

The S • T • R • U • C • T • U • R • E of Clinical Translation: *Efficiency, Information, and Ethics*

BY JONATHAN KIMMELMAN AND ALEX JOHN LONDON

The so-called drug pipeline is not really about drugs and is not much like a pipeline. It is really about the production and dissemination of information, and it is much more like a web. The misunderstanding leads to a poor understanding of what's wrong with clinical translation and how it can be improved.

*All things are poison, and nothing is without poison;
only the dose permits something not to be poisonous.*

—Paracelsus

The last two decades have witnessed a crescendo of allegations that clinical translation is rife with waste and inefficiency. Patient advocates argue that excessively demanding regulations delay access to life-saving drugs, research funders claim that too much basic science languishes in academic laboratories, journal editors allege that biased reporting squanders public investment in biomedical research,¹ and drug companies (and their critics) argue that far too much is expended in pharmaceutical development.

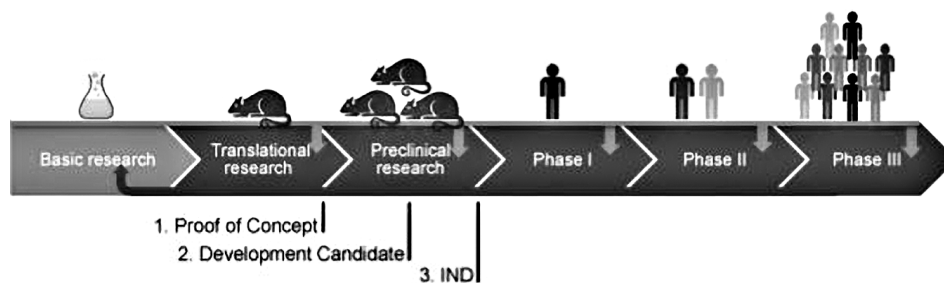
But how should stakeholders evaluate the efficiency of translation and proposed reforms to drug development? Effective reforms require an accurate model of the systems they aspire to improve—their

components, their proper functions, and their pathologies. However, there is currently no explicit and well-developed model of translation for evaluating such criticisms.

In what follows, we offer an explicit model of clinical translation. Many discussions of clinical translation and its pathologies presume that its main output is hardware: new drugs, vaccines, devices, and diagnostics. We disagree. We argue that the principal output of clinical translation is information²—in particular, information about the coordinated set of materials, practices, and constraints needed to safely unlock the therapeutic or preventive activities of drugs, biologics, and diagnostics. Developing this information is far different from a simple linear progression of clinical trials (as shown in figure 1); it requires exploratory sampling of many different elements in this set. Our model highlights the importance of such information for the development and clinical application of new interventions and identifies factors that limit the rate at which new treatment strategies are vindicated in trials. It thus links the efficiency of translation to questions

Jonathan Kimmelman and Alex John London, "The Structure of Clinical Translation: Efficiency, Information, and Ethics," *Hastings Center Report* 45, no. 2 (2015): 27-39. DOI: 10.1002/hast.433

Figure 1.
A Standard Model of Clinical Translation



(“IND” refers to an investigational new drug application, which would be submitted to the FDA.)

Source: California Institute of Regenerative Medicine, “Progress toward Therapies: Path to the Clinic,” <http://www.cirm.ca.gov/path-clinic>.

of fairness and equity regarding when exploratory activities are pursued and how their costs and burdens are distributed. Our model requires further elaboration and refinement, but the form presented here already points to some limitations and liabilities of influential proposals for reforming research. It also reveals some underrecognized opportunities for improving the efficiency of clinical translation.

Why Models Matter

To say that a process is inefficient is to suggest that there is some way of altering it so that the same set of inputs could be used to produce more or better quality outputs. Evaluating claims about translation efficiency—and interventions designed to improve performance—requires an explicit account of the relevant outputs. Any attempt to improve the efficiency of a machine requires knowing something about how the mechanism works, of course, and an adequate model of translation should identify key processes that will generate these outputs. It should also show how these processes relate to the delivery of health services and the advancement of larger public health goals.

Many criticisms of current translation practices are founded on the assumption that the primary outputs of clinical translation are small and large molecules, biologics, devices,

and diagnostics—which we shall refer to collectively as “drugs” or “hardware.” Further, many criticisms of translation imply that almost all unsuccessful efforts in clinical development contribute little or nothing to the practice of medicine. These assumptions imply that an efficient translation enterprise should minimize failed development trajectories (especially trajectories that fail in late stages) and reduce the time and the number of trials needed to prove a new drug’s clinical utility.

One extreme proposal has been championed by the head of the U.S. Food and Drug Administration, Andrew von Eschenbach. Noting that “American patients wait as much as 60% longer than they did in 2005 for new and life-enhancing medical devices . . . to reach the market,” Von Eschenbach advocates that the FDA promote “rapid and efficient” translation by approving drugs based on safety alone, leaving the evaluation of efficacy to postmarket studies.³ Intel founder Andy Grove has advocated a similar proposal, arguing that all patient outcomes in care settings should be collected and fed into databases for the kinds of analyses used in electronic retail to discover factors predictive of “real-life efficacy.”⁴ Some patient advocates and ethicists favor more aggressive, patient-centered trials to enhance efficiency. One set of commentators advocates a “best guess” approach in phase 1 studies,

whereby the dose of a new drug offered to patients would be dictated not by a protocol but rather by what the investigator and patient believe are best doses of a drug.⁵ On the premise that research resources are wasted when downstream users are not interested in the results, another set of commentators favors incorporating patient preferences into decisions about which investigations to pursue.⁶ Still others have advocat-

ed a suite of new research techniques for reducing the number of failed drug development trajectories and the time needed to prove efficacy. Lamenting that “the triple frustrations of long timelines, steep costs, and high failure rates bedevil the translational pathway,” National Institutes of Health director Francis Collins has championed the National Center for Advancing Translational Sciences,⁷ which is chartered to “speed the delivery of new drugs . . . to patients.”⁸ One prominent component of this and many other initiatives is the use of trial designs that restrict participation to patients who harbor biomarkers predicting drug response. In several instances, such approaches have greatly reduced the time and number of patients needed to license drugs.

As intuitive as it may seem, though, the view that new hardware is the primary fruit of translation is inadequate, and conflating the rate of new drug approval with efficiency in translation obscures the importance of a range of activities that are integral to successful translation. In particular, a focus on hardware obscures the fact that exploratory activities in the translation process are critical to developing information that guides later stages of translation and provides the warrant for the therapeutic use of interventions at the bedside. It also obscures the liabilities that some proposals for improving the efficiency

of translation may present for the patients and health care systems that rely on medical innovation.

The Intervention Ensemble

The first problem with treating drugs as the primary output of translation is that drugs alone are not therapeutic agents. Drugs on their own are substances—chemical or biological entities that have the capacity to interact with particular physiological systems. When they are used in the absence of the knowledge and capacity for applying them properly, they are likely to be toxic. One of the key challenges of translation is to discover how they can be clinically useful.

