An Argument for Fewer Clinical Trials

BY KIRSTIN BORGERSON

The quantity of published clinical research, much of it of poor quality, is out of hand. Physicians cannot keep up with it, and patients are likely suffering from it. Research ethics committees need to rein it in, and they can do so by drawing on the principle that clinical research must be justified by its social value.

The volume of clinical research is increasing exponentially—far beyond our ability to process and absorb the results. Given this situation, it may be beneficial to consider reducing the flow at its source. In what follows, I will motivate and critically evaluate the following proposal: researchers should conduct fewer clinical trials. More specifically, I consider whether researchers should be permitted to conduct only clinical research of very high quality and, in turn, whether research ethics committees (RECs) should prohibit all other, lower-quality research, even when it might appear to meet some minimal ethical standard.[1] Following a close analysis of the social-value requirement of ethical clinical research, I argue that this proposal is defensible.

The Sorting Problem

Quantity. The problem identified in this paper has two parts. The first part has received considerable attention lately, with critics remarking, "Every day there are now 11 systematic reviews and 75 trials, and there are no signs of this slowing down: but there are still only 24 hours in a day,"[2] or, more directly, "[T]he quantity of new data exceeds the field's ability to process it appropriately."[3] According to an article recently published in *BMJ*, medical research output doubles every seven years.[4] An analysis by Ming-yueh Tsay and Yen-hsu Yang indicated that the publication rate of randomized control trials (RCTs) has not only grown since 1965 but grown exponentially.[5] And there is no plateau in sight.[6] Tsay and Yang identified 4,600 medical journals worldwide, as of 2002, and this number is also increasing. In another study from 2005, An-Wen Chan and Douglas Altman estimated the number of human subjects in clinical research at more than two million a year.[7] Approximately one million papers from clinical trials have been published to date.[8] And publications are longer, more detailed, and contain considerably more data than in the past.[9] Dramatic statements about the "flood" of biomedical research data seem entirely reasonable given these figures.

As Tammy Hoffmann and colleagues point out, physicians see the explosion of research as both a blessing and a curse: the potential benefits are "inhibited by the information overload experienced by clinicians struggling to keep abreast of new

Kirstin Borgerson, "An Argument for Fewer Clinical Trials," *Hastings Center Report* 46 (2016): 25-35. DOI: 10.1002/hast.637

research."[10] The authors point out that while the pace of research has accelerated, the time available to read the research has not seen a similar increase. Finding the information scattered across a variety of journals adds to the challenge: "To find even half of the papers published in one year . . . a clinician would need to read an impracticable number of journals . . . for example, an estimated 39 journals for randomised trials on diabetes and 23 journals for systematic reviews on myocardial infarction."[11] Information overload involves being "swamped with information, but starved of data."[12]

Most physicians won't need to stay on top of all of the medical research literature, just the literature in their area of specialization. Unfortunately, this doesn't help as much as one might hope in dealing with the problem. In 1995, the estimated number of articles a physician would have to read to stay on top of his or her field was seventeen per day.[13] A 2004 study estimated the volume of literature relevant to one area of medicine (primary care) at 7,287 articles per month. Physicians trained in the methods of clinical epidemiology would require approximately twenty-nine hours per workday to evaluate the evidence on their own.[14] Given what we know about the increasing rate of publication, these figures would almost certainly be an underestimate of the time required today.

As one would expect, these numbers map poorly onto the time available to health care professionals for reading and critical analysis of the literature. One study, which surveyed the reading habits of physicians working in internal medicine, found that internists reported spending 4.4 hours per week reading articles in medical journals.[15] One survey respondent said, "It is unrealistic to expect that, even if you have the skills, you will have time to critically review all the literature that is out there."[16] The authors of the study indicate that internists face intense pressure both to stay on top of the literature and to attend to increasing patient demands: "It is unlikely, therefore, that the amount of time physicians devote to continuing medical education will keep pace with the rate at which medical knowledge is growing."[17]

Quality. The exponential growth of medical research may not be, in and of itself, an insurmountable problem. If every study was of the highest quality, there may be some hope that systems of knowledge synthesis could be developed to process all the data (an approach discussed below). Unfortunately, quantity is not matched by quality in the literature. In fact, the quality of much of the clinical research is judged to be poor.[18] This is a widely recognized problem in the medical literature, with deep historical roots. Before James Lind published his review of the medical literature on scurvy in 1753, he wrote that "it was necessary to remove a great deal of rubbish."[19] In 1994, Douglas Altman described the problem in stark terms: "[H]uge sums of money are spent annually on research that is seriously flawed through the use of inappropriate designs, unrepresentative samples, small samples, incorrect methods of analysis, and faulty interpretation."[20] Altman didn't hold back in his assessment of the situation, stating unequivocally, "We should be appalled. . . . This is surely a scandal."[21] Hilda Bastian, Paul Glasziou, and Iain Chalmers write eloquently of the modern version of this challenge: "[T]he problem of having to trawl through and sift vast amounts of data has grown . . . [M]ountains of unsynthesized research evidence accumulate."[22] They cite the "overload of unfiltered information" as one of problems lingering (and worsening) in recent years.[23]

Some clarification is needed at this point, because "quality" is a highly contested term in the medical literature. When some scholars advocate for high-quality trials, they mean large-scale, simple, explanatory RCTs. Others, including myself, have defended a different characterization of high-quality research that tends more toward pragmatic trial design and the use of methods other than RCTs. Pragmatic trials aim to provide evidence that directly supports clinical decision-making in "usual" care settings. Unlike explanatory trials, which aim to abstract away from particular settings and patients, in the hopes of creating ideal conditions for the success of an intervention, pragmatic trials deliberately pursue knowledge of high applicability, through the use of representative subjects, clinically important questions, flexible treatment protocols, patient-oriented outcome measures, and so on.

One way of trying to achieve more pragmatic trials is to use a tool such as the pragmatic explanatory continuum indicator summary (PRECIS), which is intended to evaluate the fit between research intention and design. PRECIS was originally proposed by a group of international clinical researchers in 2009 and has been modified and adapted in the years since. In a 2015 paper, the group laid out the improved and validated instrument, PRECIS-2, which provides nine domains on which to assess trial design: eligibility criteria, recruitment, setting, organization, flexibility (delivery), flexibility (adherence), follow-up, primary outcome, and primary analysis.[24] In general, the more these design elements match usual care, the more pragmatic the trial. Although the designers of this tool leave it open as to how it might be used by researchers—a particular researcher might, for instance, use it to design a maximally explanatory trial—I have argued elsewhere in favor of using tools such as this to design more pragmatic trials.[25] There is growing support for this position, for instance, in trends toward comparative effectiveness and translational research, research-practice integration, and quality-improvement studies.

