The moral efficiency of clinical trials in anti-cancer drug development

Benjamin Gregory Carlisle
Department of Experimental Medicine, McGill University, Montreal

2019 April

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of doctor of philosophy

© Benjamin Gregory Carlisle 2019
# Table of Contents

Abstract ........................................................................................................................................................................ 6

   English ...................................................................................................................................................................... 6
   French .................................................................................................................................................................. 7

Acknowledgements .................................................................................................................................................. 9

Contribution to original knowledge .................................................................................................................... 10

Contribution of authors ......................................................................................................................................... 11

Introduction .......................................................................................................................................................... 13

   Literature review ............................................................................................................................................... 13
   Moral efficiency ............................................................................................................................................... 15

Methods ............................................................................................................................................................... 16

   Project overview ........................................................................................................................................... 16

Chapter 1. Benefit, Risk, and Outcomes in Drug Development: A Systematic Review of Sunitinib. 23

   Introduction .................................................................................................................................................... 24

   Methods .......................................................................................................................................................... 25

      Literature search ........................................................................................................................................ 25
      Extraction ................................................................................................................................................. 26
      Analysis and statistics .............................................................................................................................. 27

   Results .......................................................................................................................................................... 27

      Study characteristics ............................................................................................................................... 27
      Total patient benefit and burden ............................................................................................................. 27
      Patterns of research, coordination and outcomes .................................................................................. 29

   Discussion ..................................................................................................................................................... 30

   Figures ........................................................................................................................................................... 33

   Tables ............................................................................................................................................................ 37

   Notes ............................................................................................................................................................... 37

   References ..................................................................................................................................................... 38

Summary .............................................................................................................................................................. 52
Chapter 2. Cancer Precision Medicine Drug Development is Often Empirical: The Case of Imatinib, Dasatinib and Nilotinib

Introduction.............................................................................................................................................................................. 54
Methods..................................................................................................................................................................................... 54
  Literature search..................................................................................................................................................................... 55
  Study selection....................................................................................................................................................................... 55
  Extraction............................................................................................................................................................................... 55
  Analysis.................................................................................................................................................................................... 57
Results....................................................................................................................................................................................... 57
Discussion.................................................................................................................................................................................. 58
Figures........................................................................................................................................................................................ 61
Tables........................................................................................................................................................................................ 65
Notes.......................................................................................................................................................................................... 65
References.................................................................................................................................................................................. 65

Summary..................................................................................................................................................................................... 67

Chapter 3. Patient Burden and Clinical Advances Associated with Post-Approval Monotherapy

Cancer Drug Trials...................................................................................................................................................................... 68

Background................................................................................................................................................................................. 69
Methods....................................................................................................................................................................................... 69
  Literature search..................................................................................................................................................................... 70
  Extraction............................................................................................................................................................................... 70
  Analysis and statistics.......................................................................................................................................................... 71
Results....................................................................................................................................................................................... 71
  Volume and characteristics of label-extending trials testing monotherapy................................................................. 71
  Patient risk and burden..................................................................................................................................................... 72
  Advancement to randomized trials, clinical practice and FDA labels................................................................. 72
Discussion.................................................................................................................................................................................. 73
Figures........................................................................................................................................................................................ 76
Tables........................................................................................................................................................................................ 77
Notes.......................................................................................................................................................................................... 77
References.................................................................................................................................................................................. 77
Summary.............................................................................................................................................................................. 80

Chapter 4. Benefit and Burden in a Historic Cohort of Cancer Combination Trials: A Systematic Review and Meta-Analysis........................................................................................................................................................................... 81

Introduction............................................................................................................................................................................. 82

Methods.................................................................................................................................................................................. 83

Combination therapy sample.................................................................................................................................................. 83

Literature search................................................................................................................................................................. 84

Extraction................................................................................................................................................................................. 84

Analysis and statistics.......................................................................................................................................................... 85

Results................................................................................................................................................................................... 85

Volume and characteristics of trials testing combination-indication pairings after FDA approval........................................ 85

Patient risk and burden.......................................................................................................................................................... 86

Patient benefit and the clinical impact of participation...................................................................................................... 87

Discussion............................................................................................................................................................................... 87

Figures.................................................................................................................................................................................... 90

Tables................................................................................................................................................................................... 93

Notes...................................................................................................................................................................................... 93

References............................................................................................................................................................................... 94

Summary............................................................................................................................................................................... 99

Chapter 5. Trials that say “maybe”: the disconnect between exploratory and confirmatory testing after drug approval............................................................................................................................................................................. 100

How post-license trials create clinical agnosticism............................................................................................................... 100

How unresolved clinical agnosticism harms.......................................................................................................................... 102

Toward resolving clinical agnosticism.................................................................................................................................. 103

Key messages.......................................................................................................................................................................... 105

Tables................................................................................................................................................................................... 106

References............................................................................................................................................................................... 107

Discussion............................................................................................................................................................................... 110

Ethical review of clinical trials.............................................................................................................................................. 111

Interpretation and evaluation of previously completed clinical trials.................................................................................. 112

Carlisle 4
Patient decision-making................................................................................................................................. 113
Health research policy and institutional priority setting......................................................................................... 114
Future research....................................................................................................................................................... 114
Limitations............................................................................................................................................................ 119
Conclusion............................................................................................................................................................. 120
References for introduction, summaries and discussion.......................................................................................... 121
Abstract

Drug development research programmes can be very extensive, costly and risky. In many cases, clinical trials of a drug continue after regulatory approval to test new indications, combinations, populations, doses and schedules. The efficiency of drug development is usually considered in economic terms that only take into account the cost of bringing a drug to market, and often neglect costs that are incurred after licensure is achieved. In what follows, we describe approaches for measuring and evaluating the moral efficiency of drug development, considering inputs such as the welfare of human subjects, and outputs such as scientific value and clinically useful information. This may better capture the ethical landscape of drug development, identify inefficiencies and bring to light dynamics that are invisible at the single-trial level. Here, we conduct a series of analyses, moving from particular cases to general principles. First, systematic review and meta-analytic methods are used to characterize patterns of patient risk and benefit, and to examine measures of clinical impact among clinical trials of sunitinib in indications that were not approved by the FDA at the time the trial was launched. Our methods are then applied to a comparison of imatinib with two of its “me too” successors, dasatinib and nilotinib. These projects reveal a large volume of research after a drug is approved by the FDA and cases of worsening risk and benefit over the course of drug development. Next, we conduct two retrospective cohort studies capturing first monotherapy and then combination therapy trials of anti-cancer drugs first approved by the FDA between 2005-2007. These quantify the clinical payoff and investment of human research subjects among indications and combinations that were first launched into testing after a drug’s approval. Finally, these insights are used to define and develop the concept of “clinical agnosticism,” the uncertainty of clinical efficacy that arises following positive hypothesis-generating clinical trials that lack confirmatory follow-up. This evaluation of the moral efficiency of drug development has implications for the practice of independent ethics review, monitoring of ongoing trials, interpretation of clinical trial results, regulatory approval and policy setting.
Les programmes de recherche en développement de médicaments peuvent être très volumineux, coûteux et risqués. Dans de nombreux cas, les essais cliniques d’un médicament se poursuivent après l’approbation réglementaire pour tester de nouvelles indications, combinaisons, populations, doses et calendriers. L’efficacité du développement d’un médicament est généralement considérée en termes économiques, en ne prenant en compte que le coût de la mise sur le marché d’un médicament et en négligeant souvent les coûts encourus après l’obtention de l’homologation. Dans ce qui suit, nous décrivons des approches permettant de mesurer et d’évaluer l’efficacité morale du développement de médicaments en tenant compte d’intrants tels que le bien-être des sujets humains, d’outils tels que la valeur scientifique et les informations utiles sur le plan clinique. Cela permettrait de mieux cerner le paysage éthique du développement des médicaments, d’identifier les inefficacités et de mettre en lumière des dynamiques invisibles au niveau du test unique. Nous effectuons ici une série d’analyses allant de cas particuliers à des principes généraux. Tout d’abord, une analyse systématique des méthodes méta-analytiques est utilisée pour caractériser les modèles de risque et avantages pour le patient et pour examiner les mesures d’impact clinique entre les essais cliniques de sunitinib dans des indications non approuvées par le FDA au moment du lancement de l’essai. Nos méthodes sont ensuite appliquées à une comparaison de l’imatinib avec deux de ses successeurs «moi aussi», le dasatinib et le nilotinib. Ces projets révèlent un volume important de recherche après l’approbation d’un médicament par le FDA et des cas d’aggravation du risque et des avantages au cours du développement du médicament. Ensuite, nous conduisons deux études de cohorte rétrospectives capturant d’abord des essais en monothérapie, puis en association de médicaments anticancéreux approuvés pour la première fois par le FDA entre 2005 et 2007. Celles-ci quantifient le bénéfice clinique et l’investissement des sujets de recherche humains parmi les indications et les combinaisons qui ont été introduites pour la première fois dans les tests après l’approbation d’un médicament. Enfin, ces informations sont utilisées pour définir et développer le concept de «agnosticisme clinique», l’incertitude quant à l’efficacité clinique résultant des essais cliniques positifs générant une hypothèse et dépourvus de suivi de confirmation. Cette évaluation de l’efficacité morale du développement de médicaments a des implications pour la pratique de l’évaluation éthique indépendante, de la surveillance des essais...
en cours, de l'interprétation des résultats des essais cliniques, de l'approbation règlementaire et de la définition des politiques.
Acknowledgements

I am indebted to my supervisor, Dr. Jonathan Kimmelman, for the wisdom, support and guidance he provided throughout the course of my doctoral work. I am truly grateful for his care for my education, commitment to my development and excitement for our work together. It has been an honour and a pleasure.

My colleagues and co-authors have also been invaluable help to me, assisting with data collection, curation, feedback, editing, and supporting me in countless other ways. I am grateful for the help of Brianna Barsanti-Innes, Danny Benjamin, Nadine Demko, Samantha Dolter, Adelaide Doussau, Carole Federico, Aden Feustel, Georgina Freeman, Silviya Ganeshamoorthy, Peter Grabitz, Yasmina Hachem, Amanda Hakala, Valerie Henderson, Spencer Hey, Alex John London, Nathalie MacKinnon, Amanda MacPherson, Vince Madai, James Mattina, Michael Pratte, Tim Ramsay, Holly Sarvas, Esther Vinarov, Taiji Wang, Sandy Wong, Sean Zhang, Tiger Zheng and all the other members of the STREAM (Studies of Research, Ethics and Medicine) research group.

I am also very thankful for the efforts of the members of my thesis committee, Harry Atkins, Hugh Bennett, Dean Fergusson and Abe Fuks, who have been supportive and helpful at every stage of my doctoral work.

Alain Desroches, who translated my abstract into French, is also my greatest support and partner, and I love him very much.

This work was supported by a grant from the CIHR (EOG111391) and the Rabinowitch Fellowship for Biomedical Ethics.
**Contribution to original knowledge**

Chapter 1 is an original systematic review of the monotherapy trials of the anti-cancer drug sunitinib, as well as a cross-section of sunitinib combination therapy, in which we provide a novel ethical analysis of benefit, risk and outcomes within a successful drug's development.

Chapter 2 is an original systematic review of imatinib, dasatinib and nilotinib, in which we provide a novel ethical analysis of the moral efficiency of their development.

Chapter 3 is an original systematic review of monotherapy trials launched within 5 years after the approval of the 12 anti-cancer drugs that were first approved by the FDA between 2005-2007, in which we provide a novel ethical analysis of a retrospective cohort study of monotherapy drug development trajectories.

Chapter 4 is an original systematic review and meta-analysis of the combination therapy trials launched within 5 years after the approval of the 12 anti-cancer drugs that were first approved by the FDA between 2005-2007, in which we provide a novel ethical analysis of a retrospective cohort study of combination therapy drug development trajectories.

Chapter 5 is a definition and exploration of the original concept, “clinical agnosticism,” the uncertainty generated by poor hand-off between exploratory and confirmatory clinical trials.
Contribution of authors

Chapter 1

- Funding: Jonathan Kimmelman
- Conception and development of methods and protocol: Jonathan Kimmelman, Alex John London, Spencer Hey, Benjamin Gregory Carlisle, Georgina Freeman
- Data collection, curation and editing: Nadine Demko, Georgina Freeman, Amanda Hakala, Nathaline MacKinnon, Spencer Hey, Benjamin Gregory Carlisle
- Statistical analysis and data presentation: Benjamin Gregory Carlisle and Tim Ramsay
- Writing and editing: Benjamin Gregory Carlisle and Jonathan Kimmelman

Chapter 2

- Funding: Jonathan Kimmelman
- Conception, development of methods and protocol: Jonathan Kimmelman, Tiger Zheng, Benjamin Gregory Carlisle
- Data collection, curation: Tiger Zheng and Benjamin Gregory Carlisle
- Statistical analysis and data presentation: Benjamin Gregory Carlisle
- Writing and editing: Benjamin Gregory Carlisle, Tiger Zheng and Jonathan Kimmelman

Chapters 3 and 4

- Funding: Jonathan Kimmelman
- Conception, development of methods and protocol: Jonathan Kimmelman, Benjamin Gregory Carlisle
- Statistical analysis and data presentation: Benjamin Gregory Carlisle
- Writing and editing: Benjamin Gregory Carlisle, Jonathan Kimmelman, Adelaide Doussau

Chapter 5

Carlisle 11
• Funding: Jonathan Kimmelman

• Data collection: Carole Federico and Benjamin Gregory Carlisle

• Conception, development of methods, writing and editing: Benjamin Gregory Carlisle, Carole Federico and Jonathan Kimmelman
Introduction

Literature review

Drug development is a lengthy, extensive, costly and error-prone process, in which patients are exposed to risky treatment regimens.\textsuperscript{1–3} In some clinical trials, the efficacy of the drug under consideration is entirely unproven.\textsuperscript{4} Uncertainties regarding the efficacy of the drug puts patient benefit in question.\textsuperscript{5} The profile of adverse events is likewise uncertain.\textsuperscript{6,7} The likelihood of a serious adverse event varies by drug type, but in oncology, the rates of treatment-related death can be very high.\textsuperscript{8,9}

Clinical trials in drug development often occur within large programmes of research that do not necessarily stop when a drug receives approval. Drugs with well-characterized efficacy are often tested in expansive clinical trial portfolios against speculative new indications, combinations, populations, schedules or doses.\textsuperscript{10–13} We define a “portfolio” to be a family of related clinical trials testing hypotheses that are related. A portfolio can be defined in terms of the drug being tested, a mechanism of interest, or even a drug class.\textsuperscript{14} A drug that was developed for one indication may pose a different set of risks when tested in another indication,\textsuperscript{15,16} and the effectiveness of a drug in one cancer indication will be correlated with its effectiveness in another cancer indication. In many attempts to extend the use of a drug to new indications, the nature of the adverse events is not known, to say nothing of their frequency.

Risks in drug development also extend beyond the personal risk and benefit experienced by patients enrolled on clinical trials.\textsuperscript{17,18} Future medical practice relies on valid inference from trustworthy evidence. A clinical trial represents an opportunity cost to investigators, administrators, institutions, funders and entire research programmes, who will be unable to allocate time, effort or other resources to another trial once it has been invested. A poorly run trial may shake the trust that we have in the enterprise of scientific research, making future recruitment of patient-subjects more difficult, and decreasing the political impetus for advancing medical knowledge through research.\textsuperscript{19}
Given the risky and costly nature of drug development, the efficiency of drug development efforts is rightly questioned. However, analyses of the efficiency of drug development often focus on the economic costs of bringing a successful drug to market, with the cost of drug development often being quoted in terms of dollar amounts per drug. Despite the fact that many of the risks in drug development are borne by patient-subjects, analyses that focus on the economic efficiency of drug development fail to consider the expenditure of human welfare that is involved. Further, a focus on bringing a drug to market may fail to recognize inefficiencies to drug development that occur in clinical trials of a drug that are launched after the drug has already received regulatory approval. A purely economic evaluation of the cost of bringing a drug to market misses many of the relevant moral dimensions of clinical translation.

Clinical trials are, of course, ethically evaluated before their launch through the process of prospective review, during their conduct by a Data and Safety Monitoring Board and after their conclusion in publication, peer review, meta-analysis and uptake into practice. Modern clinical trials are generally scrutinized by an independent committee to ensure that they meet a number of ethical requirements, including: a favourable balance of risks and benefits, justice in patient selection, scientific validity and value to society. These norms have been taken up by medical journals, codified by codes of research ethics, applied by institutional review boards, required by funding agencies, and even enshrined in law. However, prospective review generally focuses on the ethical properties of a single trial. It is an often-overlooked assumption that the unit of ethical analysis for drug development is the individual clinical trial.

This thesis will explore the relationship between scientific value of clinical trials and the balancing of risk and benefit to patient-subjects, from a meta-level perspective on human research. Scientific value and a favourable balance of risks and benefits have been recognized since the Nuremberg Code, the Belmont Report and the Declaration of Helsinki, as necessary for ethical research in human subjects. Here, we look for patterns within human research that may be invisible at the single-trial level, but become apparent when larger programmes of research are looked at in the aggregate.

This is not, of course, the first attempt to conduct meta-research on drug development. There has already been substantial meta-research on clinical trials revealing insights on research waste, duplicative research, multiple testing and many cancer-specific research trends.
Risk and benefit are estimated and weighted prospectively by ethics review, and measured retrospectively by systematic reviews and meta-analysis. However, the efficiency of drug development—the amount of clinical impact derived from human research and how that relates to the human welfare expended to achieve it—is usually evaluated in economic terms, and ignores many morally relevant inefficiencies. In this thesis, I will apply meta-research techniques to capture dynamics that extend beyond the individual trial, and evaluate the morally relevant inputs, outputs and potential residuals of drug development.

Moral efficiency

An efficient process is one that makes good use of its inputs to generate a desirable yield of outputs, while minimizing the amount of residuals. I will define the moral efficiency of drug development to be the relationship between useful scientific value accrued from clinical trials and the human welfare expended to attain it. Useful scientific knowledge can come in many forms, including directly actionable information produced by a clinical trial, instructive negative findings, uptake of a therapy regimen into clinical practice guidelines, or regulatory approval of a drug or combination of drugs for use in a particular indication. There are likewise many ways that human welfare may be expended over the course of a clinical trial. Patient-subjects give up time, undergo potentially burdensome research procedures, experience opportunity costs, and take on risk of incurring treatment-related adverse events or death.

In this thesis, I will take a broad, meta-level look at entire drug development programmes in cancer to examine the moral efficiency of anti-cancer drug development. I will quantify both the scientific value of a drug development programme, and the risk and benefit to patient-subjects. I will compare the efficiency of the development of different drugs, and highlight potential inefficiencies. I will identify less commonly appreciated sources of inefficiency, such as changes in risk and benefit profiles over the course of a drug’s development, large programmes of post-approval exploration of new indications and combinations, and threats to the scientific value of clinical trials.

This project will inform prospective patient-subjects, the practice of ethical review, research policy, and the interpretation of clinical trial results. It will also guide further investigations into the nature and potential pitfalls of clinical trial analysis.

Carlisle 15
Methods

Many of the methods used in this thesis have been re-purposed from the field of systematic reviews and meta-analysis and applied to meta-research. While meta-analysis is often an attempt to mathematically pool the effect sizes measured in two or more clinical trials of the administration of a single treatment regimen or drug to a single population,\textsuperscript{49,50} we are not attempting to estimate a treatment effect size for clinical practice. We use systematic review methods to capture a reproducible sample of clinical trials for study,\textsuperscript{51} and in some cases, use meta-analysis\textsuperscript{52} or other mathematical models or data presentation methods\textsuperscript{53,54} to clarify dynamics within and across drug development programmes.

