
CLINICAL TRIAL PORTFOLIOS

A CRITICAL OVERSIGHT IN HUMAN RESEARCH ETHICS, DRUG REGULATION, AND POLICY

BY ALEX JOHN LONDON AND JONATHAN KIMMELMAN

Clinical trials cannot be reviewed in isolation from each other. Trial portfolios—series of trials interrelated by a common set of objectives—must be considered. The acceptability of a trial can change depending on the characteristics of the portfolio in which it is embedded. And different ways of structuring trial portfolios have different risks, benefits, and prospects for generating socially valuable information.

Evidence generated from clinical trials is critical to a wide range of stakeholders in making decisions and fulfilling their moral obligations. Regulators rely on clinical trials for drug approval and labeling decisions. Health systems, medical societies, and expert committees rely on the evidence from trials to determine treatment and utilization policy or to set treatment guidelines. Clinicians use this evidence to support treatment recommendations, and patients rely on it to decide which courses of care to undertake. Many of these stakeholders presume that the careful review of individual stud-

ies is enough to address the ethical and scientific questions and problems that arise in clinical trials. For example, new cancer drugs are routinely granted regulatory approval on the basis of a single trial showing large effects. Nowhere is this presumption more apparent than in the current system of research ethics and oversight.

The fields of research ethics and oversight presume that nearly all relevant ethical issues in research involving human participants can be identified and dealt with by the careful review of individual study protocols or their components. Its core institution, the institutional review board, and its central documents—such as the U.S. Common Rule (the federal regulations governing research with human subjects), *The Belmont Report*, the Declaration of

Alex John London and Jonathan Kimmelman, “Clinical Trial Portfolios: A Critical Oversight in Human Research Ethics, Drug Regulation, and Policy,” *Hastings Center Report* 49, no. 4 (2019): 31-41. DOI: 10.1002/hast.1034

Helsinki, and Council for International Organizations of Medical Sciences guidelines—provide moral standards for evaluating individual study protocols. With rare exceptions,¹ ethical and policy debates focus on moral dimensions of individual research procedures (such as sham surgery² or research biopsy³), particular study designs (such as cluster randomized trials⁴), and the ethics of particular contested trials (such as the Surfactant Positive Airway Pressure and Pulse Oximetry Trial [SUPPORT]⁵). Focusing ethical analysis and oversight on individual trials presumes that information reported in individual trial protocols is sufficient to render a sound ethical assessment of a trial or its results and that, if each study protocol meets an ethically acceptable standard, then the entire enterprise of human research will meet that standard as well. The problem, however, is that both of these presumptions are false.

In what follows, we demonstrate that explicit consideration of trial portfolios—series of trials that are interrelated by a common set of objectives—is crucial for two distinct but related reasons. First, the ethical acceptability and evidentiary probity of individual trials can change depending on the characteristics of the portfolios in which they are embedded. Second, how trial portfolios are composed, how well they are coordinated, and how efficiently they use information determines the balance of risks and benefits they present as well as their different prospects for generating socially valuable information; these three factors also raise distinct questions of justice.

Our analysis has implications for many stakeholders in research. We show that a set of what are currently treated as private decisions of study sponsors raise ethical questions that require explicit justification and that make them legitimate targets for policies that encourage fairer and more efficient portfolios. Oversight and regulatory bodies may need to adjust how they evaluate research

claims. Clinicians, health systems, policy-makers, and other consumers of research information may need to broaden the scope of information they use to evaluate treatments and services. And bioethics and research ethics need to better facilitate discussions about the fairness and economy with which the costs and burdens of medical uncertainty are distributed across health care and research systems.

The Concept of the Drug Trial Portfolio

One of the main goals of clinical translation is to identify clinically useful interventions (from now on, we will refer to these simply as “drugs”) and to generate sufficient evidence to warrant or discourage an intervention’s clinical use. Establishing the clinical utility of a drug for a particular indication requires a sequence of studies. We call the sequence of studies in which a drug is tested in a particular indication a “research trajectory.”

Research trajectories involve a division of labor between different types of studies. Typically, a research trajectory begins with a hypothesis that a drug may have clinical utility in a particular indication. Early studies aim to explore hypotheses about how features of the drug’s use—such as dose, schedule, co-interventions, and so on—might modulate its clinical effects. The goal of these exploratory trials is to identify the ensemble of practices⁶ most likely to result in clinical utility. Once this has been identified, late-phase, confirmatory trials subject the deployment of a drug within that package of practices to testing that provides a more reliable estimate of treatment effects.