The shift to pragmatic trials may seem at first like an extension of efforts to include particular underrepresented groups in research, as occurred in the 1990s for women and as seems

to be occurring today for pregnant women. While it is consistent in some ways with these developments, there are at least two key differences worth noting. First, the eligibility criteria for enrollment in clinical research constitute only one of nine factors that make a trial more (or less) pragmatic. A PRECIS-2–based approach goes far beyond narrow representation of particular groups in making trials more generally representative (although it should improve representation in this narrow sense). Concern extends to the setting of care, the flexibility of treatment regimens, and so on, as described above.

Second, the approach I favor shifts the burden of justification to researchers who want to add idealizations of any kind to their trial designs. I have argued that the context in which research is meant to be applied should be the context in which new interventions are evaluated. In other words, I see applicability as a marker of high-quality research. As Kirsty Loudon and colleagues put it, "Applicability (the ability for a trial result to be applied or used in a particular situation) is the outcome of these choices, which affect the ease with which the trial results can be applied to and by the usual community of users of the intervention in the settings in which the trial designers envisioned it being used."[26] I do not argue that certain groups need to be "added" to trials or targeted for inclusion, which would require that we name and identify these groups. Steven Epstein's excellent book, *Inclusion: The Politics of Difference in Medical Research*, has carefully documented the challenges of this sort of "niche-standardization."[27] Rather, I suggest that the default should be the standard-care context and all exclusions and idealizations should be carefully justified across all nine domains above. This understanding of quality shapes some of what follows, in that I regard trials with limited applicability as lower quality, but I have also attempted to highlight elements of quality that defenders of explanatory trials can support, such as adequate sample size, trial designs appropriate to the question being asked, appropriate methods of analysis, minimal bias, and so on. I will also return to discussions of quality at the end of the paper.

Returning to the current situation, then, one undisputed indication of the low quality of research is the relatively small sample size of many RCTs. Small trials may be appropriate in some cases—early-phase research or research on rare diseases, for instance—but are generally thought to be undesirable once researchers are investigating effectiveness (and comparative effectiveness) of new treatments in phase III trials. Robert Califf and colleagues found that, between 2007 and 2010, 62 percent of trials registered on ClinicalTrials.gov enrolled one hundred or fewer participants;[28] 50 percent enrolled fewer than seventy participants. Many of the trials were originally designed to enroll more participants but fell short at recruitment, leading the authors to speculate that there may be a widespread problem with underpowered trials. In addition to these problems, many trials failed to report details of randomization, blinding, and use of data-monitoring committees.

Another way in which the quality of research is compromised is the significant presence of biases in research. At this point, the list of known tactics used to manipulate clinical trials would number in the dozens and include extensive and unjustified use of exclusion criteria (about 40 percent of potential subjects were excluded from the 280 most influential trials from 2002 to 2010), suboptimal dosing of comparison treatments, and manipulated data analysis.[29] Many of these tactics are prevalent in the medical literature, and an extensive body of research tracks their widespread use.[30] Any measure of quality that takes objectivity seriously will find fault with many contemporary clinical trials.

In a 2012 *Nature Medicine* article, Peter Humaidan and Nikolaos Polyzos report that the problem extends to meta-analyses. Because meta-analyses are, among other things, increasingly seen as providing "an easy way to get published," the number of such publications has increased fivefold (from 849 in 2000 to 4,720 in 2011).[31] Many of these meta-analyses are unnecessary (reporting no new research since previous meta-analyses or based on few or no studies in the area) or are of poor general quality (for instance, because they were performed on the basis of only a few small trials). And guidelines do not fare any better. Two recent studies indicate that guideline recommendations in clinical practice are based on high-quality evidence less than 15 percent of the time.[32] The pace of research is so quick that some commentators wonder whether it will ever be possible for critical syntheses of the research to catch up. Bastian, Glasziou, and Chalmers express this skeptical view: "There is nevertheless a risk that the increasing burden placed on the methods of systematic reviewing could make the goal of keeping up-to-date with the knowledge won from trials recede ever more quickly into the distance."[33]

To sum up, whether one's measure of research quality tracks applicability, as mine does, or sticks with more widely shared markers, such as objectivity and sample size, we can agree that there is much low-quality research being produced. The problem faced by those wanting to use the results of research in practice today

The problem faced by those wanting to use the results of research in practice today is one of sorting good evidence from bad evidence and, given the overload of research studies, doing so efficiently.

is one of sorting good evidence from bad evidence and, given the overload of research studies, doing so efficiently. I will refer to this as the "sorting problem."

The Harm

It seems likely that the overproduction of low-quality research would be harmful to patients. The harms of research may be either direct or indirect. Direct harms are those experienced by research subjects as a result of their participation in research, for example, an infection acquired as a result of IV placement. Indirect harms are those experienced by future patients whose care is affected by the results of that research.[34] Those harms may be experienced differently by members of different groups, for instance, vulnerable groups, because harms can have a multiplicative effect, and any harm calculation would have to be responsive to these differences. Similarly, direct benefits are those that affect the health and well-being of individual research subjects, while indirect benefits accrue to future patients as a result of greater knowledge of "what works" in medicine.

Of course, tracking the harm arising as a result of the sorting problem is extremely difficult. There may be historically controlled or observational studies capable of doing so, but I have yet to encounter them.[35] In the absence of clear data, we can make some progress by identifying harms that could be reasonably anticipated.

Perhaps most obviously, there are direct harms associated with participation in low-quality trials, including any of the net risks to subjects of research arising from blood draws and other monitoring tests, as well as inconveniences and wasted time, which are not balanced by benefits to society. Research subjects may also be deprived of the standard of care and thus be harmed by being deprived of an effective treatment.