Two of the tools used in preparation of this thesis represent original contributions to methods for meta-research on clinical trials. Over the course of this project, I developed and refined The Drug Trials Visualiser,\textsuperscript{55} a tool for automatically generating a chronological chart of clinical trials based on registry entries in the National Library of Medicine (NLM) database\textsuperscript{56} and cross-referencing publicly available U. S. Food and Drug Administration (FDA) regulatory documentation. This was used to rapidly graph entire drug development programmes, assisting in hypothesis generation, estimation of project feasibility and data collection. Numbat Systematic Review Manager\textsuperscript{57} was also developed and refined throughout the course of this project. Numbat is software that enables the coordination of extraction of data from clinical trial records or publications by multiple users, and then merging these data into a single final copy. This systematic review management software was used extensively throughout the projects that make up this thesis. See Appendix G for details.

Project overview

This thesis progresses from the particular to the general, beginning by defining the methods that we will use in a study of one particular drug’s development and then applying them to a case series of related drugs. These methods are then used to quantify, for a larger set of drugs, a less appreciated source of inefficiency—indications and combinations launched into testing after a drug has already received regulatory approval. This thesis concludes with a concept paper that highlights one potential inefficiency in drug development. In chapter 1, I will consider the successful anti-cancer drug sunitinib. This study is an analysis of the evolution of risk within a
drug’s development. Chapter 2 will compare imatinib, widely considered to be the “poster child” for rational drug development against two of its “me too” drug successors, dasatinib and nilotinib. These case studies will allow us to compare the efficiency of development among a small set of related drugs. In chapters 3 and 4, I consider the trials within the first five years of post-licensure clinical trials for all of the novel anti-cancer drugs that received first FDA approval between 2005-2007 in monotherapy and combination therapy, respectively. Chapter 5 defines and describes the generation of “clinical agnosticism,” and how this relates to the scientific value of clinical trials.

In chapter 1, “Benefit, Risk, and Outcomes in Drug Development: A Systematic Review of Sunitinib,” I characterize the risk and benefit for patients enrolled in clinical trials testing sunitinib, a successful anti-cancer drug. The primary outcome of this study was to capture the entire portfolio of clinical trials for a single drug and characterize patient benefit and risk.

We searched Medline and Embase for clinical trials exploring sunitinib in trials that were launched at a time when the drug had not yet received FDA licensure for that indication. For example, because sunitinib was never approved for breast cancer, a trial of sunitinib in breast cancer that was launched at any time would be included. However, sunitinib was approved for renal cell carcinoma (RCC) and gastrointestinal stromal tumour (GIST) in 2006 as well as pancreatic neuroendocrine tumours (pNET) in 2011, and so a clinical trial that was launched before the approval date for that indication would be included in the sample, whereas clinical trials launched in approved indications after licensure would not be included. We considered all monotherapy trials and a cross-section of combination therapy, namely studies of combination therapy with sunitinib in RCC.

This literature search returned 115 clinical trial publications, 103 in monotherapy and 12 in combination therapy. In this sample, we found that risk and benefit to patients worsened over the course of the drug’s development, and there were several instances of very similar studies with negative results. There were very few examples of positive results after the first responding indications were identified.

The potentially duplicative negative trial results that were found, for example, represent instances of clinical trials that were, when considered in isolation, perfectly ethical. They were each reviewed by their respective institutional review board and approved. However by the lights
of an analysis of the moral efficiency of the entire set of drug development activities, their ethical justification becomes suspect. The risks imposed by clinical trial participants are justified, in part, by the scientific value of the study and the direct potential benefits to participants. Negative trial results suggest a low rate of response to the drug, and a replicated negative result is very low in scientific value, as the drug had already been tested in that indication and found to be ineffective or unfeasible for therapeutic use. Hence, these trials represent potential threats to moral efficiency.

In chapter 2, “Cancer Precision Medicine Drug Development is Often Empirical: The Case of Imatinib, Dasatinib and Nilotinib,” we extend the methods from the previous chapter to three anti-cancer drugs. Imatinib, first approved in 2001, is widely regarded as the “poster child” for rational drug development and precision medicine. Dasatinib and nilotinib are two “me too” drugs that were developed shortly thereafter. They are unique drugs, but they employ the same mechanism of action, targeting the Philadelphia chromosome and working as multi-tyrosine kinase inhibitors. This similarity allowed us to investigate a potential moral efficiency across the development programmes of these drugs. We examine the learning that can occur both within a single drug’s development portfolio and from one drug to other similar drugs, as evidenced by the exploration of hypotheses in clinical trials.

Our sample included 179 translational trials of imatinib, dasatinib and nilotinib. We defined a “trajectory” to be a set of trials that test a particular drug against a particular indication. For example, the set of clinical trials testing imatinib in breast cancer would be an example of a drug-indication trajectory. The efficiency of drug development is often considered in terms of the costs of bringing a drug to market, which potentially overlooks residuals, namely the costs incurred by translational research after a drug is already approved. Our methods allowed us to examine the ratio of patients enrolled in first-approval seeking trajectories (trajectories launched up to and including the first drug-indication pairing to receive licensure from the FDA) as compared to patients enrolled in label-extending trajectories (drug-indication pairings launched after the first trajectory to receive FDA licensure). We were also able to consider the proportion of patients enrolled in trajectories that would ultimately be vindicated with FDA approval.

While the reputation that imatinib has as a “magic bullet” seems to be deserved on the basis of the small volume of research activity that we found to be required to reach first licensure, the
post-approval trial activities are much less efficient in terms of patient enrolment required for subsequent FDA approvals. The balance of risk and benefit, which was very favourable toward the beginning of the imatinib’s development, shifted dramatically toward being similar to the risk and benefit balance of other non-precision medicine drugs.

Dasatinib and nilotinib testing was much more conservative than imatinib in terms of the number of trajectories explored. While there was broad indication exploration within the imatinib portfolio, a possible explanation for the restraint in label-extending exploration may have been learning from empirical insights gleaned from failures in imatinib testing.

These dynamics that we describe—a worsening balance of risk and benefit within the development of imatinib, and the potential efficiency of learning from another drug’s negative results—are, of course, invisible at the individual trial level. The popular understanding of imatinib development is that it was an unqualified success, starting with its rational development and culminating in a effective and safe treatment that was swiftly taken up into clinical practice. The methods and analysis applied here allow us a greater degree of nuance when considering drug development efficiency and give us means of comparing one drug to another.

In chapter 3, “Patient burden and clinical advances associated with post-approval monotherapy cancer drug trials,” I move from comparisons within and across single drugs to a retrospective cohort study of development trajectories launched after a drug’s first FDA approval. In this project, we apply the same systematic review methods developed in the previous projects described to all 12 novel anti-cancer drugs approved by the FDA between 2005-2007. These methods allow us to determine the investment in terms of patient involvement, and the payoff in terms of clinical impact for indications that were launched after approval.

This timeframe was chosen to capture five years of clinical trials after each drug’s first approval, and still allow sufficient follow-up time for every drug-indication “trajectory” within our sample to have eight years to be approved by the FDA or taken up into clinical practice guidelines. Clinical impact was measured in terms of FDA approvals, uptake into National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines and advancement to randomized controlled trials (RCT). The expenditure of human welfare was measured in terms of grade 3 or higher drug-related adverse events, and grade 5 drug-related adverse events.
Our sample included 104 clinical trials testing 69 drug-indication trajectories, enrolling 4699 patients. The drugs represented by our sample ranged in the number of trajectories from 0 in the case of nelarabine to 25 trajectories explored in the case of sunitinib. None of the drugs received an additional FDA approval for a trajectory that was launched after its first license, 5 were taken up into NCCN clinical practice guidelines, and only 9 trajectories had a RCT with a survival endpoint registered within 8 years of the launch of the first trial of the trajectory. The pooled grade 3-4 serious adverse event rate among all patients was 19.6% [95% CI 17-22%] and the rate of grade 5 serious adverse events (death) was 2.8% [2.3-3.4%]. Only one of the trajectories that were taken up into the NCCN clinical practice guidelines was also advanced to RCT testing.

This project reveals extensive monotherapy testing, significant patient burden, and limited advancement to confirmatory testing. There is limited clinical payoff in terms of FDA approvals or RCT testing, and clinical guideline uptake is often based on lower-level evidence. Resolving clinical uncertainty regarding a new drug’s potential applications to other indications certainly does have value, and this is a necessary part of the drug development process. However the clinical trials captured by this cohort study represent a very large investment expended for a modest amount of clinical uncertainty that is actually resolved.

These methods allow us to characterize how many trajectories are explored on a per-drug basis, and how often the pattern of extensive post-licensure testing observed in the case of sunitinib is recapitulated in other drugs. Two other drugs in our cohort, sorafenib and dasatinib, also had comparable programmes of label-extending testing in which 10 and 12 trajectories were launched. The approval by the FDA of most of the other drugs that we captured was still followed up with label-extending testing, but on a much more modest level.

Chapter 4, “Burden, Benefit, and Clinical Impact Associated with Exploring Cancer Combination Therapies After Drug Approval: A Systematic Review and Meta-Analysis” is parallel to chapter 3 in its methods, the drugs being sampled, and the follow-up time allowed. However, rather than characterizing the investment and clinical payoff for monotherapy in anti-cancer drug development, we apply our methods to post-approval combination therapy exploration.

This sample includes 323 clinical trials of anti-cancer combination therapy in “pairings” defined by a combination that included one of the original 12 index drugs that defined our cohort, with an
indication in which it is being tested. There were 266 unique pairings tested, enrolling 29,835 patients. Combination therapy exploration was much more extensive than monotherapy for the same cohort of drugs and the same amount of follow-up, representing a much higher investment of patient enrolment and human research resources.

Because the extent of trialling was greater than in monotherapy, we were able to meta-analyse several measures of patient benefit and burden, namely overall survival, progression-free survival and serious adverse event rates among trials where patients were randomized between an arm testing a combination therapy and a standard-of-care comparator. The pooled risk ratios for grade 3-4 and grade 5 serious adverse events for trials randomized between a combination arm and a comparator were 1.54 [1.33-1.79] and 1.51 [1.16-1.97] respectively. The pooled hazard ratios for OS and PFS were 0.99 [0.92-1.05] and 0.85 [0.79-0.93], respectively. None of the combination-indication pairings launched after initial drug approval received FDA approval, and 12 pairings (4.5%) were recommended by the NCCN within five years of the first trial within that pairing.

Similarly to the dynamics observed in monotherapy, combination therapy exploration represents a huge investment of patient involvement and clinical trial resources, with very limited payoff. Further, patient benefit in terms of overall survival is not significantly better in the combination arm as compared to the standard-of-care comparator, while serious adverse event rates are significantly higher.

Chapter 5, “Trials that say ‘maybe’: the disconnect between exploratory and confirmatory testing after drug approval” is a synthesis of many of the insights earned by the previous chapters. In this chapter, we define and explore the concept of “clinical agnosticism,” which we define to be the “suspicion of clinical value” that is often left unresolved by confirmatory testing. Clinical agnosticism is a threat to the scientific value of clinical trials, which is a source of research inefficiency. A state of clinical agnosticism brought about by an unconfirmed positive result from an hypothesis-generating trial is one that provides little guidance for clinicians who require good evidence on which to base their practice.

A large volume of lower-level evidence produced by hypothesis-generating trials, typically in phase 1 or 2, is generated by extensive programmes of indication or combination exploration that
are often launched after a drug receives its first regulatory approval. While these smaller, exploratory trials are a necessary part of the process of drug development, we find that there is often a disconnect between these trials and the confirmatory testing that these trials should lead to. To be sure, conditions of uncertainty are necessary in order to establish clinical equipoise, which is considered by many to be a condition for ethical research, however if this uncertainty is not resolved with confirmatory testing, it can be harmful in the following ways:

Prolonged periods of clinical agnosticism can lead to off-label prescription of drugs in indications or combinations where evidence is weak. Because of the nature of hypothesis-generating clinical trials, early signals of efficacy from smaller trials may be found to be much smaller than originally thought, or even completely overturned in confirmatory testing. Because of the extent of post-approval exploratory testing, this fails to redeem the patient welfare expended in hypothesis-generating trials; it may represent significant off-label prescription of drugs in indications or combinations where they are ineffective; and it may also shift incentives within healthcare systems away from producing good evidence on which to advance medical practice.

The following chapters examine the moral efficiency of anti-cancer drug development in the aggregate by the lights of systematic review. Every trial in our cohort was duly reviewed by an institutional review board, and found to have a favourable balance of risks and benefits. Every trial in our cohort also represents the exploration of a hypothesis that, at the time the trial was launched, was considered to be of sufficient scientific value to warrant a clinical trial. There is, of course, value in exploring other indications and combinations of anti-cancer drugs after a drug’s approval. And yet, when viewed in the aggregate, questions arise about the moral efficiency of the programme as a whole.

The moral inefficiencies found in anti-cancer drug development have implications for the practice of independent clinical trial review and for the monitoring of ongoing trials. There are also implications for the interpretation of clinical trial results, and the scientific validity of clinical trials. It is also possible that these insights can also be extended beyond the realm of anti-cancer drug development to other areas of drug development. I will explore some of these issues in the discussion section of my thesis.
Chapter 1. Benefit, Risk, and Outcomes in Drug Development: A Systematic Review of Sunitinib

Benjamin Gregory Carlisle, Nadine Demko, Georgina Freeman, Amanda Hakala, Nathalie MacKinnon, Tim Ramsay, Spencer Hey, Alex John London, Jonathan Kimmelman

Abstract

Background: Little is known about the total patient burden associated with clinical development and where burdens fall most heavily during a drug development program. Our goal was to quantify the total patient burden/benefit in developing a new drug.

Methods: We measured risk using drug-related adverse events that were grade 3 or higher, benefit by objective response rate and trial outcomes by whether studies met their primary endpoint with acceptable safety. The differences in risk (death rate) and benefit (ORR) between industry and non-industry trials were analyzed with an inverse-variance weighted fixed effect meta-analysis implemented as a weighted regression analysis. All statistical tests were two-sided.

Results: We identified 103 primary publications of sunitinib monotherapy, representing 9092 patients and 3991 patient-years of involvement over 10 years and 32 different malignancies. In total, 1052 patients receiving sunitinib monotherapy experienced objective tumor response (15.7% of intent-to-treat population, 95% CI 15.3%–16.0%), 98 died from drug-related toxicities (1.08%, 95% CI 1.02%–1.14%), and at least 1245 experienced grade 3–4 drug-related toxicities (13.7% 95% CI 13.3%–14.1%). Risk/benefit worsened as the development program matured, with several instances of replicated negative studies and almost no positive trials after the first responding malignancies were discovered.

Conclusions: Even for a successful drug, the risk/benefit balance of trials was similar to phase 1 cancer trials in general. Sunitinib monotherapy development showed worsening risk/benefit and the testing of new indications responded slowly to evidence that sunitinib monotherapy would not extend to new malignancies. Research decision-making should draw on evidence from whole research programs, rather than a narrow band of studies in the same indication.
Introduction

Numerous analyses have chronicled high costs, attrition and lengthy delays in drug development.\textsuperscript{1-3} In one recent study of FDA approved drugs, the median time between start of clinical development and regulatory licensure was 6.5 years.\textsuperscript{4} In another, only one in ten drugs entering clinical development is ultimately licensed for clinical application. The fraction of drugs receiving licensure in areas where clinical need is especially pressing—cancer and neurological drugs—is lower still.\textsuperscript{5,6} Such costs and attrition exact a heavy toll on health care systems, as expenses associated with drug development are ultimately absorbed into the cost of pharmaceuticals. They also impose burdens on patients participating in trials.

Little is known about the total patient burden associated with clinical development and where burdens fall most heavily during a drug development program. Clinical development is characterized by a range of different investigations: drug developers conduct multiple trials in parallel with each other, continue testing long after a drug is licensed, and search for responding secondary clinical indications alongside lead indications. Thus, many of the aggregate figures about translation success and cost obscure patterns of success and trends in risk and benefit within a whole drug development program.

Many recent policy and research initiatives aim at accelerating the pace and improving success rates in clinical translation.\textsuperscript{7-10} Understanding patterns of research activity and burden in drug development can help identify opportunities for directing research resources more efficiently. It can also identify activities where burden is greatest, thus providing an evidence base for human protections.

To better understand the total risk/benefit associated with developing a drug, and trends in risk/benefit spanning clinical development, we conducted a systematic review of monotherapy cancer trials for a successful drug, sunitinib, across 10 years of clinical development. We used objective response rate as a measure of benefit, and grade 3 or above drug-related adverse events as a measure of harm.
Methods

The primary aim of this study was to capture a portfolio of published clinical trials for a single drug, sunitinib, and to quantify the total amount of patient burden and benefit encountered in clinical development. We defined a portfolio as all monotherapy trials testing sunitinib in oncology. Our secondary goals were to track the evolution of risk/benefit over the course of development and the relationship between risk/benefit with funding. Because sunitinib is licensed only as monotherapy and to limit the scope of this research effort, we focused our analysis on monotherapy trials.

Literature search

We selected the drug sunitinib for analysis because it afforded ten years of published trials in a variety of disease indications, and a large but manageable number of trials. Trials were captured by searching Embase and Medline on March 11, 2015 for trials using the terms “sunitinib,” “Sutent,” and variations on “SU11248” and combined results with the exploded MeSH terms, including “randomized controlled trials,” “controlled clinical trials,” “random allocation,” “double-blind,” “single-blind,” “placebo,” “feasibility,” “pilot,” “proof of principle,” “open label,” “non-random,” “clinical trials” and “phase 1” to “phase 4.” (Appendix A)

Trial publications were then screened by BC for the following inclusion criteria: 1) primary report, 2) final report (where a trial's results were published more than once), 3) interventional trials, 4) human subjects, 5) phase 1 to phase 4 and 6) monotherapy trials. We excluded publications that were 1) secondary reports, 2) interim results, 3) meta-analyses, 4) retrospective or observational studies, 5) laboratory analyses of ex vivo human tissues, 6) reviews, 7) preclinical studies; 8) letters, editorials, guidelines, interviews, etc. During full-text analysis, abstract-only publications and poster presentations were excluded.

To explore a second dimension of drug development, we repeated the above for a second portfolio of sunitinib studies: combination studies involving the most intensively researched indication in our sample, renal cell carcinoma (RCC).
We devised an extraction template that captured variables in the following domains: trial demographics, methods, hypothesis, trial design, measures of patient risk / burden and benefit (including adverse events, treatment duration, response, and investigator conclusions). Criteria for extraction were pre-specified in a codebook, and coders underwent training before data collection.

We scored the number of grade 3 or 4 adverse events (defined by Common Terminology Criteria for Adverse Events (CTCAE) criteria\textsuperscript{11}) in a trial conservatively, based on the highest number of events in a single category per trial. We based median duration of treatment on the length of time a patient must be enrolled in the trial to collect the primary endpoint. When not specified, duration of treatment was imputed by the product of the median number of cycles and the period of each cycle. Toxicities were only scored if they were defined as probably or definitely treatment-related. Benefit was calculated in terms of the overall response rate (ORR, sum of complete and partial responses using RECIST criteria\textsuperscript{12}) in individuals receiving sunitinib on an intent-to-treat basis. We used ORR because it is a widely used surrogate endpoint in cancer trials, and allowed us to compare the magnitude of effect across diverse indications and trial phases. Indications were divided into meaningful categories identified in consultation with an oncologist (the exact indication for any node can be reviewed by reference to the numbered citation in our bibliography).

The success or failure of a study was determined based on whether the study reached its pre-specified primary endpoint (usually ORR) and whether the regime was deemed tolerable by the authors. All studies were extracted by two independent coders using Numbat software—an open source meta-analysis management tool developed by BC (Numbat available from: http://bgcarlisle.github.io/Numbat/). Disagreements were reconciled by discussion between coders. In cases where opening and closure dates were unavailable, we contacted the corresponding authors for these details with a reply rate of 35%. Our codebook is available upon request.
Analysis and statistics

The differences in risk (death rate) and benefit (ORR) between industry and non-industry trials were analyzed in R version 3.1.3\textsuperscript{13} using an inverse-variance weighted fixed effect meta-analysis implemented as a weighted regression analysis. The outcome was regressed on a 0/1 indicator variable to identify industry trials. The individual trial results were weighted by sample size, as the variance of a rate estimate is directly proportional to sample size. We defined $p<0.05$ as statistically significant. All statistical tests were two-sided.

