Varieties of Community Uncertainty and Clinical Equipoise

Alex John London, Patrick Bodilly Kane, Jonathan Kimmelman

Kennedy Institute of Ethics Journal, Volume 33, Number 1, March 2023,
pp. 1-19 (Article)

Published by Johns Hopkins University Press

For additional information about this article
https://muse.jhu.edu/article/899457

[Access provided at 9 Jun 2023 12:42 GMT from Carnegie Mellon University]

Alex John London, Patrick Bodilly Kane, and Jonathan Kimmelman

Varieties of Community Uncertainty and Clinical Equipoise

ABSTRACT. The judgments of conscientious and informed experts play a central role in two elements of clinical equipoise. The first, and most widely discussed, element involves ensuring that no participant in a randomized trial is allocated to a level of treatment that everyone agrees is substandard. The second, and less often discussed, element involves ensuring that trials are likely to generate social value by producing the information necessary to resolve a clinically meaningful uncertainty or disagreement about the relative merits of a set of interventions. The distribution of judgments in expert communities can take many forms, each with important implications for whether a trial satisfies one or both elements of clinical equipoise. In this article we use a graphical approach to represent three ways in which expert community uncertainty can vary: by spread, modality, and skew. Understanding these different distributions of expert judgment has three important implications: it helps to make operational the requirement of social value, it shows that some conditions for initiating studies to promote social value diverge from common assumptions about clinical equipoise, and it has important implications for how trials should be designed and monitored, and what patients should be told during informed consent.

1. INTRODUCTION

The concept of equipoise refers to a state of uncertainty that is supposed to play at least two important roles in research ethics. First, when experts are uncertain about the relative merits of alternative strategies to diagnose, prevent, or treat disease, randomizing participants to those strategies is ethically permissible. Second, research designed to reduce or eliminate such uncertainty is likely to produce social value. How we operationalize the concept of equipoise thus has far-reaching implications for trial design and research ethics.

Kennedy Institute of Ethics Journal Vol. 33, No. 1, 1–19 © 2023 by Johns Hopkins University Press

In its earliest and most intuitive formulation, equipoise was understood as a function of the subjective beliefs of a single expert (Fried 1974). As the name "equipoise" suggests, this kind of uncertainty was crudely modeled as a "50–50" split in which the relative merits of two interventions are equally balanced, reflecting the proposition that the expert has no reason to prefer one treatment to the other. Despite its close connection to the name "equipoise" and its intuitive appeal, this model of uncertainty is unworkable (London 2021). As early critics noted, it is unlikely that all relevant information will leave the individual expert completely indifferent between the available options (Gifford 1986; Hellman 2002). In the unlikely event that evidence did point this way, this state of indifference would likely vanish in the face of new evidence about the relative merits of these interventions. As a result, in the rare event that a researcher could ethically initiate a randomized controlled trial (RCT), their individual uncertainty would likely evaporate long before the trial reached statistical significance.

In response to these problems, Benjamin Freedman (1987) argued that equating equipoise with the belief state of a single expert fails to capture an important type of uncertainty in medicine, conflicting expert judgment. Several experts considering the same interventions for the same medical condition might each regard one as superior to the rest but differ in which intervention they favor. No expert is uncertain in the sense above, but the divisions within the expert community reflect uncertainty about the best strategy to advance patient interests, and reducing or eliminating this conflict in judgment and treatment practices could have significant social value. Freedman's key insight was that the uncertainty that research must be designed to reduce or eliminate—which he called "clinical equipoise"—should be understood as a function of the judgments of informed experts in the medical community rather than as a state of the individual investigator.[1]

Less well appreciated, however, is the extent to which clinical equipoise is consistent with a diverse mix of views or belief states on the part of the experts who constitute this community. Various critics of clinical equipoise have noted challenges in terms of characterizing expert community uncertainty, as well as operationalizing the concept (Gifford 1995; London 2020). The present essay in part responds to these challenges. In contrast to the idea of equal balance implied by the term equipoise, there are many ways in which a community of experts can be divided that constitute a state of clinical equipoise. In what follows, we provide a taxonomy for classifying different distributions of expert judgment and

a visual representation of those distributions. We argue that, although this representational system is not exhaustive, it supports three valuable insights. First, our framework provides concrete, operational content to an important aspect of social value. This is the idea that studies have social value if they produce the knowledge needed to reduce unwarranted treatment diversity or efficiently shift practice in a direction that improves the quality of care for patients (London 2018; 2021; Wendler and Rid 2017; Wenner 2017). Second, it allows us to show that one way research can be organized to promote social value hinges on initiating studies under conditions that diverge from common assumptions about clinical equipoise. This underscores the limitations of common conceptions of equipoise and the importance of attending to the diverse ways in which expert beliefs might be distributed in a community. Finally, although our framework makes important simplifying assumptions, interrogating the reasons for divergent expert opinion can have significant implications for informed consent, trial design, and monitoring.