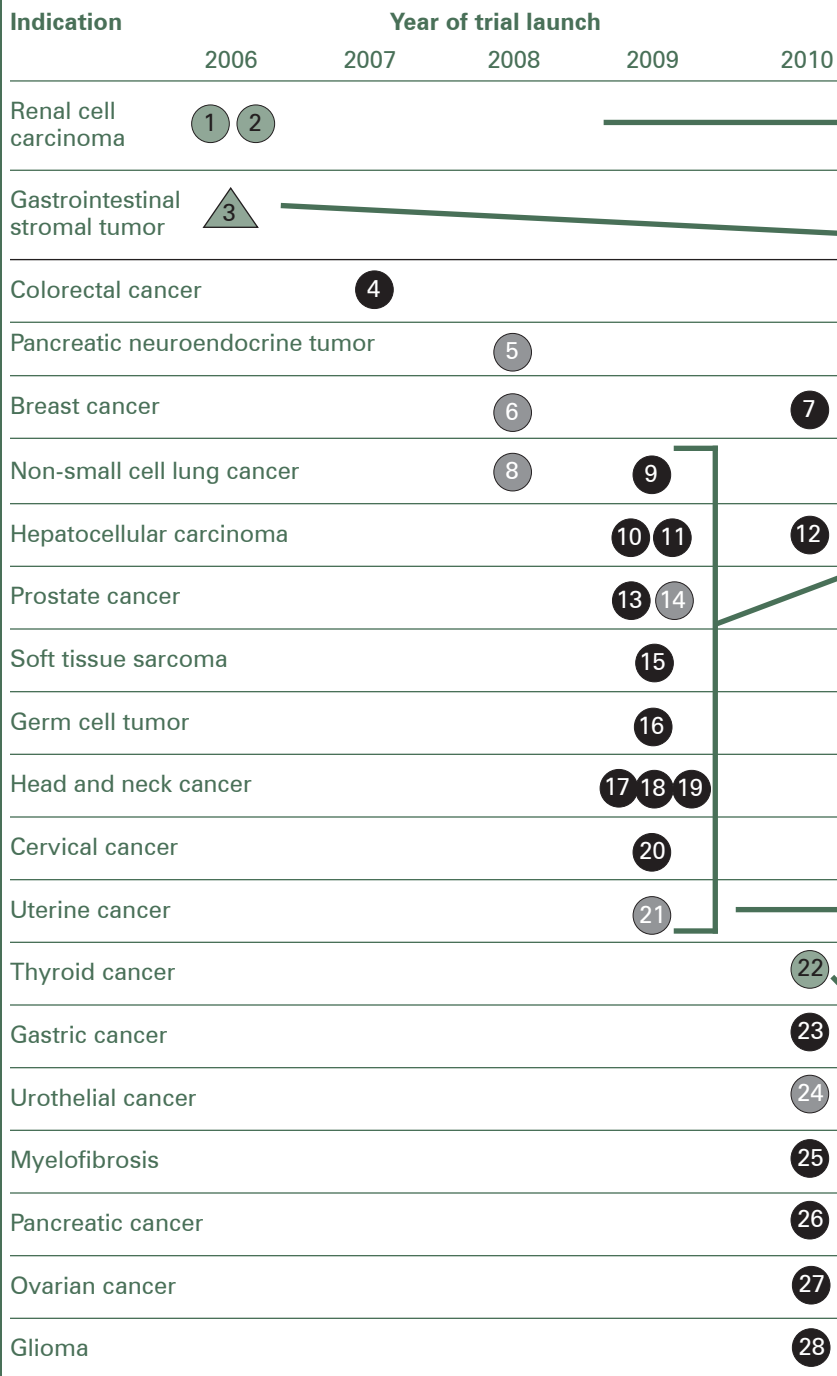
For example, the trajectory of development of sunitinib as a treatment for renal cell carcinoma began when patients with this malignancy showed promising responses to it in a phase I trial.⁷ The hypothesis that sunitinib could be effective for renal cell carcinoma was then tested directly in a

single-armed phase II study.⁸ After this study was positive, researchers conducted a phase III study aimed at testing a more defined hypothesis, namely, that sunitinib could be effective as first-line therapy if patients who were at a higher risk for serious side effects due to the drug’s cardiotoxicity were excluded.

A drug trial portfolio consists of the set of trials in various research trajectories in which the same drug is tested against a range of indications. These studies are linked by a network of evidentiary connections such as assumptions about the mechanism of action of a drug and the pathophysiology of disease. The development of sunitinib for renal cell carcinoma thus represents one research trajectory within the larger portfolio of sunitinib research. While researchers were pursuing the sunitinib-renal cell carcinoma trajectory, other researchers were pursuing trajectories testing sunitinib for gastrointestinal stromal tumor, breast cancer, and lung cancer.⁹ In cancer, these distinct research trajectories often develop out of common exploratory studies that test the same drug in multiple indications looking for signals of promise. What we are calling drug trial portfolios are distinct from “indication trial portfolios”—trial portfolios in which a range of drugs are tested against the same indication.¹⁰ The figure in this article graphically represents the completed drug trial portfolio for sunitinib monotherapy as of 2010, as well as moral dynamics that we will discuss below.

For many drugs, trial portfolios consist of a small number of trials and trajectories. But for drugs that are considered breakthroughs, trial portfolios can be enormous. For the “silver bullet” anticancer drug imatinib, the first ten years of testing resulted in a portfolio consisting of thirty-seven trajectories and 128 trials.¹¹ For sorafenib, one of the first multityrosine kinase inhibitors, the first thirteen years of testing resulted in a portfolio consisting of twenty-six trajectories and 203 trials. The

Sunitinib Monotherapy Trial Portfolio, 2006-2010



Each row represents a trajectory, a sequence of studies. As of 2010, twenty indication trajectories had published results.

This portfolio contained only one confirmatory trial. Trials 15 and 22 led to recommendations in clinical practice guidelines. To date, no trials confirming effects in these indications have been launched.

In 2009, seven new trajectories published results, with three trajectories involving more than one exploratory trial. Might some of these, as well as those published in 2010, have been avoidable had studies been pursued in a staggered fashion?

At about the eleventh trajectory published after initial trajectories showing efficacy, it was clear that sunitinib was unlikely to show activity as a single agent against other indications. This information might have discouraged the launch of new indication trajectories in 2010 or beyond.

Study 22 stands out as a single positive trial amid a sea of negative and inconclusive studies. Might this estimate reflect a false positive? Since 2010, two more thyroid cancer trials showed lower response rates.