Results

Study characteristics

Our literature searches captured 115 primary publications of interventional trials of sunitinib (see Figure 1 for a PRISMA flow diagram and Table 1 for per-trial details). There were 103 monotherapy studies\textsuperscript{14-113} and 12 combination therapy trials\textsuperscript{117-128} spanning 11 different combination therapies. Properties of trials in our sample are described in Table 1.

Total patient benefit and burden

The 103 trials included in the sunitinib monotherapy analysis represent 9092 patients and 3991 patient-years of involvement over 10 years and with corresponding authors based in 17 different countries. In the portfolio we identified, 7 different doses, 7 schedules, and 32 separate malignancies were tested. In total, 1052 patients in sunitinib monotherapy experienced objective tumor response (15.7% of intent-to-treat population, 95% confidence interval 15.3%–16.0%) and 98 died from drug-related toxicities (1.08%, 95% confidence interval 1.02%–1.14%). There was a minimum of 1245 grade 3–4 drug-related toxicities (13.7%, 95% confidence interval 13.3%–14.1%). In Figure 2a we present the sequence of monotherapy and select combination therapy trials using an Accumulating Evidence and Research Organization (AERO) diagram.\textsuperscript{129} Based on start dates for these trials, an average of 10 new monotherapy trials were launched each year.

The phase 1 acute myeloid leukemia (AML) trial is white despite not being followed up in phase 2 because it reports 5 (33%) partial responses, but these were short-lived and the authors of this
report recommended against further development. There are three phase 2 trials, in pancreatic adenocarcinoma, prostate cancer and hepatocellular carcinoma (HCC) that met pre-specified primary endpoints other than ORR (e.g. disease control rate (DCR) or 12-month progression-free survival (PFS-12)). However, the authors of the pancreatic adenocarcinoma trial “argue against any usefulness of sunitinib in this patient population.” The prostate cancer trial authors qualified their use of PFS-12 as a “soft endpoint,” and the authors of the HCC trial admit that PFS-12 “should no longer be used as a primary endpoint for trials testing new compounds in HCC.” In each of these cases, the trials had very few objective responses (1 (1.4%), 2 (5.6%) and 1 (2.2%), respectively) and no licensure was obtained for these indications. Later trials in mesothelioma, adrenocortical carcinoma, oligometasteses, uveal melanoma, endometrial cancer, meningioma and thymoma in Figure 2A appear to show evidence of success in identifying new responding indications. However, the positivity represents a switch in primary endpoint from response rate to progression-free survival. With the exception of endometrial cancer, all of these “positive” trials would have been considered negative, had response rate been chosen for primary endpoint (i.e. ORR was 5% or lower).

Figure 2B shows the exploration of 11 combination therapy regimes for RCC—an indication that responded and received licensure as monotherapy. In all, 241 patients and 111 patient-years of involvement in combination therapy studies over 5 years in 4 different countries, testing 11 different combination therapies, were sampled. In total, 64 patients experienced objective tumor response (32.2% of intent-to-treat population, 95% confidence interval 29.5%–34.8%) and 4 patients died from treatment-related toxicity (1.7%, 95% confidence interval 1.4%–1.9%). There was a minimum of 69 grade 3–4 drug related toxicities (28.6%, 95% confidence interval 25.2%–32.1%).

Sunitinib was first licensed by the FDA in 2006 for second-line gastrointestinal stromal tumor (GIST) and RCC, and was licensed for pancreatic neuroendocrine tumors (pNET) in 2011. The relationship between achievement of milestones in sunitinib development and patient burden as measured by treatment-related deaths in the whole portfolio is depicted in Figure 3. Very few patient deaths occurred before appropriate dose, schedule, and initial responding indications were defined (3 deaths before FDA approval). However, burden accumulated before a third indication was established—81 treatment-related patient deaths in total.

Carlisle 28
The above analysis suggests that most of the clinically useful applications of sunitinib were discovered very early on in clinical development, and our AERO diagram suggests a worsening risk/benefit balance with further testing of new malignancies. To characterize the evolution of risk and benefit, we plotted cumulative rates of objective response and drug-related life threatening morbidity and mortality for all monotherapy trials testing indications that were not yet FDA approved (Figure 4). Cumulative death rates increased, while cumulative ORR rates diminished with time.

**Patterns of research, coordination and outcomes**

Our AERO diagram reveals 30 phase 2 trials testing indications that had been previously tested in phase 1. In 3 instances, positive signal in phase 1 (defined as an objective response) led to positive phase 2 studies; in 8 instances, absence of signal in phase 1 was followed by negative phase 2 studies of the same indication in a larger population. In 18 instances, indications were explored in phase 2 without prior reported testing of the same indication in phase 1 trials. We found many instances where single indications were tested in monotherapy studies before the results of earlier studies in the same indication were available. For RCC and GIST, phase 2 studies were initiated before closure to enrollment in phase 1, and phase 3 trials were initiated before closure to enrollment in phase 2. In 2 instances, inconclusive initial phase 2 studies in a given indication were followed by negative trials in the same indication. We found 16 instances where negative phase 2 trials in an indication were followed by other, negative phase 2 trials in the same indication. We calculate that 22 patients died and 155 grade 3–4 events occurred due to drug toxicity in these replicative studies. Based on inception and closure dates in trial reports, however, no replicative phase 2 trials were initiated after the first closed to recruitment. The mean time between closure to recruitment and publication was 26 months. Mean time to publication for negative studies in our cohort of trials was similar to that for positive studies (26 vs 27 months).

Industry would be expected to vigorously pursue malignancies that show the greatest prospect of approval. Since industry is the single largest funder, it would also be better positioned to synthesize the totality of evidence and coordinate investigations accordingly. We therefore tested whether industry-funded studies showed a more favorable risk/benefit profile for trials of
indications not yet FDA approved. Industry-funded studies showed no statistically significant
difference in ORR (10.0% vs. 8.5%, p=0.62), or in drug-related death rate (1.54% vs. 1.53%,
p=0.98) when compared with non-industry.

Discussion

We found striking discontinuities as sunitinib advanced toward post-licensure trials, with success
rates and patient benefit diminishing as the research program matured. Our findings suggest
potential deficiencies in the way information was used for planning studies, and have
implications for the planning and review of trials for new interventions.

Our analysis reveals patterns of risk and outcomes that would not be apparent from aggregate
figures regarding success rates in translation. Though sunitinib clearly counts as a success story
for drug development, the number of negative trials for indications not yet FDA approved far
exceeded the number of positive trials, and risk/benefit worsened with time. Indeed, the total
risk/benefit for the whole monotherapy portfolio, which included phase 2 and 3 studies leading
to licensure, was similar to that reported for phase 1 monotherapy cancer studies in general.130

Key elements for unlocking the clinical utility of sunitinib (responding indications, schedule and
dose) were established early on—and at a very modest burden for patients. The almost uniform
inability to detect clinically promising signal after the first three indications suggests that after
initial discovery of responding indications, researchers were unable to marshal preclinical
evidence, knowledge of pathophysiology or biomarkers to select new indications for further
testing.

Policy discussions of clinical translation often tacitly portray the process of clinical translation as
an orderly and methodical process, whereby early phase studies beget safer late phase studies
that match test conditions.131 Our analysis of the whole portfolio of sunitinib monotherapy trials
—and a slice of combination therapy studies—reveals a pattern that runs contrary to this
portrayal. Later phase and same phase trials are launched before earlier trials in the same
indication are completed, and trials exploring new indications are run concurrently and in rapid
succession; the latter makes it difficult for researchers to use outcomes to plan subsequent
studies. Exploration of new indications continued despite mounting evidence that responding
indications had been saturated, or that criteria for launching tests of new indications were not bearing fruit. The patient burden associated with this approach to indication exploration—at least for sunitinib—was considerable.

Our identification of replicative studies, and a rapid fire approach to exploring indications, suggest that some burdens of translation, in particular, adverse events for participants and costs associated with unsuccessful primary endpoint attainment, arise not merely from inherent uncertainties, but also, from poor coordination and information flow across the full research portfolio.

The patterns we observed—if they hold for other drugs—have implications for ethical review and data monitoring of trials, and what is disclosed to prospective subjects during informed consent. For instance, patients might be told whether they are participating in a study that is preceded by studies with positive signal or not, and what this might entail in terms of risk. Data monitoring committees should remain abreast of and respond to the outcomes of other new indication studies in a translation trajectory. When a study in an indication fails, institutional review boards and data safety monitoring boards should require explicit justification for initiating new or continuing underway studies in the same indication.

Our findings also have implications for initiatives addressing inefficiencies in drug development. Studies in our sample were funded and pursued by various actors and sponsors, and no single entity is synthesizing findings across different indications and coordinating further testing. The highly favorable risk/benefit profile of studies early in development represents the effect of incentives for drug developers to sprint to market with an agent that can begin to recoup the costs of development. Industry quickly capitalized on low-hanging fruit. After clear demonstration of activities in several different malignancies, we see a process of exploration in which the search for additional indications appears to have been poorly coordinated and may have involved a relaxation of evidentiary criteria.

These results should be interpreted in light of the following limitations. First, trends for sunitinib monotherapy may not generalize to other drugs. Sunitinib is a multikinase inhibitor; the prospect of many potential applications for a relatively novel drug—and strong signal of activity early on—may have driven an unusual level of indication exploration; many off-target effects likely drove
the heavy burden. Second, to our knowledge, all of the studies enrolled patients who had exhausted established effective therapy. Given the absence of effective therapy for these life-threatening diseases, there are good reasons for vigorous research programs testing possible therapies. Nevertheless, patients with advanced disease are entitled to having their interests protected by proper planning, coordination and risk minimization. Moreover, studies of second-line therapy and beyond still entail opportunity costs and draw on scarce research resources. Third, any single malignancy may have seemed—and may continue to appear—a plausible candidate for sunitinib monotherapy. Nevertheless, knowledge of how sunitinib monotherapy had fared against a host of other malignancies might have tempered the rationale for further exploration of indications and combination therapies. Fourth, some apparent redundancy reflected in our AERO diagram reflects important differences in trial hypotheses. For example, variation in trial outcomes in the RCC stratum in part reflects that some trials tested RCC subtypes, such as non-clear cell RCC, that are less responsive to sunitinib. Fifth, had investigators chosen different endpoints (e.g. progression-free survival), the pattern of success and failure and evolution of risk/benefit might appear different. Last, our study relied on fully published research reports. According to one analysis, 61% of phase 2 studies are never published, and we found 36% registered sunitinib trials that were not published within 4 years of the trial’s registry closure date. However, we suspect that unpublished studies, if added to our analysis, would not improve the risk/benefit balance of the entire portfolio.

Our findings reinforce the suggestion, offered by others, that systematic review and research decision-making should draw on evidence from whole research programs, rather than particular tranches within them. Any one protocol may appear to have a sound basis, but when the sunitinib monotherapy development portfolio is viewed as a whole, there was a trend of worsening risk/benefit, and the testing of new indications responded slowly to accumulating evidence that sunitinib monotherapy would not extend to new malignancies. Most major codes of ethics prescribe the minimization of risk; a view of the entire portfolio and an assessment of cumulative risk/benefit offer a resource for achieving these ends.
Figures

Figure 1. PRISMA flow diagram
Figure 2. AERO graphs for sunitinib therapy. A) monotherapy stratified by indication. Nodes represent all available sunitinib cancer trials testing anti-cancer activity, arranged according to first publication date horizontally and stratified by indication. Larger square nodes are phase 1 trials in the designated indication. Small squares represent presence of patients with a particular indication in a “mixed malignancy” phase 1 trial. Circular nodes are phase 2 trials. Triangles are
phase 3 trials. White nodes indicate acceptable toxicity and positive effect; grey represents inconclusive results; and black nodes indicate negative results for their prespecified primary endpoint. Number of nodes may not sum to figures represented in table 1 due to different classificatory schema used in AERO graph. B) AERO graph for sunitinib combination therapy. Nodes represent sunitinib cancer trials, stratified by combination therapy, arranged according to publication year. Number of nodes may not sum to data in table 1 due to different classificatory schema used in AERO graph.

Figure 3. Cumulative treatment-related deaths in trials of sunitinib for key milestones. The total number of deaths in the whole portfolio is charted against time. Dates are based on trial closure.
Figure 4. Cumulative risk and benefit proportions for studies in non-FDA approved indications. All events were classified by report authors as probably or definitely drug related. Dates are based on trial closure.
### Tables

<table>
<thead>
<tr>
<th>Trial Characteristics</th>
<th>Monotherapy trials(^{14-116}) (n=103)</th>
<th>Combination therapy(^{117-128}) (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1</td>
<td>15 (15%)</td>
<td>8 (67%)</td>
</tr>
<tr>
<td>Phase 1–2</td>
<td>1 (1%)</td>
<td>2 (17%)</td>
</tr>
<tr>
<td>Phase 2</td>
<td>80 (78%)</td>
<td>2 (17%)</td>
</tr>
<tr>
<td>Phase 3</td>
<td>7 (7%)</td>
<td></td>
</tr>
<tr>
<td>Sponsor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry only</td>
<td>43 (42%)</td>
<td>8 (67%)</td>
</tr>
<tr>
<td>Non-industry</td>
<td>54 (52%)</td>
<td>3 (25%)</td>
</tr>
<tr>
<td>Not stated</td>
<td>6 (6%)</td>
<td>1 (8%)</td>
</tr>
<tr>
<td>Number of centers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single center</td>
<td>36 (35%)</td>
<td>3 (25%)</td>
</tr>
<tr>
<td>Multi center</td>
<td>63 (61%)</td>
<td>8 (67%)</td>
</tr>
<tr>
<td>Not stated</td>
<td>4 (4%)</td>
<td>1 (8%)</td>
</tr>
<tr>
<td>Randomization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized</td>
<td>19 (18%)</td>
<td>1 (8%)</td>
</tr>
<tr>
<td>Not randomized</td>
<td>84 (82%)</td>
<td>11 (92%)</td>
</tr>
<tr>
<td>Trials with a non-sunitinib comparator</td>
<td>15 (15%)</td>
<td>2 (17%)</td>
</tr>
<tr>
<td>arm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average duration (weeks)</td>
<td>20.0</td>
<td>18.5</td>
</tr>
<tr>
<td>Results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>48 (47%)</td>
<td>1 (8%)</td>
</tr>
<tr>
<td>Inconclusive</td>
<td>14 (14%)</td>
<td>2 (17%)</td>
</tr>
<tr>
<td>Negative</td>
<td>41 (40%)</td>
<td>9 (75%)</td>
</tr>
<tr>
<td>Mean sample size</td>
<td>88.3</td>
<td>20.1</td>
</tr>
<tr>
<td>Location of corresponding author</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N America</td>
<td>58 (56%)</td>
<td>9 (75%)</td>
</tr>
<tr>
<td>Europe</td>
<td>32 (31%)</td>
<td>3 (25%)</td>
</tr>
<tr>
<td>Asia</td>
<td>11 (11%)</td>
<td></td>
</tr>
<tr>
<td>S America</td>
<td>1 (1%)</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1 (1%)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Properties of extracted trials in our sample

### Notes

This work was funded by CIHR (EOG111391). Without intending to suggest their endorsement of our analysis, we thank Dean Fergusson, Abe Fuks and Charles Weijer for various discussions; we also thank Janet Dancey for consultations.

Conflict of Interest Statement: The authors declare no competing interests related to the content of this paper.
References


Carlisle 42


[59] Mahmood S Tariq, Agresta Samuel, Vigil Carlos E, et al. Phase II study of sunitinib malate, a multitargeted tyrosine kinase inhibitor in patients with relapsed or refractory soft tissue...


[78] Baldazzi Valentina, Tassi Renato, Lapini Alberto, Santomaggio Carmine, Carini Marco, Mazzanti Roberto. The impact of sunitinib-induced hypothyroidism on progression-free survival

Carlisle 45


[133] Ioannidis John PA, Karassa Fotini B. The need to consider the wider agenda in systematic reviews and meta-analyses: breadth, timing, and depth of the evidence BMJ. 2010;341.
Summary

The previous chapter, “Benefit, Risk, and Outcomes in Drug Development: A Systematic Review of Sunitinib” introduces our methods and provides an example of their use to illuminate the evolution of risk and benefit over the course of the successful anti-cancer drug, sunitinib. We apply systematic review methods to capture all of the pre-approval trials of sunitinib monotherapy, and a cross-section of combination therapy within our inclusion criteria. This allows us to report the volume of post-approval trial activity and quantify changes in objective tumour response as well as serious adverse event rates over the course of the drug’s development. We introduce the concept of a drug development portfolio, document a broad programme of indication exploration, and highlight research inefficiencies such as potentially replicative negative phase 2 trials.

In the next chapter, “Cancer Precision Medicine Drug Development is Often Empirical: The Case of Imatinib, Dasatinib and Nilotinib” we apply these methods to three related but distinct drugs, imatinib, dasatinib and nilotinib. Imatinib is a successful anti-cancer drug that was first licensed for use against chronic myeloid leukaemia. Dasatinib and nilotinib were developed after success of imatinib, making use of the same mechanism of action.

We applied the methods developed for sunitinib to imatinib, dasatinib and nilotinib in order to determine whether the patterns of risk, benefit and clinical payoff in sunitinib development would be recapitulated in the development of imatinib. We were interested to see whether the high development efficiency in early imatinib development would continue into its later stages of development, and to examine whether there were efficiencies to be gained as investigators build on evidence from earlier generations of drugs in the same class.
Chapter 2. Cancer Precision Medicine Drug Development is Often Empirical: The Case of Imatinib, Dasatinib and Nilotinib

Benjamin Gregory Carlisle, Tiger Zheng and Jonathan Kimmelman

Abstract

Introduction: Imatinib has a well-deserved reputation as a precision-medicine drug that was efficiently translated from concept into clinical applications. However, there has been little discussion about the efficiency of efforts aimed at extending this drug and second-generation BCR-ABL inhibitors to new indications.

Methods: We searched Medline and Embase for primary reports of clinical trials of imatinib, dasatinib and nilotinib aimed at validating the drugs against indications that were not FDA approved at the time of trial launch ("translational trials"). We extracted data on indications tested, patient enrolment and serious adverse events, and whether drug-indication development trajectories were productive (as defined by an ultimate FDA approval or positive randomized trial).

Results: We identified 179 translational trials imatinib, dasatinib and nilotinib. These drugs received FDA approvals for the first indication tested, enrolling 899, 1011 and 1480 patients, respectively. However, 83.6% of all indication-development trajectories pursued among the three drugs (comprising 10,203 patients or 71.0% of those receiving imatinib, dasatinib or nilotinib in translational trials) did not lead to an FDA approval or a positive randomized trial. Whereas 31 unproductive translational trajectories were pursued for imatinib (enrolling 6227 patients), 12 and 3 unproductive translational trajectories (enrolling 2652 and 1324 patients) were pursued for dasatinib and nilotinib, respectively. Precision medicine drug development strategies accounted for only 52% of trials; when used, however, development trajectories were far more likely to be successful, with 26% of these trials in indications that received FDA licensure vs 7% among non-precision medicine trials.

Discussion: Though imatinib received its first FDA approval with very few patients, further exploration of activity in other indications was much less productive and often used non-
precision medicine approaches. Dasatinib and nilotinib testing was more limited, and also more efficient in terms of patients enrolled per productive trajectory.

**Introduction**

Drug development in precision medicine is motivated by a molecular understanding of disease processes. The paradigm for this approach was established with imatinib, which was discovered as an inhibitor of the defining molecular marker for chronic myeloid leukaemia (CML), BCR-ABL. Imatinib was translated into an effective treatment for CML within 3 years of its first human trial.\(^1\)

Molecular insights have since enabled the rapid extension of imatinib to other malignancies, including C-Kit+ gastrointestinal stromal tumour (GIST),\(^2\) Philadelphia chromosome-positive acute lymphocytic leukaemia (Ph+ ALL), and three other cancer indication categories.\(^3\) Even in precision-medicine however, drug developers and clinicians are often uncertain about a drug’s range of activity—for example, whether the drug will show activity in diseases that share similar molecular characteristics or pathophysiological features.

