than control (e.g., placebo). In criminal trials the default assumption is the claim that the accused is innocent. The similar functional element is an epistemic one: just as in clinical trials, the failure to reject the hypothesis (i.e., type II error, a false negative) that ‘the intervention does no better than placebo,’ is not a proof that the intervention is no better than placebo, so the failure to disprove the accused is innocent is not proof that the accused did not commit the crime. Conversely, erroneously rejecting the null-hypothesis (i.e., type I error, a false positive) is no proof that the intervention does better than placebo, so falsely disproving the accused is innocent, is no proof that the accused committed crime.

²⁴ According to the U.S. FDA *Guideline for the Format and Content of the Clinical and Statistical Sections of New Drug Applications*, the following paragraph on the full disclosure of DSMB decision-making is mentioned:

The process of examining and analyzing data accumulating in a clinical trial, either formally or informally, can introduce bias. Therefore, all interim analyses, formal or informal, by any study participant, sponsor staff member, or DSMB should be described in full, even if the treatment groups were not identified. The need for statistical adjustments because of such analyses should be addressed. Minutes of meetings of a DSMB may be useful and may be requested by the FDA.

²⁵ The public articulation of the reason for the decision is also present in civil law. Civil law imposes a statutory requirement to provide reason(s) for the judgment. This is certainly true of the Canadian legal system.

²⁶ According to a recent report from the U.S. Department of Health and Human Services “too many researchers are not adhering to standards of good clinical practice. The FDA has identified cases in which researchers failed to disqualify subjects who did not meet the criteria for a study, failed to report adverse events as required, failed to ensure that a protocol was followed, and failed to ensure that study staff had adequate training. These were not isolated incidents on the fringes of science. Instead, these troubling problems occurred at some of our most prestigious research centers and involved leaders in their fields of study. There can be no shortcuts when it comes to the protection of human subjects. Good clinical practices are neither esoteric nor frivolous. When adverse events are properly reported, the FDA and other bodies charged with the oversight of research can assess the safety of a particular study, as well as similar studies, and look for trends. In that way, subjects are better protected.” (“Protecting research subjects—what must be done” *New England Journal of Medicine*, vol.343, number 11, p.809)

²⁷ See Stanev (2012b) for a presentation of the model in greater details.