Each numbered node represents one trial, color coded based on whether the results were positive (green with a black number), nonpositive (black), or inconclusive (gray with a white number). Triangles indicate trials that are confirmatory and phase III, and circles indicate exploratory and phase II trials. This graph does not include trials pursued in a trajectory after a given indication received FDA approval. The data in this figure derive from B. Carlisle et al., "Benefit, Risk, and Outcomes in Drug Development: A Systematic Review of Sunitinib," *Journal of the National Cancer Institute* 108, no. 1 (2016): doi:10.1093/jnci/djv292. (Graphical representation of trial portfolios is described further in S. P. Hey, C. M. Heilig, and C. Weijer, "Accumulating Evidence and Research Organization [AERO] Model: A New Tool for Representing, Analyzing, and Planning a Translational Research Program," *Trials* 14 [2013]: doi:10.1186/1745-6215-14-159.)

trial portfolio for the blockbuster antiepilepsy drug pregabalin produced a portfolio of twenty-four trajectories and seventy-three trials.¹² Checkpoint inhibitors have rapidly transformed cancer care in the last five years. According to one report, there are 803 open trials testing checkpoint inhibitors for treatment of cancer, with over 166,000 patient slots.¹³

Many institutions concerned with clinical research—such as institutional review boards, funding bodies, and regulatory agencies—view clinical trials as the primary mechanism for generating knowledge about treatments. The methods and design of clinical trials are scrutinized for a variety of scientific characteristics, including the steps taken to guard against various forms of bias or confounding (for example, a difference between groups of participants within a study with respect to characteristics that affect the association between the study intervention and the outcome measures). How a trial is designed and executed and how well information is used within that trial affect the risks to which participants are exposed, the prospect that those risks are offset by direct medical benefit to participants, and the prospect that those risks are offset by the production of medical information that has scientific and social value.

Trial portfolios raise two sets of related but distinct challenges for ethics, policy, and decision-making. First, the ethical acceptability of individual trials and the strength of the evidence they produce, considered in isolation, can change when considered in the larger context of a trial portfolio.¹⁴ As a result, the assessment of individual studies is incomplete if not carried out at least partly with consideration of the context of the trial portfolio to which those studies belong. Second, trial portfolios themselves can present a more or less optimal balance of risks and benefits, present different prospects for generating socially valuable information, and raise distinct questions of justice depending on what trials they contain

and how well trials within them are coordinated with each other. Decisions and policies that affect trial portfolio properties and composition therefore warrant explicit ethical and policy consideration.

Portfolio Composition, Risk, and Expected Benefits

Currently, the assessment of risk in research involves balancing the burdens and potential harms within individual studies against the likelihood of direct benefit to participants and against the value of the information that the investigations are expected to produce. Because trial portfolios are interrelated sets of studies, the composition of a portfolio can affect both the merits of individual studies in that set and the overall risk-benefit associated with the portfolio.

The composition of trials in a portfolio matters because of the aforementioned division of scientific labor between studies in a research trajectory. Exploratory studies (typically, phase I and II trials) use surrogate end points in small samples of participants over relatively short periods to identify and define the ensemble of practices most likely to result in clinical utility. Confirmatory trials (typically, phase III trials) test whether a drug, delivered according to this ensemble of practices, has clinical value by enrolling larger populations of patients and often targeting clinical end points. As a result, these trials require more time and resources to complete. In the absence of confirmatory trials, the results of exploratory studies have asymmetric value. When these studies are negative (in other words, they fail to support the hypothesis around which the trial was designed), they generate information that is valuable to a range of stakeholders: drug developers, clinicians, policy-makers, and patients learn that this drug is unlikely to have a beneficial effect when delivered as tested. However, the information from positive exploratory trials is often unreliable. Because they lack specificity for detecting clinical

promise, treatment effects on surrogate end points in small studies conducted over a short time might not translate into beneficial effects in the clinical setting. As a result, the information from such studies is most useful to researchers who can subject such findings to confirmatory trials.

To appreciate the ethical consequences of different compositions of studies in alternate trial portfolios, consider a drug development portfolio in which a prior trajectory has resulted in regulatory approval for the use of the drug in a first indication. Knowledge of the drug's pharmacology, preclinical evidence, and an understanding of disease mechanism suggest strong promise in two indications, although there is a range of other indications that might respond to the drug as well. For simplicity, now imagine two alternative strategies for expanding this portfolio, each potentially involving a thousand patients.

The first strategy expands the portfolio by adding two small trials (a and b) enrolling one hundred patients each, exploring a drug's activity in the two indications of promise. If either of these studies shows a signal of promise, a large confirmatory trial involving eight hundred participants is carried out. The second strategy expands the portfolio by initiating ten small trials in ten new trajectories (trials a and b plus eight other exploratory trials), each enrolling one hundred patients, aimed at exploring the potential of the drug against ten different indications.¹⁵

The composition of a trial portfolio affects the merits of the individual studies in it. Trials a and b have greater social value in the context of the first way of expanding the portfolio because the expected value of an exploratory study depends, in part, on whether it is a member of a trial portfolio in which signals of promise are likely to be subject to confirmatory testing. In this portfolio, trials a and b perform the task to which they are best suited—supplying information to researchers that can be subjected to

confirmatory testing. In the second way of expanding the portfolio, these studies have less value because their results, on their own, are unreliable and unsuited to guiding clinical practice. Positive findings from exploratory studies that are not followed by confirmatory testing can entice patients and providers to consider off-label use of the drug for the promising indication. These stakeholders, along with health systems, policy-makers, and third-party payers, are left without sufficient evidence to warrant using that intervention in the relevant patient population. The result is that potentially large populations of patients are exposed to drugs that are possibly ineffective or harmful.¹⁶

Ethical review practices do not usually consider the composition of studies in a trial portfolio when evaluating individual protocols. Nor do they necessarily contemplate the prospect of follow-up trials that would be necessary to redeem the burdens and social investments for an exploratory trial. As a result, ethical review practices often involve tacit assumptions that exploratory studies, if positive, are likely to feed into confirmatory testing.¹⁷ At the very least, such considerations should be placed in the foreground and made the subject of explicit ethical assessment, if not regulatory evaluation.

Trial portfolios also have properties that should be subject to ethical assessment in their own right. In the choice between alternative trial designs, if all else is equal, an approach that reduces the number of people harmed without detracting significantly from the quality of the evidence produced is ethically preferable to one that results in a larger number of people harmed for roughly the same gain in information. This principle applies at the level of trial portfolios as well.