Doing so requires that researchers identify a coordinated set of materials (the substance and any necessary cointerventions), practices (for example, dose, schedules, diagnostic tests, and methods for managing likely side effects), and constraints (for example, of treatment populations, contraindications, and likely side effects) that can be deployed in a treatment setting. Elsewhere, we have called this package of materials, practices, and constraints an “intervention ensemble.”⁹ At a minimum, a clinically useful intervention ensemble consists of the drug, the populations that are likely to benefit from its administration, doses, timing of administration relative to the course of a disease, and cointerventions or medical countermeasures necessary for maintaining clinical benefit. Although the substance may be the most visible element of an intervention ensemble, what is ultimately validated, and what is reflected on an FDA label once a drug is approved, is not merely a substance but a compendious set of coordinated elements for unlocking its clinical utility.

There are an infinite number of intervention ensembles into which a given drug can be integrated, but only a small number, if any, will prove useful. A key goal of clinical translation is to explore various dimensions

of intervention ensembles in order to generate two kinds of information. The first kind is the optimal values of various variables in an intervention ensemble. Many elements in an ensemble—such as dose, timing of drug administration, or diagnostic scores—can be represented as continuous variables. These variables often have values at which a drug achieves the most favorable risk-benefit balance. To maximize the prospect for achieving therapeutic outcomes in subsequent trials or care settings, researchers have to identify the set of approximately optimal values—that is, the dose, diagnostic criteria, schedule, indication, and so on—that promote desirable outcomes and minimize adverse effects.

The principal output of clinical translation is information—about the coordinated set of materials, practices, and constraints needed to safely unlock the therapeutic or preventive activities of drugs, biologics, and diagnostics.

A second, complementary goal in translation is defining the boundaries on dimensions beyond which an intervention ensemble ceases to be clinically useful. This demarcation consists of clarifying the minimal effective and maximum tolerated doses, the earliest and latest a drug can be applied in disease course, and so on. Information on boundaries is crucial for subsequent trials because trials may have to employ suboptimal conditions when testing a drug. For instance, trials of a neuroprotective drug that optimally confers protection three hours after a stroke might be difficult to implement because of the challenges in routing victims into trials; clarity on the upper boundary of the time delay helps researchers design practical trials that are more likely representative of the practice setting. Similarly, information on boundaries is crucial for clinical practice. If stroke patients arrive at the

hospital later than is optimal, should they be given the drug? If patients require dose reductions of an anticancer drug because of underlying medical frailties, is it still worth giving them the drug?

Discovering the set of values that would constitute an intervention ensemble capable of therapeutic activity generally requires sampling within various dimensions in order to locate relevant boundaries.¹⁰ Often, identifying a relevant boundary is an explicit goal of a trial, as when doses of a drug are escalated in a phase 1 trial until a cohort of patients receives a dose that exceeds the limit of tolerability. But boundaries are also inferred from the results of whole trials, such as phase 2 studies showing

that a drug used in advanced disease is ineffective in patients with recent onset disease or that a drug that can be used against one disease is not effective with a different, closely related disease.

If translation is the process of producing the information necessary to use a substance for therapeutic effect, then, given the role in this process of identifying optima and boundaries, there is a clear link between exploratory activities undertaken in the early stages of translation and the use of intervention ensembles in clinical practice. In particular, when an intervention ensemble receives regulatory approval, what is being validated is the cumulative information about how to produce clinical utility with a particular drug for a specific indication in a rigidly defined population. But in the care setting, clinicians will routinely encounter patients who differ in some way from those in the trial

population (perhaps, for example, patients will be younger than those enrolled in trials, or taking other drugs that were excluded in the trials, or presenting later in the course of the disease). At the bedside, therefore, clinicians must adjust the values of relevant dimensions in order to treat such patients. Generating information about boundaries can help guide this process in ways that may be difficult to appreciate when our focus is restricted to the rate at which research is delivering drugs.

As a result, we argue, any attempt to measure or improve the process of clinical translation will need to consider not merely the number of safe and effective new drugs brought to market but also whether the translation process furnishes health care providers with the information they need for adjusting and delivering safe and effective intervention ensembles. The boundaries on intervention ensembles have to be clarified at some point in order to be used to produce clinical utility. What may appear to be a tax or a burden on the translation process—namely, cohorts of patients receiving inactive intervention ensembles or trials that disappoint with negative results—can be integral to demarcating the operational dimensions of useful intervention ensembles. We suggest, therefore, that the operative question for policymakers is not how to eliminate these burdens, but whether to encounter them early in development—under controlled conditions that are designed for evidence collection—or later—in the context of care delivery.

Exploration and Confirmation

If the output of translation is not simply new drugs but a body of information about how to use a drug in order to produce a beneficial effect, then how can this information be generated with minimal burden for patients and cost for drug developers and health care systems? How we answer this question is critical to the way we evaluate proposals to accel-

erate translation and open access to new drugs at earlier stages of research.

At the heart of the translation process is a division of epistemic labor. Since the number of intervention ensembles into which a given drug could be integrated is virtually infinite, testing them exhaustively would demand huge quantities of scarce material and human resources. At best, only a few will prove clinically useful. Efficient clinical translation therefore requires a process in which the effects of adjusting different dimensions can be sampled so that configurations that show signs of clinical promise can be identified and then promoted into further development. For example, one dimension that drug developers commonly explore is disease indication: drug developers often have only a vague understanding of which disease indication will respond to their drug, and surprisingly frequently, a drug developed initially for one disease emerges as a licensed drug for a different disease. Once promising indications are discovered, intervention ensembles involving them are directed into confirmatory testing.

The first task—exploration—involves identifying the dimensions that are relevant to constructing a promising ensemble and searching for optimal values and boundaries on those dimensions. This task is ethically fraught because it involves exposing patients to intervention ensembles that are unproven and may be toxic. In addition, only a fraction of the combinations explored will be promising. Accordingly, sponsors and researchers seek to maximize the number of intervention ensembles sampled while minimizing the resources and number of patients allocated to testing any one of them. This balance is struck by running short, inexpensive trials in small patient populations. These studies typify early-phase trials.

The role of late-phase studies, by contrast, is to decisively test the clinical utility of intervention ensembles produced in early phases of development. They subject explicitly defined

hypotheses about an ensemble's clinical utility to critical testing as a means of ascertaining whether the relevant optima have been found and the coordinated set of materials, practices, and knowledge confers a benefit to patients. The findings from prior exploration help to structure late-phase trials by indicating the parameters on permissible variation of drug delivery and inclusion and exclusion criteria and by defining the clinically relevant endpoints that the study is designed to detect.

This epistemic division of labor—which itself is motivated by the goal of efficiency—has two important consequences. The first is that it imposes predictable limits on the extent to which failures in confirmatory trials can be reduced or eliminated. This is important because critics of the present system point to high failure rates in confirmatory trials as a clear signal of inefficiency. But some failures are predictable, given the inherent limitations of the evidence produced by exploratory studies. Early-phase trials produce large random variation because they are not statistically powered to detect many clinically meaningful effects; an underappreciated flip side of low statistical power is a tendency to produce false positives due to random variation.¹¹ These problems are compounded by systematic errors associated with small and inexpensive trials. For example, to conserve resources, exploratory studies often eschew comparators, thus introducing a risk of experimenter bias—that is, of a tendency to treat or perceive patients differently because of beliefs about the drug they are receiving. They also generally use surrogate endpoints that are observable within shorter frames than the clinically relevant endpoints that are used in late-phase studies but that often produce inflated estimates of clinical utility. As a consequence, even in a perfectly efficient system of clinical development, a predictable number of intervention ensembles will continue to fail confirmatory studies with regularity.