2. REPRESENTING DIVERSITY IN EXPERT JUDGMENT

Because clinical equipoise refers to the distribution of judgments in a community of experts, this concept can be represented with a histogram that plots the distribution of expert opinion regarding the relative clinical merits of a set of interventions. For simplicity, we focus on a case in which there are two interventions for a specific medical condition: *A* represents one treatment (in most cases below, it will refer to the more novel of the treatments), and *B* the current standard of care. Imagine further that the considered and informed judgment of each individual expert can be plotted as a point on a scale from 0 to 100%, reflecting the level of confidence that expert has in the superiority of *A*. A score of 0% indicates that an expert believes the evidence unequivocally favors *B*, and therefore reflects 100% confidence in the superiority of *B*. A score of 100% indicates that an expert believes the evidence unequivocally favors *A*, and therefore represents 100% confidence in the superiority of *A*. A score of 50% refers to the case where an expert believes existing evidence provides no basis for preferring *A* or *B* and so represents a 50–50 split in confidence between *A* and *B*. Thus, scores lower than 50% represent an expert favoring *B*, while values higher than 50% represent an expert favoring *A*. Imagine we compile the judgments of a large group of experts and position each expert on our x-axis. Now we have a histogram reflecting the state of expert belief about the clinical value of *A* relative to *B*.

This approach to representing community uncertainty about the comparative advantage of a new treatment involves several simplifying assumptions, two of which are worth noting here. First, to represent each expert's confidence as a point in this space, we are limited to each expert's all-things-considered judgment of the relative clinical merits of *A* and *B*. This elides important respects in which experts might disagree about the various attributes of interventions that contribute to their all-things-considered judgment. For example, experts might have equivalent confidence in the superiority of *A* relative to *B*, but have different views on the attributes of efficacy, side effects, or ease of use that contribute to their overall assessment. Second, representing each expert's opinion as a point elides complex ways in which different experts might each have different sets of beliefs even regarding the same attribute. For example, imagine two experts who have 90% confidence in *A* relative to *B* and who are in total agreement about safety, ease of administration, and so on. One expert's 90% confidence might reflect that they are nearly certain that *A* will offer small benefits to patients. Another expert's 90% confidence might reflect that they have a small amount of confidence that *A* might offer substantial benefits, like a cure. Despite these simplifying assumptions, the approach we take here is sufficient to capture an important element of the diversity in belief states that can constitute clinical equipoise. Throughout the rest of the article, we will assume in the discussion that follows that we are always dealing with the all-things-considered judgments of informed and conscientious experts, and so we will use terms like the "judgments," "assessments," or "preferences" of experts as shorthand.

Against this background, we can describe the diversity of expert judgment about the relative merits of *A* and *B* using three features: spread, modality, and skew. First, expert preferences can be more or less spread across the continuum of belief. At one limit, there might be situations where there is practically no spread of preferences: all experts are bunched up close to a single number, reflecting little difference of opinion among them. At the other limit, there are roughly equal numbers of experts at every position between 0 and 100%, resulting in a flat distribution.

The second feature of the distribution of expert judgment in a community is modality. Although some distributions might have a single peak or "mode," there might be circumstances in which expert assessments are bimodal, with groups of experts clustered on either side of 50%. In such bimodal cases, a community of experts is divided into "camps," with one camp having a clear preference for *A* while the other has a clear preference for *B*.

The third feature is skewness. Instead of a symmetrical or bell curve, distributions might be asymmetric. Small numbers of experts might hold extreme preferences for one treatment while the rest of the experts favor the alternative treatment with varying degrees of intensity. Distributions illustrating these three features are depicted in Figure 1. Some studies suggest that expert community uncertainty about the comparative merits of a new treatment frequently varies on each of these attributes (Benjamin et al. 2020).

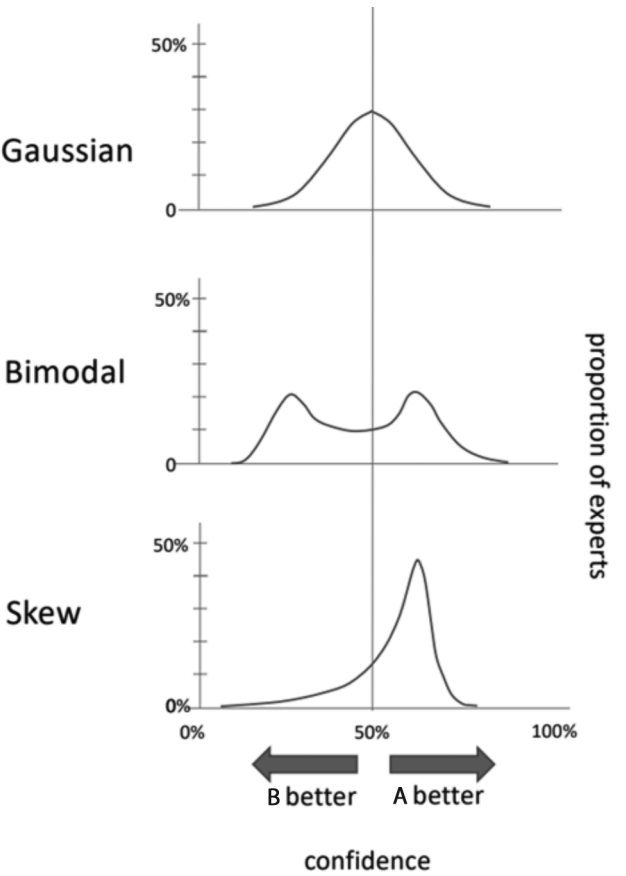

Figure 1. Histograms Representing Varieties of Expert Community Uncertainty. In all cases graphed, enrollment of patients might not necessarily violate the welfare condition of clinical equipoise.