If we consider the likelihood that risks from study participation in any given trial will be justifiable in light of the prospect of direct benefit to participants, then the first portfolio is ethically preferable. That portfolio

concentrates on indications in which there is prior signal of promise and enrolls additional patients only if those signals are borne out in subsequent studies. The second portfolio allocates patients to a range of trajectories for which evidence of promise is weak, increasing the proportion of study participants unlikely to receive direct medical benefit.

If we consider the value of the information that these portfolios are expected to produce, we also see that the first portfolio, as a whole, is ethically preferable to the second. The first portfolio has the prospect of producing evidence sufficient for guiding clinical practice because any

in different indications, generally because some aspect or aspects of its activity might be useful against some set of pathophysiological mechanisms shared by different diseases. As a consequence, evidence from trials testing a drug against one disease is relevant to the probability that a different but related disease might respond to the drug. The pacing and coordination of studies in a portfolio determines the extent to which these evidentiary linkages are exploited to reduce the burdens necessary to generate reliable medical evidence.

“Pacing” refers to the timing with which new trials and trajectories are initiated. When trajectories

As portfolios expand—as additional trials are added—it becomes more difficult to avoid false positives or inaccurate estimates of treatment effects.

signal of promise will be followed up with confirmatory testing. Moreover, the second portfolio is less able to guide clinical practice because it does not subject any positive results that emerge from exploratory trials to confirmatory testing.

The composition of studies included in a drug trial portfolio thus affects the value of the individual trials included in that portfolio, the number of people placed at risk in a group of studies, the extent to which those harms are likely to be offset by direct benefits to participants, and the expected value of the information resulting from the series of tests.

Evidentiary Linkages and Efficient Knowledge Production

In a trial portfolio, trials in different trajectories pursue hypotheses that are related to each other. A drug trial portfolio features a drug that is tested

are launched simultaneously, lessons learned in one trajectory about toxicities, optimal dosing, scheduling, and response that affect the window of clinical utility cannot be applied in other trajectories.¹⁸ There are no opportunities to absorb emerging insights into the planning and design of new trajectories so that hazards can be avoided and inquiries can be concentrated on promising avenues.

“Coordination” refers to the degree to which information from studies in a portfolio is incorporated into or influences the conduct of other studies in the portfolio. Recently, trial designs have been proposed that evaluate a larger portion of a trial portfolio under a uniform statistical and methodological framework. Basket, umbrella, platform, and some expansion cohort trials¹⁹ represent an effort to subsume many trials within a unified design that integrates evidence across studies, allowing unpromising

trajectories to be quickly identified and terminated so that resources can be shifted to more promising indications. Trial designs that efficiently use evidentiary linkages between studies of a drug in different indications can generate reliable medical information using fewer participants.²⁰

Our aim is not to advocate for specific trial designs, but to use these examples to illustrate two points about research ethics. First, both pacing and coordination can alter the balance of risks and benefits in individual trials. Whether unnecessary risks and burdens have been eliminated from individual studies cannot be determined unless researchers or other stakeholders consider how information from other studies in the portfolio could be used to increase the efficiency of the study design. Second, whether portfolios make efficient use of evidentiary connections affects the number of patients that are burdened or harmed in the process of generating the same medical evidence. Because rapidly paced and poorly coordinated portfolios make an inefficient use of resources and generate risks and burdens that can be eliminated within alternative portfolios, their risks and burdens are not necessary to generate reliable medical evidence. The same moral principles that support eliminating unnecessary risks and burdens within individual trials support an ethical preference for trial portfolios that make a more efficient use of evidentiary linkages between trials in a portfolio. Nevertheless, decisions about how studies in a drug trial portfolio are paced and coordinated are not the focus of explicit policy, oversight, or review. Even though these decisions affect the health and welfare of study participants and the use of scarce resources, they are left to the discretion of private parties pursuing their own interests.

Portfolio Expansion and Inferential Power

Given the linkages between trials in a portfolio, the evidentiary

value of any of its individual studies cannot be evaluated without considering the other studies conducted in the portfolio. There is another way in which studies in a trial portfolio are interlinked even when they explore radically different hypotheses. Adding new trials to a portfolio expends a portion of that portfolio's ability to detect true treatment effects. The more trials are added to a portfolio, the more resources are needed to estimate the efficacy of a treatment accurately. This effect derives from two features of trial portfolios that become increasingly important as portfolios grow in size: random variation in measured effects and heterogeneity in populations or diseases tested.

When testing a hypothesis in a randomized trial, researchers power their studies based on a prespecified tolerance for declaring differences between a treatment and comparator to be "real" even though they are due to chance alone (the value of this tolerance is called an "alpha"). For many clinical trials, researchers use an alpha of 0.05, meaning that they are willing to tolerate a 1 in 20 chance that, because of random variation, they will wrongly accept the hypothesis that a drug has a bigger or smaller treatment effect than the comparator in a randomized trial. Often, however, data monitors wish to probe whether treatment effects are emerging early on so that, if a study is futile or if it is showing a huge treatment effect, the trial can be stopped early. Yet, unless the alpha in a trial is adjusted, the statistical testing of an additional hypothesis increases the probability of a false positive.²¹ Researchers therefore often adjust their alpha so that their overall tolerance for a false positive is still 1 in 20. They might do this by using a very small alpha for interim analysis (say, 0.01) and then a slightly adjusted alpha for the overall trial (such as 0.04 instead of 0.05). This adjustment is called an "alpha spending function."²² Effective trial conduct requires stewarding a tolerance for false-positive results by testing as few hypotheses as possible, thus

minimizing the spending of a trial's alpha.