The second consequence is that efficient production of reliable information requires proper coordination of this division of labor. At one extreme, the desire to reduce failure in late stages of translation can itself produce inefficiencies. Because confirmatory studies are large, take longer to run, and are very expensive, they should be used sparingly. But too much time spent exploring dimensions of intervention ensembles exposes patients to harm—especially given that the vast majority of intervention ensembles will prove clinically useless if not harmful. Protracted exploration also consumes resources and delays the confirmatory trials needed to vindicate the therapeutic value of an intervention.

At the other extreme, prematurely advancing an intervention ensemble into confirmatory testing also produces inefficiencies. Consider first the case of negative findings when intervention ensembles are advanced into confirmatory testing although their constituent dimensions have not been explored adequately enough to locate optima and boundaries. Such trials are mostly uninformative.¹² All that has been done is to eliminate one of a virtually infinite number of intervention ensembles that might be created by adjusting the values of relevant dimensions. Negative findings are informative only when the ensemble being tested is believed to capture optimal values on relevant dimensions.

Second, consider the case where researchers happen on an optimal intervention ensemble and a drug is licensed for clinical application. Because researchers have not sampled closely related intervention ensembles, physicians do not know how far they can extend the label and still produce therapeutic effects with the drug. Will a higher dose bring a stronger response? Will a lower dose still be effective but with fewer side effects? What disease variants are candidates for the drug? Can a drug be applied earlier in a disease course or in combination with other drugs? If

such questions are not resolved early on, risks and costs associated with resolving them are likely to be redistributed to health care systems and patients. We return to the implications of this redistribution below.

Information and Theory

To this point, we have emphasized a fairly narrow sense in which information should be seen as the principal output of translation. In particular, we have described the trajectory of translation as characterized by a particular division of epistemic labor. Exploratory studies seek to identify the dimensions of use that will give a drug therapeutic effect and to explore these dimensions in search

conservatism,¹³ and this perspective features prominently in discussions of the efficiency of translation. For instance, the routine focus on the length of a translation trajectory assumes that the relevant evidence in translation is exhausted by what is produced in the trials that lead up to the approval of a particular intervention. It should be clear by now that we believe that the length of a translation trajectory is not an adequate measure of efficiency, in part because a longer trajectory may provide better information about the optima and the boundaries of a given intervention ensemble.

Our view, however, is that focusing on the narrow band of information produced within a translational

Any attempt to improve clinical translation will need to consider not merely the new drugs on the market but also whether the translation process furnishes clinicians with the information they need for treating patients who differ from the participants in drug trials.

of boundaries and optimal values. Once researchers have assembled a coordinated set of materials, practices, and constraints that approximates the optimal values on the relevant dimensions, the therapeutic promise of the intervention ensemble is put to confirmatory testing in late-phase studies. This division of labor produces information that is useful not simply in subsequent stages of development but also in eventual clinical use. We have also suggested that how this division of labor is structured, including when ensembles are advanced into confirmatory testing, influences the value of the information that results from a sequence of studies.

The position articulated so far is “narrow” in the sense that it focuses on the information generated within a translational trajectory about a particular, still-evolving intervention ensemble. Elsewhere, we have called such a position “evidential

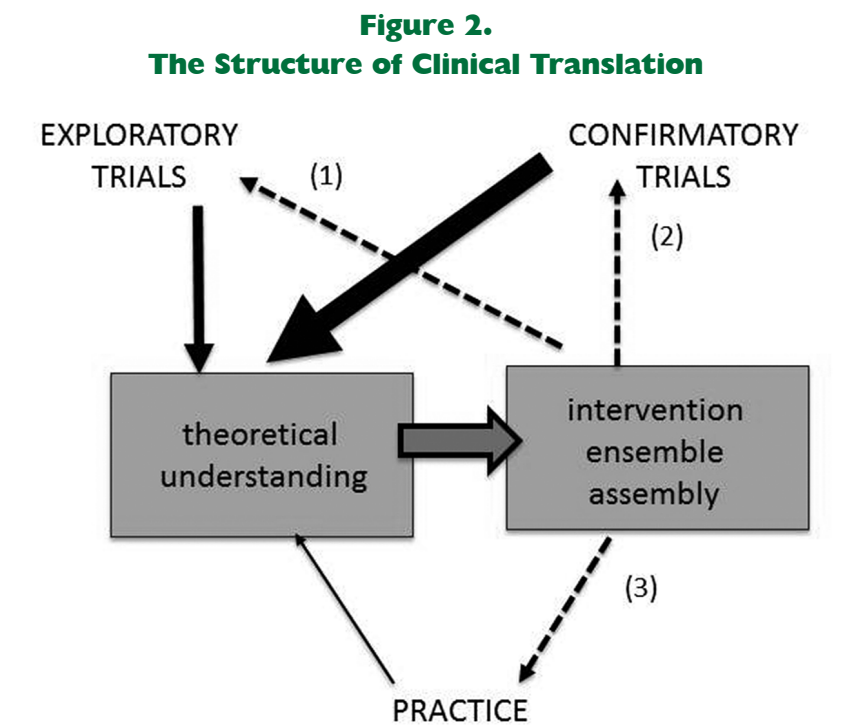
trajectory is too conservative. It misses the broader bandwidth of information that is generated in clinical testing. In particular, it overlooks how information from other trials and from medical practice contributes to theories that guide the decision-making of researchers and caregivers.

To simplify a bit, two kinds of theory are important in this process. Both contemporary drug development and clinical practice are heavily informed by theories of pharmacology (concerning how drugs work) and pathophysiology (concerning the mechanisms underlying disease processes). In particular, theories of pharmacology and pathophysiology increasingly guide the process of identifying the dimensions that are likely to be relevant to creating an effective intervention ensemble. For example, suppose developers discover that a compound has a particular effect on an enzyme in a specific pathway. If

that enzymatic pathway is implicated in the pathophysiology of a set of diseases, then developers may hypothesize that the drug can be useful for those indications. Here, background information is used to postulate a dimension (in this case, related to a similarity in mechanism between various possible indications) that subsequent trials can then explore.

The results of well-designed trials provide information not just about the effects of the specific intervention ensembles they test but also about the larger theories that guide their development and intended use.¹⁴ Patterns of outcomes observed across a series of trials provide information that influences the plausibility of or warrant for different claims and can be used to refine these theories. For example, a range of strategies might be explored in an effort to exploit the role of a particular mechanism in a disease process. Failure across this field of trials might lead to the revision of pathophysiological models of the target condition, help to identify common failure modes, or prompt the postulation of new treatment approaches. Alternatively, a particular treatment strategy might be explored in a set of indications believed to be linked by a common pathophysiology. Success in some indications but not others might lead researchers to posit distinctive pathophysiological features. These in turn give rise to hypotheses about similarities between what were otherwise believed to be dissimilar indications. Such refinements may lead to new avenues of exploration and testing.