Exploring the activity of a new drug in other indications or drug combinations can require a large number of trials, each of which entails patient burdens and opportunity costs for research systems. Previous studies suggest that the yield of clinically useful information from indication exploration after drug approval is often limited.\(^4,5\) Despite this, in principle, precision-medicine approaches should limit patient and trial burdens by narrowing the selection of indications for testing. Further efficiencies should be possible if developers of precision medicine drugs build on evidence from earlier generation drugs in the same class.

In what follows, we estimate the total number of trials and patient burden associated with discovering the range of applications for three precision-medicine drugs sharing very similar mechanisms of action: imatinib (approved by FDA May 2001), dasatinib (approved June 2006) and nilotinib (approved October 2007).

**Methods**

Our primary objective was to describe the portfolio\(^6\) of published translational trials (which we define as trials testing a drug-indication pairing or drug combination-indication pairing that are
initiated prior to or including a trial leading to an FDA license in that pairing, and excluding post-licensure on-label trials) for imatinib, dasatinib, and nilotinib (hereafter called "BCR-ABL inhibitors" based on a key shared mechanism of action) with the goal of determining how much patient burden was associated with identifying clinically useful applications of first- and second-generation BCR-ABL inhibitors.

**Literature search**

We conducted a Medline and Embase search for publications and published abstracts for trials of each the three drugs using the following search terms: drug name (and variants), and MeSH terms including variations of "clinical trial," or "randomized controlled trial," or other key words associated with clinical trial design. Our complete search strategy is in appendix A. Because the time period of this search straddled policies requiring clinical trial registration, and because our goals included a measure of patient burden, our searches were limited to publications and did not include searches of clinical trial registries. Our data cutoff allowed at least 7 years of follow-up since the enrolment date of the last trial in our study.

**Study selection**

Trial publications were screened for inclusion using the following criteria: 1) primary data, 2) full-text publication, 3) English language, 4) final report, 5) interventional trials, 6) administered imatinib, nilotinib or dasatinib as monotherapy or in a combination treatment regimen in 7) patients with a cancer diagnosis.

Exclusion criteria were: 1) laboratory study of ex vivo human tissues 2) case reports, 3) expansion cohorts of previously published trials, 4) adjuvant or maintenance therapy trials, 5) expanded access, and 6) trials not aimed at drug development (i.e. trials of drug-indication pairings or combination-indication pairings that were reflected on the FDA label at the time of trial launch). The date for search cut-off was January 2016.

**Extraction**

We extracted articles using a previously described template that captures the following: 1) purpose and investigator conclusions, 2) study design and funding, 3) patient characteristics, 4)
treatment procedures and duration, and 5) measures of patient risk and benefit—including measures of response, survival and adverse events.

Trial publications were extracted independently by two coders using Numbat meta-analysis management software and disagreements were resolved by discussion. Where possible, missing enrolment dates were inferred from the associated clinicaltrials.gov registration record. Otherwise, when data were not available, we emailed corresponding authors. We received responses for 19% of inquiries.

The most frequent single category of treatment-related grade 3 to 5 adverse events (G3–5 AEs) according to the National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) were captured on a per-trial level as specified in our extraction codebook. Trials using an efficacy primary endpoint were categorized as 1) positive, 2) inconclusive, or 3) negative on the basis of whether a pre-defined primary endpoint was attained with statistical significance. In cases where trials used multiple primary endpoints and results were inconsistent or no pre-defined endpoints were provided, trials were deemed inconclusive.

Trials were classified as pursuing a precision medicine strategy or a non-precision-medicine strategy based on whether the authors reported a biomarker being used to select patients for treatment or; if at least 50% of patients with a given malignancy are known to harbor the relevant biomarker, based on a previously-used definition. Combinations were considered to use a precision-marker approach so long as trial subjects were enriched with biomarkers for one of the drugs in the combination in question.

We define a "development trajectory" as a set of trials pursued in a drug-indication pairing preceding FDA approval or until a first positive randomized trial of that pairing (if any). Trajectories were "first approval-seeking" where they were initiated up to and including the first drug-indication pairing that received licensure from the FDA. Subsequent trajectories were considered to be "label-extending." We defined a "productive" trajectory to be one that resulted in an FDA approval or a positive phase 3 trial. We did not consider trials of BCR-ABL inhibitor combination therapies in indications that received licensure in BCR-ABL inhibitor monotherapy to be "productive" for the purposes of Figure 1, as no combination regimens received licensure or led to a positive phase 3 trial.

Carlisle 56
Analysis

Extractions were performed using Numbat Meta-Analysis Extraction Manager. Analysis and graphing was done using R v. 3.4.1. The heatmap in Figure 3 was plotted using the stats and gplots packages. Because the purpose of this analysis is descriptive, we did not perform any inferential statistics.

Results

Our literature search captured 179 translational trials of first- and second-generation BCR-ABL inhibitors. Of these, 128, 34, and 17 trials were conducted using imatinib, dasatinib, and nilotinib, respectively. BCR-ABL inhibitors were tested against at least 39 different disease indications and in 49 discrete drug combinations. A minority of trials (25.1%) were positive based on their primary endpoint. Regulatory licensure was ultimately granted for 9 cancer indication categories (16.4% of all drug-indication pairings tested) and 0 combination therapies. Of all patients enrolled in translational trials of BCR-ABL inhibitors, 29% participated in trajectories that were "clinically productive."

The first indication selected for imatinib testing went on to receive an FDA approval. Therefore, all patients participating in imatinib first approval-seeking trajectories participated in clinically productive trials. Thereafter, drug developers were less successful in discovering additional applications of imatinib. Of 38 cancer indications tested after the first CML trial of imatinib, 5 led to an FDA approval or positive randomized trial, and only 11.2% of patients participated in clinically productive label-extending trajectories (Figure 1). Especially unproductive were combination trials—none of which led to an FDA label or positive randomized trials.

The first indications selected for testing with the second generation drugs, dasatinib and nilotinib went on to receive FDA approval—and as such 100% of patients in first approval-seeking trajectories participated in clinically productive trajectories. Also, dasatinib and nilotinib developers published fewer studies aimed at extending these drugs to new indications or combinations (see Figure 4). For nilotinib in particular, the proportion of drug-indication pairings that went on to receive an FDA approval was higher (25%) than for its two predecessor drugs, imatinib (16.2%) or dasatinib (14.3%).

Carlisle 57
For dasatinib, the number of patients needed to obtain a first FDA approval, 1011, was comparable as compared with imatinib (899). Also, the proportion of efficacy trials that were positive for dasatinib (24%) was similar to imatinib (24%) and less than nilotinib (35%). Whereas imatinib developers were able to extend the drug to a total of 6 cancer indications, efforts at extending dasatinib and nilotinib to new indications were not as successful.

Precision medicine strategies were employed for 93 trials enrolling 8583 patients (59.7% of all patients among the 3 drugs in our sample). Monotherapy trials accounted for 67.4% of non-precision medicine strategy trials (48.8% of patients). Among the patients enrolled in trials applying a precision medicine approach, 4053 (47.2%) received a drug in an indication that was ultimately approved by FDA licensure. Among the 5795 patients enrolled in trials that did not use a precision medicine strategy, 122 (1.4%) received a drug in an indication that was clinically productive. (See Figure 2.)

A minimum of 2,927 patients experienced grade 3 to 5 serious adverse events or laboratory abnormalities by CTCAE criteria in BCR-ABL inhibitor translational trials. The majority of these adverse events, 1,527 (52.2%), occurred in interventional arms of imatinib trials; while 985 and 411 were experienced in dasatinib and nilotinib trials. As indicated in Figure 4, the kinetics of events for imatinib was similar for dasatinib, whereas nilotinib was able to attain FDA approvals with far fewer safety events.

**Discussion**

Imatinib is often presented as the "poster child" for precision medicine and has sometimes been portrayed as a "magic bullet." This status is deserved, given the responses observed in its first phase 1 trial, the small volume of clinical research activity needed to achieve a first regulatory approval (see Figure 1), and the drug’s impact on different diseases with certain molecular signatures.

Yet further efforts to extend this precision medicine treatment to new indications and combinations were preponderantly unproductive. Though six regulatory approvals in cancer indications were achieved for imatinib, the vast majority of patients enrolled in label-extending trials of imatinib (88.8%) participated in trials of indications or combinations that have not—to
date, with at least 7 years of follow-up since last trial captured in this study—been productive in terms of FDA licensure or leading to a positive phase 3 trial.

To put imatinib development in perspective, previous analyses of two non-precision medicine drugs, sunitinib and sorafenib, showed that 169 (4%) and 1455 (21%) patients enrolled in first approval-seeking trials, as compared to 899 (11%) for imatinib. For sunitinib and sorafenib, 47% and 22% of monotherapy trials were positive on their primary outcome as compared to 17% for imatinib monotherapy. Sunitinib and sorafenib trajectories met with FDA approval in 9% and 12% of cases, whereas 16% of cancer trajectories tested in imatinib met with FDA approval. While imatinib development did result in a greater number of FDA approved indications, the number of patients required to reach first license, and the number of trials with a positive outcome were comparable to two non-precision medicine drugs.

Drug developers appear to have been more selective with indication exploration for dasatinib and nilotinib. Several indications against which imatinib was tested were not tested—at least in publication—for dasatinib and nilotinib. Nilotinib trials were more likely to be positive on their primary endpoint than trials of the other two drugs. Though dasatinib developers and regulators could draw on outcomes from earlier imatinib testing, greater efficiencies were not observed for receiving a first approval of Ph+ CML.

One possible explanation for the greater efficiencies observed in dasatinib and nilotinib development is that developers were able to learn from the successes and failures of previous similar drugs and build on empirical findings. Alternatively, dasatinib and nilotinib developers—which include both commercial sponsors as well as public funding bodies—may have, for strategic, scientific, or commercial reasons, opted for a more restrictive approach to exploring secondary indications for these drugs.

Our findings should be considered in light of the fact that our analysis was based on published trials only. We cannot rule out that nonpublication might alter our findings somewhat. Given the well-established bias against publishing negative results, we suggest our report presents a somewhat more favorable picture of BCR-ABL inhibitor development. Moreover, some trajectories we described as "unproductive" may have involved drug-indication pairings that have utility in certain patient subgroups. If so, the level of evidence supporting those pairings is low. Finally, our
analysis is premised on the view that dasatinib and nilotinib are very similar to imatinib. Though all three target the BCR-ABL tyrosine kinase, each has different properties. As a consequence, the drugs are not fully interchangeable and some amount of exploration and label extension help clarify the drugs’ respective value. It is possible that

For the case of imatinib and its first imitators, precision medicine drug development approaches enabled rapid progression to its first approval, with very limited exposure of patients in trials. However, the effectiveness with which precision medicine can guide drug development must be considered in light of the totality efforts, not merely the first ones in a research programme. Despite sophisticated knowledge of molecular targets, many of the label-extending activities we identified would appear to have been empirically driven or not motivated by a precision-medicine hypothesis. The majority of patients involved in development of these trajectories—indeed the majority of patients in development trials in general—participated in trajectories that have not led to practice changing applications. Our study suggests limits on the efficiency of drug development in cancer precision medicine.
Figure 1. Patient enrolment in first approval-seeking vs label-extending trajectories for imatinib, dasatinib and nilotinib. One dark square represents 100 patients enrolled in a trajectory that received FDA approval; one light square represents 100 patients in a trajectory that did not receive FDA approval.
Figure 2. Patient enrolment in trials testing a precision medicine hypothesis with biomarker enriched patients (top) non-enriched for biomarker positive patients (bottom). One dark square represents 100 patients enrolled in a trial where the trajectory received FDA approval. One light square represents 100 patients enrolled in a trajectory that did not receive FDA approval.
Figure 3. Heat map of patient involvement. Warmer colours indicate greater level of patient enrolment. Initials in parentheses indicate FDA license for that indication in monotherapy.

Carlisle 63
Figure 4. Cumulative grade 3-4 SAE per drug development trajectory as a function of trial start date. Vertical lines indicate FDA approval dates for drug indication pairings as labeled. Note that plateaus of event volume for each curve probably is partly influenced by the unavailability of publications for trials launched since that date.
### Tables

<table>
<thead>
<tr>
<th></th>
<th>Imatinib (n=128)</th>
<th>Dasatinib (n=34)</th>
<th>Nilotinib (n=17)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phase</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1</td>
<td>3 (2.3%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Phase 1-2</td>
<td>11 (8.6%)</td>
<td>3 (8.8%)</td>
<td>1 (5.9%)</td>
</tr>
<tr>
<td>Phase 2</td>
<td>92 (72%)</td>
<td>30 (88%)</td>
<td>13 (76%)</td>
</tr>
<tr>
<td>Phase 3</td>
<td>2 (1.6%)</td>
<td>1 (2.9%)</td>
<td>3 (18%)</td>
</tr>
<tr>
<td>Not stated</td>
<td>20 (16%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td><strong>Sponsor</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry only</td>
<td>35 (27%)</td>
<td>16 (47%)</td>
<td>10 (59%)</td>
</tr>
<tr>
<td>Non-industry</td>
<td>65 (51%)</td>
<td>17 (50%)</td>
<td>6 (35%)</td>
</tr>
<tr>
<td>Not stated</td>
<td>28 (22%)</td>
<td>1 (2.9%)</td>
<td>1 (5.9%)</td>
</tr>
<tr>
<td><strong>Randomization</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomised</td>
<td>10 (7.8%)</td>
<td>3 (8.8%)</td>
<td>3 (18%)</td>
</tr>
<tr>
<td>Not randomised</td>
<td>118 (92%)</td>
<td>31 (91%)</td>
<td>14 (82%)</td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>31 (24%)</td>
<td>8 (24%)</td>
<td>6 (35%)</td>
</tr>
<tr>
<td>Inconclusive</td>
<td>31 (24%)</td>
<td>10 (29%)</td>
<td>6 (35%)</td>
</tr>
<tr>
<td>Negative</td>
<td>66 (52%)</td>
<td>13 (38%)</td>
<td>3 (18%)</td>
</tr>
<tr>
<td><strong>Mean patients enrolled per study</strong></td>
<td>61.8</td>
<td>108</td>
<td>165</td>
</tr>
<tr>
<td><strong>Location of corresponding author</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N America</td>
<td>71 (55%)</td>
<td>26 (76%)</td>
<td>5 (29%)</td>
</tr>
<tr>
<td>S America</td>
<td>1 (0.78%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Europe</td>
<td>46 (36%)</td>
<td>7 (21%)</td>
<td>8 (47%)</td>
</tr>
<tr>
<td>Asia</td>
<td>10 (7.8%)</td>
<td>1 (2.9%)</td>
<td>4 (24%)</td>
</tr>
</tbody>
</table>

Table 1. Sample demographics for trials in our sample of imatinib, dasatinib and nilotinib

### Notes

We gratefully acknowledge Dr Harry Atkins for his insight into leukaemia indications and surrogate measures.

### References


Summary

The previous chapter, “Cancer Precision Medicine Drug Development is Often Empirical: The Case of Imatinib, Dasatinib and Nilotinib” explores the attempts to extend the utility of the precision medicine imatinib to new indication areas, as well as comparing the imatinib portfolio with two of its “me too” successors, dasatinib and nilotinib. The post-approval development of imatinib is met with six subsequent regulatory approvals, however the vast majority of patients enrolled in label-extending trials of imatinib (88.8%) participated in indication or combinations that did not lead to FDA licensure or a positive phase 3 trial. Dasatinib and nilotinib indication exploration was much more conservative, testing fewer indications, and requiring fewer patients enrolled to reach their first FDA approval. This suggests decreasing efficiency over the course of the development of imatinib, but a gain in efficiency in the development of two closely related drugs.

In the next chapter, “Patient burden and clinical advances associated with post-approval monotherapy cancer drug trials,” we apply the methods from chapters 1 and 2 to five years of post-approval clinical trials in monotherapy for the 12 novel anti-cancer drugs that were approved between 2005-2007. We quantify the patient burden and benefit, as well as the clinical impact. This project provides insight into the volume of testing in post-approval monotherapy exploration, and gives some insight into whether the dynamics observed in chapters 1 and 2 are generalizable to a larger set of drugs. We describe patient burden in terms of serious adverse events and clinical impact in terms of FDA approvals, clinical guideline recommendations and advancement to randomized controlled testing.
Chapter 3. Patient Burden and Clinical Advances Associated with Post-Approval Monotherapy Cancer Drug Trials

Benjamin Gregory Carlisle, Adélaïde Doussau, Jonathan Kimmelman

Abstract

**Background:** After regulatory approval, drug companies, public funding agencies and academic researchers often pursue trials aimed at extending the uses of a new drug by testing it in new non-approved indications. The patient burden and clinical impact of such research is not well understood.

**Methods:** We conducted a retrospective cohort study of post-approval clinical trials launched within five years after the drug’s first approval, testing anticancer drugs in monotherapy in indications that were first pursued after a drug’s first FDA license for all 12 anticancer drugs approved between 2005-2007. Each trial was sorted into a trajectory defined by the drug and cancer indication. Risk was operationalized by proportions of grade 3-4 and grade 5 serious adverse events. Clinical impact was measured by estimating the proportion of patients participating in trajectories that resulted in an FDA approval, uptake into National Comprehensive Cancer Network (NCCN) clinical practice guidelines or advancement to randomized controlled trials (RCT) within eight years.

**Results:** Our search captured 104 published trials exploring monotherapy, including 69 unique trajectories. In total, trials in our sample enrolled 4699 patients. Grade 3-4 serious adverse events were experienced by 19.6% of patients; grade 5 events were experienced by 2.8% of patients. None of the trajectories launched after initial drug approval received FDA approval. Five trajectories were recommended by the NCCN within eight years of the first trial within that trajectory. Eleven trajectories were advanced to randomized controlled testing.

**Conclusions:** The challenges associated with unlocking new applications for drugs that first received approval 2005-2007 were similar to those for developing new drugs altogether. Our findings can help inform priority setting in research, and provide patients and physicians a basis for calibrating expectations when considering enrollment in label-extending trials.
Background

After new cancer drugs receive regulatory approval, researchers often pursue trials testing the drug in indications that would extend the use of the drug beyond the indication for which it was initially approved.\textsuperscript{1} Such "label-extending" trials are facilitated by the fact that newly approved cancer drugs are less likely to elicit unexpected safety issues, regulatory approval standards are weaker for post-approval trials,\textsuperscript{2} and the ease of testing an already approved and manufactured drug are likely to be much less as compared with a drug that has not yet received approval.\textsuperscript{1,3,4}

In numerous cases, label-extending trials have uncovered new uses for approved drugs—especially in the setting of rare malignancies. For example, imatinib was approved for hypereosinophilic syndrome, chronic eosinophilic leukemia and dermatofibrosarcoma, indications that were not launched into clinical trials until after the initial approval of imatinib.\textsuperscript{5} However, where much is known about the volume and clinical impact of pre-license clinical trials in cancer,\textsuperscript{6} there is much less systematic evidence about label-extending research activities. A previous study suggested that label-extending trials can sometimes be very extensive and failure-prone.\textsuperscript{7} A 2016 report found that 550 trials of the approved checkpoint inhibitor drugs pembrolizumab and nivolumab were open.\textsuperscript{8} Such activities, if not leading to regulatory approvals or major advances, can deprive other potentially productive lines of cancer drug development of eligible patients and resources. It can also lead to excess patient burden.

In the following, we use the research activities within a sample of trials of drugs used in monotherapy to estimate the patient burden, magnitude of investment, and success associated with label-extending monotherapy research efforts for drugs that were newly approved in years 2005-2007, inclusive.