What is true about false positivity and spending within trials is also true for trial portfolios. The more trials in a trial portfolio that test an intervention in different settings, populations, or subgroups, the greater the odds that some trials will produce false positive results. Because there is random variation across multiple trials within a portfolio, estimates of treatment effect from any one trial must be adjusted in light of effects observed in other trials within the portfolio. Imagine that a drug that has no effect on any disease is tested in a trial prespecifying a tolerance for false-positive results of 5 percent. If that drug is tested in a portfolio consisting of only one trial, then the probability that the portfolio will produce a false-positive result favoring the drug is 2.5 percent (assuming a two-tailed test is used). Now imagine that the same drug is tested in a portfolio consisting of twenty trials. The probability that the portfolio will produce at least one false positive result is more than 40 percent. If the portfolio had forty trials, this probability would jump to 87 percent. As this example makes clear, the greater the number of trials in a portfolio, the greater the probability of erroneously concluding that drug works against a disease for which it is tested.

Additionally, trials within portfolios show variability in treatment effects due to underlying heterogeneity in populations or diseases tested. As a consequence of this heterogeneity, outcomes in each trial in a portfolio also vary randomly around a central effect, the "portfolio mean." Because the variability in treatment effects that are estimated in trials exceed the true variability, trials that show unusually large effects are likely to overestimate efficacy unless they are adjusted downward toward the portfolio mean using a statistical technique known as "shrinking."²³ Similarly, trials showing unusually small effects should be "shrunk" upward toward a portfolio mean. The idea that estimates from

one trial should be adjusted in light of estimates from another trial testing a different disease is highly counterintuitive and hence called “Stein’s paradox.”²⁴ Problems of overestimation are compounded if outcomes of trials within portfolios correlate with each other (for example, when outcomes in breast cancer trials provide information on the probability of detecting efficacy in lung cancer).

Therefore, as portfolios expand—as additional trials are added—it becomes more difficult to avoid false positives or inaccurate estimates of treatment effects. As a result, more resources are needed to avoid these errors, including larger numbers of participants who must be exposed to the burdens and inconveniences of clinical investigation to test a given claim of clinical efficacy. A corollary of this observation is that the risk-benefit ratio for a trial under review is potentially diminished by the launch of other new trials pursuing different hypotheses within the trial portfolio. To accurately assess the inferential power of a trial and whether it is sufficient to offset risks to participants that are not offset by the prospect of direct benefit, individual trials have to be evaluated in light of all other trials in the drug trial portfolio.

Expanding trial portfolios has significant implications for the value of the evidence produced by individual studies and for the number of participants who must be exposed to research risk in order to generate reliable medical evidence. Adding exploratory trials that are not supported by a strong signal of promise increases the probability of spurious positive results. If these results are not subject to confirmatory testing or to correcting in light of the entire portfolio of research, they can mislead a range of stakeholders into undertaking treatments or dedicating resources to interventions that lack clinical utility. These defects in the value of information undermine the justification for exposing study participants to the associated risks from these added exploratory trials.

Adding studies to a portfolio requires using larger numbers of participants in subsequent trials, thereby increasing the number of participants exposed to research risks. However, these problems cannot be identified, let alone addressed, if the fields of research ethics and regulatory oversight limit their attention to the assessment of individual study protocols.

These properties of trial portfolios also have important implications for decision-making in policy and regulation. When companies submit trial results to regulatory agencies like the U.S. Food and Drug Administration for approval, they select among the drug-indication pairings that show

The gatekeeping function of regulators involves establishing evidentiary thresholds for safety and efficacy that balance the need for timely access to medical innovation with the importance of ensuring a sound evidence base for the many stakeholders who rely on medical evidence in their decision-making.²⁶ Together, the considerations addressed above influence whether studies in a trial portfolio are likely to reduce or amplify medical uncertainty, where that uncertainty is addressed, and who bears the cost of dealing with such uncertainties.

For example, drug developers cannot earn revenue from a novel

Decisions about how to pace and coordinate studies in a drug trial portfolio affect the health and welfare of study participants, yet they are left to the discretion of private parties pursuing their own interests.

the greatest efficacy. As noted above, these trials are likely to have overestimated treatment effects and are at elevated risk of generating false positives. Unless the FDA (or guideline developers) adjusts effects observed in trials based on the risk of false positives and by shrinkage, the trial results used in regulatory decisions (or clinical practice guidelines) are likely to be biased—especially when trial portfolios are large.

Portfolios, Medical Uncertainty, and Justice

Part of the scientific and social value of individual studies resides in the prospect that their successful completion will reduce medical uncertainty and contribute to improvements in clinical practice.²⁵

drug until regulators grant a license based on positive confirmatory trials. This provides strong incentive to construct drug portfolios that concentrate on indications with prior signals of promise and that include confirmatory trials. Once a drug is approved, however, companies and academic researchers often expand portfolios by launching many small exploratory studies.²⁷ The incentives for drug companies to run large and expensive confirmatory trials are attenuated when physicians are free to use a drug off label, and many clinical practice guidelines offer recommendations based on exploratory trial evidence.²⁸ Public funding is far more limited for academic researchers wishing to conduct expensive confirmatory trials. As a result, the threshold for initiating exploratory

trials is low, while revenue can be earned (or careers advanced) simply by showing a signal of promise in small, and less reliable, exploratory trials. This creates an incentive for companies and academic researchers to explore a wide range of indications, and even to explore indications where the evidence base for success is dubious but the potential market revenues from a positive signal are sufficiently high.²⁹

Promising but incomplete research trajectories shift the burden of evidence generation from drug developers to patients, clinics, hospitals, and health systems. When these parties use drugs off label, they expend resources to purchase and implement interventions of unproven value and then shoulder the costs of investigating their clinical merits—if such investigations are even carried out. In such cases, drug developers potentially reap a double windfall—they enjoy revenues from expanded sales of drugs without having to cover the costs of validating their efficacy. Taxpayers also pay a double burden, because they foot the bill for publicly funded research while also paying for the reimbursement of off-label medical interventions that are motivated by exploratory but inconclusive studies. Because this windfall comes at the cost of information that patients, providers, policy-makers and others rely on to make momentous decisions, it raises questions about the justice of the system of incentives currently used to align the interests of stakeholders with the production of medical evidence.