In clinical care, reasonably anticipated indirect harms would also include the continued prescription of drugs shown to be inferior to others in published but not yet recognized trials (in other words, "suboptimal prescribing"). Positive results in trials of new treatments have been shown to influence practice only very slowly, and sometimes negligibly. Lack of awareness of the results is one of the reasons offered in explanation for this phenomenon.[36] This lack of awareness can easily arise when high-quality trials must first be sorted and identified among the vast number of low-quality trials. Further, the astonishingly high rates of adverse events in clinical care may be partly a result of a mismatch between the research evidence produced and the realities of clinical care. When physicians don't have the evidence they need to make individualized patient-care decisions—because, for instance, their patients don't resemble those on whom the therapies were tested—those patients may be harmed.

Similarly, indirect harms could arise when new evidence emerges about the side effects of a popular drug but this evidence takes years to reach physicians and affect prescribing habits, subjecting patients to significant harms in the meantime. We might use the Vioxx scandal as an example.[37] Of course, in the Vioxx case, the delay was the result of a decision to withhold information from the public, whereas the time lag we are imagining would occur as a result of the time needed to perform an adequate research synthesis. But the case does help us to see the harms possible with delayed action: it was estimated that tens of thousands of patients were harmed by the delay in releasing data about the harms of the drug. When drugs are prescribed to millions of patients, as many blockbuster drugs are today, any delay in changing prescriptions in light of dangerous side effects can lead to tremendous harm to patients. These harms often have a ripple effect on the families and communities of the people affected.

Further harms could reasonably be expected to arise from physicians acting on the basis of heavily biased research results and industry-sponsored meta-analyses designed to ignore or exclude evidence not in support of a drug. It isn't always clear which evidence synthesis to trust, and there are many such syntheses competing for physicians' allegiance. Trusting the wrong source can be harmful to patients, who once again are exposed to treatments that are not best for their conditions. But when the research is too overwhelming to assess as an individual with far fewer than twenty-nine hours per day to devote to the task, someone has to be trusted to perform independent critical analyses. The frustration and uncertainty generated by this situation are also likely to have harmful effects on physicians.

There may also be more abstract harms associated with physicians' failure to discharge their professional responsibility to provide evidence-based care. These include moral harms to physicians themselves, who face a duty to provide treatments supported by the best available evidence but are impeded in discharging this duty by the overproduction of variable-quality research. This failure may also have implications for the trust-based physician-patient relationship and, by extension, the well-being of patients. If a patient stumbles on a recent excellent study on a new treatment for her condition and brings it to the clinical encounter, the physician's failure to stay up to date may well come to light. Perhaps this can be managed judiciously, but it could also lead to a (justified!) lack of confidence in the physician, with predictable depreciation in trust.

A related systemic problem is that poor-quality research can damage the social trust required by the research enterprise as a whole.[38] This can make recruitment of research subjects even more difficult and lead to even more underpowered studies. In sum, there are many possible, and even likely, harms of overproducing low-quality research evidence.

The Evidence Synthesis Solution and Its Shortcomings

The problem outlined above has attracted considerable attention in the medical literature. Almost all commentators identify some range of synthesis solutions as if this exhausts the options available for addressing this problem.[39] I will provide a brief description of one such representative solution and summarize the problems identified with the proposal. Because my interest is not in replicating this debate over knowledge synthesis but in advancing a new solution to the problem it was designed to address, this section will be relatively brief.

The evolution of what I will refer to as the "S hierarchy" approach—"4S" (from 2001), "5S" (2006), and "6S" (2009)—to evidence-based clinical decision-making effectively illustrates this shift to knowledge synthesis.[40] The original hierarchy proposed by evidence-based medicine (EBM) identified small-scale and observational trials as lower-quality methodologies and prized randomized controlled trials and meta-analyses of RCTs as the most reliable sources of research evidence. Each addition to the original hierarchy by the S hierarchy movement (syntheses, synopses [of studies and of syntheses], summaries, and systems) was proposed in order to make clinical research evidence more digestible and easier to use. Developers of the approach seem interested in getting to a point where all of the information needed by clinical decision-makers would be contained in the titles of short synopses or summaries.[41] This is done with an awareness of the information overload present in the literature and an appreciation for the relatively limited time health care professionals have to search through and critically appraise research data.

The S hierarchy approach has been criticized on a number of fronts, including its continuity with the original—now widely discredited—EBM hierarchy of evidence; overreliance on computer systems without appreciation for the necessary limits of such systems; undue conservatism about what counts as evidence (the exclusion of observational research, case studies, and other nonrandomized research from higher-level synopses); the division of communities into separate groups of synthesizers, researchers, and practicing physicians, resulting in decreased communication across these groups; contamination of the evidence by an endless supply of loopholes in trial methodologies; de-emphasis on critical thinking; cultivating a false sense of certainty (and the challenges this creates for full informed consent); and a lack of appreciation for detail and context.[42]

This last criticism has been bolstered by empirical evidence from Paul Glasziou and colleagues, suggesting that syntheses of clinical research rarely contain the sorts of information required in clinical decision-making.[43] Syntheses may include reference to "behavioral interventions," "salt reduction," or an "exercise program" without providing any further information about the intervention, and while at the same time obscuring differences in the terms as used by the trials being analyzed. The challenges of implementing any such intervention, however positive the results, should be immediately obvious. For evidence on pharmaceutical treatments to be usable by physicians, a description should include "the dose, titration, route, timing, duration, and any monitoring used."[44] Yet many syntheses fail to provide precisely these pieces of information. Systematic reviews fared especially poorly when it came to providing details essential to the implementation of research results. The authors acknowledge that "[i]n systematic reviews, the high level of abstraction used in selecting 'similar' treatments causes a problem."[45] Abstraction may make it easier to access research results, but this is often at the cost of making the results far more difficult, if not impossible, to implement in practice.

This study provides support for my claim that the problems of evidence synthesis are not easily or decisively remedied, since it highlights the tension between any attempt to provide clear, decisive, simple answers to clinical questions and the complex and detailed evidence required from physicians on, for instance, variable dosage, the time required for a drug to take effect, interactions between drugs, side effects, and so on. With respect to the evidence synthesis services available in 2012, Hoffman and colleagues conclude that "few current systems seem adequate."[46] Perhaps it is time to consider other options, however unattractive they may appear at first. Of course, these solutions aren't exclusive: we can continue to try to improve our system of knowledge synthesis (if we think there is value to this enterprise) while exploring other possibilities. If the synthesis solution is less than complete, as it surely is, we have reason to look for other solutions to the problem (particularly when they are complementary).