In many cases, the vindication of successful intervention ensembles adds support for a series of hypotheses about the role of certain mechanisms in disease processes, the action of therapeutics, and how these are modulated by cointerventions or by varying features of the disease or treatment population. So, too, the failure of diverse efforts to assemble effective ensembles for a particular indication can reveal the inadequacies of the knowledge base that guided those



This model depicts necessary elements in the process of clinical translation, highlighting the central role of theoretical understandings of issues such as drug properties, mechanism, and pathophysiological process in the assembly of intervention ensembles. Dashed arrows trace the path of translation from (1) a series of exploratory trials that enable further assembly and refinement of intervention ensembles for (2) confirmatory trials and, ultimately, (3) inclusion in clinical practice. Solid arrows represent the flow of information, with width representing strength of evidence. A key point of our model is that all decisions in early- and late-stage research, as well as practice, are mediated by theoretical understandings, and these theoretical understandings are heavily informed by various research activities—and, occasionally, practice.

efforts. One question for any assessment of the efficiency of translation, therefore, is how well all of this information is captured and exploited.

So far, we have focused on the role of pathophysiological and pharmacological theories in the sphere of research and drug development. Caregivers, too, rely on such theories to guide their use of intervention ensembles at the bedside. Trials produce evidence of activity in populations, but medicine is practiced on individual patients. Caregivers must therefore judge the extent to which population-level findings—obtained in fastidious settings—apply generally to patients or to settings that have not been sampled, and that judgment inevitably draws on theory. For example, trials frequently exclude patients

who have multiple illnesses or are taking multiple drugs or, of course, those who are older than the maximum eligible age for trials;¹⁵ and caregivers must then adjust the ensemble for these patients. Caregivers might also administer a drug in ways that are not indicated on a drug label; an estimated 50 percent of all cancer drugs are given to patients off-label.¹⁶ Theories of pharmacology and pathophysiology guide this process, and better theories guide it better.

The relationship between exploratory studies, confirmatory studies, clinical practice, and theory must be understood as a complex web of interconnections, as shown in figure 2, not just the linear drug pipeline shown in figure 1. Findings from early-phase, exploratory investigations

are absorbed by drug developers into theories about a drug's pharmacology and an indication's pathophysiology. These theories are used to build intervention ensembles for confirmatory testing or for application at the bedside. The arrows here are bidirectional, however, because sometimes information from care or confirmatory trials can inform theories that prompt trials or exploratory studies. For instance, a case report showing significant response for a novel intervention ensemble can prompt systematic investigations of those intervention ensembles in clinical trials. Key questions for those concerned about efficiency in translation are whether this mechanism is well oiled. Are research systems adequately establishing optima and boundaries on key dimensions with intervention ensembles before drugs are taken up into practice? Are they harnessing the capacity of exploratory and confirmatory investigations for refining theories of pathophysiology and pharmacology? Are exploratory and confirmatory studies being coordinated with each other? Are observations in case reports being advanced promptly into systematic investigation?

Sunitinib: An Illustration

This model of clinical translation is illustrated by the development of the cancer drug sunitinib,¹⁷ which is licensed for the treatment of three different malignancies—renal cell carcinoma, gastrointestinal stromal tumor, and pancreatic neuroendocrine tumor. Sunitinib was originally discovered in a biotechnology start-up, Sugen, which had been founded in 1991 to develop drugs targeting several classes of enzymes known to be implicated in different cancers, including tyrosine kinases. Sunitinib precursors known as SU6668 and SU5416 had been identified in screenings for drugs that inhibited tyrosine kinases.¹⁸ Both failed clinical development, with the later reaching phase 3 trials. But promising findings from these efforts, combined

with further characterization of the interaction of these drugs with their target, were used to produce a closely related chemical compound. This drug, SU11248, had similar activity against tyrosine kinases but more favorable pharmacological properties.¹⁹ It is what eventually was renamed sunitinib.

Preclinical studies of sunitinib predicted activities similar to those of its predecessors—in particular, the drug was active against numerous tumor types, including acute myelogenous leukemia, melanoma, renal cell carcinoma, and colorectal and breast cancers. At the point where it was first administered to patients, however, sunitinib was also known to present a risk of bone marrow and adrenal

and it sometimes caused tumors to collapse so quickly that fistulas formed. Patients with other malignancy types, like colorectal and breast cancers, did not respond. Our model of clinical translation emphasized that a key task for early-phase clinical development is identifying elements that can be combined with a drug to tap into its therapeutic activity. The first two extended dosing phase 1 studies did just that, clarifying at least four dimensions: malignancy type, dose, schedule, and several risk management strategies. They also established boundaries on two dimensions: dose and schedule.

From there, Pfizer (which had bought Sugen) initiated dozens of phase 2 trials testing sunitinib mono-

Caregivers might administer a drug in ways that are not indicated on a drug label; an estimated 50 percent of all cancer drugs are given to patients off-label. Theories of pharmacology and pathophysiology guide this process, and better theories guide it better.

therapy against a variety of tumors. Among these were phase 2 trials of two malignancies that had responded in the phase 1 trial. These produced large, positive responses and led to the accelerated approval of sunitinib for renal cell carcinoma and gastrointestinal stromal tumor.²² The utility of these intervention ensembles was later decisively confirmed in large, randomized trials using clinical endpoints, though response rates were somewhat diminished in confirmatory trials (predictably, given what we stated above about the inherent error tendencies of exploratory studies).

To resolve these uncertainties, investigators conducted two phase 1 trials in advanced cancer patients, administering several doses and schedules. In one, sunitinib was applied against acute myelogenous leukemia.²⁰ However, this trial was a bust, with toxicity at high doses and no sustained responses. In another trial, sunitinib was applied in a shotgun approach to patients with various solid tumor types. Six patients with three malignancy types—renal cell carcinoma, gastrointestinal stromal tumor, and neuroendocrine tumor—responded.²¹ These same patients also encountered life-threatening toxicities: sunitinib caused hypertension,

The other phase 2 trials were driven by a combination of theory and empirical exploration. The role of theory was manifest in the way drug developers pursued several intervention ensembles involving colorectal cancer, an indication that had not shown activity in phase 1 monotherapy studies. Pursuit of this indication

was largely driven by observations that sunitinib was a potent inhibitor of kinases involved in tumor angiogenesis (the process by which tumors recruit a blood supply) and that a different angiogenesis inhibitor, bevacizumab, had proven effective in colorectal cancer.²³ This simple theory that angiogenesis drives progression of colorectal cancer, however, proved flawed. Drug developers were never able to integrate sunitinib into intervention ensembles that were effective against colorectal cancer (and, as of this writing, attempts to apply other angiogenesis inhibitors as monotherapies in colorectal cancer have also proven unsuccessful).²⁴

The role of empirical exploration is suggested in the way Pfizer tested numerous other malignancies in phase 2 trials—ovarian cancer, urethral cancer, prostate cancer, head and neck cancer, and so on. The vast majority of these efforts were negative. Indeed, phase 2 monotherapy trials testing malignancies that did not respond in phase 1 produced tumor response rates of 4.6 percent and a death rate of 1.6 percent—figures that are more adverse than the general objective response rate and death rate for monotherapy cancer drugs in phase 1 testing (3.8 percent and 0.54 percent, respectively).²⁵