Even when developers plan to pursue promising results with large-scale confirmatory trials, they frequently face choices about study pacing and coordination. To maximize the duration of their exclusive right to sell a drug, developers launch multiple studies in parallel. This decision effectively trades an increase in speed and profit against an increase in the number of participants likely to be harmed or burdened in the process. Similarly, sponsors may

be reluctant to include their drug in study designs that maximize the comparability of results from testing different drugs against common indications if this involves disclosing comparative effectiveness information earlier in the life cycle of development. For example, I-SPY 2 is a phase II breast cancer drug trial designed to compare multiple investigational drugs to a common control and to one another.³⁰ Although six drugs have “graduated” from the trial and others are still being evaluated, no direct comparisons of investigational drugs to other investigational drugs have been reported to date. Decisions about how to use the full range of information available in a drug trial portfolio can pit the interests of health care systems, clinicians, and patients in having access to compendious evidence about the relative clinical merits of available treatments against the parochial interests of drug developers.

Similarly, the decision to expand portfolios by running many exploratory studies that are not supported by strong prior evidence of promise expends inferential capital in a way that increases the likelihood of obtaining false-positive results. When such results are not subject to confirmatory testing, spurious findings can drive the decision-making of patients, providers, and policy-makers, increasing costs without improving patient outcomes or health system efficiency. Portfolios containing positive exploratory trials without confirmatory testing therefore have questionable social value at best and, because they can distort the decision-making of many stakeholders, potentially have negative social value.

Because decisions about portfolio expansion and composition take place outside the frame of individual trial protocols, they are not subjected to scrutiny within research ethics or regulatory review. This means that there is frequently no public accountability for these decisions. Their rationale is not known, and how they balance important ethical

values like reducing risk, ensuring social value, and promoting clinical utility over private considerations, such as companies’ financial goals or researchers’ professional interests, remains largely outside the scope of oversight. Treating such decisions as purely private matters for firms or academic investigators fails to account for the social implications of such decisions. The current narrow focus on protocol-level evaluations permits a range of morally relevant inefficiencies without public debate, let alone oversight.

Policy Implications and Possible Responses

The analysis presented here has implications for many stakeholders in the research enterprise. First, if the scientific and ethical merits of an individual trial cannot be reliably assessed in isolation from the larger portfolio of studies to which it is connected, then current practice within research ethics, oversight, and regulation is inadequate. Within the research enterprise, the assessment of the reasonableness of research risks, the distribution of research costs and burdens, and the value of information likely to be produced by individual trials will have to be made with reference to a much larger base of information. Outside of research, stakeholders who rely on evidence generated from individual clinical trials will have to evaluate findings in light of a similarly broadened information base. This includes a reassessment of the adequacy of regulatory procedures for approving new drugs and additional indications.

Second, this analysis suggests that traditional values of research ethics related to risk assessment, the social value of studies, and the justice of the way benefits and burdens of research are distributed should be applied at the level of trial portfolios. Because decisions that are traditionally seen as the private prerogatives of study sponsors or investigators can impinge on each of these values,

these decisions are legitimate targets for ethical assessment and policy-making. In particular, this analysis highlights the ethical issues involved in decisions about how studies are paced and whether to employ comprehensive study designs that more efficiently capture and use information generated from clinical studies. Additionally, some of these decisions involve other important values, such as respecting intellectual property, fostering innovation, enabling freedom of inquiry, and promoting competition in appropriate areas of drug development.

To address the issues we raise here, research ethics, policy, and regulation require mechanisms to evaluate trials or influence their planning in light of the larger portfolios in which they are embedded. The goal of these mechanisms should be to encourage portfolio composition, coordination, and pacing in a manner that minimizes risk, makes efficient use of medical information, promotes social value and facilitates an equitable distribution of the costs and burdens of research.

Many institutions charged with human protections, research policy, and drug approval have limited traction on various aspects of trial portfolios. Institutional review boards (IRBs) and data monitoring committees, for example, are authorized to consider only individual protocols. Funding bodies and drug companies might have control over some—but not all—trials in a portfolio. Drug regulators typically oversee individual trials, and ultimately evaluate single trajectories when making regulatory decisions. As a result, addressing the challenges presented here may require alterations to the current approaches to research ethics, oversight, drug regulation, and health care policy.

Research ethicists and policy-makers need to consider how oversight practices, public funding, drug approval, health care reimbursement, and perhaps other policy instruments like tax law or drug pricing can be altered to be more sensitive to the issues

we have raised here. In the immediate term, IRBs and regulatory authorities can use their existing power to influence the organization of portfolios or to leverage the information contained in them by requiring researchers and funding agencies to submit comprehensive assessments of prior and ongoing studies along with individual protocol submissions.

Permissive ethical approval of clinical trials enables some of the problematic coordination and inefficiencies in trial portfolios. IRBs can play a role in promoting portfolios

that reduce patient burden by requiring sponsors to submit information about the portfolio in which a trial is embedded, alongside supporting evidence.

have been launched, IRBs can withhold approval to encourage at least a more staggered pacing of portfolio expansion. IRBs can also use information on the number of unsuccessful trajectories launched to assess the probability that a new trajectory will lead to clinically actionable evidence. If dozens of trajectories have been launched without leading to the discovery of new responding indications that are well on their way toward confirmation, IRBs should demand especially compelling evidence before approving a

IRBs can play a role in promoting portfolios that reduce patient burden by requiring sponsors to submit information about the portfolio in which a trial is embedded, alongside supporting evidence.

that reduce patient burden by requiring sponsors to submit information about the portfolio in which a trial is embedded, alongside supporting evidence in a trial brochure. For drugs that are not yet approved, this information would specify the composition of studies planned in a drug development trajectory and outline methods being employed to coordinate studies to increase efficient use of evidentiary linkages and reduce unnecessary risk and burden to study participants. For trials being added to trial portfolios involving an already approved drug, IRBs can ask sponsors to present information from public trial registries like ClinicalTrials.gov to show how many other exploratory trajectories within a portfolio have been launched. This can be supplemented with information on how many trajectories have led to results that are clinically actionable. If a large number of poorly coordinated trials

new trajectory. Data safety monitoring bodies should be similarly apprised of parallel trajectories within a portfolio and should use more stopping rules when a trial is testing hypotheses that will be partly addressed in parallel investigations.

