Both the abundance of new therapeutic strategies in the drug-development pipeline and the high rate of attrition of medical products during clinical trials place extraordinary pressure on stages of drug development in which clinical activity is first evaluated—typically phase 2. The abundant pipeline demands that such trials quickly evaluate candidates, whereas the prospect of heavy human subject burdens and costs of late attrition demands that phase 2 trials accurately predict results of subsequent confirmatory trials.

Concerns about phase 2 predictivity—both in terms of accurate and efficient pipeline screening (that is, negative predictivity) and reduction of late-phase attrition (that is, positive predictivity)—have prompted a series of innovations in phase 2 trial design. However, many of the contemplated trial reforms aimed at boosting phase 2 positive predictivity have important repercussions for human subjects and for the capacity of the research enterprise to discharge its social mission. Here, we articulate four factors that should guide the level of positive predictivity sought in middle stages of clinical development.

**QUEST FOR PREDICTIVE TRIALS**

As many as two-thirds of the interventions entering phase 3 fail to reproduce success observed in phase 2 trials (1). Ostensibly, this poor rate of translation betrays an inefficient use of research resources and needless burdens imposed on patient-subjects. Concerns about the number of negative confirmatory trials have prompted a series of innovations in phase 2 trial design. However, many of the contemplated trial reforms aimed at boosting phase 2 positive predictivity have important repercussions for human subjects and for the capacity of the research enterprise to discharge its social mission. Here, we articulate four factors that should guide the level of positive predictivity sought in middle stages of clinical development.

**PREDICTIVITY AND PROTECTION**

However, trial designs aimed at reducing false positives also have costs for human subjects. Some of the gains in subject welfare described above are offset by greater burden. Introducing randomization in phase 2 studies, for example, roughly doubles the number of patients in trials because they now require comparator arms. Using enrichment designs, pharmacodynamics, or real-time pharmacokinetics all entail more frequent (and often invasive) tissue collection from volunteers. In areas of vaccine development, the quest for predictive phase 2 designs has kindled interest in phase 2 “challenge studies,” which deliberately infect healthy volunteers with a manageable form of disease.

These extra burdens in phase 2 are not morally equivalent to those typically encountered in phase 3. Risks of drug administration in confirmatory trials are ethically justified by clinical equipoise—a state of collective uncertainty about whether experimental treatment is preferable to standard care—and hence can plausibly claim therapeutic value for subjects. In contrast, the case for clinical equipoise is far weaker...
in initial tests of efficacy because evidence of clinical utility is lacking at the outset, and base rates for discovering a useful intervention are low at that stage (4). The risk of more intensive tissue collection or disease challenge are morally justified by ends that are predominantly external to the volunteer.

Minimizing phase 2 false positives also threatens clinical equipoise for subsequent confirmatory trials (5). If the pretest likelihood of successful outcomes are too high, patients randomized to comparators are systematically disadvantaged. Further, knowledge gained per patient enrolled is diminished because less is learned from the successful prosecution of confirmatory trials.

THE RESEARCH ENTERPRISE
Predictive phase 2 trial designs also entail costs for the integrity of the research enterprise. The social mission of clinical research is to furnish health care decision-makers with information and evidence for addressing unmet health needs. The delivery of this social good is threatened by any process that destabilizes the types of stakeholder collaborations that are required by clinical research (6), or where the scarce resources of a research enterprise are not directed in accordance with social priorities. Highly predictive phase 2 designs have potential unintended consequences for each.

First, there are social costs associated with contemplated design innovations. Gains in reducing false positives in phase 2 are potentially at the expense of greater false negatives (that is, eliminating truly useful agents in phase 2 trials). In areas with few therapeutically candidates and pressing clinical need, such loss is associated with large opportunity costs. Further, more predictive designs can strain the capacity of research systems to vet new drug candidates because larger sample sizes, lengthened observation periods, or extensive real-time laboratory analyses demand greater resources. Although preempted confirmatory studies free up resources for vetting other candidates, some resources cannot be redirected in this way. For example, the financial resources that pharmaceutical firms save by avoiding confirmatory testing will be distributed according to company priorities rather than toward upstream investigations of other candidates for the disease.