3. REPRESENTING CLINICAL EQUIPOISE

With the above elements in place for representing the diversity of community opinion, we now consider how these possible distributions of belief in the community might relate to clinical equipoise. It is useful to begin by considering cases in which clinical equipoise does *not* exist. For example, if every expert in the community has a confidence below 50%, then every expert favors the standard of care (*B*) and there is no disagreement in the community (Figure 2a). Likewise, if every expert has a confidence above 50%, then every expert favors *A* and there is no disagreement in the community. In both cases, allowing study participants to be randomized to either *A* or *B* would be morally problematic for two reasons. First, such a study would involve knowingly assigning some study participants to an intervention that is not regarded as the best treatment option by even a reasonable minority of experts. This violates the first component of clinical equipoise, which we will refer to as "concern for welfare"; this states that it is impermissible to knowingly expose a person to interventions, practices, or procedures that are known or credibly believed to be worse than another available option (London 2021; 2018). Second, randomizing patients to *A* or *B* in this case is unlikely to contribute to improving medical practice. Individual experts have clear preferences for one option over the alternatives and there is no dissensus in the community. As a result, such studies would violate the second component of clinical equipoise, which we will refer to as the "social value requirement." This requires that research produce information that is likely to enhance the capability of health systems to understand, identify, and intervene on important health problems (London 2021; 2018; Wendler and Rid 2017; Wenner 2017).

Within this representational scheme, a necessary condition for clinical equipoise is that at least a reasonable minority of informed and conscientious experts should fall below the 50% mark, while at least a reasonable minority falls above the 50% mark. This condition is necessary to satisfy the condition of concern for welfare. From now on, therefore, we will only consider distributions that have this necessary feature. Even so, distributions that satisfy this condition can take a wide range of shapes and, importantly, these shapes can have implications for the second aspect of clinical equipoise, its relationship to the social value requirement.

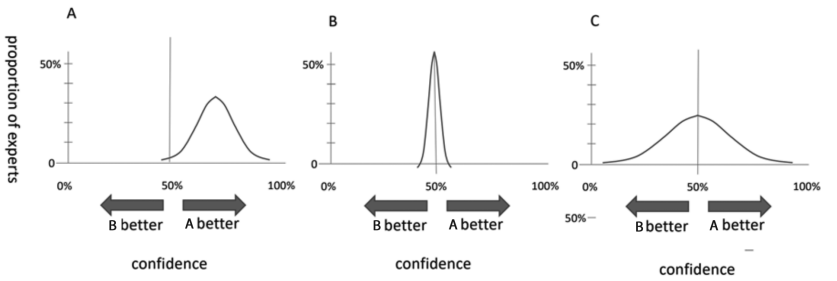

Figure 2. Different Ways a Gaussian Distribution of Expert Preference Might Not Fulfill Clinical Equipoise. In the first panel, almost all experts favor A. In the middle, experts potentially regard the interventions as interchangeable. In the third, experts harbor a wide variety of views and theories, such that randomized trials may not be the best approach for shaping clinical uncertainty and convincing most experts.

3.1 Spread

Because the concept of equipoise seems to imply the equal distribution of some quantity, probably the most intuitive and stereotypic representation of clinical equipoise would be represented within our framework as a symmetrically distributed, bell-shaped curve centered at 50% (Figure 1, top panel). Since at least a meaningful minority of experts are positioned below 50% and at least a meaningful minority are positioned above 50%, this distribution would satisfy the concern for welfare condition of clinical equipoise. Whether it satisfies the social value condition likely depends on several factors including the degree of divergence or spread in expert judgment.

For example, consider the case where the judgments of experts are narrowly distributed around 50% (Figure 2b). Such a distribution might arise under one of two conditions. In one condition, all experts recognize that clear evidence supports the judgment that *A* and *B* have roughly equivalent value. Although a randomized trial would not violate concern for welfare, it would fail the social value requirement because it would consume scarce resources without the expectation of generating knowledge needed to enable health systems to better meet the needs of patients. In this case, the narrow degree of spread among expert judgments provides an indication that there is little dissensus in the community, and so altering expert judgment is unlikely to translate into significant benefits to patients or health systems.