At many points in the development of sunitinib, negative and positive trial outcomes were integrated with knowledge of pathophysiology and pharmacology to inform both research and care. Consider the use of sunitinib for metastatic thyroid cancer. The RET proto-oncogene is a receptor tyrosine kinase long known to be implicated in familial and sporadic medullary thyroid cancers. By 2005, researchers had demonstrated that sunitinib potently inhibited the RET proto-oncogene *in vivo*.²⁶ Further, sunitinib was observed to damage thyroid tissues in trials—an unwanted side effect that might be harnessed for therapeutic ends as a thyroid cancer treatment.²⁷ Combining this molecular and clinical evidence, a team of physicians in Ireland offered

sunitinib to a patient with refractory metastatic thyroid carcinoma in late 2005 and observed a clinical response.²⁸ Another positive case report followed.²⁹ Spurred by this evidence, phase 2 trials investigating single-agent sunitinib for metastatic thyroid cancer had begun enrolling patients in July 2007.³⁰ Before any of these trials had even completed enrollment, the American Thyroid Association issued guidelines recommending clinicians consider using tyrosine kinase inhibitors such as sunitinib in refractory patients otherwise ineligible for trials.³¹ The National Comprehensive Cancer Network followed suit in 2010.³²

Under the conventional way of looking at it, the development of sunitinib would seem inefficient. Two prior drugs had failed development. The first phase 1 trial in a disease indication—acute myelogenous leukemia—proved a false start. Most preclinical studies showing promise against various indications proved discordant with subsequent clinical trials. A large number of seemingly haphazard, exploratory trials testing new indications proved ineffective or toxic, as has practically every trial testing sunitinib in combination with other cancer drugs. Lack of clinical utility for malignancies like breast and colorectal cancers was concluded only after sunitinib had been put into several large confirmatory trials.

Our account opens up a finer-grained assessment of the efficiency with which sunitinib was translated. On the one hand, significant pharmacological findings from the two failed sunitinib precursors were quickly absorbed into several successful translation trajectories. In only a few phase 1 studies, researchers identified roughly optimal intervention ensembles for two different indications as well as boundaries on dose and schedule. Sunitinib was thoroughly tested against indications believed, on the basis of sometimes inchoate theories of pathophysiology, to be candidates for the drug. And negative findings—some of which were incorporated as

advisories in clinical practice guidelines—helped practitioners learn about limits of the drug's application. Finally, in several instances, an understanding of the pharmacology gained from trials was combined with knowledge of pathophysiology to develop novel intervention ensembles that are currently in trials. All in all, clinical research activities appear to have furnished health care systems and caregivers with a rich evidence and theory base for adjusting intervention ensembles at the bedside.

On the other hand, a string of failures in later trials suggests that the research system may well have “over-sampled” new malignancies and drug combinations. The late-phase failures in breast and colorectal cancer—two relatively prevalent and hence lucrative indications—might show how commercial considerations can eclipse scientific rationale in launching confirmatory testing. Perhaps theories of pathophysiology and pharmacology for these malignancies were not updated as rapidly as they should have been. The trial leading to the licensure of sunitinib for renal cell carcinoma was later found to have underreported adverse events.³³ A review undertaken by our research group of citation patterns among the trials of sunitinib suggests that negative trials were almost never cited in subsequent trials and reviews (in contrast to positive and inconclusive studies), suggesting that pharmacological and pathophysiological insights may not have been effectively captured. The record shows that some information from these trial activities appears to have been lost.

Reconsidering Translation Efficiency

We are now in a position to relate our model of clinical translation to claims about the efficiency of clinical translation. First, many accounts lament the very high rate at which new drugs fail rigorous tests of safety and efficacy in clinical development. In our view, these

failures are deeply problematic only if they did not explore dimensions of application that would contribute information about relevant optima or boundaries, or if the information from these attempts was not effectively captured and integrated into attempts to build other intervention ensembles. Many accounts also are critical of the lengthy time periods and number of trials needed for clinical development. However, more trials allow for greater sampling of intervention ensemble dimensions and greater statistical power for evaluating various cause-and-effect relationships. These, in turn, improve the prospects for good clinical judgment when adjusting and delivering intervention ensembles at the bedside.

We are also in a position to use our model to evaluate various scientific and policy proposals aimed at improving translation efficiency. We previously noted several influential proposals aimed at improving the efficiency of clinical translation. In one, commentators urge further liberalization of drug regulations so that drugs for life-threatening illness are provisionally approved after phase 1 testing. This proposal appears plausible if we define the goal of translation as the licensure of a safe drug. Our account, however, has emphasized that drugs are, by themselves, inert or toxic entities. The key output of translation is information about how and how not to integrate them into intervention ensembles. Under this reading, the proposal to license new drugs immediately on successful completion of safety testing looks like a category error. When a drug has completed phase 1 testing, researchers have not yet discovered the necessary or sufficient components of an intervention ensemble. Rather, they have a schema for many potential ensembles into which a drug might be incorporated. On completion of phase 1 testing, sunitinib was believed to be a promising candidate for a variety of different malignancies, including breast and colorectal cancers. All but three indications tested have proven

unresponsive under conditions used in phase 1 trials. The idea that sunitinib would show clinical promise against breast and colorectal cancers had strong preclinical support and pathophysiological plausibility, but subsequent trials proved highly unfavorable. Granting market approval at the completion of phase 1 would likely have exposed far more patients to toxic intervention ensembles than were needed to discover the clinically useful intervention ensembles that involve this potent drug. That is the opposite of efficiency.

A second set of proposals would compress drug development timelines by incorporating patient preferences. One proposal involves aggressive trial designs, where patients are given the choice of designing their own inter-

The proposal to license new drugs immediately on successful completion of safety testing looks like a category error. When a drug has completed phase 1 testing, researchers have not yet discovered how to deploy it in a treatment setting.

vention ensemble by selecting a drug dose based on their “best guess.” Proponents of this approach assert that, “in the case of monoclonal antibodies, it is known that most evaluated products are eventually administered at doses between 1 and 10 mg/kg (or equivalent) every 1 to 4 weeks. This best guess approach to choosing initial doses might be associated with increased chances of toxicity, but some patients with cancer would accept this.”³⁴ Leaving aside the debacle involving TGN1412 (where a monoclonal antibody administered at a dose 1/500th of what caused effects in nonhuman primates caused life-threatening toxicities in six healthy volunteers),³⁵ and other failed attempts at employing a similar approach,³⁶ this proposal seems to assume that theories of pharmacology and pathophysiology are mostly reliable at the point where drugs are

advanced into clinical testing. This may be the case for drugs and indications for which there is very extensive experience. But very often, clinical development is concerned with drugs or indications for which prior experience is likely to be unreliable. Even so, the “best guess” approach would leave dose to be explored in a desultory way through subsequent trials or monitoring of clinical experience.

Another patient-centered reform proposes to incorporate patient preferences into research planning, starting with preclinical research: “discussions with users of research . . . provide opportunities to abort unpromising efforts early.”³⁷ Though we are receptive to democratizing approaches to priority setting in research, our account of translation

highlights some important liabilities of this approach. Drug developers often know very little about the application of novel drugs until they have pursued a series of exploratory investigations. Too strong a “demand pull” from end users may stymie that exploration. It might also prioritize confirmatory testing before evidentiary warrant is established, leading to new inefficiencies.