Methods

Our primary objective was to measure the burden and clinical impact associated with participating in monotherapy trials aimed at extending the label of 12 anti-cancer drugs receiving first FDA approval 2005 and 2007 inclusive. This timeframe allowed us to capture publications all for trials launched within five years after first approval, affording at least an additional 8 years to account for publication delay after trial closure, uptake in National Comprehensive Cancer
Network (NCCN) guidelines, or FDA approval. All trials were considered to be part of a "trajectory," which was defined by the drug under study and the indication in which it is being tested (e.g., all trials testing temsirolimus against prostate cancer would be part of the same trajectory). Our analysis included only trials involving trajectories that were not listed on an FDA label at the time of trial launch.

We operationalized burden with two indicators: the rates of grade 3 or above drug-related adverse events in a single Common Terminology Criteria for Adverse Events (CTCAE) criterion, and the rates of grade 5 drug-related serious adverse events. We used patient enrolment and number of trials as a proxy for magnitude of research investment. We used advancement of a trajectory to a randomized controlled trial (RCT), FDA approvals and recommendations in the NCCN guidelines within 8 years of the launch of a trajectory as proxies for clinical impact.

**Literature search**

We searched Medline and Embase (see Appendix A) and manually screened for trials of each drug. Our inclusion criteria were: 1) primary data; 2) full-text; 3) English language; 4) final report; 5) interventional trials; 6) administered a drug in our sample in monotherapy in 7) patients with cancer 8) with an enrolment date within 5 years of the drug’s first FDA approval.

We excluded studies that were 1) laboratory studies only; 2) case reports; 3) expansion cohorts of published trials; 4) expanded access; 5) mixed malignancy and 6) FDA-approved on-label studies where enrolment began after regulatory approval for the same indication. Adjuvant, maintenance therapy, or palliative therapy trials were excluded to focus on monotherapy and curative research efforts.

Once trials were classified into trajectories, we classified them by whether there was a registry entry in clinicaltrials.gov for at least one randomized clinical trial with a survival endpoint (progression-free survival, PFS or overall survival, OS) testing that trajectory (drug and indication) with start dates within 8 years of the trajectory’s launch.

**Extraction**

Extractions were completed according to a previously published template, using Numbat Meta-Analysis Extraction Manager. Our extraction form (see Appendix B) captured a) drug under

Carlisle 70
study; b) study indication; c) study characteristics (e.g. phase, sponsor); d) trial demographics (e.g. number enrolled); e) enrolment date; d) primary outcome; e) treatment-related serious adverse events (SAE).

Trajectories were defined by the anticancer drug being studied and the broad cancer indication explored. Solid tumour indications were classified by anatomical site. In consultation with a hematologist-oncologist after extraction but before analysis, we grouped hematological malignancies into five indication categories (see Appendix B).

Because the goal of this project was to quantify the burdens and costs expended in label-extending research, the total patient-subject \( n \) per trial was extracted as the total number of patients enrolled, regardless of whether they were enrolled on an arm that received the drug in question, or were ultimately included in the final safety or efficacy analysis. Trials were marked as having a "positive" result if they met their prespecified primary efficacy outcome with statistical significance and toxicity was reported as acceptable, according to the authors of the publication. SAEs were counted conservatively, capturing only the reported CTCAE\(^9\) category with the highest number of adverse events.

**Analysis and statistics**

Analysis of results was conducted using \( R \) version 3.4.1.\(^{12}\) Graphics were prepared using the \texttt{ggplot2} package.\(^{13}\) Pooled serious adverse event rates were calculated using the random-effects inverse variance meta-analysis method for proportions from the \texttt{meta} package, version 4.8-2.\(^{14}\) Unless explicitly stated otherwise, ranges given after pooled proportions or hazard ratios represent 95% confidence intervals.

**Results**

**Volume and characteristics of label-extending trials testing monotherapy**

Our search captured 104 label-extending trials of monotherapy, with 69 drug-indication trajectories that were launched within five years of a drug’s first FDA approval.

One or more label-extending trials were pursued for 11 of the 12 drugs in our cohort (ranging from 1 to 25 trajectories per drug, see Table 1 and Appendix D for details).\(^{15}\) Label-extending

Carlisle 71
trials were predominantly single arm studies (k=99; 95%), phase 2 (k=96; 92%), followed by phase 1 trials (k=7; 6.7%) and phase 3 (k=1; 1.0%). Industry funded 39 trials (38%) in our sample and 57 trials (55%) had non-industry sponsors (the remaining trials did not report a funding source). In 18 trajectories (26%), the first trial was sponsored by industry, whereas there was a non-industry sponsor for the first trial of 45 trajectories (65%; the remaining 6 trajectories had first trials that did not report a funding source). Label-extending trials enrolled 4699 patients.

**Patient risk and burden**

Toxicity was reported as acceptable by the authors of the report in 71 trials (68%) in our sample. Among the trajectories represented by our sample, at least 875 patients, 19.6% [17-22%] experienced a treatment-related grade 3-4 serious adverse event, and 59 patients, 2.8% [2.3-3.4%] experienced a treatment-related grade 5 serious adverse event (see Appendix J for per-trial details and Appendix K for per-trajectory details).

**Advancement to randomized trials, clinical practice and FDA labels**

Figure 1 shows the relationship between launch of new label-extending trajectories, their testing in randomized trials, and uptake into NCCN clinical practice guidelines and/or FDA approval.

Label-extending trials with an efficacy endpoint were positive in 37 instances (36%). Nevertheless only 11 trajectories (1.4%) progressed to randomized testing using a survival primary endpoint within 8 years of the first trial in the trajectory. Of the 3 that had posted results to clinicaltrials.gov as of this writing, one was terminated for insufficient accrual, one was terminated early due to serious adverse events, and one had a PFS hazard ratio that did not show a significant difference between arms. There were 24 trajectories in our sample in which more than half of the trials had a positive primary endpoint. Among these trajectories, only 4 (16%) led to a confirmatory RCT with a survival primary endpoint registered within 8 years of the launch of the trajectory.

No trajectories led to a new FDA license within eight years of the launch of the trajectory. Five trajectories (7.2%) involving two drugs (sunitinib and sorafenib) led to recommendations in NCCN clinical practice guidelines within 8 years of the trajectory's first clinical trial. Only one of

Carlisle 72
new NCCN recommendations was based on an RCT registered within 8 years of its launch (sunitinib for pancreatic adenocarcinoma).\textsuperscript{19} However, the overall survival endpoint of the trial cited in the NCCN guideline was not significant. See Figure 1 for a plot of the dynamics of launching of new trajectories as compared with advancement to RCT and take-up by NCCN clinical practice guidelines.

**Discussion**

In our cohort of twelve newly approved drugs, 69 label-extending trajectories were launched within 5 years of FDA approval. Of these, 11 (16\%) advanced to randomized phase 3 trials, 5 (7.2\%) led to recommendations in clinical practice guidelines, and none resulted in new FDA approvals within eight years of trajectory launch. These activities involved 104 trials and enrolled 4699 patients and resulted in 875 patients (19.6\%) experiencing drug-related adverse events that were grade 3 or higher.

To put these estimates in perspective, the probability that a new cancer treatment will advance from entry into phase 1 testing to FDA approval has been reported to range from 10.4-11\%.\textsuperscript{6,20} The frequency with which unapproved cancer drugs in phase 2 advance to phase 3 testing was estimated to be 28.3\%. Estimates of drug-related grade 3 or above adverse events in phase 1 studies have been reported to be 10.3\%.\textsuperscript{21} Therefore, for the label-extending trial activities we studied, rates of clinical advancement were lower than for pre-approval research while toxicities were higher than those that have been reported for phase 1 trials.

Our findings have potential implications for human protections, research policy, and clinical decision-making. The moral basis for trials rests on burdens being redeemed by benefits to volunteers (if any) and new knowledge regarding potential uses of a drug.\textsuperscript{22} Our analysis provides a baseline for considering the clinical value and patient burden associated with attempted repurposing of new cancer drugs for other malignancies. Such data can be used to explain to patients considering label-extending trials historic rates with which patient contributions have led to major changes in clinical practice. Though it might be objected that our drugs are from an era that predates immunotherapy and precision medicine, all but one of our twelve drugs are targeted agents, and two are precision medicine drugs. At the time of the trials in our sample, molecularly targeted agents were taken to be fundamentally different from previous generations...
of oncology drugs, and there was an expectation that they would lead to lower rates of attrition in drug development.\textsuperscript{23}

Our study also has implications for clinical practice. The volume of label-extending research is such that false positive signals of treatment activity due to chance are inevitable for some drug-indication trajectories. On the assumption that the null hypothesis in one efficacy trial in each of the trajectories in our sample were true, that statistical tests were two-sided, that trial outcomes were independent, and that researchers pre-specified an alpha of 0.05, the probability that there would be at least one false positive among them is greater than 96%.\textsuperscript{24} In our sample, 5 trajectories were recommended for off-label use by the NCCN for treatment of various malignancies. Given that only one of these trajectories has advanced to an RCT, it seems plausible that some of the evidence on which such recommendations are based may rest on false positives and/or overestimates of effect.

Our analysis has several limitations. First, we considered only published monotherapy trials launched within 5 years of first license. However, given the tendency to selectively report positive trials, we suspect our findings would be even less favorable for label-extending research had we been able to capture the results of unpublished studies. Second, a more recent sample of drugs might show a different pattern. As noted above, previous generations of cancer drugs were also viewed with optimism regarding their anticipated effect on attrition in drug development. We note that clinicaltrials.gov lists numerous active trials testing drugs in our cohort. The registry also lists a large volume of label-extending trials for newer drugs. For example, for pembrolizumab, there are at least 30 active label-extending trajectories (search date 2018 October). Last, our analysis is not equipped to morally evaluate the relationship between risk and benefit in label-extending trials. One successful label-extending trajectory—if it entailed a substantial impact on disease—could justify the efforts associated with label-extending research. Many label-extending trials involve rare malignancies.

Label-extending trials offer opportunities to build on safety data accumulated in pre-license research, as well as emerging mechanistic knowledge about a new drug. Despite this, our findings suggest that the challenges associated with unlocking new applications for drugs that first received approval 2005-2007 were similar to those for developing new drugs altogether. Our
findings can help inform priority setting in research, and provide patients and physicians a basis for calibrating expectations when considering enrollment in label-extending trials.
Figures

Figure 1. Cumulative plot of the number of drug-indication trajectories launched into testing within 5 years of initial FDA approval, trajectories advanced to randomized controlled trials within 8 years of trajectory launch, and trajectories recommended in NCCN clinical practice guidelines within 8 years of trajectory launch. There were no new FDA approvals in trajectories launched after the initial FDA approval. Grey lines reflect the last value of the specified follow-up period carried forward.
Tables

<table>
<thead>
<tr>
<th>Drug</th>
<th>Trajectories launched</th>
<th>Trajectories recommended by advanced to RCT NCCN</th>
<th>Trials (k)</th>
<th>Patients (n)</th>
<th>Grade 3-4 SAE (n, %, 95% CI)</th>
<th>Grade 5 SAE (n, %, 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dasatinib</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>763</td>
<td>96, 14.1% [9.7-20%]</td>
</tr>
<tr>
<td>Decitabine</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>59</td>
<td>18, 33.9% [8.7-73%]</td>
</tr>
<tr>
<td>Isabeplone</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>85</td>
<td>13, 15.3% [9.1-25%]</td>
</tr>
<tr>
<td>Lapatinib</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>18</td>
<td>1, 5.6% [0.78-31%]</td>
</tr>
<tr>
<td>Lenalidomide</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>241</td>
<td>73, 30.1% [16-49%]</td>
</tr>
<tr>
<td>Nelarabine</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Nilotinib</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>73</td>
<td>6, 9.4% [4.2-19%]</td>
</tr>
<tr>
<td>Panitumumab</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>68</td>
<td>8, 14.0% [4-39%]</td>
</tr>
<tr>
<td>Sorafenib</td>
<td>10</td>
<td>3</td>
<td>0</td>
<td>12</td>
<td>394</td>
<td>62, 16.7% [11-25%]</td>
</tr>
<tr>
<td>Sunitinib</td>
<td>25</td>
<td>2</td>
<td>4</td>
<td>47</td>
<td>2652</td>
<td>524, 21.7% [18-26%]</td>
</tr>
<tr>
<td>Temsirolimus</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>214</td>
<td>37, 18.7% [11-29%]</td>
</tr>
<tr>
<td>Vorinostat</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>132</td>
<td>37, 21.8% [7.6-49%]</td>
</tr>
<tr>
<td>All drugs</td>
<td>69</td>
<td>5</td>
<td>11</td>
<td>104</td>
<td>4699</td>
<td>875, 19.6% [17-22%]</td>
</tr>
</tbody>
</table>

Table 1. Sample demographics per drug

Notes

We gratefully acknowledge the contributions of Harry Atkins, Gina Freeman, Spencer Hey, Amanda Hakala, Tiger Zheng, Taiji Wang, James Mattina, Yasmina Hachem, Nadine Demko, Natalie McKinnon, Sylviya Ganeshamoorthy, Holly Sarvas, Amanda MacPherson and the other members of the STREAM research group. This work was funded by CIHR (EOG111391).

References


Carlisle 78


Summary

The previous chapter, “Patient burden and clinical advances associated with post-approval monotherapy cancer drug trials,” and the next chapter, “Burden, Benefit, and Clinical Impact Associated with Exploring Cancer Combination Therapies After Drug Approval: A Systematic Review and Meta-Analysis” are parallel projects using similar methods, drugs, data sources, and timeframes, but in the former, we look at the moral efficiency of post-approval monotherapy and in the latter, we look at combination therapy.

In the previous chapter, we examined 5 years of post-approval monotherapy clinical trials for the 12 anti-cancer drugs approved by the FDA between 2005-2007. We measured patient involvement and risk, and quantified the clinical impact of trial activity in terms of FDA approvals, clinical guideline recommendations and advancement to randomised controlled trials. This project finds a high degree of patient investment and numerous clinical trials, but no new FDA approvals. Only five drug-indication trajectories (7.2%) led to recommendations in the National Comprehensive Cancer Network (NCCN) clinical practice guidelines within 8 years of follow-up, and only one of the NCCN recommendations was based on a randomized clinical trial. These results suggest a high level of investment for a low amount of return.

The next chapter is an application of the same methods to the clinical trials of combination therapy for the same drugs and the same timeframe. Combination therapy exploration is much more extensive than monotherapy, and so there is a large enough sample of randomized trials with survival endpoints to meaningfully meta-analyze patient-level data. We quantify the patient risk as pooled risk ratios for grade 3-4 and grade 5 serious adverse events. We quantify patient benefit as pooled hazard ratios for overall survival and progression-free survival. The clinical impact of combination testing is measured in terms of FDA approvals and NCCN recommendations.
Chapter 4. Benefit and Burden in a Historic Cohort of Cancer Combination Trials: A Systematic Review and Meta-Analysis

Benjamin Gregory Carlisle, Adélaïde Doussau, Jonathan Kimmelman

Abstract

Introduction: After approval, drug developers often pursue trials aimed at extending the uses of a new drug by combining it with other drugs. Little is known about the risk and benefits associated with such research.

Methods: To establish a historic benchmark of risk and benefit, we searched Medline and Embase for clinical trials testing anti-cancer drugs in combination within five years of FDA approval of 12 anticancer “index” drugs first licensed 2005-2007 inclusive. Risk was assessed based on grade 3 or above drug related adverse events; benefit was assessed based on efficacy outcomes and advancement of combinations into clinical practice guidelines or FDA approval.

Results: We captured 323 published post-approval trials exploring combinations, including 266 unique combination-indication pairings and enrolling 29,835 patients. The pooled risk ratios for grade 3-4 and grade 5 serious adverse events for trials randomized between a combination arm and a comparator were 1.54 [1.33-1.79] and 1.51 [1.16-1.97] respectively. The pooled hazard ratios for OS and PFS were 0.99 [0.92-1.05] and 0.85 [0.79-0.93], respectively. None of the combination-indication pairings launched after initial drug approval received FDA approval, and 12 pairings (4.5%) were recommended by the NCCN within five years of the first trial within that pairing. The proportion of patients in our sample who participated in trials leading to an FDA approval or NCCN guideline recommendation was 12%.

Discussion: Patients were just as likely to benefit in the treatment arm as the control arm in terms of overall survival, but they were more likely to experience a treatment-related serious adverse event in post-approval trials of combination therapy.
Introduction

Cancer drugs are often tested in combination with other drugs in order to explore additive or synergistic effects.\(^1\) Such research tends to occur after FDA approval, in part because of the need to first establish single-agent safety of the new drug.\(^2\) Many combination therapies have shown superior efficacy to monotherapies and have become a mainstay of cancer care.\(^3\)

Patients pursue trial participation primarily for the prospect of accessing superior treatments, and combination therapy trials generally show higher response rates than those reported in monotherapy studies.\(^3\)–5\) For example, several meta-analyses of phase 1 trials have shown higher rates of tumor response for combination therapy trials as compared with monotherapy trials.\(^6,7\)

However, combination therapy trials in cancer often present higher rates of drug-related adverse events as compared with monotherapy.\(^2,8\)–10\) Moreover, prioritizing combinations for testing is challenging. The volume of possible combination therapies that could be tested is enormous, given the number of cancer drugs that are approved.\(^1\) Some studies suggest that the yield of validated strategies is small compared with the quantity of combination therapies tested.\(^11,12\) Such findings point to the need for finding better ways to maximize the impact of patient participation in trials on future practices.

To our knowledge, no single study has systematically evaluated the risk, direct benefit, and clinical impact of participating in post-approval combination therapy trials in cancer. As a result, patients considering such trials—and oncologists who might counsel them—have limited information to help them weigh the expected burden of trial participation against the potential impact on their disease and future cancer care. The absence of systematic evidence to provide a baseline historical level of risk, benefit, and clinical impact for post-approval combination therapy trials also limits the ability of policy makers, regulators, and others to track progress in improving success rates in clinical translation.

Assessment of the clinical impact of trial participation requires looking at a historic cohort of combination trials, since the lag between a trial and regulatory approval or guideline recommendations can take years. In what follows, we use a historical cohort of combination therapy research activities to estimate the benefit and risk associated with participating in combination therapy trials that are launched after new cancer drugs are approved. We further
estimate the clinical impact of such research participation in terms of the proportion of patients whose participation leads to treatment strategies that are approved by the FDA or taken up by clinical practice guidelines within 5 years of testing.

**Methods**

Our primary objective was to measure the burden, direct benefit, and clinical impact associated with participating in trials testing combination therapies after an FDA approval for a historic cohort of all 12 anti-cancer drugs that were newly approved by the FDA between 2005-2007 inclusive (“index drugs”). To capture all of these properties adequately and allow a consistent follow-up time for clinical trial data to be submitted to the FDA or taken up into clinical practice guidelines, we used a longitudinal retrospective cohort study design. We only included newly approved drugs; drugs that only received a label revision for a new formulation within our time period, such as nab-paclitaxel were excluded. We operationalized burden with two indicators: the rates of grade 3 or above drug related adverse events in a single Common Terminology Criteria for Adverse Events (CTCAE) criterion, and the rates of grade 5 drug-related serious adverse events. We operationalized direct patient benefit by overall survival (OS) or progression-free survival (PFS) hazard ratios between combination arms and non-combination comparators for randomized controlled trials. We defined a “combination-indication pairing” to be a set of trials that: a) tested an index drug, b) in combination with another drug (or cocktail) in c) a particular cancer indication. For example, all trials testing temsirolimus combined with bortezomib in multiple myeloma would be defined as a combination-indication pairing. Clinical impact of participation was operationalized as the proportion of patients who participated in trials of a combination-indication pairing that resulted in either an FDA approval or a recommendation in National Comprehensive Cancer Network (NCCN) clinical practice guidelines within five years.

**Combination therapy sample**

We chose our timeframe for index drugs in order to create a cohort of combination therapies involving newly approved cancer drugs that would have sufficient follow up time since testing to enable our measure of the clinical impact of participation. This timeframe was chosen so that we could capture combination-indication pairings launched into testing up to 5 years after an index drug’s first approval and still have another 5 years of maturation in that cohort for combination
trial results to be published and be incorporated into FDA labels or clinical practice guidelines. For example, temsirolimus with bortezomib in multiple myeloma was first tested in June of 2007, so for this combination-indication pairing, we only considered FDA approvals or NCCN recommendations that occurred before June of 2012.