In another condition, a narrow symmetrical distribution over 50% might obtain when expert communities lack the information necessary to

formulate clear preferences. Such a distribution is likely to be relatively rare since it presupposes that experts are unable to form a considered view of the relative merits of the current standard of care for a condition in comparison to some novel alternative. Such a state is most likely to arise in the face of unfamiliar interventions or unfamiliar medical conditions. Under such circumstances, however, a small amount of new information is likely to produce dramatic shifts in expert preferences. In this case, questions might arise about the level of maturity of the relevant science—whether there is sufficient information about the relevant scientific questions to support the conduct of a randomized, controlled trial. Studies in this case might violate the social value condition because there are alternative ways of generating relevant information at lower cost (e.g., additional preclinical research), while imposing fewer risks on study participants.

In both of the above cases, narrow spread might be an indication that an RCT might not produce sufficient social value to satisfy the equipoise condition. What about cases where expert judgments are symmetrically centered over 50% but there is a very wide distribution of expert assessments (Figure 2c)? A second limiting case is a flat distribution in which there is at least one expert that occupies every position between 0% and 100%. This is a case of radical diversity in the expert community. Unlike the cases in the previous paragraphs, where every clinician was so uncertain that they had a difficult time forming a preference for *A* or *B*, in this case most experts have a clear preference for *A* or *B*, but there is little agreement between experts on those assessments. Such a case is also likely to be rare, but it would also seem to indicate a level of immaturity in the knowledge base surrounding the relative merits of *A* and *B*. Each expert seems to be taking a different interpretation of the available evidence, potentially drawing on competing theories about a disease process, intervention strategy, or measurement. On the one hand, a well-designed clinical trial might have significant social value in this case if its results were likely to move expert judgments from a flat line to a more Gaussian or skewed distribution centered around 50%. In this sense, greater spread might be an indication of the value a trial might have in terms of the gains to patients and health systems from reducing unwarranted diversity in practice. On the other hand, it might be difficult to design a trial that would have this effect if there is this much dissensus in the expert community. It also seems likely that there are other modes of investigation that might involve fewer or no human participants, like

more basic or preclinical research, or additional early phase trials, to recruit some experts away from extreme views, thus paving the way to a more efficient randomized trial.

As a result, one might think that the paradigm of clinical equipoise would be a symmetrical distribution centered above 50% with a degree of spread that indicates a meaningful disagreement in the expert community that might be sufficiently mature that a trial or small set of trials would be likely to reduce unwarranted diversity in practice. Although such a distribution would constitute an example of clinical equipoise, requiring this distribution of uncertainty in the expert community would be highly inefficient. The reason is that it requires that roughly half of the experts in the clinical community favor an intervention other than the standard of care.

3.2 Bimodality

Bimodal distributions reflect an expert community polarized by disagreement. These camps need not be of equivalent size, and their judgments need not be symmetrical or centered over the 50–50 point (Figure 1, middle panel). This might occur where medical oncologists and radiation oncologists have rival approaches for managing a particular cancer (Sheehan et al. 2014). What matters from the moral point of view is simply that the experts in each camp represent at least a "reasonable minority" of the expert community.

When a community of informed and conscientious experts is divided in this way, trials designed to evaluate the relative merits of the relevant interventions satisfy the condition of concern for welfare while also having a strong claim to producing social value. First, as we noted earlier, research in such cases is consistent with concern for welfare because every participant is assigned to an intervention that would be recommended for them by at least a reasonable minority of clinical experts. In that sense, the treatment standard in this trial does not fall below what would be regarded as acceptable care outside of the trial.

Second, resolving disagreements of this kind can have considerable social value because clarifying the relative clinical value of these various interventions can improve patient outcomes by eliminating unwarranted diversity in practice. In contrast to high spread, where every expert can be thought of as in their own camp, the existence of a small number of camps indicates that differences of opinion have crystalized around a

small set of interpretations that could reasonably be pitted against one another in a well-designed trial. Resolving such disputes allows health systems to shift scarce human and material resources to the provision of safer and more effective practices and procedures. Trials will often be the most efficient way of generating the evidence needed to resolve such clinical controversies.

At the same time, bimodal distributions are most likely to emerge when key scientific questions have gone unanswered for too long and theory has been allowed to develop past the available evidence (Figure 3, upper panel). This issue is more common in areas where there is a weaker tradition of clinical trials, such as surgery or transplant medicine. High-dose chemotherapy with autologous bone marrow transplant (HDC-ABMT) for end-stage breast cancer in the 1990s (Rettig et. al., 2007) likely reflected such a situation. Because the standard of care for patients in this class was inadequate, it was comparatively easy to recruit experts and patients to a more aggressive alternative approach despite the lack of confirmatory testing available. Once the expert community becomes increasingly divided like this, it can be difficult to mount needed trials and more difficult to convince adherents that an intervention they have championed and worked to deliver to patients lacks sufficient clinical value relative to a standard of care that they regard as inadequate.