A third set of proposals would accelerate translation through a suite of scientific innovations in clinical trial design. One celebrated example is the integration of personalized medicine into drug development and, in particular, the use of molecular markers that are predictive of drug response to select patients for clinical trials—so-called enrichment trials. Such proposals build on the success of drugs like trastuzumab, imatinib, and vemurafenib, which were licensed after

a very small number of trials employing enrichment. Enrichment designs are very attractive because they leverage deep pathophysiological insights to zero in on approximately optimal indications for trials. These confer efficiencies because they reduce the amount of exploration needed to find approximately optimal intervention ensembles. But our analysis suggests that, if they are not deployed properly, enrichment designs harbor important dangers for efficient clinical translation. First, if all prelicensure trials of a new drug employ enrichment designs, caregivers have little basis for inferring the drug's activity in patients who are marker negative or who score intermediate values on marker levels. Again, because drug development often occurs at the cutting edge of theory, researchers are unable to rule out the possibility that marker-negative patients will not benefit. In the case of cetuximab, a drug approved using only certain marker-positive patients turned out to be just as effective in patients negative for that marker. Much of the information showing this was gathered, not in trials, but, less efficiently, in care settings.

Our analysis also reveals subtle moral implications of proposals to reform the translation process. Prelicensure, most costs and risks associated with clinical translation are borne by drug companies and relatively small populations of study participants. Because the burdens on participants from trial participation can be quite high, reducing the number of studies needed to approve a drug can seem like a good goal. Our analysis reveals, however, that in some cases these proposed reforms would not eliminate but, rather, shift the costs and risks to other parties. For example, once drugs are licensed, drug companies have greatly diminished incentives to conduct trials. This means that many further refinements and extensions of intervention ensembles (and theories guiding their use) are underwritten by public research agencies or by health care systems, whose experience

with the drugs generates the evidence necessary to warrant these extensions. Because manufacturers reap the profits of expanded use, there are legitimate questions of fairness about the extent to which they should be able to offload the costs of generating this evidence base.

Similarly, reducing the number of patients in potentially burdensome clinical trials is a laudable goal, but extreme care must be taken to ensure that such reforms do not simply shift the associated risks onto the shoulders of much larger groups of patients. As they are currently configured, health care settings are neither budgeted nor structured to collect reliable information. As a consequence, resolving the uncertainties that remain after a drug is licensed is likely to require greater patient exposure and cost. How to fairly apportion this work between clinical trials and treatment surveillance requires more thought.

Improving Efficiency

Finally, our model of clinical translation suggests additional opportunities for addressing inefficiencies and highlights some natural limits on these efforts. First, consider the relationship between exploratory and confirmatory investigations. In our account, an important inefficiency arises when confirmatory investigations are poorly coordinated with exploration. This can (and does) happen in several ways: drug developers might pursue excessive exploration, exploratory studies might find approximately optimal intervention ensembles that are never advanced to confirmatory trials, inadequately refined intervention ensembles might be rushed into confirmatory testing, or intervention ensembles might be taken up in practice before they are confirmed as useful. Conventional mechanisms, like review by an institutional review board, have a role to play in preventing some of these misuses. For example, when reviewing confirmatory trials, ethics committees should examine whether prior stud-

ies have adequately explored key dimensions of intervention ensembles; they can also require evidence that confirmatory studies are testing approximately optimal ensembles. Nevertheless, the remit of IRBs is limited to individual protocols, and many of the concerns we have raised apply at the level of large sets or portfolios of studies. Institutions such as the FDA are better positioned to evaluate issues that arise at the portfolio level, but their remit is limited to prelicensure research, and it is unclear that they have the regulatory authority to coordinate exploratory and confirmatory testing.³⁸ Other mechanisms may therefore be needed.

Second, consider our claim that clinical translation is primarily about generating information and assimilating it to innovation and practice. This motivates a search for better ways of capturing and absorbing information. In exploratory investigations and unsuccessful development trajectories, a wealth of theory-enriching information is produced that vanishes into thin air. For instance, no regulations compel deposition of even confirmatory trial results unless they involve licensed agents within the FDA-approved label.³⁹ At least a third of exploratory trials are never published,⁴⁰ and experiments embedded within exploratory investigations that test pharmacological premises—so-called pharmacodynamics studies—are often withheld from publication.⁴¹ Trials that are published, moreover, are often reported in ways that are misleading or that frustrate valid inference. In one recent study, a third of randomized trials in top-tier journals showed evidence of modifying the primary endpoint between registration and publication.⁴² In another, 8 percent of phase 3 trials of cancer drugs—a category known for its toxicity—did not report drug-related deaths.⁴³ Much of the information that is published is never absorbed into theories guiding translation and practice. Several reports show that “positive” findings tend to receive more citations

than “negative” findings, leading to the amplification and propagation of erroneous beliefs.⁴⁴ For research ethics to address these failings in an effective way, it will need to strengthen its influence over activities that occur *after* trials are completed.

Finally, our analysis reveals some natural limits on the ambition of eliminating unsuccessful translation trajectories. In domains such as structural engineering, where relevant causal systems are well characterized, optimal designs or solutions to engineering problems can be identified without producing a series of failed exploratory efforts. In the medical domain, however, where knowledge of the underlying causal systems is significantly less developed, empirical failure plays a crucial role in clarifying the boundaries of useful intervention ensembles and refining inchoate theories of pharmacology and pathophysiology. Until our understanding of the underlying causal systems in medicine matures, “negative cohorts” (cohorts of patients receiving unfavorable intervention ensembles) and negative confirmatory trials (trials that fail to demonstrate the clinical utility of a novel drug) cannot be eliminated without there also being an elimination of information critical to the development not just of new interventions but also of the theories on whose development the maturation of medicine as a science depends. One implication of our analysis, therefore, is that the near-term focus should not be on eliminating such failures (because this is impossible), but on making sure that they illuminate dimensions of intervention ensembles that guide further development or clinical application and that this information is captured and used to refine the theories and strategies that justified undertaking such studies.

A Complex View of Translation Efficiency

Debates about the efficiency of translation implicate the inter-

ests of a wide range of stakeholders: researchers and study participants who produce scientific evidence, clinicians and patients who rely on that evidence, private and public bodies that fund research and health care, and the larger community in whose name medical progress is often justified.⁴⁵ It is surprising, therefore, that there is currently no explicit model of translation that these stakeholders can use to focus debate and evaluate the epistemic, ethical, and policy dimensions of reforms. The model that we outline here is offered as a first step in filling this gap.

If nothing else, we have exposed serious deficiencies in the ways many advocates of research reform think about translation efficiency. Where most commentators view negative trials, late-phase failures, and abandoned drugs as clear signs of inefficiency, we have argued that some failures are integral to demarcating the appropriate application of new drugs and to developing the theoretical basis needed to guide further development and clinical practice. We have also suggested that there are inherent limitations on the degree to which failure in the late stages of development can be reduced, given the current state of medical science and the statistical properties of exploration.

Our model also connects a range of ethical and epistemological issues that are typically treated in isolation from one another. For instance, there is mounting support for registering trials and requiring the publication of “negative” findings, yet the rationale for these practices is disconnected from issues of decision-making at the bedside. Even where our model of translation connects with established debates, it highlights synergies and tensions between issues and casts narrow debates in a broader light.

Nevertheless, the model that we sketch here is tentative in several respects. Although it casts issues of fairness into stark relief, it does not yet offer guidance about the proper balance between exploratory and

confirmatory trials or the most efficient division of labor between formal clinical studies and post-approval surveillance activities. Additional work will be needed to clearly delineate the point beyond which failure cannot be eliminated without compromising the evidence base needed for translation or treatment or without shifting risk and burdens onto other parties. And while we believe common metrics of translation efficiency—dollars per approved drug, years from basic science discovery to bedside application—fail to capture many of the morally relevant aspects of efficiency, further work will be required to develop alternative metrics. These issues are grist for further refinement and extension of the model. It is sufficient for our present purposes to have demonstrated the moral and policy rationale for additional work in this area.