Poor-quality research can damage the social trust required by the research enterprise as a whole, leading to even more underpowered studies.

In spite of extensive commentary on how best to synthesize knowledge and translate it to practicing clinicians, there is near silence on the alternative considered below. The only researchers bold enough to mention, if briefly, the possibility of conducting fewer research trials write, "First, we need to prioritise effectively and reduce avoidable waste in the production and reporting of research evidence."[47] Of course, they are careful to say, "although funding for evaluative clinical research internationally

remains a priority," before offering any such suggestions.[48] I will proceed with a related disclaimer: my account is neutral on matters of funding. Fewer large-scale trials, if that is what is required, may well turn out to be the same cost as a series of smaller trials. This is an empirical matter, and it turns on the debates over quality in clinical research.

Moreover, my suggestion that we ought to pursue fewer clinical trials should be understood as defending the position that many of the low-quality trials being proposed right now should not be conducted. In other words, if we take a snapshot of the trials being reviewed by RECs today, around the world, I am arguing that fewer of them should be approved than are currently being approved. If, over time, researchers adjusted to higher standards by turning all low-quality trials into high-quality trials, it is possible we would not end up with fewer trials in the long run; however, I think there are efficiency-related reasons (discussed below) to doubt that this would be a likely outcome. In any case, since it is probably best to avoid speculating about the future, I will state as clearly as I can: I argue for fewer trials relative to the status quo in research.

In sum, then, given the overload of low-quality biomedical research data published every day and the fact that there are serious problems with knowledge synthesis as a full solution to this predicament, how might research be responsibly pursued?

The Social-Value Solution

Clinical trials should be socially valuable. In the list of seven ethical requirements of clinical research identified by Ezekiel Emanuel, David Wendler, and Christine Grady, social or scientific value is first.[49] The most recent draft of the CIOMS guidelines mentions social value in its opening sentence: "The ethical justification of health-related research is its social value: the prospect of generating the knowledge and/or the means necessary to protect and promote people's health."[50] What does it mean for research to be socially valuable? Among other things, "[t]he social value of this research is ultimately grounded in the quality of the information that it produces."[51] This is consistent with the discussion above, in which I suggested that one key marker of the quality of clinical research is the relevance or applicability of results. The beneficiaries of clinical research are future patients within the same health care context.[52] In what follows I take a closer look at the social-value requirement, consider why it has not prevented the sorting problem outlined above, and evaluate whether it might play a more active role in restricting the overflow of low-quality research.

There are two standard arguments supporting the social-value requirement. These are the "responsible use of finite resources" or efficiency argument and the avoidance of exploitation argument.[53] In a recent paper, Alan Wertheimer devoted considerable effort to debunking the exploitation argument, drawing on his own account of mutually advantageous exploitation to make the case that social value is not as robust or universal a requirement as it might seem.[54] Though I think that alternative accounts of exploitation would likely provide a better defense for the social-value requirement, for the purposes of this paper I am more interested in examining a particular version of the efficiency argument, as well as drawing attention to (and correcting) an oddly blinkered, or one-sided, reading of the requirement within the bioethics literature.[55]

One reason social value is identified as an ethical requirement of research is because clinical trials should fairly and efficiently evaluate new medical interventions. The social-value requirement connects the conduct of research to the ultimate goal of clinical research: better health for humans. It does so by requiring that research aim to benefit society, in the long run. It can do so by, among other things, ensuring that finite social resources are used responsibly in pursuit of that goal. One of the resources shared by all clinical research, whether publicly or privately funded, is human subjects. And, as the frequency of trial delays at recruitment suggests, this resource is in short supply. Given how few people are both willing and able to participate in (often quite burdensome) clinical research, it is not surprising that there are tremendous challenges with recruitment.[56] The social-value requirement exists partly to ensure that we are doing the best we can with the limited collective resources at our disposal, given our shared interest in advancing human health.[57] When collective resources are limited, some form of principled rationing is surely defensible. The social-value requirement can serve as a kind of rationing tool. This is one of many topics within bioethics that would benefit from closer attention from political philosophers.

An even more problematic feature of the requirement as it is understood today concerns the risk-benefit calculation. For most proposed clinical trials, the risks to subjects outweigh the benefits to those same subjects. This shouldn't be surprising: the enterprise of clinical research involves venturing into the unknown; in this sort of situation all kinds of risks will present themselves. Researchers try to contain those risks and learn from what doesn't work by systematically tracking what takes place over the course of research and comparing outcomes across groups, but the context of research does not, on its own, eliminate risks. The residual risks to subjects, then, demand attention. The social-value requirement exists in part to enable us to find a measure of benefit that will balance out the risks taken by human research subjects. This has led to a focus on the potential positive effects of prospective trials. After all, the task of evaluating the ethics of a proposed trial appears to require that researchers seek out potential social benefits in order to deal with the

pesky problem of residual risk (even when risks are minimized).

There seems to be some general support for the view that it is acceptable to include only the social effects that "do the work" we want them to do in balancing risks of harm to individuals (in other words, social benefits).[58] An otherwise excellent forty-page paper on risk-benefit calculations, for instance, relies throughout on discussion of social benefit: "Whether a given level of research risk is excessive depends on the magnitude of the risks, the level of corresponding potential benefits for the participant, if any, *and the level of potential social benefit from performing the intervention and the study*" (emphasis added).[59] But this common position is transparently unprincipled. After all, we don't generally make important decisions by considering only the benefits of particular choices while ignoring any harm.[60] The requirement is an assessment of the importance of research—of *social value*, not social benefits. Full consideration of social value includes both social harms and social benefits. Benjamin Freedman's classic analysis of social value makes this point: "Exogenous factors may be relevant . . . . On [the] positive side, high priority research; on [the] negative side, potential for abuse of knowledge gained."[61] There will be a measure of speculation required, unquestionably, in assessing social harm, as there always is in assessing social benefit, but refusing to venture an estimate of effects on grounds of difficulty just sidesteps the issue, and negligently so. So, whatever argument we offer in support of the social-value requirement, it should support the assessment of both potential benefits and risks of harm, for any investigation.