When the "camps" of a bimodal distribution represent divergent theories, ancillary theories often emerge within camps that buttress their interpretation of a treatment's effects. For example, one camp may come to regard a pathophysiological mechanism as central to disease progression, or that a surrogate outcome is a strong indicator of a treatment's benefits. Once such ancillary theories have strong support within a camp, a well-run trial with decisive results may fail to dispel controversy. Perhaps such entrenched ancillary theories help explain why, for example, the University Group Diabetes Program trial testing a widely used treatment of pre-diabetes, tolbutamide, failed to persuade many physicians and endocrinologists when results were announced in 1970. In this case, a well-run randomized trial strongly showed that tolbutamide was ineffective and possibly harmful (Schwartz and Meinert 2004). Many physicians were unpersuaded (Greene 2006). Had trials been performed earlier in the development of tolbutamide, when only a few experts had become "true believers" in it, prolonged controversy about the drug's value might have been avoided, with the pro-tolbutamide camp persuaded to pursue different sulfonylurea drugs instead. This is discussed in the next section.

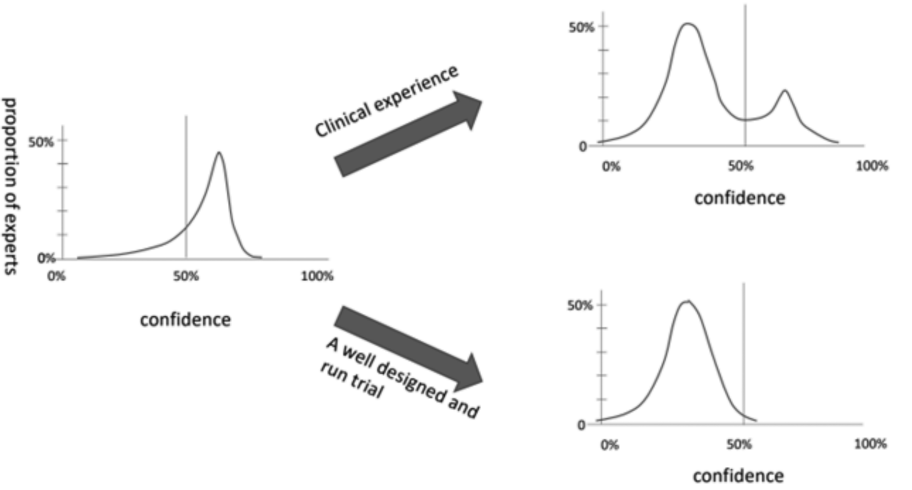

Figure 3. How Community Uncertainty Changes. In the upper panel, clinical experience and observational studies lead some experts to favor treatment B, while a substantial minority of experts, who subscribe to different theories and standards of evidence, remain unconvinced. In the bottom panel, a well-run trial is potentially sufficient to persuade almost all of the community of the superiority of B.

3.3 Skew

Skewed distributions represent cases where there are small numbers of experts who hold strong preferences that are contrary to consensus (Figure 1, bottom panel). This might occur early in the development of a novel treatment approach. Before the development of vemurafenib for late-stage melanoma, the standard of care was dacarbazine—a drug that offered small survival benefits at the expense of substantial treatment burden. Vemurafenib targeted tumors with mutations in the gene BRAF, and the theoretical rationale for the drug's action along with compelling safety data raised the expectation among some experts that targeting BRAF would lead to qualitative improvements in survival with much less toxicity (Flaherty et al. 2010). A histogram of expert community preference likely would have reflected a distribution in which most experts had small preferences for a standard of care but a small number of "early adopters" leaned strongly toward the novel approach of targeting BRAF. In our representation, this distribution of belief would be depicted by a histogram in which the majority of the community sits to the right of the 50–50 point, but a few early supporters of the novel intervention create a distribution skewed to the left.

Skewed distributions would fulfill the concern for welfare condition of clinical equipoise, since there is at least a reasonable minority of experts who favor either option. Importantly, they would also generally fulfill the condition of social value. The reason is that skewed distributions of this kind are likely to represent a situation in which the information produced from well-designed and well-run trials has the largest impact on expert judgment (Figure 3, bottom panel). In the case where the early adopters are correct, a well-designed and run trial is likely to expeditiously recruit adherents away from the current standard of care for the medical condition in question. Alternatively, if early adopters are mistaken, a well-designed and run trial might dampen the convictions of those advocating the novel approach (perhaps sending them back to the drawing board with a different drug targeting the same mechanism) and make it less likely that the novel approach will recruit new adherents until its relative merits have been clearly demonstrated. However, the failure to conduct well-designed clinical trials at this optimal point can result in the kind of polarization described above in our account of bimodal distributions.

4. IMPLICATIONS

In the previous section, we described how different clinical questions can result in different states of community uncertainty. We also showed that various states of expert uncertainty might fulfil the concern for welfare condition of clinical equipoise but fail the social value requirement. In this section, we consider the practical implications of our exposition for trial justification, design and monitoring, and informed consent.