**Literature search**

We searched Medline and Embase for clinical trials of combination therapies that tested at least one of our 12 index drugs in combination therapy in cancer. Our inclusion criteria were: 1) primary data; 2) full-text; 3) English language; 4) final report; 5) interventional trials; 6) administered a drug in our sample in combination therapy; 7) cancer; 8) beginning enrolment no later than 5 years after the drug’s first FDA approval; and 9) testing a combination-indication pairing that was launched after the first approval of the index drug. We excluded studies that were 1) expansion cohorts of published trials; 2) expanded access drug trials; 3) non-drug therapeutic combinations (e.g. monotherapy with radiation or surgery); 4) mixed malignancy; 5) FDA-approved on-label studies where enrolment began after regulatory approval for the same indication. Our full search strategy, as well as our full extraction codebook is included in Appendices A and B.

**Extraction**

Extractions were completed according to a previously published template, using Numbat software to manage multiple coding. We captured: a) index drug; b) combination; c) indication; d) patient enrolment; e) enrolment date; f) funding (industry vs other); g) primary outcome; h) overall survival (OS) and progression-free survival (PFS) hazard ratios among trials where patients were randomized to receive an additional drug in a combination therapy or not. We also captured per-trial and per-arm incidences of treatment-related grade 3-4 serious adverse events in a single CTCAE category as well as grade 5 serious adverse events.

Combination-indication pairings were classified by whether they received FDA approval within 5 years of the first trial testing this pairing, and the individual drugs that made up the combination were classified by whether they received FDA approval individually. We also classified

Carlisle 84
combination-indication pairings by whether they were recommended in the NCCN clinical practice guidelines within 5 years of the first trial testing this pairing.

In cases where these data points were reported, OS and PFS hazard ratios as well as per-arm grade 3-4 serious adverse events and grade 5 serious adverse events were extracted from trials where patients were randomized between a therapy regimen (monotherapy or a drug cocktail) vs that regimen with an additional drug in combination. In some cases the index drug was the standard of care comparator, and in some cases, the index drug was the drug added in combination.

To ensure consistent follow up time until FDA approval or NCCN recommendation for each combination-indication pairing, we limited our analysis to pairings for which we had at least five years of follow up since enrolment in the first trial of that combination-indication pairing. As a sensitivity analysis to account for combination-indication pairings that might take longer to get an FDA approval or NCCN recommendation, we re-performed analyses for those combination-indication pairings for which we had eight years of follow up.

Analysis and statistics

Analyses of results were conducted using \textit{R} version 3.4.4.\cite{16} Pooled outcomes were calculated using random-effects inverse variance meta-analysis methods for proportions, hazard ratios and risk ratios between arms of randomized trials, using \textit{metaprop}, \textit{metagen} and \textit{metabin} functions, respectively, from the \textit{meta} package, version 4.8-2.\cite{17} Unless explicitly stated otherwise, ranges given after pooled proportions or hazard ratios represent 95\% confidence intervals calculated using the specified pooling techniques.

Results

Volume and characteristics of trials testing combination-indication pairings after FDA approval

Our search captured 323 published trials exploring combinations with the 12 newly approved index drugs in our sample. (See Appendix E for our PRISMA flow diagram.) In total, combination trials in our sample enrolled 29,835 patients. This included 266 unique combination-indication pairings that had been launched into testing. (See Appendix M) Between zero and 62

Carlisle 85
Combination-indication pairings were explored for each index drug (median 16). (See Table 1 for a breakdown of our sample). For one of the index drugs (nelarabine), there were no new combination-indication pairings launched into testing within five years after its first approval by the FDA.

Our sample included 96 phase 1 trials (30%), 46 phase 1-2 trials (14%), 156 phase 2 trials (48%), 20 phase 2-3 or phase 3 trials (6.2%) and 5 trials in which the phase was not stated (1.5%). Of the 222 trials in phase 2 or higher, 23% of the phase 1-2 or phase 2 trials were randomized, and 90% of the phase 2-3 or phase 3 trials used randomization.

There was a positive result on the primary endpoint in 173 trials (54%) in our sample. Toxicity was reported as acceptable by the authors of the report in 244 trials (76%). Industry funded 172 (53%) of the trials in our sample, and 123 trials (38%) had non-industry sponsors (the remaining 8.7% of the trials did not report a funding source).

For 30% combination-indication pairings (79 of 266 combination-indication pairings), the index drug had been approved as monotherapy in the indication. For 88 combination-indication pairings (33%), neither the index drug, nor the treatment it was combined with, was approved in monotherapy for the indication in question. Among trials in our sample, 10 combination-indication pairings involved 7 drugs that, as of the five-year follow up period, had not been approved by the FDA for any indication—in monotherapy or in combination therapy.

**Patient risk and burden**

There were 8764 grade 3-4 serious adverse events in our entire sample (33% of patients [30-36%]) and 263 grade 5 serious adverse events that were attributed to the study drug (2% of patients, [1.8-2.2%]). (See Appendix L for details.)

Our sample included 39 randomized clinical trials reporting grade 3-4 serious adverse events on a per-arm basis, and 29 randomized clinical trials reporting grade 5 serious adverse events on a per-arm basis. Among these trials, the pooled risk ratio for grade 3-4 serious adverse events was 1.54 [1.33-1.79] (see Figure 3; Appendix P). The pooled risk ratio for grade 5 serious adverse events among randomized trials was 1.51 [1.16-1.97] (see Figure 4; Appendix Q).
Patient benefit and the clinical impact of participation

Our sample included 26 randomized clinical trials enrolling a total of 12,495 patients reporting overall survival hazard ratios between between a standard of care comparator or an additional combination therapy, and 25 trials enrolling 12,406 patients reporting progression-free survival hazard ratios. The pooled hazard ratios for OS and PFS were 0.99 [0.92-1.05] and 0.85 [0.79-0.93], respectively (see Figures 1 and 2).

Among combination-indication pairings that were launched post-licensure, none received FDA approval, and 12 pairings (4.5%) among 4 index drugs (panitumumab, lapatinib, dasatinib and nilotinib) were recommended by the NCCN within five years of the first trial within that pairing. The proportion of patients in trials testing these combination-indication pairings was 12%. (See supplementary for details.) For the 210 combination-indication pairings for which we had eight years of follow-up since initial tests (79% of our total sample), an additional 2 pairings, both combinations of panitumumab in colorectal cancer, were recommended by the NCCN, and 3 combinations of lenalidomide in multiple myeloma were approved by the FDA (1% of all pairings in our sample).

Discussion

After FDA approval, drug companies and academic researchers often attempt to extend the impact of a new drug by testing it in combination with other drugs. Based on this historic sample of post approval trials for 12 drugs, assignment to combination therapy arms was not associated with survival advantage but was associated with greater risk. Though two drugs (ixabepilone and panitumumab), were approved as combination therapies, no post-approval combination-indication pairings in our cohort received FDA approval within five years of testing. Five combinations were recommended in clinical practice guidelines, including panitumumab with bevacizumab and FOLFIRI in colorectal cancer, and lapatinib with trastuzumab for colorectal cancer.18–20

Assignment to combination therapy arms in randomized trials showed benefit in terms of progression-free survival. However, the clinical value of this surrogate endpoint has been questioned in the context of combination therapy.21 To contextualize the magnitude of PFS benefit observed in our analysis a hazard ratio of 0.86 and a median PFS of 6 months in the control
group, the gain in PFS for the combination arm would be approximately 1 month. That patients in combination arms were just as likely to experience an overall survival benefit as the patients in the comparator arms is consistent with what others have observed when studying large populations of cancer studies.\textsuperscript{22,23}

Using different methods and assumptions, others have shown that 5\% of cancer drugs entering clinical development ultimately advance to FDA approval.\textsuperscript{24} Our study suggests that the proportion of post-approval combination therapies that advance to an FDA label has historically been lower.

Our findings also suggest that some combination therapy testing efforts may be more productive than others. For example, combination regimens where neither drug was approved as monotherapy for a given indication were uniformly unproductive in terms of FDA approvals or NCCN recommendations.\textsuperscript{25}

Our findings may be useful for benchmarking present day activities exploring drug combinations. The contemporary era is marked by vigorous activities exploring combination immunotherapies.\textsuperscript{26} Several have shown dramatic advantages over monotherapy.\textsuperscript{27} Whether, on balance, combination therapies explored in the contemporary post-approval setting have an improved risk / benefit balance, or whether a greater proportion lead to regulatory approvals, remains to be seen. In the meantime, our findings may be helpful for informed consent discussions surrounding post-approval trials of cancer combinations therapies. Patients participate in cancer trials in part to access possible innovative treatments.\textsuperscript{4,5,28,29} Our findings, that 12\% of patients who participate in post-approval combination therapy trials contribute to the development of applications that are recommended for practice, can help patients and physicians gauge their potential contribution to science when considering trial participation.

Our findings should be considered against several limitations. First, as we emphasized, we only considered combination therapy trials pursued within 5 years of approval for 12 index drugs. This reflects our need to maintain a long maturation time in order to allow the combination-indication pairings to be approved or taken up into clinical practice guidelines. This study is longitudinal in nature, in order to capture the long-term outcomes of the pairings we captured, and so we only captured trials involving one of our index drugs, and thus did not include every

Carlisle 88
trial of combination therapy that occurred within this timeframe. Our findings may also have been driven by properties particular to our sample of index drugs; a preponderance of trials involved sunitinib, sorafenib and lenalidomide. Nevertheless, in interpreting our findings, one should avoid the temptations of hindsight bias: at the time when these trials were launched, expectations were high that angiogenesis inhibition would remake the landscape of cancer care.30

Second, our results should be interpreted in the context of other reports. Some meta-analyses (but not others9,31) have suggested survival advantages for combination therapies as compared with monotherapies.8,10 Third, we cannot rule out that some combinations tested in our sample will receive recommendations in clinical practice guidelines in the future. Our sensitivity analysis suggested that at least some combination-indication pairings might be further refined and advance into clinical practice guidelines and FDA approvals. Fourth, our analysis reflected only published trials. Because the index drugs in our sample were approved before legally mandated clinical trial registration, we relied on published results. Moreover, published reports often contain incomplete records of adverse events.32,33 Our estimates may therefore be biased toward overestimating benefit and the clinical impact of participation and underestimating risk. At the same time, crossover in combination therapy trials may have limited our ability to detect a survival advantage for patients receiving novel combination therapies.34–36

The development of combination therapies in cancer presents a distinct set of scientific, social and commercial challenges.37 Among these is the task of helping patients make informed decisions about study enrolment. Though the risk, benefit and potential clinical impact of any given combination therapy will depend on a variety of factors, the analyses provided above provide historical baselines for estimating risk, benefit, and translational success associated with post-approval testing of combination therapies in cancer.
Figures

<table>
<thead>
<tr>
<th>Study</th>
<th>Index drug</th>
<th>Combination</th>
<th>Indication</th>
<th>Hazard Ratio</th>
<th>HR</th>
<th>95%-CI</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>J. C. Arnao, 2008</td>
<td>Dasatinib</td>
<td>Docetaxel</td>
<td>Prostate</td>
<td>0.99 (0.87; 1.13)</td>
<td>8.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. R. Johnston, 2013</td>
<td>Lapatinib</td>
<td>Paclitaxel</td>
<td>Breast cancer</td>
<td>0.91 (0.50; 1.65)</td>
<td>1.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stewart AK, 2015</td>
<td>Lenalidomide</td>
<td>Carfilzomib Dexamethasone</td>
<td>Multiple myeloma</td>
<td>0.79 (0.63; 0.99)</td>
<td>5.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Petrylak DP, 2015</td>
<td>Lenalidomide</td>
<td>Docetaxel Prednisone</td>
<td>Prostate</td>
<td>1.53 (1.17; 2.00)</td>
<td>4.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Durin BO, 2017</td>
<td>Lenalidomide</td>
<td>Bortezomib Dexamethasone</td>
<td>Multiple myeloma</td>
<td>0.71 (0.52; 0.98)</td>
<td>3.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vermorken, 2013</td>
<td>Panitumumab</td>
<td>Cisplatin Fluorouracil</td>
<td>Head and Neck</td>
<td>0.87 (0.73; 1.05)</td>
<td>6.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seymour, 2013</td>
<td>Panitumumab</td>
<td>Irinotecan</td>
<td>Colorectal</td>
<td>1.01 (0.83; 1.23)</td>
<td>6.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liu, 2015</td>
<td>Panitumumab</td>
<td>Bevacizumab FOLFIRI</td>
<td>Colorectal</td>
<td>0.79 (0.57; 1.09)</td>
<td>3.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mesia, 2015</td>
<td>Panitumumab</td>
<td>Cisplatin</td>
<td>Head and Neck</td>
<td>1.63 (0.88; 3.02)</td>
<td>1.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wirth, 2016</td>
<td>Panitumumab</td>
<td>Cisplatin Docetaxel</td>
<td>Head and Neck</td>
<td>1.10 (0.71; 1.72)</td>
<td>2.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leore, 2016</td>
<td>Panitumumab</td>
<td>Gemcitabine Oxaliplatin</td>
<td>Biliary tract</td>
<td>0.83 (0.53; 1.30)</td>
<td>2.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G. Scazziali, 2010</td>
<td>Sorafenib</td>
<td>Carboplatin Paclitaxel</td>
<td>NSCLC</td>
<td>1.15 (0.84; 1.54)</td>
<td>6.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. R. Spiegel, 2011</td>
<td>Sorafenib</td>
<td>Erlotinib</td>
<td>NSCLC</td>
<td>0.89 (0.59; 1.34)</td>
<td>2.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. G. Paz-Ares, 2012</td>
<td>Sorafenib</td>
<td>Cisplatin Gemcitabine</td>
<td>NSCLC</td>
<td>0.98 (0.83; 1.16)</td>
<td>7.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H. Serve, 2013</td>
<td>Sorafenib</td>
<td>Cisplatin Gemcitabine</td>
<td>NSCLC</td>
<td>0.98 (0.83; 1.16)</td>
<td>7.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. Tabernero, 2013</td>
<td>Sorafenib</td>
<td>nFOLFOX6</td>
<td>Colorectal</td>
<td>1.13 (0.78; 1.61)</td>
<td>2.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W. J. Gradishar, 2013</td>
<td>Sorafenib</td>
<td>Paclitaxel</td>
<td>Breast cancer</td>
<td>1.02 (0.71; 1.46)</td>
<td>2.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G. Cascino, 2014</td>
<td>Sorafenib</td>
<td>Cisplatin Gemcitabine</td>
<td>Pancreatic</td>
<td>0.95 (0.61; 1.53)</td>
<td>2.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. H. Yi, 2012</td>
<td>Sunitinib</td>
<td>Docetaxel</td>
<td>Gastric</td>
<td>0.94 (0.60; 1.48)</td>
<td>1.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. Bergh, 2012</td>
<td>Sunitinib</td>
<td>Docetaxel</td>
<td>Breast cancer</td>
<td>1.21 (0.91; 1.60)</td>
<td>4.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. P. Crown, 2013</td>
<td>Sunitinib</td>
<td>Cephalosporin</td>
<td>Breast cancer</td>
<td>0.99 (0.76; 1.29)</td>
<td>4.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H. J. Groen, 2013</td>
<td>Sunitinib</td>
<td>Erlotinib</td>
<td>NSCLC</td>
<td>1.07 (0.70; 1.61)</td>
<td>2.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Carano, 2013</td>
<td>Sunitinib</td>
<td>FOLFIRI</td>
<td>Colorectal</td>
<td>1.17 (0.94; 1.47)</td>
<td>5.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. D. Michaelson, 2014</td>
<td>Sunitinib</td>
<td>Prednisone</td>
<td>Prostate</td>
<td>0.91 (0.76; 1.09)</td>
<td>6.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N. E. Ready, 2015</td>
<td>Sunitinib</td>
<td>Cisplatin Eltopsid</td>
<td>SCCL</td>
<td>0.78 (0.48; 1.27)</td>
<td>1.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramalingam, 2010</td>
<td>Vorinostat</td>
<td>Carboplatin Paclitaxel</td>
<td>NSCLC</td>
<td>0.68 (0.43; 1.07)</td>
<td>1.9%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Random effects model

<table>
<thead>
<tr>
<th>Favours combination</th>
<th>Favours non-combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 1. Overall survival hazard ratios for clinical trials of label-extending combination therapy exploration.
<table>
<thead>
<tr>
<th>Study</th>
<th>Index drug</th>
<th>Combination</th>
<th>Indication</th>
<th>Hazard Ratio</th>
<th>HR</th>
<th>95%-CI Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>J. C. Araujo, 2008</td>
<td>Dasatinib</td>
<td>Docetaxel</td>
<td>Prostate</td>
<td>0.92</td>
<td>[0.81: 1.04]</td>
<td>6.5%</td>
</tr>
<tr>
<td>S. R. Johnston, 2013</td>
<td>Lapatinib</td>
<td>Pazopanib</td>
<td>Breast cancer</td>
<td>1.30</td>
<td>[0.76: 2.23]</td>
<td>1.9%</td>
</tr>
<tr>
<td>Stewart AK, 2015</td>
<td>Lenvlolidnide</td>
<td>Carfilzomib</td>
<td>Multiple myeloma</td>
<td>0.69</td>
<td>[0.57: 0.83]</td>
<td>5.5%</td>
</tr>
<tr>
<td>Petrylak DP, 2015</td>
<td>Lenvlolidnide</td>
<td>Docetaxel</td>
<td>Prednisone</td>
<td>1.32</td>
<td>[1.05: 1.68]</td>
<td>4.9%</td>
</tr>
<tr>
<td>Durie BS, 2017</td>
<td>Lenvlolidnide</td>
<td>Bortezomib</td>
<td>Dexamethasone</td>
<td>0.71</td>
<td>[0.56: 0.91]</td>
<td>4.7%</td>
</tr>
<tr>
<td>Vermerken, 2013</td>
<td>Panitumumab</td>
<td>Cisplatin</td>
<td>Fluorouracil</td>
<td>0.78</td>
<td>[0.66: 0.92]</td>
<td>5.8%</td>
</tr>
<tr>
<td>Seymour, 2013</td>
<td>Panitumumab</td>
<td>Imitotecan</td>
<td>Colorectal</td>
<td>0.78</td>
<td>[0.64: 0.95]</td>
<td>5.4%</td>
</tr>
<tr>
<td>Liu, 2015</td>
<td>Panitumumab</td>
<td>Bevacizumab</td>
<td>FOLFIRI</td>
<td>0.45</td>
<td>[0.26: 0.77]</td>
<td>4.9%</td>
</tr>
<tr>
<td>Messia, 2015</td>
<td>Panitumumab</td>
<td>Cisplatin</td>
<td>Head and Neck</td>
<td>1.15</td>
<td>[0.68: 1.95]</td>
<td>1.9%</td>
</tr>
<tr>
<td>Wirth, 2016</td>
<td>Panitumumab</td>
<td>Cisplatin</td>
<td>Docetaxel</td>
<td>0.63</td>
<td>[0.39: 1.00]</td>
<td>2.3%</td>
</tr>
<tr>
<td>Leone, 2016</td>
<td>Panitumumab</td>
<td>Gemcitabine</td>
<td>Oxaliplatin</td>
<td>0.78</td>
<td>[0.51: 1.20]</td>
<td>2.5%</td>
</tr>
<tr>
<td>G. Scaglioni, 2010</td>
<td>Sorafenib</td>
<td>Carboplatin</td>
<td>Paclitaxel</td>
<td>0.99</td>
<td>[0.84: 1.16]</td>
<td>6.0%</td>
</tr>
<tr>
<td>D. R. Spiegel, 2011</td>
<td>Sorafenib</td>
<td>Erlotinib</td>
<td>NSCLC</td>
<td>0.86</td>
<td>[0.60: 1.23]</td>
<td>3.2%</td>
</tr>
<tr>
<td>J. Baselga, 2012</td>
<td>Sorafenib</td>
<td>Capcitabine</td>
<td>Breast cancer</td>
<td>0.58</td>
<td>[0.41: 0.83]</td>
<td>3.4%</td>
</tr>
<tr>
<td>L. G. Paz-Ares, 2012</td>
<td>Sorafenib</td>
<td>Cisplatin</td>
<td>Gemcitabine</td>
<td>0.83</td>
<td>[0.71: 0.97]</td>
<td>6.6%</td>
</tr>
<tr>
<td>J. Tabernerio, 2013</td>
<td>Sorafenib</td>
<td>mFOLFOX8</td>
<td>Colorectal</td>
<td>0.88</td>
<td>[0.63: 1.22]</td>
<td>3.6%</td>
</tr>
<tr>
<td>W. J. Gradishar, 2013</td>
<td>Sorafenib</td>
<td>Paclitaxel</td>
<td>Breast cancer</td>
<td>0.78</td>
<td>[0.56: 1.11]</td>
<td>3.3%</td>
</tr>
<tr>
<td>S. Cassiru, 2014</td>
<td>Sorafenib</td>
<td>Cisplatin</td>
<td>Gemcitabine</td>
<td>0.91</td>
<td>[0.62: 1.34]</td>
<td>2.9%</td>
</tr>
<tr>
<td>J. Bergh, 2012</td>
<td>Sunitinib</td>
<td>Docetaxel</td>
<td>Breast cancer</td>
<td>0.92</td>
<td>[0.72: 1.20]</td>
<td>4.5%</td>
</tr>
<tr>
<td>J. P. Crown, 2013</td>
<td>Sunitinib</td>
<td>Capscitabine</td>
<td>Breast cancer</td>
<td>1.22</td>
<td>[0.95: 1.57]</td>
<td>4.5%</td>
</tr>
<tr>
<td>H. J. Groen, 2013</td>
<td>Sunitinib</td>
<td>Erlotinib</td>
<td>NSCLC</td>
<td>0.90</td>
<td>[0.67: 1.20]</td>
<td>4.0%</td>
</tr>
<tr>
<td>A. Cattane, 2013</td>
<td>Sunitinib</td>
<td>FOLFIRI</td>
<td>Colorectal</td>
<td>1.09</td>
<td>[0.89: 1.34]</td>
<td>4.4%</td>
</tr>
<tr>
<td>M. D. Michaelson, 2014</td>
<td>Sunitinib</td>
<td>Prednisone</td>
<td>Prostate</td>
<td>0.72</td>
<td>[0.59: 0.89]</td>
<td>5.3%</td>
</tr>
<tr>
<td>N. E. Reedy, 2015</td>
<td>Sunitinib</td>
<td>Cisplatin</td>
<td>Etoposide</td>
<td>0.62</td>
<td>[0.39: 0.99]</td>
<td>2.3%</td>
</tr>
<tr>
<td>Ramalingam, 2010</td>
<td>Vorinostat</td>
<td>Carboplatin</td>
<td>Paclitaxel</td>
<td>0.84</td>
<td>[0.54: 1.31]</td>
<td>2.4%</td>
</tr>
</tbody>
</table>