If indirect harms have a place in the social-value assessment performed by researchers and RECs, as I argue they do, then the harms outlined above should tip the balance of favor against otherwise marginally acceptable clinical trials. In effect, this means that the social-value requirement will be harder to meet. Researchers will need to work to ensure that the benefits brought about by a particular trial overcome the higher risk and harm threshold arising out of a more balanced calculation. The current proliferation of low-quality trials would be less rampant according to this more complete understanding of the social-value requirement. At the very least, if researchers are pressed by RECs to explain why their trials aren't better than they are—aren't perhaps even the best they could be under the circumstances—this might be a much-needed nudge. If this is paired with a robust understanding of research quality, of the sort outlined earlier, neither researchers nor RECs will have an excuse for inaction.

The idea that it is acceptable to include only the social effects that "do the work" we want them to do in balancing risks of harm to individuals seems to be commonly held. But this is transparently unprincipled.

Support for the Social-Value Solution

The status quo in research ethics is the view that researchers and research ethics committees need not consider—and are perhaps even prohibited from considering—the potential indirect harms of research. In what follows, I demonstrate that, contrary to popular interpretations of current regulations, the Canadian Tri-Council Policy Statement and the American Common Rule permit, and in the Canadian case may even be said to encourage, consideration of social harms in the ethical assessment of research. In what follows, I outline my interpretation of these guidance documents in the hopes of convincing skeptics that the position I have defended is more plausible than it might seem at first. Of course, even if all regulatory documents took a different position on the matter, I would argue that my argument as presented above is still sound.

Canada: Tri-Council Policy Statement. Two passages in the most recent version of the TCPS (TCPS-2) provide general guidance on social-value assessments. The first provides support for my suggestion that social value—which adds benefit to the risk-benefit calculation—be enhanced where possible for each proposed trial: "Researchers and REBs must attempt to minimize the risks associated with answering any given research question. They should attempt to achieve the *most favourable balance of risks and potential benefits* in a research proposal" (emphasis added).[62] I take this to support my proposal (and perhaps even take it one step further).[63] A second passage makes it clear that the welfare of society—where that includes assessment of social benefit *and social harm*—is of considerable importance in the design of research: "Groups may benefit from the knowledge gained from research, but they may also suffer from stigmatization, discrimination or damage to reputation."[64] In other words, social harms matter to the ethical assessment of research. Together, these passages support my proposal as outlined above. In fact, as with the common rule, I take it to be the case that the TCPS already endorses my position, and members of ethics committees and researchers who fail to consider indirect harms are failing in their existing ethical responsibilities.

United States: Common Rule. In response to a recent proposal by Alan Fleischman and colleagues that called for a national advisory group to assess the potential social harm of particular research projects, legal scholar John Lunstroth clarified the status of claims made in the Common Rule regarding the obligations of institutional review boards in

the assessment of the potential social value of proposed research trials.[65] Lunstroth draws attention, in particular, to a mandatory rule and subrule expressed in 46.111, "Criteria for IRB approval of research": "(2) Risks to subjects are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of knowledge that may reasonably be expected to result" (the mandatory rule), and "[t]he IRB should not consider possible long-range effects of applying knowledge gained in the research (for example, the possible effects of research on public policy) as among those research risks that fall within the purview of its responsibility"[66] (the subrule). Because the subrule seems to suggest that IRBs must not consider the possible long-term effects of the results of research, most scholars have believed that consideration of the long-term harms to society of particular research trials is beyond the purview of ethics review.

What Lunstroth meticulously points out is that the mandatory rule, which refers to the importance of the knowledge resulting from research, could not properly be assessed if half the information (the potential harms to society) were ignored. Further, proper legal analysis demands a more open reading of the "should not" claim in the subrule than might seem to be required by a common language analysis of the passage. I quote him at length:

> There are no limitations on the kinds of importance that the IRB must consider, and therefore if there is the possibility the knowledge may be politically or socially important, the IRB is required to consider those possibilities. *Furthermore, if the IRB were restricted from considering the risks of sociopolitical effects, then the advisory rule would have the effect of putting a metaphorical thumb on the benefit side of the scale, inasmuch as the risk side of the scale would be artificially lightened by removal of a risk factor.*[67]

According to Lunstroth, it is "indisputable [that] the general mandatory rule requires consideration of all aspects of the outcome."[68]

Lunstroth asks what would happen if the results of research were highly socially beneficial. Surely the committee would take this favorable outcome into consideration. Since the full range of potential risks and benefits is rightly within the purview of what is reviewed by IRBs, these committees fail to fulfill their responsibilities when they don't take such factors into account. Further, he chastises those who resist the suggestion that ethics review is the right place to consider the sociopolitical consequences of research.[69] How might we think of providing any assessment of the importance of research without considering these factors? The proposal he is critiquing, which puts forward a national advisory group as the appropriate body for such assessments, would weaken the current system of review by removing one of its central responsibilities, and "the beneficiaries of this weakening would be scientists and firms who engage in controversial research."[70] I would add that they might also be the researchers conducting low-quality trials.

While the revisions proposed in 2015 to the Common Rule offer helpful clarifications of many of the original rules, no such illumination is provided for the rule cited above.[71] It is hard to say whether other developments in the latest version, such as calibrating the level of review to the level of risk to individual subjects, will enhance or diminish discussion of indirect harms by IRBs. Only time will tell.

Objections and Replies

One anticipated objection to my position runs as follows. If the problem is that there is too much low-quality research, shouldn't we raise our standards of scientific validity rather than our standards of social value? If current research is of such low quality, it should be rejected based on its failure on the second listed requirement from Emanuel, Wendler, and Grady: scientific validity.[72] There is no need to appeal to social value. This is an important objection because it reminds us of the complex relationship between the requirements of social value and scientific validity.

I am sympathetic to an argument against low-quality trials on grounds of scientific validity. Such an argument might well be complementary to the one offered here. The problem as I see it is that the requirement of scientific validity has been interpreted so narrowly that it only captures our interest in methodological rigor.[73] But it may well be the case that methodologically rigorous trials are the *least* socially valuable trials. Benjamin Freedman called attention to this when he wrote that "a useless study is more likely to be valid than a useful study."[74] At the end of the day, the social-value requirement, which presupposes scientific validity, goes further in attending to concerns beyond those narrowly in the methodological domain. It is not only scientific validity, narrowly construed, but also clinical relevance or applicability that matters to the assessment. It requires that research fairly and efficiently advance human health. As noted, it also provides an opportunity to include social harms in the assessment of proposed research.