Acknowledgments

This work was funded by the Canadian Institutes of Health Research (EOG111391). We wish to thank Spencer Hey, Charles Weijer, and members of the Studies of Translation, Ethics and Medicine research group for feedback on our analysis. We also thank Benjamin Carlisle and Gina Freeman for research assistance. All faults in the manuscript remain our own.

Notes

1. *The Lancet*, “What Is the Purpose of Medical Research?,” *The Lancet* 381 (2013): 347.
2. We use the term “information” to refer to data that have been collected, organized, aggregated, and analyzed in a way that enables application towards a set of ends, or the evaluation of such application.
3. A. Von Eschenbach and R. Hall, “Von Eschenbach and Hall: FDA Approvals Are a Matter of Life and Death,” *Wall Street Journal*, June 17, 2012. See also A. Von Eschenbach, “Medical Innovation: How the U.S. Can Retain Its Lead,” *Wall Street Journal*, February 14, 2012; P. D. Jacobson and W. E. Parmet, “A New Era of Unapproved Drugs: The Case of *Abigail Alliance v. Von Eschenbach*,” *Journal of the American Medical Association* 297 (2007): 205-08.
4. A. Grove, “Rethinking Clinical Trials,” *Science Magazine* 333 (2011): 1679.

5. D. D. Rosa, J. Harris, and G. C. Jayson, "The Best Guess Approach to Phase I Trial Design," *Journal of Clinical Oncology* 24, no. 1 (2006): 206-08.
6. I. Chalmers et al., "How to Increase Value and Reduce Waste When Research Priorities Are Set," *The Lancet* 383 (2014): 156-65.
7. F. S. Collins, "Reengineering Translational Science: The Time Is Right," *Science Translational Medicine* 3, no. 90 (2011): 90cm17.
8. National Center for Advancing Translational Sciences, *About NCATS: Organization and Background*, 2013, <http://www.ncats.nih.gov/about/org/organization.html>, accessed February 13, 2014.
9. J. Kimmelman, "A Theoretical Framework for Early Human Studies: Uncertainty, Intervention Ensembles, and Boundaries," *Trials* 13, no. 1 (2012): 173; S. P. Hey and J. Kimmelman, "Ethics, Error, and Initial Trials of Efficacy," *Science Translational Medicine* 5, no. 184 (2013): 184fs16.
10. Because intervention ensembles combine many different dimensions, the process of early phase development can be likened to searching a multidimensional landscape for optima and boundaries on key dimensions.
11. This point is vividly demonstrated with the following thought experiment: imagine small trials of various intervention ensembles are powered at 50 percent and use $p < .05$ to declare an intervention ensemble clinically useful (meaning that, 5 percent of the time, ineffective ensembles are erroneously declared promising due to random variation alone). Imagine further that, for every ten intervention ensembles tested, only one is clinically useful. According to simple laws of probability, 50 percent of intervention ensembles that are actually useful will produce a "negative" trial when tested. And 50 percent of the intervention ensembles producing "positive" trial findings will, in fact, later prove clinically useless.
12. "Mostly" because, for complex interventions, "negative trials" may still validate aspects of the intervention. For instance, negative trials involving surgical delivery of cells to a designated tissue might nevertheless produce valuable knowledge concerning whether delivery methods are effective and whether cells survive implantation.
13. J. Kimmelman and A. J. London, "Predicting Harms and Benefits in Translational Trials: Ethics, Evidence, and Uncertainty," *PLoS Medicine* 8, no. 3 (2011): e1001010.
14. *Ibid.*; Hey and Kimmelman, "Ethics, Error, and Initial Trials of Efficacy."
15. M. Hordijk-Trion et al., "Patients Enrolled in Coronary Intervention Trials Are Not Representative of Patients in Clinical Practice: Results from the Euro Heart Survey on Coronary Revascularization," *European Heart Journal* 27, no. 6 (2006): 671-78.
16. P. G. Casali, "The Off-Label Use of Drugs in Oncology: A Position Paper by the European Society for Medical Oncology (ESMO)," *Annals of Oncology* 18, no. 12 (2007): 1923-25; M. Soares, "'Off-Label' Indications for Oncology Drug Use and Drug Compendia: History and Current Status," *Journal of Oncology Practice* 1, no. 3 (2005): 102-05; C. D. Mullins et al., 2012. "Recommendations for Clinical Trials of Off-Label Drugs Used to Treat Advanced-Stage Cancer," *Journal of Clinical Oncology* 30, no. 6 (2012): 661-66.
17. We chose this drug at random from a sample of cancer drugs that were introduced into clinical development between 2000 and 2003 and that went on to receive licensure. This time window was selected because it enabled us to follow a decade of subsequent trials.
18. T. Fong et al., "SU5416 Is a Potent and Selective Inhibitor of the Vascular Endothelial Growth Factor Receptor (Flk-1/KDR) That Inhibits Tyrosine Kinase Catalysis, Tumor Vascularization, and Growth of Multiple Tumor Types," *Cancer Research* 59, no. 1 (1999): 99.
19. J. Schlessinger, "SU11248: Genesis of a New Cancer Drug," *The Scientist*, April 11, 2005, 17; L. Sun et al., "Discovery of 5-[5-fluoro-2-oxo-1,2-dihydroindol-(3Z)-ylidenemethyl]-2,4-dimethyl-1H-pyrrole-3-carboxylic acid (2-diethylaminoethyl) amide, A Novel Tyrosine Kinase Inhibitor Targeting Vascular Endothelial and Platelet-Derived Growth Factor Receptor Tyrosine Kinase," *Journal of Medicinal Chemistry* 46, no. 7 (2003): 1116-19.
20. W. Fiedler et al., "A Phase 2 Clinical Study of SU5416 in Patients with Refractory Acute Myeloid Leukemia," *Blood* 102, no. 8 (2003): 2763-67.
21. 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.
22. R. J. Motzer et al., "Sunitinib versus Interferon alfa in Metastatic Renal-Cell Carcinoma," *New England Journal of Medicine* 356 (2007): 115-24; G. D. Demetri, "Efficacy and Safety of Imatinib Mesylate in Advanced Gastrointestinal Stromal Tumors," *New England Journal of Medicine* 347 (2002): 472-80.
23. A. Carrato et al., "Fluorouracil, Leucovorin, and Irinotecan Plus Either Sunitinib or Placebo in Metastatic Colorectal Cancer: A Randomized, Phase III trial," *Journal of Clinical Oncology* 31, no. 10 (2013): 1341-47; L. B. Saltz et al., "Phase II Trial of Sunitinib in Patients with Metastatic Colorectal Cancer after Failure of Standard Therapy," *Journal of Clinical Oncology* 25, no. 30 (2007): 4793-99.
24. R. K. Jain et al., "Lessons from Phase III Clinical Trials on Anti-VEGF Therapy for Cancer," *Nature Clinical Practice Oncology* 3, no. 1 (2006): 24-40.
25. T. G. Roberts et al., "Trends in the Risks and Benefits to Patients with Cancer Participating in Phase 1 Clinical Trials," *Journal of the American Medical Association* 292 (2004): 2130-40.
26. D. W. Kim et al., "An Orally Administered Multitarget Tyrosine Kinase Inhibitor, SU11248, Is a Novel Potent Inhibitor of Thyroid Oncogenic RET/Papillary Thyroid Cancer Kinases," *Journal of Clinical Endocrinology and Metabolism* 91, no. 10 (2006): 4070-76.
27. J. Desai et al., "Hypothyroidism after Sunitinib Treatment for Patients with Gastrointestinal Stromal Tumors," *Annals of Internal Medicine* 145, no. 9 (2006): 660-64.
28. F. C. Kelleher and R. McDermott, "Response to Sunitinib in Medullary Thyroid Cancer," *Annals of Internal Medicine* 148, no. 7 (2008): 567.
29. M. J. Bugalho, R. Domingues, and A. Borges, "A Case of Advanced Medullary Thyroid Carcinoma Successfully Treated with Sunitinib," *Oncologist* 14, no. 11 (2009): 1083-87.
30. L. L. Carr et al., "Phase II Study of Daily Sunitinib in FDG-PET-Positive, Iodine-Refractory Differentiated Thyroid Cancer and Metastatic Medullary Carcinoma of the Thyroid with Functional Imaging Correlation," *Clinical Cancer Research* 16, no. 21 (2010): 5260-68; ClinicalTrials.gov, U.S. National Institutes of Health, "Sunitinib Malate in Treating Patients with Iodine-Refractory Recurrent or Metastatic Thyroid Cancer," <http://www.clinicaltrials.gov/ct2/show/NCT00519896?term=NCT00519896&rank=1>, accessed November 6, 2012. If this one indication is removed, the relative response rate for sunitinib monotherapy trials involving indications that were nonresponsive in phase 1 studies falls from 4.6 to 2.6.
31. American Thyroid Association Guidelines Taskforce on Thyroid, Nodules, Cancer Differentiated Thyroid, "Revised American Thyroid Association Management Guidelines for Patients with Thyroid Nodules and Differentiated Thyroid Cancer," *Thyroid* 19, no. 11 (2009): 1167-214.
32. R. M. Tuttle et al. and Network National Comprehensive Cancer, "Thyroid Carcinoma," *Journal of the National Comprehensive Cancer Network* 8, no. 11 (2010): 1228-74.
33. M. L. Telli et al., "Cardiotoxicity Associated with the Cancer Therapeutic Agent Sunitinib Malate," *Annals of Oncology* 19, no. 9 (2008): 1613-18.