Random effects model

0.85 [0.79; 0.93] 100.0%

Figure 2. Progression-free survival hazard ratios for clinical trials of label-extending combination therapy exploration
Figure 3. Risk ratios between experimental and control arms for grade 3-4 serious adverse events in randomized clinical trials of label-extending combination therapy exploration

<table>
<thead>
<tr>
<th>Study</th>
<th>Index Drug</th>
<th>Combination</th>
<th>Indication</th>
<th>Experimental Events Total</th>
<th>Events Total</th>
<th>Control Events Total</th>
<th>Risk Ratio</th>
<th>RR</th>
<th>95% CI</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>J. C. Amus, 2008</td>
<td>Docetaxel</td>
<td>Docetaxel</td>
<td>Prostate</td>
<td>93</td>
<td>761</td>
<td>44</td>
<td>757</td>
<td>2.10</td>
<td>(1.49, 2.97)</td>
<td>3.5%</td>
</tr>
<tr>
<td>A. Spinaface, 2010</td>
<td>Docetaxel</td>
<td>Cediranib</td>
<td>Prostate</td>
<td>3</td>
<td>11</td>
<td>3</td>
<td>11</td>
<td>1.00</td>
<td>(0.26, 3.91)</td>
<td>0.9%</td>
</tr>
<tr>
<td>J. P. Issa, 2015</td>
<td>Docetaxel</td>
<td>Cediranib</td>
<td>MDSC/AML</td>
<td>9</td>
<td>79</td>
<td>10</td>
<td>70</td>
<td>0.80</td>
<td>(0.54, 1.19)</td>
<td>3.1%</td>
</tr>
<tr>
<td>G. Trosino, 2015</td>
<td>Docetaxel</td>
<td>Cediranib</td>
<td>Breast cancer</td>
<td>15</td>
<td>37</td>
<td>12</td>
<td>40</td>
<td>1.15</td>
<td>(0.73, 2.15)</td>
<td>2.5%</td>
</tr>
<tr>
<td>Stewart, 2015</td>
<td>Docetaxel</td>
<td>Cediranib</td>
<td>Colorectal cancer</td>
<td>37</td>
<td>395</td>
<td>19</td>
<td>396</td>
<td>1.95</td>
<td>(1.14, 3.32)</td>
<td>2.9%</td>
</tr>
<tr>
<td>Petral, 2015</td>
<td>Docetaxel</td>
<td>Cediranib</td>
<td>Colorectal cancer</td>
<td>114</td>
<td>533</td>
<td>85</td>
<td>526</td>
<td>1.32</td>
<td>(1.02, 1.71)</td>
<td>3.8%</td>
</tr>
<tr>
<td>Dahi, 2017</td>
<td>Docetaxel</td>
<td>Cediranib</td>
<td>Colorectal cancer</td>
<td>114</td>
<td>364</td>
<td>104</td>
<td>261</td>
<td>1.68</td>
<td>(0.88, 3.20)</td>
<td>0.4%</td>
</tr>
<tr>
<td>Okines, 2018</td>
<td>Panitumumab</td>
<td>Cetuximab</td>
<td>Panitumumab</td>
<td>17</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td>1.42</td>
<td>(0.53, 3.81)</td>
<td>1.5%</td>
</tr>
<tr>
<td>Vernet, 2013</td>
<td>Panitumumab</td>
<td>Cetuximab</td>
<td>Head and Neck</td>
<td>103</td>
<td>325</td>
<td>106</td>
<td>325</td>
<td>0.97</td>
<td>(0.79, 1.22)</td>
<td>3.5%</td>
</tr>
<tr>
<td>Seymour, 2013</td>
<td>Panitumumab</td>
<td>Cetuximab</td>
<td>Colorectal</td>
<td>63</td>
<td>219</td>
<td>38</td>
<td>218</td>
<td>1.46</td>
<td>(1.10, 1.96)</td>
<td>3.5%</td>
</tr>
<tr>
<td>Dei, 2014</td>
<td>Panitumumab</td>
<td>Cetuximab</td>
<td>Colorectal</td>
<td>133</td>
<td>399</td>
<td>137</td>
<td>396</td>
<td>1.64</td>
<td>(0.94, 2.80)</td>
<td>4.4%</td>
</tr>
<tr>
<td>Liu, 2015</td>
<td>Panitumumab</td>
<td>Cetuximab</td>
<td>Colorectal</td>
<td>32</td>
<td>57</td>
<td>10</td>
<td>69</td>
<td>3.87</td>
<td>(1.13, 12.68)</td>
<td>2.6%</td>
</tr>
<tr>
<td>Merx, 2015</td>
<td>Panitumumab</td>
<td>Cetuximab</td>
<td>Head and Neck</td>
<td>18</td>
<td>33</td>
<td>17</td>
<td>63</td>
<td>2.04</td>
<td>(1.31, 3.12)</td>
<td>31%</td>
</tr>
<tr>
<td>Schuelke, 2015</td>
<td>Panitumumab</td>
<td>Cetuximab</td>
<td>Colorectal</td>
<td>10</td>
<td>49</td>
<td>10</td>
<td>47</td>
<td>0.95</td>
<td>(0.54, 1.67)</td>
<td>2.6%</td>
</tr>
<tr>
<td>Wirth, 2015</td>
<td>Panitumumab</td>
<td>Cetuximab</td>
<td>Head and Neck</td>
<td>10</td>
<td>58</td>
<td>3</td>
<td>55</td>
<td>3.27</td>
<td>(0.96, 11.24)</td>
<td>1.1%</td>
</tr>
<tr>
<td>Test, 2015</td>
<td>Panitumumab</td>
<td>Cetuximab</td>
<td>Fluorouracil</td>
<td>11</td>
<td>37</td>
<td>7</td>
<td>39</td>
<td>1.56</td>
<td>(0.72, 2.61)</td>
<td>1.8%</td>
</tr>
<tr>
<td>Leone, 2015</td>
<td>Panitumumab</td>
<td>Cetuximab</td>
<td>Panitumumab</td>
<td>9</td>
<td>45</td>
<td>6</td>
<td>44</td>
<td>1.47</td>
<td>(0.57, 3.78)</td>
<td>1.6%</td>
</tr>
<tr>
<td>Wieg, 2015</td>
<td>Panitumumab</td>
<td>Cetuximab</td>
<td>Panitumumab</td>
<td>45</td>
<td>159</td>
<td>33</td>
<td>147</td>
<td>3.39</td>
<td>(1.91, 6.32)</td>
<td>2.6%</td>
</tr>
<tr>
<td>G. Sciglietti, 2015</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Proteinuria</td>
<td>39</td>
<td>436</td>
<td>27</td>
<td>459</td>
<td>1.52</td>
<td>(0.95, 2.44)</td>
<td>3.0%</td>
</tr>
<tr>
<td>D. R. Sarge, 2012</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>17</td>
<td>11</td>
<td>7</td>
<td>55</td>
<td>1.20</td>
<td>(0.53, 2.71)</td>
<td>1.8%</td>
</tr>
<tr>
<td>G. Prontino, 2011</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>RCC</td>
<td>8</td>
<td>56</td>
<td>6</td>
<td>62</td>
<td>1.25</td>
<td>(0.49, 3.41)</td>
<td>1.5%</td>
</tr>
<tr>
<td>J. Depra, 2012</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Breast cancer</td>
<td>49</td>
<td>112</td>
<td>16</td>
<td>112</td>
<td>3.08</td>
<td>(1.80, 5.20)</td>
<td>2.9%</td>
</tr>
<tr>
<td>L. G. Pau-Ans, 2012</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>38</td>
<td>365</td>
<td>24</td>
<td>384</td>
<td>1.58</td>
<td>(0.97, 2.54)</td>
<td>2.9%</td>
</tr>
<tr>
<td>H. Sever, 2013</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>58</td>
<td>102</td>
<td>50</td>
<td>96</td>
<td>1.18</td>
<td>(0.84, 1.68)</td>
<td>2.9%</td>
</tr>
<tr>
<td>J. Toberres, 2013</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>47</td>
<td>97</td>
<td>22</td>
<td>101</td>
<td>2.22</td>
<td>(1.46, 3.39)</td>
<td>3.2%</td>
</tr>
<tr>
<td>W. J. Gittens, 2013</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>36</td>
<td>115</td>
<td>9</td>
<td>118</td>
<td>4.10</td>
<td>(2.07, 8.13)</td>
<td>2.9%</td>
</tr>
<tr>
<td>S. Kim, 2014</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>5</td>
<td>40</td>
<td>7</td>
<td>49</td>
<td>0.88</td>
<td>(0.30, 2.35)</td>
<td>1.2%</td>
</tr>
<tr>
<td>S. Cagirci, 2014</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>12</td>
<td>58</td>
<td>12</td>
<td>56</td>
<td>0.97</td>
<td>(0.47, 1.97)</td>
<td>2.2%</td>
</tr>
<tr>
<td>E. L. Mayer, 2010</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>10</td>
<td>23</td>
<td>2</td>
<td>23</td>
<td>5.60</td>
<td>(2.31, 10.32)</td>
<td>0.6%</td>
</tr>
<tr>
<td>J. H. Yi, 2012</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>48</td>
<td>56</td>
<td>10</td>
<td>49</td>
<td>1.68</td>
<td>(0.81, 3.43)</td>
<td>2.3%</td>
</tr>
<tr>
<td>J. Sepp, 2012</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>137</td>
<td>395</td>
<td>128</td>
<td>393</td>
<td>1.06</td>
<td>(0.99, 1.14)</td>
<td>4.2%</td>
</tr>
<tr>
<td>J. P. Green, 2013</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>69</td>
<td>217</td>
<td>52</td>
<td>215</td>
<td>1.11</td>
<td>(0.97, 1.27)</td>
<td>3.7%</td>
</tr>
<tr>
<td>H. J. Green, 2013</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>11</td>
<td>64</td>
<td>2</td>
<td>64</td>
<td>5.66</td>
<td>(2.71, 10.93)</td>
<td>0.8%</td>
</tr>
<tr>
<td>A. Cianci, 2013</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>104</td>
<td>394</td>
<td>80</td>
<td>379</td>
<td>2.37</td>
<td>(1.00, 5.91)</td>
<td>4.0%</td>
</tr>
<tr>
<td>Thordarson, 2014</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>39</td>
<td>140</td>
<td>8</td>
<td>13</td>
<td>0.93</td>
<td>(0.50, 1.23)</td>
<td>2.9%</td>
</tr>
<tr>
<td>M. D. Michaelison, 2014</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>117</td>
<td>581</td>
<td>32</td>
<td>285</td>
<td>1.79</td>
<td>(1.25, 2.58)</td>
<td>3.4%</td>
</tr>
<tr>
<td>N. E. Fray, 2015</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>5</td>
<td>43</td>
<td>4</td>
<td>41</td>
<td>1.91</td>
<td>(0.87, 3.12)</td>
<td>1.3%</td>
</tr>
<tr>
<td>L. R. Haaga, 2015</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>24</td>
<td>52</td>
<td>13</td>
<td>54</td>
<td>0.82</td>
<td>(0.54, 1.27)</td>
<td>1.8%</td>
</tr>
<tr>
<td>Renkling, 2010</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>Sorafenib</td>
<td>27</td>
<td>62</td>
<td>15</td>
<td>32</td>
<td>0.93</td>
<td>(0.50, 1.78)</td>
<td>3.0%</td>
</tr>
</tbody>
</table>

| Random effects model | 6572 | 6175 | 1.54 | (1.32, 1.79) | 100.0% |
Figure 4. Risk ratios between experimental and control arms for grade 5 serious adverse events (treatment-related death) in randomized clinical trials of label-extending combination therapy exploration.

### Tables

<table>
<thead>
<tr>
<th>Study</th>
<th>Index drug</th>
<th>Combination</th>
<th>Indication</th>
<th>Experimental Events Total</th>
<th>Control Events Total</th>
<th>Risk Ratio</th>
<th>RR</th>
<th>95% CI</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>J. C. Anjou, 2008</td>
<td>Dasatinib</td>
<td>Diclofenac</td>
<td>Prostate</td>
<td>50 761</td>
<td>40 757</td>
<td>1.24 [0.83; 1.86]</td>
<td>43.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Spechtel, 2010</td>
<td>Dasatinib</td>
<td>Ketorolac</td>
<td>Prostate</td>
<td>0 11</td>
<td>0 11</td>
<td>0.33 [0.02; 7.36]</td>
<td>0.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O. Tian, 2015</td>
<td>Dasatinib</td>
<td>Ketorolac</td>
<td>Breast cancer</td>
<td>0 37</td>
<td>0 40</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang ES, 2012</td>
<td>Lenalidomide</td>
<td>Corticosteroids</td>
<td>MDS</td>
<td>0 29</td>
<td>0 9</td>
<td>0.56 [0.06; 4.24]</td>
<td>0.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stray-Teck, 2015</td>
<td>Lenalidomide</td>
<td>Corticosteroids</td>
<td>Myeloma</td>
<td>6 396</td>
<td>6 396</td>
<td>0.75 [0.26; 2.14]</td>
<td>0.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parikh D, 2015</td>
<td>Lenalidomide</td>
<td>Doxorubicin</td>
<td>Prostate</td>
<td>7 533</td>
<td>6 526</td>
<td>1.15 [0.92; 3.44]</td>
<td>6.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donn DG, 2017</td>
<td>Lenalidomide</td>
<td>Doxorubicin</td>
<td>Multiple myeloma</td>
<td>2 254</td>
<td>0 283</td>
<td>4.94 [3.24; 10.57]</td>
<td>0.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Onken, 2020</td>
<td>Pantamutamib</td>
<td>Captopril + Epinephrine</td>
<td>Oxaliplatin</td>
<td>1 38</td>
<td>0 13</td>
<td>2.45 [1.11; 5.32]</td>
<td>0.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vermeulen, 2013</td>
<td>Pantamutamib</td>
<td>Captopril</td>
<td>Head and neck</td>
<td>14 325</td>
<td>8 325</td>
<td>1.75 [0.74; 4.11]</td>
<td>9.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sujan, 2013</td>
<td>Pantamutamib</td>
<td>Hexitol</td>
<td>Cardiac</td>
<td>1 213</td>
<td>3 218</td>
<td>0.13 [0.03; 0.57]</td>
<td>1.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mears, 2015</td>
<td>Pantamutamib</td>
<td>Captopril</td>
<td>Head and neck</td>
<td>1 87</td>
<td>1 63</td>
<td>0.72 [0.09; 11.36]</td>
<td>0.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winth, 2015</td>
<td>Pantamutamib</td>
<td>Captopril</td>
<td>Doxorubicin</td>
<td>9 58</td>
<td>4 55</td>
<td>2.24 [0.72; 7.40]</td>
<td>9.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leon, 2016</td>
<td>Pantamutamib</td>
<td>Gemcitabine</td>
<td>Oxaliplatin</td>
<td>1 45</td>
<td>0 44</td>
<td>2.93 [0.12; 70.12]</td>
<td>0.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G. Sagar, 2020</td>
<td>Sorafenib</td>
<td>Captopril</td>
<td>Pseudocarcinoid</td>
<td>14 436</td>
<td>4 459</td>
<td>3.68 [1.22; 11.12]</td>
<td>5.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. R. Sigler, 2011</td>
<td>Sorafenib</td>
<td>Etorphine</td>
<td>NSCLC</td>
<td>0 111</td>
<td>1 55</td>
<td>0.17 [0.01; 1.70]</td>
<td>0.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. Basara, 2012</td>
<td>Sorafenib</td>
<td>Captopril</td>
<td>Breast cancer</td>
<td>0 112</td>
<td>2 112</td>
<td>0.60 [0.01; 4.12]</td>
<td>0.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. G. Faz-Azet, 2012</td>
<td>Sorafenib</td>
<td>Captopril</td>
<td>Gemcitabine</td>
<td>NSCLC</td>
<td>5 386</td>
<td>2 384</td>
<td>2.49 [0.49; 12.71]</td>
<td>2.6%</td>
<td></td>
</tr>
<tr>
<td>J. Tabak, 2013</td>
<td>Sorafenib</td>
<td>Captopril</td>
<td>Cardiac</td>
<td>2 97</td>
<td>1 101</td>
<td>2.68 [0.19; 23.65]</td>
<td>1.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W. J. Gredl, 2013</td>
<td>Sorafenib</td>
<td>Pancreatitis</td>
<td>Breast cancer</td>
<td>2 110</td>
<td>0 118</td>
<td>5.12 [0.25; 106.74]</td>
<td>0.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. Knapke, 2014</td>
<td>Sorafenib</td>
<td>Captopril</td>
<td>gemcitabine</td>
<td>Urothelial</td>
<td>0 40</td>
<td>1 40</td>
<td>0.41 [0.02; 9.73]</td>
<td>0.7%</td>
<td></td>
</tr>
<tr>
<td>E. L. Meyer, 2015</td>
<td>Sorafenib</td>
<td>Decitabine</td>
<td>Pancreatitis</td>
<td>Breast cancer</td>
<td>0 23</td>
<td>0 23</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. Bergh, 2012</td>
<td>Sorafenib</td>
<td>Decitabine</td>
<td>Breast cancer</td>
<td>2 596</td>
<td>0 293</td>
<td>4.97 [0.24; 103.09]</td>
<td>0.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. P. Green, 2013</td>
<td>Sorafenib</td>
<td>Captopril</td>
<td>Breast cancer</td>
<td>2 217</td>
<td>1 215</td>
<td>1.99 [0.18; 21.69]</td>
<td>1.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H. G. Law, 2013</td>
<td>Sorafenib</td>
<td>Etorphine</td>
<td>NSCLC</td>
<td>1 64</td>
<td>0 64</td>
<td>3.60 [0.12; 72.28]</td>
<td>0.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Cowley, 2013</td>
<td>Sorafenib</td>
<td>Etorphine</td>
<td>Cardiac</td>
<td>12 364</td>
<td>4 379</td>
<td>2.96 [0.96; 9.36]</td>
<td>5.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. D. Michaelson, 2014</td>
<td>Sorafenib</td>
<td>Pancreatitis</td>
<td>Prostate</td>
<td>12 581</td>
<td>1 285</td>
<td>5.99 [0.77; 45.02]</td>
<td>1.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N. E. Reddy, 2015</td>
<td>Sorafenib</td>
<td>Captopril</td>
<td>Etoposide</td>
<td>SCLC</td>
<td>0 43</td>
<td>0 41</td>
<td>0.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. Bergmann, 2015</td>
<td>Sorafenib</td>
<td>Gemcitabine</td>
<td>Plasmatic</td>
<td>3 52</td>
<td>0 54</td>
<td>7.27 [0.36; 173.37]</td>
<td>0.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remmelsma, 2010</td>
<td>Temsirolimus</td>
<td>Captopril</td>
<td>Pancreatitis</td>
<td>NSCLC</td>
<td>2 62</td>
<td>0 92</td>
<td>2.60 [0.13; 52.57]</td>
<td>0.9%</td>
<td></td>
</tr>
</tbody>
</table>

**Random effects model \[ \text{Weight} = 5.796 \text{ SAE n, % [95% CI] Weight} \]

Table 1. Sample demographics by index drug. Serious adverse event (SAE) rates pooled by inverse variance weighted random-effects model meta-analysis of proportions.