A second objection concerns the breadth of indirect harm as a category. If RECs are encouraged to consider potential indirect harms in their evaluations, they may end up rejecting too many research studies. This is because many important and fruitful scientific discoveries have resulted in technologies that caused significant harm to humans. Consider, for instance, research in nuclear physics in the 1930s and 1940s, which permitted both the development of life-saving medical treatments and the construction of deadly nuclear bombs. It may be better for scientists simply to pursue knowledge for its own sake, without limitation. Key to

this objection is an assumption about the limits of foreknowledge: we never really know, in advance, how knowledge will be put to use.

It will be helpful to clarify my position, by way of initial response: I am not arguing that we should stop or prohibit any research with the potential to lead to harm. That would be to tip the balance too far in the direction of weighing harm, thus ignoring the need to balance benefit. This sort of engineered imbalance, whichever way it tips, is precisely what I want to avoid. If we keep in mind that assessments include social benefits as well as social harms, it isn't clear that most research will fail this test, since there will be potentially beneficial applications to weigh in the balance. To go a bit further, though, it is not always the case that we are deeply uncertain about the harmful effects or applications of research. For instance, some research will be aimed primarily at bad ends: think here of research done to enhance the transmissibility or virulence of a pathogen. While people may argue about the permissibility of research that is ambiguous in its applications (and I am in favor of precisely this sort of deliberation by RECs), most people also agree that research with only or primarily harmful effects or applications should be restricted on ethical grounds. I argue in this paper that low-quality research cannot produce useful knowledge with positive applications because, for instance, it is too biased or it is statistically incapable of answering the question it poses. Useless research will have no benefit and some potential harm because it contributes to the sorting problem.

A third anticipated objection goes as follows: why does this responsibility to assess the full set of harms and benefits fall primarily to research ethics committees rather than, for instance, funding agencies? After all, funders—particularly public funders—are in a unique position to assess social benefit and harm. If they set out a research agenda, in which certain topics, methods, and questions are prioritized, they can shape the direction of research much more easily than members of RECs can by rejecting individual proposals.

In reply, I would return to the widely accepted view that social benefits are necessarily a part of the assessment done by RECs. This is necessarily of concern to RECs because many trials have net risks to subjects and would not otherwise achieve a favorable risk-benefit ratio. If RECs are assessing social benefit, there is no principled reason for ignoring social harm.

Contrary to popular interpretations, the Canadian Tri-Council Policy Statement and the American Common Rule permit, and in the Canadian case may even be said to encourage, consideration of social harms in the ethical assessment of research.

A fourth, related objection concerns the limited capacity of RECs to carry out the sorts of robust social-value assessments required. It is no secret that RECs in most international jurisdictions have limited resources—whether time, expertise, or training—to discharge their many responsibilities. We have evidence of this from many sources, including an IOM report in 2003.[75] More recently, in a survey of NIH-funded researchers, "many researchers viewed the IRB as cumbersome and slow; some viewed their IRB as not competent to review the research, described as not understanding the protocols or analytic methods; IRBs were criticized for applying regulations inconsistently or for 'over-protecting' subjects."[76] My proposal seems to further burden an already overburdened regulatory system.

I want to acknowledge the force of this practical concern. There is a serious capacity shortage on RECs in most jurisdictions. I don't dispute this. But let us keep in mind three things. First, the obligation I have identified and argued for in this paper is one that is already existing. I am not suggesting we add an entirely new obligation, as, for instance, was arguably the case with the addition of the "responsiveness" criterion following an increase in ethically dubious international research. Rather, I am helping to develop our understanding of an existing obligation. Second, resources exist to assist researchers and RECs with assessments of quality; the PRECIS-2 tool outlined earlier is one such resource. Third, the fact that there is a crisis in a system of regulation is not in itself reason to reject efforts to make that regulatory system more rigorous. Ethical obligations weigh on us whether we like it or not, and they don't go away simply because they seem to be too demanding.

In sum, harms associated with the overproduction of low-quality research evidence are rarely included in the social-value calculations conducted during the ethical review of proposed clinical trials.[77] This happens for (at least) two reasons: first, because of a failure to recognize the need to preserve limited research resources, for instance, human subjects, for high-quality trials and, second, because social-value calculations—when they are conducted—focus on positive outcomes of potential trials. But the overproduction of low-quality clinical research is very likely to be harmful to patients. On ethical grounds there are persuasive reasons to endorse the position that we should conduct fewer clinical trials. Researchers and research ethics committees should work together to ensure that trials truly benefit society, as they are meant to do.

Acknowledgments

Special thanks to the reviewers of the *Hastings Center Report* for helpful

comments and to colleagues at Dalhousie University and members of the New Scholarship in Bioethics research network for ongoing scholarly support. Thanks also to my mentors in the American Society for Bioethics and Humanities Mentor-Pairing Workshop—Rebecca Dresser, Scott Kim, and David Wendler—for offering detailed feedback on the first draft of this paper.

Notes

1. Where possible, I use the neutral term "research ethics committees" or "RECs" to cover such variations as, for instance, "institutional review boards" ("IRBs"), used in the United States, and "research ethics boards" ("REBs"), used in Canada. By "clinical research," I mean all human-subjects research that aims to produce generalizable knowledge that will advance human health and well-being.

2. H. Bastian, P. Glasziou, and I. Chalmers, "Seventy-Five Trials and Eleven Systematic Reviews a Day: How Will We Ever Keep Up?," *PLoS Medicine* 7 no. 9 (2010): e10000326, doi:10.1371/journal.pmed/10000326.s001.

3. S. Sieber, L. M. Machesky, and R. H. Insall, "Overflow in Science and Its Implications for Trust," *eLife* 4 (2015): e10825, doi:10.7554/eLife.10825.

4. T. Hoffman et al., "The Scatter of Research: Cross Sectional Comparison of Randomised Trials and Systematic Reviews across Specialties," *BMJ* 344 (2012): e3223, doi:10.1136/bmj.e3223.

5. M. Tsay and Y. Yang, "Bibliometric Analysis of the Literature of Randomized Controlled Trials," *Journal of the Medical Library Association* 93, no. 4 (2005): 450-58.