First, our framework for visualizing the diversity of expert judgment helps operationalize the concept of social value. The social value of a trial is a function of its ability to reduce unwarranted diversity in medical practice and thereby improve outcomes for patients or streamline the provision of health services. Second, our framework illustrates how two key components of clinical equipoise can diverge in practice. Traditional criticisms of clinical equipoise have focused on cases in which it is alleged that it is not possible to generate sufficient evidence to guide drug approval or policy decisions while satisfying the concern for welfare condition. Although we do not find these arguments persuasive, it is important to note that our analysis reveals something different—that studies that satisfy the concern for welfare condition can fail to satisfy the social value requirement.[2] In several of the examples we discussed above, this stemmed from the fact that there might be means of generating relevant

information other than randomized trials in humans. In areas where the science remains immature, other research methods, like preclinical or early phase studies,[3] can shift expert judgments in ways that might reduce the appeal of novel approaches or recruit enough experts to their side, with less participant exposure to risk and burden, to create a form of clinical equipoise (e.g., skew or a more Gaussian distribution) that a randomized trial might resolve efficiently.

Our framework may also help with assessing the efficiency of the clinical research enterprise, where efficiency is understood in terms of the knowledge gain per participant randomized. We noted above that the optimal time to initiate trials of novel therapies is when skewed distributions fulfilling the concern for welfare condition of clinical equipoise begin to emerge. This is because skewed distributions are most likely when evidence supporting the clinical merits of a new intervention is sufficiently mature that it leads a reasonable minority of conscientious and informed experts to regard the novel intervention as at least as good or better than the standard of care. Health systems that launch well-designed and run trials at this point avert unwarranted diversity that snowballs into the polarization typical of bimodal distributions while avoiding the waste of initiating large, resource-intensive studies when alternative approaches to learning could shift expert assessments. They are also most likely to move large portions of the clinical community to a superior intervention when those are present. As a result, the paradigm case that best satisfies both conditions of clinical equipoise is neither symmetric nor centered over the 50–50 point.

The framework we propose for visualizing clinical equipoise might also inform the design of trials. As noted, one circumstance where the distribution of expert assessments might skew to the right is where a small group of experts has strong beliefs that a new therapeutic strategy will dramatically change the treatment landscape. Another setting is a randomized noninferiority trial testing an approved dosage of an effective drug against a lower dose that some believe will be equally effective. Given that social value is a function of the ability of a trial to reduce unwarranted diversity, it would be important to design these trials in a way that has favorable prospects of dissuading confident proponents of an approach if it lacks merit, or nudging mild doubters toward support for the new approach if its merits are confirmed. Such a trial will be most impactful if negative results are highly diagnostic (i.e., they cannot easily be dismissed as due to bias or a false negative by those experts who strongly favor

the approach). Design and monitoring provisions that might enhance the diagnosticity of negative results include higher statistical powering, a higher threshold for declaration of futility during data monitoring, pharmacodynamic secondary outcomes that can corroborate whether an intervention is affecting a molecular target, or (if an intervention involves surgical delivery) optimizing administration procedures according to guidance from proponents of the approach.

The varieties of uncertainty and equipoise we describe above have implications for informed consent as well. Consider a trial where expert preferences are bimodally distributed. Many prospective participants are likely to learn that some specialists, far from being uncertain themselves, strongly favor one treatment over another. This may be unsettling for research participants. Evidence from other areas suggests that people tend to be averse to expert disagreement, with expert division engendering doubts about expert competence or trustworthiness (Smithson 1999; 2015; 2013). Such doubts may translate into distress, low accrual, and drop out. When clinical trials involve divisions among experts, researchers may want to directly address these doubts and anxieties among potential participants and emphasize the value of generating the evidence necessary to reduce conflicts in judgment and practice.

The case of right skew might require a different approach during informed consent. As noted, in many cases skew will reflect that some experts prefer a novel treatment because they anticipate a small prospect of substantial gains. Other experts may prefer a standard of care to lock in smaller but more certain gains. Patients are likely to vary in their willingness to take risks for major gains. Some may be risk averse, in which case declining enrollment in such a trial and receiving standard of care may be the preferable option. For those who are more drawn to taking risks for the prospect of big gains, trials reflecting skewed distributions of preference may be attractive. A proper consent disclosure will allow patients to match a trial offer to their risk tolerance.

5. LIMITATIONS AND CONCLUSIONS

To this point we have argued that clinical equipoise can exist not only when distributions are Gaussian, but also when the distribution of expert judgments is bimodal or asymmetrically distributed between a set of interventions. Contrary to the language of equal balance, the optimal time to conduct large-scale confirmatory trials for novel interventions is when the distribution of expert judgments begins to skew toward the

novel alternative. We also noted that randomized controlled clinical trials may not be the ethically appropriate response to distributions of expert judgment that meet the formal conditions of clinical equipoise (e.g., very low or very high spread) as we are representing it here.

The above analysis, however, should be considered against several limitations. First, we have already noted that our framework for analyzing equipoise necessarily elides some of the complexity in an individual expert's beliefs. Simplifications in our representation make it difficult to distinguish between high confidence in small benefits vs. low confidence in substantial benefits. It also elides complexity in which benefits different stakeholders might expect to result from each intervention. One expert's opinion that *A* is superior to *B* could be based on a different set of criteria from the one used by a different expert who shares the same view or who believes that *B* is likely superior to *A*. For example, some experts might regard the clinical merits of *A* and *B* as roughly equivalent while preferring *A* for reasons relating to ease of administration, cost, simplifying formularies, and other programmatic considerations, while some might regard *B* as preferable to *A* because they have similar views about efficacy but regard *B* as having a more benign side-effect profile (Shah et al. 2021). Such complexities do not alter the main conclusions of our analysis but they would need to be taken into account when evaluating the merits of any particular trial.