34. Rosa, Harris, and Jayson, "The Best Guess Approach to Phase I Trial Design."

35. G. Suntharalingam et al., "Cytokine Storm in a Phase 1 Trial of the Anti-CD28 Monoclonal Antibody TGN1412," *New England Journal of Medicine* 355 (2006): 1018-28.

36. C. K. Daugherty et al., "Quantitative Analysis of Ethical Issues in Phase I Trials: A Survey Interview of 144 Advanced Cancer Patients," *IRB: Ethics & Human Research* 22, no. 3 (2000): 6-14.

37. Chalmers et al., "How to Increase Value and Reduce Waste When Research Priorities Are Set."

38. A. J. London, J. Kimmelman, and B. Carlisle, "Rethinking Research Ethics: The Case of Postmarketing Trials," *Science* 336 (6081): 544-45.

39. Food and Drug Administration, Food and Drug Administration Amendments Act

(FDAAA) of 2007, Pub. L. 110-85, Title VIII, Section 801.

40. L. H. Camacho et al., "Presentation and Subsequent Publication Rates of Phase I Oncology Clinical Trials," *Cancer* 104, no. 7 (2005): 1497-504; R. T. Hoeg et al., "Publication Outcomes of Phase II Oncology Clinical Trials," *American Journal of Clinical Oncology* 32, no. 3 (2009): 253-57; R. W. Scherer, P. Langenberg, and E. von Elm, "Full Publication of Results Initially Presented in Abstracts," *Cochrane Database of Systematic Reviews* 2 (2007): MR000005; A. T. Turer et al., "Publication or Presentation of Results from Multicenter Clinical Trials: Evidence from an Academic Medical Center," *American Heart Journal* 153, no. 4 (2007): 674-80.

41. G. A. Freeman and J. Kimmelman, "Publication and Reporting Conduct for Pharmacodynamic Analyses of Tumor Tissue in Early Phase Oncology Trials,"

Clinical Cancer Research 18, no. 23 (2012): 6478-84.

42. S. Mathieu et al., "Comparison of Registered and Published Primary Outcomes in Randomized Controlled Trials," *Journal of the American Medical Association* 302 (2009): 977-84.

43. S. Niraula, "The Price We Pay for Progress: A Meta-analysis of Harms of Newly Approved Anticancer Drugs," *Journal of Clinical Oncology* 30, no. 24 (2012): 3012-19.

44. S. A. Greenberg, "How Citation Distortions Create Unfounded Authority: Analysis of a Citation Network," *BMJ* 339 (2009): b2680.

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

Another Voice

Translational Research May Be Most Successful When It Fails

by JOHN P. A. IOANNIDIS

In this issue of the *Hastings Center Report*, Jonathan Kimmelman and Alex London argue that in assessing the success of clinical translation, it is narrow-minded to focus only on how many new drugs get licensed and how quickly they achieve licensure.¹ I fully agree that this simplified view of clinical translation tends to increase the temptation to cut corners, lower a bar that is already low, and encourage the adoption of new treatments without sufficiently reliable data on efficacy—we almost never have sufficiently reliable data on safety at the time when drugs get licensed anyhow. It is disappointing that leaders at both regulatory and public research funding organizations are under such pressure to portray biomedical research as a success story—producing machine. Tenuous success stories are indispensable for companies to make money from and for the media to sensationalize on, but not for science. Kimmelman and London show that clinical translation should be judged on its ability to generate as comprehensive an intervention ensemble as possible for the tested interventions. This may include many "negative" studies and other aspects of trial-and-error in the tortuous nonlinear process of trying to understand what works and what does not. "Negative" results

are very informative. They can correct mechanistic and other "basic" science misconceptions and help define the optima for new interventions in terms of dose, setting, population, and other parameters that shape best use at the clinical and population level.

I would like to extend Kimmelman and London's position in two ways. First, I would argue that in the current environment, failures should be seen not just as acceptable, but probably as the most useful outcomes that translational research efforts can offer. Failures are probably more important than successes. Among failures I include studies with "negative" results that show that large lines of preclinical and early clinical investigation are not fruitful and should be abandoned or at least radically modified. I also have in mind later-stage clinical trials with "negative" results that modulate our understanding about which among several already-known interventions may not have as much merit as we thought, such as when we discover that interventions that are already licensed and widely adopted should actually be used in a more limited fashion or should be totally discarded.

Hype is rampant nowadays both in the "basic" biomedical sciences and in clinical research. Several investigative fields are fueled with resources mostly because of inertia, expressed by self-promoting study sections whose members do not want to admit that they should better quit their uninformative minutiae. Well-done research

John P. A. Ioannidis, "Translational Research May Be Most Successful When It Fails," *Hastings Center Report* 45, no. 2 (2015), 39-40. DOI: 10.1002/hast.429