6. Bastian, Glasziou, and Chalmers, "Seventy-Five Trials and Eleven Systematic Reviews a Day."

7. A. W. Chan and D. G. Altman, "Epidemiology and Reporting of Randomized Trials Published in PubMed Journals," *Lancet* 365 (2005): 1159-62. See also Bastian, Glasziou, and Chalmers, "Seventy-Five Trials and Eleven Systematic Reviews a Day."

8. J. Ioannidis, "Why Most Clinical Research Is Not Useful," *PLOS Medicine* 13, no. 6 (2016): e1002049, doi:10.1371/journal.pmed.1002049.

9. Sieber, Machesky, and Insall, "Overflow in Science."

10. Hoffman et al., "The Scatter of Research."

11. Ibid.

12. Sieber, Machesky, and Insall, "Overflow in Science."

13. F. Davidoff et al., "Evidence-Based Medicine: A New Journal to Help Doctors Identify the Information They Need," *British Medical Journal* 310, no. 6987 (1995): 1085.

14. B. S. Alper et al., "How Much Effort Is Needed to Keep Up with the Literature Relevant for Primary Care?," *Journal of the Medical Library Association* 92, no.4 (2004): 429-37.

15. S. Sanjay et al., "Journal Reading Habits of Internists," *Journal of General Internal Medicine* 15 (2000): 881-84.

16. Ibid., 883.

17. Ibid., 883.

18. R. M. Califf et al., "Characteristics of Clinical Trials Registered in ClinicalTrials.gov, 2007-2010," *Journal of the American Medical Association* 307 no. 17 (2012): 1838-47; see also J. P. A. Ioannidis, "Why Most Published Research Findings Are False," *PLoS Medicine* 2 no. 8 (2005): e124, doi:10.1371/journal.pmed.0020124.

19. J. Lind, "A Treatise of the Scurvy" [1753], cited in Bastian, Glasziou, and Chalmers, "Seventy-Five Trials and Eleven Systematic Reviews a Day."

20. D. Altman, "The Scandal of Poor Medical Research," *BMJ* 208 (1994): 283-84, at 283.

21. Ibid., 283.

22. Bastian, Glasziou, and Chalmers, "Seventy-Five Trials and Eleven Systematic Reviews a Day."

23. Ibid.

24. K. Loudon et al., "The PRECIS-2 Tool: Designing Trials That Are Fit for Purpose," *BMJ* 350 (2015): h2147, doi:10.1136/bmj.h2147.

25. K. Borgerson, "Are Explanatory Trials Ethical? Shifting the Burden of Justification in Clinical Trial Design," *Theoretical Medicine and Bioethics* 34, no. 4 (2013): 293-308.

26. Loudon et al., "The PRECIS-2 Tool."

27. S. Epstein, *Inclusion: The Politics of Difference in Medical Research* (Chicago: University of Chicago Press, 2007). Niche standardization is "a general way of transforming human populations into standardized objects available for scientific scrutiny, political administration, marketing, or other purposes that eschews both universalism and individualism and instead standardizes at the level of the social group" (p. 135). Certain vulnerable populations have been both overincluded and underincluded in medical research, and niche standardization, combined with explicit targeting of such groups for inclusion, has been one (fraught) approach to ensuring representation.

28. Califf et al., "Characteristics of Clinical Trials Registered in ClinicalTrials.gov, 2007-2010," 1838.

29. K. Humphreys, "Extent and Reporting of Patient Nonenrollment in Influential Randomized Clinical Trials, 2002 to 2010," *JAMA Internal Medicine* 173 no. 11 (2013): 1029-31, doi:10.1001/jamainternmed.2013.496.

30. See K. Borgerson, "Why Reading the Title Isn't Good Enough: An Evaluation of the 4S Approach to Evidence-Based Medicine," *International Journal of Feminist Approaches to Bioethics* 2 no. 2 (2009): 152-75.

31. P. Humaidan and N. Polyzos, "(Meta)analyze This: Systematic Reviews May Lose Credibility," *Nature Medicine* 18, 1321 (2012), doi:10.1038/nm0912-1321.

32. D. H. Lee and O. Vielemeyer, "Analysis of Overall Level of Evidence behind Infectious Disease Society of America Practice Guidelines," *Archives of Internal Medicine* 171, no. 1 (2011): 18-22; P. Tricoci et al., "Scientific Evidence Underlying the ACC/AHA Clinical Practice Guidelines," *Journal of the American Medical Association* 301, no. 8 (2009): 831-41.

33. Bastian, Glasziou, and Chalmers, "Seventy-Five Trials and Eleven Systematic Reviews a Day."

34. A familiar account of direct and indirect benefit is as follows: "A *direct benefit* is the benefit obtained by those who need and receive a resource (as a result of their needing and receiving that resource), and an indirect benefit is a non-direct benefit obtained by a third party as a result of the fact that the resource is given to a direct beneficiary" (J. Du Toit and J. Millum, "Are Indirect Benefits Relevant to Health Care Allocation Decisions?," *Journal of Medicine and Philosophy* 41, no. 5 (2016): 540-57.

35. Any such empirical evidence would be most welcome since it is directly relevant to the assessments made in this paper. A speculative list similar to my own appears (in a somewhat different context) in A. J. London, J. Kimmelman, and B. Carlisle, "Rethinking Research Ethics: The Case of Postmarketing Trials," *Science* 336, no. 6081 (2012): 544-45.

36. P. Glasziou et al., "What Is Missing from Descriptions of Treatment in Trials and Reviews?," *BMJ* 336 (2008): 1472-74.

37. A. Berenson et al., "Despite Warnings, Drug Giant Took Long Path to Vioxx Recall," *New York Times*, November 14, 2004, http://www.nytimes.com/2004/11/14/business/14merck.html.

38. A. J. London, "A Non-Paternalistic Model of Research Ethics and Oversight: Assessing the Benefits of Prospective Review," *Journal of Law, Medicine and Ethics* (2012): 930-44.

39. For instance, in a 2015 paper, Sabina Sieber and colleagues say, "Reducing overflow is hardly a solution. It is widely accepted that funding scientific research leads to many social and economic benefits, and calls to limit science participation are rarely supported by either governments or most scientists" (S. Sieber, L. M. Machesky, and R. H. Insall, "Overflow in Science and Its Implications for Trust"). They go on to

recommend changes to peer-review and journal acceptance policies as means of addressing the problem.