Regulatory authorities like the FDA can ask drug companies submitting applications for regulatory approval to also describe all launched trajectories, as well as estimates from completed trials, within a trial portfolio. If regulators state that they will shrink estimates and adjust inferential tests based on portfolio size, drug companies will have incentives to limit testing only to indications supported by a higher level of evidence. Such a proposal is less radical than it sounds, since the pharmaco-epidemiology division of FDA already collects and analyzes safety information for a drug across many different drug development trajectories.

We do not for a moment underestimate the policy challenges of addressing research inefficiencies and inequities that emerge from poor trial portfolio management. For example, it may be difficult for companies and academic researchers to anticipate the way portfolios might grow. Pressures like intellectual property issues will continue to influence the willingness of developers to exploit the full range of emerging information in trial portfolios. Another challenge concerns the illiquidity of research resources: an academic's decision to forgo an exploratory trial does not entail that the resources she might have expended will now be used for a confirmatory trial. Ultimately, new institutions—like portfolio-level data safety monitoring boards—might be needed to encourage better planning, coordination, and use of information generated in trial portfolios. For now, however, our point is a simple one: current systems of research ethics, drug regulation, and evidence synthesis cannot fulfill their mandates without considering how trial portfolios shape a broad range of scientific and ethical aspects of clinical research.

Acknowledgment

We thank Benjamin Carlisle, Carole Federico, Spencer Hey, Adelaide Dous-sau, and Mithat Gönem for ongoing discussions about the concept of trial portfolios, and the participants in the 2018 Carnegie Mellon Center for Ethics and Policy Workshop on Philosophical Issues in Research Ethics for their helpful feedback.

Notes

1. L. Gelinas et al., “When Clinical Trials Compete: Prioritising Study Recruitment,” *Journal of Medical Ethics* 43, no. 12 (2017): 803-9; L. Pierson and J. Millum, “Health Research Priority Setting: The Duties of Individual Funders,” *American Journal of Bioethics* 18, no. 11 (2018): 6-17.

2. W. R. Galpern et al., “Sham Neurosurgical Procedures in Clinical Trials for Neurodegenerative Diseases: Scientific and Ethical Considerations,” *Lancet Neurology* 11, no. 7 (2012): 643-50; A. J. London and J. B. Kadane, “Placebos That Harm: Sham Surgery Controls in Clinical Trials,”

Statistical Methods in Medical Research 11, no. 5 (2002): 413-27.

3. J. Kimmelman et al., “Burdensome Research Procedures in Trials: Why Less Is More,” *Journal of the National Cancer Institute* 109, no. 4 (2017): doi:10.1093/jnci/djw315.

4. C. Weijer et al., “Ethical Issues Posed by Cluster Randomized Trials in Health Research,” *Trials* 12 (2011): doi:10.1186/1745-6215-12-100.

5. J. D. Lantos and C. Feudtner, “SUPPORT and the Ethics of Study Implementation: Lessons for Comparative Effectiveness Research from the Trial of Oxygen Therapy for Premature Babies,” *Hastings Center Report* 45, no. 1 (2015): 30-40.

6. J. Kimmelman, “A Theoretical Framework for Early Human Studies: Uncertainty, Intervention Ensembles, and Boundaries,” *Trials* 13, no. 1 (2012): article 173; J. Kimmelman and A. J. London, “The Structure of Clinical Translation: Efficiency, Information, and Ethics,” *Hastings Center Report* 45, no. 2 (2015): 27-39.

7. S. Faivre et al., “Safety, Pharmacokinetic, and Antitumor Activity of SU11248, a Novel Oral Multitarget Tyrosine Kinase Inhibitor, in Patients with Cancer,” *Journal of Clinical Oncology* 24, no. 1 (2006): 25-35.

8. R. J. Motzer et al., “Sunitinib in Patients with Metastatic Renal Cell Carcinoma,” *Journal of the American Medical Association* 295, no. 21 (2006): 2516-24.

9. B. Carlisle et al., “Benefit, Risk, and Outcomes in Drug Development: A Systematic Review of Sunitinib,” *Journal of the National Cancer Institute* 108, no. 1 (2016): doi:10.1093/jnci/djv292.

10. Although we focus on drug trial portfolios here, we recognize that it may be possible to distinguish other types of portfolios organized around different considerations. An indication portfolio, for example, includes all the drugs tested on a particular indication. Portfolios might be organized around populations or other variables. The limits and implications of our ability to identify portfolios is the subject of further work.

Our use of the concept of a portfolio differs from the way drug companies use the term “portfolio” in several respects. First, drug companies often use this term to refer to the collection of different molecules in their development pipeline. We use this term to refer to the collection of studies testing a particular drug in different indications. Second, whereas drug companies use the term to mark out their intellectual property and their research plans, our use of the term does not presuppose that all of the studies within a portfolio are managed, sponsored, or conducted by a single entity. As we use the term here, portfolios reflect the sum of research activities involving a

particular drug, regardless of sponsorship. Third, “portfolio” is often used to denote a synchronic process of culling drugs that are flagging and of investing in trials of drugs that show promise. Our use of the word “portfolio” is meant to capture diachronicity as well. Trials involving a drug that were completed long ago remain within its “drug development” portfolio.