Second, our analysis is based on ideal type distributions. Expert beliefs will often reflect hybrids of the typologies we describe (indeed, previous studies of expert community forecast bear this out; Benjamin et al. 2020). We nevertheless think a focus on ideal type distributions helps surface issues that would otherwise not be appreciated. Third, our analysis leaves much unresolved: how much "mass" must be on either side of 50% to count as clinical equipoise? Should a group of experts' extreme confidence in favor of one treatment be weighed against another group of experts' tepid confidence against it, when deciding whether both conditions of clinical equipoise are fulfilled? We have also not addressed the many practical challenges associated with eliciting expert judgments that are the inputs into these distributions. We also regard clinical equipoise as a moral construct rather than a mathematical formalism, and as such, the challenges it faces in being operationalized are not much different than those for prescriptions like "favorable risk/benefit" or "valid informed consent."

Despite these limitations, our analysis has several broad implications for the ethical evaluation and implementation of randomized trials. First, classic formulations of clinical equipoise elide varieties of community uncertainty. As we have argued above, different types of distribution entail different states of existing evidence, different strengths of evidence needed to disturb equipoise, and different possibilities for patient preferences. For example, merely satisfying the criteria of uncertainty in the community does not entail that a trial is ethical. Different types of uncertainty call for different approaches in study design, consent, and monitoring.

Second, it follows logically from the above that proper design, review, conduct, and consent for clinical trials requires an assessment not merely of the existence of uncertainty in an expert community, but its form. Funders, ethics committees, and data monitoring committees should know how a given research question might map to a distribution of community uncertainty, and they should adopt design, review, monitoring, and consent standards accordingly. How the distribution of community belief should be assessed and presented—what has elsewhere been called meta-knowledge—we leave for future analysis.

NOTES

1. Here, we acknowledge that some ethicists have questioned the utility and soundness of clinical equipoise as an ethical principle that should govern randomization (see, for example, Miller 2012). However, it remains a widely accepted design principle in clinical epidemiology and has been endorsed by several human protections policies, including Canada's Tri-Council Policy Statement. Rather than rehash this discussion, the present analysis proceeds from the premise that clinical equipoise provides ethical guidance in clinical research.
2. Critics of clinical equipoise, like Fred Gifford, have argued that physicians and policy makers face very different decisions (caring for individual patients vs. policy decisions about drug approval), that they therefore have different thresholds of evidence, and equipoise for the former will be disturbed before sufficient evidence has been generated to persuade the latter. This might be interpreted as anticipating our argument that the welfare and social value elements of equipoise can diverge. Although there is insufficient space to fully rebut this criticism here, we think Gifford's view is mistaken. In part, this is because it fails to take seriously the extent to which individual clinical experts might disagree. Critics who regard a novel approach as unlikely to succeed may require more evidence, or a different kind of evidence, to

alter their assessment than champions of the new approach. Additionally, trial outcomes mature on different timelines, and trials have varying power to detect the different effects relevant to decision-makers' assessments of a drug. A drug might prove extremely effective—producing a tiny p-value on a primary efficacy outcome in a randomized trial. Yet this might fail to persuade decision-makers (clinicians or policy makers), because the drug harbors a risk of a rare and fatal toxicity that might not emerge with high frequency or in the short term. There are no principled reasons to expect that regulatory and physician decision-making would—or should—diverge on the threshold of evidence they require to decide a drug's utility, and the uncertainty they might tolerate on the various outcomes. Finally, the critique fails to acknowledge that randomized trials are not the sole modality for resolving uncertainties. Quasi- or natural experiments, observational studies, or studies in animal models have always been used to supplement evidentiary judgment in settings where randomized trials are patently unethical or infeasible, and we see no reason why policy makers need rely on randomizing patients to resolve lingering uncertainties.

3. Inclusion of early phase trials in this context raises questions about whether clinical equipoise extends to early phase trials—a discussion that will not be reviewed here. For various treatments of this question, we direct the reader to Weijer and Miller (2004), Anderson and Kimmelman (2010), and London (2021).

REFERENCES

Anderson, James A., and Jonathan Kimmelman. 2010. "Extending Clinical Equipoise to Phase 1 Trials Involving Patients: Unresolved Problems." *Kennedy Institute of Ethics Journal* 20 (1): 75–98. https://doi.org/10.1353/ken.0.0307.

Benjamin, Daniel M., David R. Mandel, Tristan Barnes, Monika K. Krzyzanowska, Natasha Leighl, Ian F. Tannock, and Jonathan Kimmelman. 2020. "Can Oncologists Predict the Efficacy of Treatments in Randomized Trials?" *The Oncologist* 26 (1): 56–62. https://doi.org/10.1634/theoncologist.2020-0054.