40. R. B. Haynes, "Of Studies, Summaries, Synopses, and Systems: The 4S Evolution of Services for Finding Current Best Evidence" [editorial], *Evidence-Based Medicine* 6 (2001): 36-38; R. B. Haynes, "Of Studies, Syntheses, Synopses, Summaries, and Systems: The '5S' Evolution of Information Services for Evidence-Based Healthcare Decisions," *Evidence-Based Medicine* 11 (2006): 162-64; A. DicCenso, M. L. S. Bayley, and R. B. Haynes, "Accessing Preappraised Evidence: Fine-Tuning the 5S Model into a 6S Model," *Evidence-Based Nursing* 12 no. 4 (2009): 99-101.

41. Sharon Straus and colleagues, for instance, claim that "in some circumstances, the title [of a review] provides enough information"; see S. E. Straus et al., *Evidence-Based Medicine: How to Practice and Teach EBM* (Toronto: Elsevier, 2005), 38.

42. Borgerson, "Why Reading the Title Isn't Good Enough," 152-75.

43. Glasziou et al., "What Is Missing from Descriptions of Treatment in Trials and Reviews?"

44. Ibid., 1472.

45. Ibid., 1474.

46. Hoffman et al., "The Scatter of Research."

47. Bastian, Glasziou, and Chalmers, "Seventy-Five Trials and Eleven Systematic Reviews a Day"; see also I. Chalmers and P. Glasziou, "Avoidable Waste in the Production and Reporting of Research Evidence," *Lancet* 374 (2009): 86-89.

48. Bastian, Glasziou, and Chalmers, "Seventy-Five Trials and Eleven Systematic Reviews a Day."

49. E. J. Emanuel, D. Wendler, and C. Grady, "What Makes Clinical Research Ethical?," *Journal of the American Medical Association* 283 (2000): 2701-11; B. Freedman, "Scientific Value and Validity as Ethical Requirements for Research: A Proposed Explication," *IRB: Ethics and Human Research* 9, no. 6 (1987): 7-10. I set aside further, more contentious, requirements at this time, simply for ease of argument.

50. Revision of CIOMS 2002 International Ethical Guidelines for Biomedical Research Involving Human Subjects. The latest draft of the guidelines was retrieved here: http://www.cioms.ch/publications/guidelines/guidelines_nov_2002_blurb.htm. The language of "social value" has been added to the first guideline of this most recent draft, accessed September 29, 2015.

51. Ibid.

52. The responsiveness criterion attempts to spell this out more explicitly for international contexts, but the social value criterion is best understood as requiring that the benefits of research are made available to the communities in which research is conducted. This is important for inner-city Canadian emergency rooms in much the same way that it is important for rural communities in Thailand.

53. Emanuel, Wendler, and Grady, "What Makes Clinical Research Ethical?," 2703.

54. A. Wertheimer, "The Social Value Requirement Reconsidered," *Bioethics* 29, no. 5 (2015): 301-08.

55. A. Rid and D. Wendler, "A Framework for Risk Benefit Calculations in Biomedical Research," *Kennedy Institute of Ethics Journal* 21, no. 2 (2011): 141-79.

56. R. Foy et al., "How Evidence Based Are Recruitment Strategies to Randomized Controlled Trials in Primary Care? Experience from Seven Studies," *Family Practice* 20 (2003): 83-92.

57. Wertheimer considers this argument but dismisses it quickly, suggesting that it extends the idea of public resources too far, and argues that there must be "demonstrable social harms" in order to justify requiring that privately funded research adhere to a strong social-value requirement. The first sections of this paper may be read as an attempt to provide provisional grounds of this sort: the abundance of low-quality research, leading to the sorting problem, harms patients, places research subjects at undue risk, and compromises public trust in the research enterprise.

58. The one exception to this, though still on the fringes of bioethics, is work on malevolent uses of biomedical research; see S. K. Green et al., "Guidelines to Prevent Malevolent Use of Biomedical Research," *Cambridge Quarterly of Healthcare Ethics* 15 (2006): 432-47.

59. Rid and Wendler, "A Framework for Risk Benefit Calculations in Biomedical Research," 148. The authors do make one cryptic reference to social risks in their discussion of "open questions" at the end of the paper: "However, the framework also makes clear that truly comprehensive risk-benefit evaluations must consider social risks from the research" (p. 168).

60. Well, perhaps we do. But we ought not to do so if we are aiming to act rationally.

61. Freedman, "Scientific Value and Validity as Ethical Requirements for Research."

62. Government of Canada, "Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans," 2014, p. 8, http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/Default/.

63. I do not defend a requirement of maximizing social value in this paper, although I believe I would endorse a qualified version of the position, if pressed on the matter.

64. Government of Canada, Tri-Council Policy Statement, 10.

65. U.S. Department of Health and Human Services, Human Subjects Research 45 CFR 46, http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html.

66. Ibid.

67. J. Lunstroth, "The Role of Controversial Research in the IRB's Risk/Benefit Analysis," *The American Journal of Bioethics* 11, no. 5 (2011): 14-16, at 16 (emphasis added).

68. Ibid., 16.

69. One reason for resistance might be that members of RECs don't have the skill set or expertise to make this assessment. In that case, perhaps there is a need to rethink the expertise required on RECs.

70. Lunstroth, "The Role of Controversial Research in the IRB's Risk/Benefit Analysis," 14-16.

71. Federal Policy for the Protection of Human Subjects: Proposed Rule, Fed. Reg. (Sept. 8, 2015).

72. Emanuel, Wendler, and Grady, "What Makes Clinical Research Ethical?," 2701-11.

73. Borgerson, "Redundant, Secretive, and Isolated," 385-411.

74. Freedman, "Scientific Value and Validity as Ethical Requirements for Research."

75. The first problem identified in the report was "significant doubt about the capacity of the current system to meet its core objectives (dissatisfaction with the current system is widespread)" (Institute of Medicine, *Responsible Research–A Systems Approach to Protecting Research Participants* [Washington, DC: National Academies Press, 2003], 5).

76. T. Straight, "Clinical Research Regulations: Challenges to the Institutional Review Board System," *Clinics in Dermatology* 27 (2009): 375-83.

77. A. Fleischman et al., "Dealing with the Long-Term Social Implications of Research," *American Journal of Bioethics* 11, no. 4 (2011): 5-9.