11. J. Kimmelman, T. Zheng, and B. Carlisle, “Cancer Precision Medicine Drug Development Is Often Empirical: The Case of BCR-ABL Inhibitors,” unpublished dataset, 2017.

12. C. A. Federico et al., “Assessment of Pregabalin Postapproval Trials and the Suggestion of Efficacy for New Indications: A Systematic Review,” *JAMA Internal Medicine* 179, no. 1 (2019): 90-97.

13. L. Brawley, “With 20 Agents, 803 Trials, and 166,736 Patient Slots, Is Pharma Investing Too Heavily in PD-1 Drug Development?,” *Cancer Letter* 42, no. 37 (October 7, 2016), https://cancerletter.com/articles/20161007_1/.

14. We show that studies that appear to be acceptable when viewed in isolation may be problematic when viewed in the larger context of their drug portfolio. Might it also be the case that studies that are viewed as ethically problematic in isolation could be seen as ethically acceptable in this larger context? Although this might be possible, conceptually, the reality of the incentives that stakeholders face to clearly elaborate the ethical merits of individual trials makes this less likely. For example, researchers proposing a drug trial in indication A will often direct reviewers to consider a positive trial result in indication B. Rarely will they direct reviewers to consider negative trials in indications C, D, and E.

15. With surprising frequency, detection of activity in positive exploratory trials is not followed up on with confirmatory trials. For example, in one study of the anticancer drug sorafenib, 11 percent of phase 2 trials produced a positive result on their primary end points but were not followed up with confirmatory clinical trials (see J. Mattina et al., “Inefficiencies and Patient Burdens in the Development of the Targeted Cancer Drug Sorafenib: A Systematic Review,” *PLoS Biology* 15, no. 2 (2017): e2000487).

16. B. Carlisle, C. A. Federico, and J. Kimmelman, “Trials That Say ‘Maybe’: The Disconnect between Exploratory and Confirmatory Testing after Drug Approval,” *BMJ* 360 (2018): doi:10.1136/bmj.k959.

17. Federico et al., “Assessment of Pregabalin Postapproval Trials.”

18. Kimmelman and London, “The Structure of Clinical Translation.”

19. L. A. Renfro and D. J. Sargent, “Statistical Controversies in Clinical Research: Basket Trials, Umbrella Trials, and Other Master Protocols: A Review and Examples,”

Annals of Oncology: Official Journal of the European Society for Medical Oncology 28, no. 1 (2017): 34-43.

20. S. M. Berry et al., "Bayesian Hierarchical Modeling of Patient Subpopulations: Efficient Designs of Phase II Oncology Clinical Trials," *Clinical Trials* 10 (2013): 720-34.

21. T. V. Pereira and J. P. A. Ioannidis, "Statistically Significant Meta-analyses of Clinical Trials Have Modest Credibility and Inflated Effects," *Journal of Clinical Epidemiology* 64, no. 10 (2011): 1060-69.

22. D. L. Demets and K. K. G. Lan, "Interim Analysis: The Alpha Spending Function Approach," *Statistics in Medicine* 13, no. 13-14 (1994): 1341-52.

23. J. Kimmelman, B. Carlisle, and M. Gonen, "Drug Development at the Portfolio Level Is Important for Policy, Care Decisions and Human Protections," *Journal of the American Medical Association* 318, no. 11 (2017): 1003-4; A. M. Lipsky et al., "The Importance of 'Shrinkage' in Subgroup Analyses," *Annals of Emergency Medicine* 55, no. 6 (2010): 544-552.e3; S. J. Kathman and M. D. Hale, "Combining Estimates from Multiple Early Studies to Obtain Estimates of Response: Using

Shrinkage Estimates to Obtain Estimates of Response," *Pharmaceutical Statistics* 6, no. 4 (2007): 297-306.

24. B. Efron and C. Morris, "Stein's Paradox in Statistics," *Scientific American* 236 (1977): 119-27.

25. A. J. London, "Learning Health Systems, Clinical Equipoise and the Ethics of Response Adaptive Randomisation," *Journal of Medical Ethics* 44, no. 6 (2018): 409-15.

26. A. J. London, "A Non-paternalistic Model of Research Ethics and Oversight: Assessing the Benefits of Prospective Review," *Journal of Law, Medicine & Ethics* 40, no. 4 (2012): 930-44.

27. Carlisle et al., "Benefit, Risk, and Outcomes in Drug Development"; Federico et al., "Assessment of Pregabalin Postapproval Trials"; Mattina et al., "Inefficiencies and Patient Burdens in the Development of the Targeted Cancer Drug Sorafenib."

28. Many recommendations in influential clinical practice guidelines are based on low-level evidence—in some cases coming from exploratory trials and, in other cases, coming from subgroup analyses within exploratory trials, observational evidence, or even case reports. Consider the National

Comprehensive Cancer Center guidelines—one of five compendia that are used to determine Medicare reimbursement in the United States (A. K. Green, W. A. Wood, and E. M. Basch, "Time to Reassess the Cancer Compendia for Off-Label Drug Coverage in Oncology," *Journal of the American Medical Association* 316 (2016): 1541-42). In one analysis, 94 percent of recommendations were based on lower-level evidence (T. K. Poonacha and R. S. Go, "Level of Scientific Evidence Underlying Recommendations Arising from the National Comprehensive Cancer Network Clinical Practice Guidelines," *Journal of Clinical Oncology* 29, no. 2 [2011]: 186-91).

29. Carlisle et al., "Benefit, Risk, and Outcomes in Drug Development"; Federico et al., "Assessment of Pregabalin Postapproval Trials and the Suggestion of Efficacy for New Indications"; Mattina et al., "Inefficiencies and Patient Burdens in the Development of the Targeted Cancer Drug Sorafenib."

30. S. Das and A. W. Lo, "Re-inventing Drug Development: A Case Study of the I-SPY 2 Breast Cancer Clinical Trials Program," *Contemporary Clinical Trials* 62 (2017): 168-74.