Second, highly predictive phase 2 studies threaten a fragile social consensus that enables investigators to randomize patients in confirmatory trials and regulators to condition distribution of new drugs on positive, replicated trials that have sufficient power to detect at least commonly occurring toxicities. In recent years, this social consensus has come under repeated attack (7, 8). The prospect that the efficacy of new drugs can be reliably inferred after phase 2 substantially weakens the moral argument for withholding market access until larger studies are completed. This, in turn, threatens the capacity of the research enterprise to rigorously evaluate new drugs before clinical uptake.

Third, negative, adequately powered phase 3 studies often have affirmative scientific and medical value. For agents that have not been licensed, decisive disconfirmation—or safety signals—can inform decision-making for the testing of other drugs in the same class. For licensed drugs that are tested in new indications or combinations, adequately powered negative trials warn caregivers against offering drugs in specified, off-label applications.

DEFINING A SOCIAL OPTIMUM
At a certain point, the gains from reducing false positives in phase 2 are exceeded by social losses. Thus, the task for study planners and funding bodies is to determine the optimal level of phase 2 false-positive results—at which the gains in avoiding negative confirmatory trials outweigh the costs—and then to deploy policies that encourage trialists to approximate this social optimum. We offer four factors that should inform the level and type of error tolerance sought in middle stages of clinical development (Fig. 1).

First, clinical equipoise in confirmatory trials establishes upper and lower boundaries for predictivity in phase 2 studies. Although the precise placement of these moral boundaries is hotly debated, we suggest that, as a general rule, when the base rate of false positives in phase 2 is high, greater effort should be invested in early- to mid-stage testing before advancing drugs into confirmatory trials; when this rate is low, researchers can scale back their efforts in early and middle stages.

Second, the rate of false positives tolerated in phase 2 should be inversely proportional to the abundance of candidates in the pipeline. Where both pipeline abundance and prior odds of finding an active agent are low, the opportunity cost of a false negative will be substantial; therefore, phase 2 studies should be relatively permissive with respect to false positives (9). Where pipeline abundance is high but prior odds of success are low, studies should aim for higher positive predictivity so that resources from preempted confirmatory testing can be redirected toward earlier phases (10). However, if too high a level of positive predictivity is attained and resources from preempted phase 3 studies cannot be rerouted upstream, more resource-intensive phase 2 studies will diminish the pool of resources available for screening new candidates, leading to similar opportunity costs.

Third, predictivity must strike a delicate moral balance. The moral justification for exposing subjects to risk in phase 2 studies differs from that for confirmatory studies. A shift toward more predictive studies entails an escalation of activities in which the medical interests of subjects conflict with research activities. As a general rule, we suggest that where individuals are least able to protect their own interests or endure harms (for example, pediatric populations and economically deprived populations), researchers should tread cautiously with implementing study designs that increase positive predictivity through increases in volunteer burden.

Fourth, the level of positive predictivity sought should be informed by the social utility of decisively negative confirmatory trials. We can envision two circumstances with large social value. The first is when confirmatory-trial outcomes enable the research community to update key theories of pharmacology or pathophysiology that are guiding drug development. Insofar as theories drive development of other drugs, disconfirmatory trials have substantial social utility because they allow drug developers to reassess priorities. In such circumstances, phase 2 designs should have greater tolerance for false positives. A second circumstance in which confirmatory trials are critical is for costly or risky interventions that have been taken up into clinical practice (for example, the off-label use of a drug). In such circumstances, only decisive disconfirmations will be adequate for altering clinical practice, putting a premium on phase 2 studies that produce fewer false positives.

REFERENCES AND NOTES
3. M. R. Sharma, M. L. Maitland, M. J. Ratain, Other paradigms: Better treatments are identified by better trials.
Funding: This work was funded by CIHR (EOG 102823).

Competing interests: The authors declare that they have no competing interests.

Citation: S. P. Hey, J. Kimmelman, Ethics, error, and initial trials of efficacy. Sci. Transl. Med. 5, 184fs16 (2013).
Ethics, Error, and Initial Trials of Efficacy
Spencer Phillips Hey and Jonathan Kimmelman

Sci Transl Med 5, 184fs16184fs16,
DOI: 10.1126/scitranslmed.3005684