Flaherty, Keith T., Igor Puzanov, Kevin B. Kim, et al. 2010. "Inhibition of Mutated, Activated BRAF in Metastatic Melanoma." *New England Journal of Medicine* 363 (9): 809–19. https://doi.org/10.1056/NEJMoa1002011.

Freedman, Benjamin. 1987. "Equipoise and the Ethics of Clinical Research." *New England Journal of Medicine* 317 (3): 141–45. https://doi.org/10.1056/NEJM198707163170304.

Fried, Charles. 1974. *Medical Experimentation Personal Integrity and Social Policy*. Amsterdam: North-Holland Pub. Co.

Gifford, Fred. 1986. "The Conflict Between Randomized Clinical Trials and the Therapeutic Obligation." *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* 11 (4): 347–66. https://doi.org/10.1093/jmp/11.4.347.

———. 1995. "Community-Equipoise and the Ethics of Randomized Clinical Trials." *Bioethics* 9 (2): 127–48. https://doi.org/10.1111/j.1467-8519.1995.tb00306.x.

Greene, Jeremy A. 2006. *Prescribing by Numbers: Drugs and the Definition of Disease*. Baltimore, MD: Johns Hopkins University Press. https://jhu.pure.elsevier.com/en/publications/prescribing-by-numbers-drugs-and-the-definition-of-disease-3.

Hellman, Deborah. 2002. "Evidence, Belief, and Action: The Failure of Equipoise to Resolve the Ethical Tension in the Randomized Clinical Trial." *The Journal of Law, Medicine & Ethics* 30 (3): 375–80. https://doi.org/10.1111/j.1748-720x.2002.tb00406.x.

London, Alex John. 2018. "Learning Health Systems, Clinical Equipoise and the Ethics of Response Adaptive Randomisation." *Journal of Medical Ethics* 44 (6): 409–15. https://doi.org/10.1136/medethics-2017-104549.

———. 2020. "Equipoise: Integrating Social Value and Equal Respect in Research with Humans." In *The Oxford Handbook of Research Ethics*, edited by Ana S. Iltis and Douglas MacKay, 1–20. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780190947750.013.13.

———. 2021. *For the Common Good: Philosophical Foundations of Research Ethics*. Oxford Scholarship Online. New York, NY: Oxford University Press.

Miller, Franklin G. 2012. "Clinical Equipoise and Risk-Benefit Assessment." *Clinical Trials (London, England)* 9 (5): 621–7. https://doi.org/10.1177/1740774512450952.

Rettig, Richard A, Peter D. Jacobson, Cynthia Farquhar, Wade M. Aubrey. 2007. *False Hope: Bone Marrow Transplantation for Breast Cancer*. Oxford: Oxford University Press.

Schwartz, Theodore B., and Curtis L. Meinert. 2004. "The UGDP Controversy: Thirty-Four Years of Contentious Ambiguity Laid to Rest." *Perspectives in Biology and Medicine* 47 (4): 564–74. https://doi.org/10.1353/pbm.2004.0071.

Shah, Seema K., Alex John London, Lynne Mofenson, James V. Lavery, Grace John-Stewart, Patricia Flynn, Gerhard Theron, et al. 2021. "Ethically Designing Research to Inform Multidimensional, Rapidly Evolving Policy Decisions: Lessons Learned from the PROMISE HIV Perinatal Prevention Trial." *Clinical Trials (London, England)* 18 (6): 681–9. https://doi.org/10.1177/17407745211045734.

Sheehan, Mark, Claire Timlin, Ken Peach, Ariella Binik, Wilson Puthenparampil, Mark Lodge, Sean Kehoe, et al. 2014. "Position Statement on Ethics, Equipoise and Research on Charged Particle Radiation Therapy." *Journal of Medical Ethics* 40 (8): 572–5. https://doi.org/10.1136/medethics-2012-101290.

Smithson, Michael. 1999. "Conflict Aversion: Preference for Ambiguity vs Conflict in Sources and Evidence." *Organizational Behavior and Human Decision Processes* 79 (3): 179–98. https://doi.org/10.1006/obhd.1999.2844.

———. 2013. "Conflict and Ambiguity: Preliminary Models and Empirical Tests." Presented at the 8th International Symposium on Imprecise Probability: Theories and Applications, Compiegne, France, 2013. https://openresearch-repository.anu.edu.au/bitstream/1885/28941/2/01_Smithson_Conflict_and_Ambiguity:_2013.pdf.

———. 2015. "Probability Judgments under Ambiguity and Conflict." *Frontiers in Psychology* 6. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00674.

Weijer, Charles, and Paul B. Miller. 2004. "When Are Research Risks Reasonable in Relation to Anticipated Benefits?" *Nature Medicine* 10 (6): 570–73. https://doi.org/10.1038/nm0604-570.

Wendler, David, and Annette Rid. 2017. "In Defense of a Social Value Requirement for Clinical Research." *Bioethics* 31 (2): 77–86. https://doi.org/10.1111/bioe.12325.

Wenner, Danielle M. 2017. "The Social Value of Knowledge and the Responsiveness Requirement for International Research." *Bioethics* 31 (2): 97–104. https://doi.org/10.1111/bioe.12316.