### Notes

We gratefully acknowledge the contributions of Harry Atkins, Gina Freeman, Spencer Hey, Amanda Hakala, Tiger Zheng, Taiji Wang, James Mattina, Yasmina Hachem, Nadine Demko, Natalie Carlisle 93.
McKinnon, Sylviya Ganeshamoorthy, Holly Sarvas, Amanda MacPherson and the other members of the STREAM research group.

References


Summary

The previous two chapters document a high level of trial activity and patient involvement in the post-approval development of anti-cancer drugs in both monotherapy and combination therapy. Five years of post-approval clinical trial activity consisted of 104 trials with a total of 4699 patients in monotherapy, and 323 trials with a total of 29,835 patients in combination therapy. Neither drug-indication trajectories in monotherapy nor combination-indication pairings in combination therapy launched post-approval yielded new FDA approvals, and recommendations by the National Comprehensive Cancer Network (NCCN) were few (7.2% and 4.5% in monotherapy and combination therapy, respectively). NCCN recommendations were often based on lower-level evidence. For many of the drugs in our sample, approval was followed by more conservative exploration, however, in the aggregate this represents a large investment, and a broad programme of post-approval testing of new indications and combinations.

The final chapter of my thesis, “Trials that say 'maybe': the disconnect between exploratory and confirmatory testing after drug approval” is a concept paper in which we explore one of the insights earned by the previous chapters, and a potential source of inefficiency in human research. The extensive post-approval testing that we observed in previous chapters generally occurs in clinical trials that are designed and statistically powered to generate hypotheses about the possible uses of a drug. These trials are a valuable part of drug development, as they provide the grounds for larger, more rigorous confirmatory trials. However, we suggest that large programmes of post-approval research may result in periods of prolonged uncertainty from hypothesis-generating clinical trials that is not resolved by confirmatory testing, which we call “clinical agnosticism.” This poses threats to the ethics of human research, as well as the interpretation of medical evidence that grounds clinical practice.
Chapter 5. Trials that say “maybe”: the disconnect between exploratory and confirmatory testing after drug approval

Benjamin Gregory Carlisle, Carole A. Federico, Jonathan Kimmelman

Contrary to what might be expected, most clinical trials are directed not toward getting new drugs approved, but rather at re-purposing already licensed drugs for new applications. Once drugs receive their first license, companies, public funders, and medical centres often mount vigorous trial programmes exploring a drug’s activity in other indications, or in combination with other therapies. Such studies generate myriad hypotheses about possible treatment options for patients. This suspicion of clinical value, which we call “clinical agnosticism,” provides grounds for rigorous, randomized trials. However, clinical agnosticism generated in exploratory trials is often not swiftly resolved by confirmatory trials. This delay has several ethical implications: it fails to redeem the sacrifice of research volunteers in exploratory studies, and it places patients and healthcare systems at risk of using ineffective treatments for many years.

How post-license trials create clinical agnosticism

The primary aim of exploratory trials (typically phase 1 and 2 trials) is to produce evidence that a new treatment might be effective. At the outset of drug development, factors like responding disease indication or effective dosing are not yet known; nor is there reliable evidence that a new approach will be competitive with standard of care. Exploratory evidence, if favorable, often establishes a moral basis for randomizing patients in larger, confirmatory trials (typically phase 3 trials). However, because exploratory evidence is based on surrogate endpoints, small sample sizes, limited follow up, or unplanned subgroup analyses, it rarely is sufficient to guide practice. Indeed, 60% of drugs that advance from exploratory trials into confirmatory trials fail to vindicate their promise. As a result, later-phase, confirmatory trials are critical for transforming clinical agnosticism generated in exploratory trials into conviction—or doubt—about a drug’s value.

For unlicensed drugs, regulators currently require confirmatory trials before they grant full marketing approval. Some critics of regulation would seek to change this, and confirmatory trials
are often slow to complete when drugs receive accelerated approval. Nevertheless, regulations provide powerful incentives for drug companies to resolve uncertainties quickly on the heels of favorable exploratory findings when drugs are not yet approved. They also make it likely that ineffective treatments are intercepted before being taken up into practice.

Once drugs are approved, however, drug companies and public sponsors continue using exploratory trials to test new indications or combinations. Physicians and clinical practice guidelines are free in the meantime to recommend off-label prescription. Because regulators lack authority to require confirmatory trials for off-label treatments, encouraging findings from exploratory trials of approved drugs are often taken up into practice without proof of value.

Consider two well-known anti-cancer drugs, sorafenib and sunitinib, that were approved by the FDA in 2005 and 2006, respectively. With over ten years of exploratory trial activity since approval, these two first-in-class drugs illustrate how the hand-off from exploratory trials to confirmation can fumble after licensure. In a previous report of all monotherapy trials launched within 5 years of their U.S. approval, we found 41 instances where exploratory trials met their primary efficacy endpoint with acceptable toxicity (31% of all exploratory studies pursued). In 5 instances, these findings were taken up as recommendations for off-label use in clinical practice guidelines for the National Comprehensive Cancer Network (NCCN, an organization whose guidelines in the U.S. are used to set reimbursement policies). In another 7 instances, negative trials reporting responding subgroups were used to justify recommendations. However, in only 1 instance among the 13 recommendations were the hypotheses generated in these exploratory studies followed by a completed randomized trial using survival endpoints (see Table 1).

Such patterns can be observed for other drugs—especially when they are novel or believed to have many potential applications. Consider the blockbuster drug pregabalin (approved in 2004 and indicated for the treatment of partial seizures, some types of neuropathic pain, and fibromyalgia). On the basis of exploratory evidence, pregabalin is widely used off-label for migraine and low back pain; indeed, off-label prescriptions account for as much as 75% of prescriptions in pain. While exploratory trials suggesting the clinical utility of pregabalin for these conditions were last published in 2015, to our knowledge, no large, randomized trials testing pregabalin in migraine or low back pain have been published to date. In fact, a meta-analysis published 7 years after the first test of pregabalin in chronic low back pain suggested
that the existing evidence is limited for gabapentinoid activity and that there is a need for large, high-quality trials to more definitely inform the issue.\textsuperscript{13}

**How unresolved clinical agnosticism harms**

Though surely some knowledge is better than none, clinical agnosticism that is not followed up rapidly with confirmatory testing raises three ethical concerns. First, the moral basis for exposing patients to unproven treatments in exploratory trials derives from the expectation that such trials—if establishing the basis for clinical equipoise—will support further, confirmatory testing. Exploratory trials use biologically active doses that present the full risk profile of a drug. Researchers and sponsors have obligations to build on encouraging findings deriving from such exposures.

A second ethical concern is the welfare of patients. Prolonged agnosticism about treatments can harm patients if physicians—with the best of intentions—adopt treatments that are ineffective. In the setting of debilitating disease where standard of care options are exhausted, suspicion of efficacy—as opposed to proof of it—is often sufficient to influence treatment. In cancer, for example, up to one third of treatments are given off-label;\textsuperscript{14} in office-based care, 20\% of prescriptions are off-label.\textsuperscript{15} Ninety-four percent of recommendations offered in NCCN guidelines are based on “lower level” evidence—generally small, non-randomised, early phase trials.\textsuperscript{16} Harms to patients accrue over years if prescription recommendations are based on spurious exploratory findings.

Finally, such practices have impacts on healthcare systems. When treatments are reimbursed for lengthy periods based on exploratory evidence, healthcare systems must either charge greater premiums, or they have fewer resources available for proven treatments (for example, palliation for advanced disease). At the same time, if healthcare systems are willing to reimburse off-label prescriptions based on exploratory evidence, or if patients are willing to pay out-of-pocket, pharmaceutical companies have little incentive to fund large trials that might disprove a drug’s value. Patients may also be less willing to participate in a confirmatory clinical trial if a therapy is available to them off-label. Pregabalin again provides an instructive example. With annual global sales approaching $5 billion US, Pfizer paid $2.3 billion in 2009 to settle claims relating to the off-
label promotion of pregabalin. These promotions often drew on exploratory evidence and cost the U.S. healthcare system millions of dollars.

To be sure, there are understandable reasons why favorable exploratory findings might sometimes go unconfirmed for extended periods. If findings concern a rare disease, it may be difficult to secure funding and recruit patients for confirmatory trials. However, drug companies are capable of meeting these challenges where there is regulatory pressure. Whereas trials supporting FDA approval of orphan drugs have a median enrollment of 96 and are randomized 30% of the time, the evidence supporting recommendations in Table 1 is considerably weaker. Another mitigating factor is that exploratory trials sometimes yield important insights about pathophysiology, biomarkers, or research techniques. We nevertheless contend this is an unsatisfactory justification for the patterns described above: all trials supporting recommendations in Table 1 stated their primary goal as testing efficacy, not exploring techniques or concepts. Finally, treatment effects in exploratory trials are occasionally very large and well-supported by mechanistic knowledge. In such instances, findings can be sufficient to prove a drug’s value without confirmatory trials.

**Toward resolving clinical agnosticism**

There are several initiatives that—if followed through on—would go some way towards dispelling the problem of prolonged clinical agnosticism. The NCI has launched an “exceptional responder” programme that aims to determine whether outliers from negative clinical cancer trials can “open up new drug development avenues.” Public investments in comparative effectiveness programs (e.g. Agency for Healthcare Research and Quality’s Effective Health Care Program) can also identify the most important clinical hypotheses to test. However, we think more should be done.

First, the ethical evaluation of post-license hypothesis-generating trials should consider that many positive findings never make good on the burdens they impose because they are not advanced into confirmatory testing (similar arguments have been made about underpowered trials in general). In keeping with the requirement that human research generate valuable information, ethics committees should consider whether there is evidence of a credible path forward for transferring positive exploratory findings into confirmatory trials. Regulators can
play a similar role, based on their obligations to protect public health. Evidence that might be considered includes commitments and financial backing from companies or research consortia to pursue randomized trials if favorable findings from exploratory trials emerge. Patients entering exploratory trials should also understand flaws in the way medical research is organized. They should be told that the trials they are entering rarely prove a treatment’s efficacy, and that even when findings are promising, sponsors do not always follow up with definitive trials.

Second, financial incentives can be used to re-balance the broader research agenda. Policy makers should consider that liberalizing drug approval standards would likely exacerbate the problem of prolonged agnosticism. They should also consider measures that might encourage companies to run confirmatory studies—especially in the context of rare disorders. Pediatric exclusivity, which grants an extra 6 months of patent or data exclusively when companies test drugs in children, is one example of how policies can encourage companies to address evidence gaps. Public funding bodies should set aside a larger budget for confirmatory trials. Confirmatory trials are expensive and public funding bodies have very limited resources. Nevertheless, we think there is a credible case to be made that federal agencies could make a bigger impact on public health if they rejigged the way they allocate funding across exploratory and confirmatory clinical trials. We recognize that such proposals would have the undesirable effect of concentrating more trial resources in fewer research groups, but think this may be worthwhile given the stakes for patients and healthcare systems. Currently, NCI and NINDS funding is allocated more strongly toward phase 2 trials than funding from NHLBI. Among clinicaltrials.gov records, confirmatory trials (as judged by phase 2-3 or phase 3 status) represented 12.4% of NCI-funded trials, 25.1% of NINDS-funded trials and 39.3% of NHLBI-funded trials. This is in contrast to early phase, exploratory studies, which represent 87.6%, 74.9% and 60.7% of phase 2 testing for NCI-, NINDS- and NHLBI-funded trials, respectively.

Third, journal editors, referees, and physicians can do a better job vetting manuscripts for “spin.” In one study, 59% of 92 trials reporting a negative primary outcome nevertheless claimed clinical promise on the basis of secondary outcomes. In a follow-up randomized trial, experts rated treatments as more promising when abstracts presented to them were “spun” in this manner. Early phase trial reports should generally be presented as generating—not resolving—clinical hypotheses.

Carlisle 104
Having a large repertoire of clinical hypotheses is key to improving outcomes for patients—especially where effective management strategies are lacking. But for this system to work, there must be tight coupling between early phase exploratory testing and late phase, confirmatory trials. Studies of clinical development\textsuperscript{8,9} suggest that drug companies and public funding agencies often generate prolonged clinical agnosticism, and such agnosticism is often sufficient to influence physicians and healthcare system decisions. Who could blame companies for not seeking to capitalize on this opportunity? Policy makers should recognize the ethical and healthcare system stakes of ensuring a reliable coordination of exploratory and confirmatory trials.

**Key messages**

- After a new drug receives approval, companies and public sponsors often run numerous small trials exploring the drug’s activity in different indications.

- Though the level of evidence produced in such trials is usually low, drug companies and public sponsors often fail to follow up on promising exploratory findings by running large, confirmatory trials.

- The poor hand-off from exploratory to confirmatory testing often results in prolonged periods where ineffective and costly drugs are used off-label; it also fails to redeem the sacrifice of patients who participated in the original small trials.

- Ethics committees and policy-makers should devise measures that encourage better coordination between exploratory and confirmatory trials after drugs are approved.
<table>
<thead>
<tr>
<th>Drug</th>
<th>Indication</th>
<th>Citation provided by NCCN</th>
<th>Randomized?</th>
<th>Primary endpoint</th>
<th>Sample size</th>
<th>Trial result on primary endpoint</th>
<th>Confirmatory RCT registered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sorafenib</td>
<td>Angiosarcoma</td>
<td>Maki et al, 2009</td>
<td>N</td>
<td>ORR</td>
<td>(37)</td>
<td>-</td>
<td>None</td>
</tr>
<tr>
<td>Sorafenib</td>
<td>Chordoma</td>
<td>Bompas et al, 2015</td>
<td>N</td>
<td>PFS-9</td>
<td>(27)</td>
<td>-</td>
<td>None</td>
</tr>
<tr>
<td>Sorafenib</td>
<td>Desmoid tumours</td>
<td>Gounder et al, 2011</td>
<td>Retrospective case series</td>
<td></td>
<td></td>
<td>NCT02066181 (ongoing, PFS primary outcome)</td>
<td></td>
</tr>
<tr>
<td>Sorafenib</td>
<td>GIST</td>
<td>Montemurro et al, 2009</td>
<td>Retrospective analysis</td>
<td></td>
<td></td>
<td>None</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kindler et al, 2011</td>
<td>N</td>
<td>ORR</td>
<td>38</td>
<td>+</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Park et al, 2012</td>
<td>N</td>
<td>DCR</td>
<td>31</td>
<td>+</td>
<td>None</td>
</tr>
<tr>
<td>Sorafenib</td>
<td>Hemangiopericytoma</td>
<td>Valentin et al, 2013</td>
<td>N</td>
<td>PFS-9</td>
<td>(5)</td>
<td>-</td>
<td>None</td>
</tr>
<tr>
<td>Sorafenib</td>
<td>Osteosarcoma</td>
<td>Grignani et al, 2012</td>
<td>N</td>
<td>PFS-4</td>
<td>35</td>
<td>+</td>
<td>None</td>
</tr>
<tr>
<td>Sunitinib</td>
<td>Alveolar soft part sarcoma</td>
<td>Stacchiotti et al, 2009</td>
<td>Retrospective case series</td>
<td></td>
<td></td>
<td>None</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stacchiotti et al, 2011</td>
<td>Retrospective case series</td>
<td></td>
<td></td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Sunitinib</td>
<td>Angiosarcoma</td>
<td>George et al, 2009</td>
<td>N</td>
<td>ORR</td>
<td>(2)</td>
<td>-</td>
<td>None</td>
</tr>
<tr>
<td>Sunitinib</td>
<td>Chordoma</td>
<td>George et al, 2009</td>
<td>N</td>
<td>ORR</td>
<td>(9)</td>
<td>-</td>
<td>None</td>
</tr>
<tr>
<td>Sunitinib</td>
<td>Hemangiopericytoma</td>
<td>George et al, 2009</td>
<td>N</td>
<td>ORR</td>
<td>(3)</td>
<td>-</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stacchiotti et al, 2012</td>
<td>Retrospective case series</td>
<td></td>
<td></td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Sunitinib</td>
<td>Meningioma</td>
<td>Kaley et al, 2015</td>
<td>N</td>
<td>PFS-6</td>
<td>36</td>
<td>+</td>
<td>None</td>
</tr>
<tr>
<td>Sunitinib</td>
<td>Thymic carcinoma</td>
<td>Thomas et al, 2015</td>
<td>N</td>
<td>ORR</td>
<td>(25)</td>
<td>-</td>
<td>None</td>
</tr>
<tr>
<td>Sunitinib</td>
<td>Thyroid</td>
<td>Carr et al, 2010</td>
<td>N</td>
<td>ORR</td>
<td>35</td>
<td>+</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cabanillas et al, 2010</td>
<td>Retrospective analysis</td>
<td></td>
<td></td>
<td>None</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Indications recommended by the NCCN for sorafenib and sunitinib monotherapy that are not approved by the FDA; citations to clinical trial data (all phase 2) or retrospective analyses used to justify recommendation. Confirmatory status was defined based on whether trials used randomized design and a survival or progression-free survival endpoint. Confirmatory trials were
sought by a search (2017 December 12) of clinicaltrials.gov for trials using the same drug in the same indication. Recommendations based on subgroup analysis are indicated by parenthetical sample size. Abbreviations: N = no; Y = yes; + = positive trial result; - = negative trial result; ORR = objective response rate; PFS = progression-free survival.

References


Discussion

The previous five chapters present a perspective on the moral efficiency of clinical trials in anti-cancer drug development that is informed by clinically relevant outputs of drug development, such as FDA approvals, recommendations in clinical practice guidelines, and morally relevant inputs, such as the enrolment of patient-subjects and measures of risk and benefit. We attempt to capture some of the lesser-appreciated inefficiencies and residuals in drug development, such as clinical trials exploring indications and combinations that are launched after a drug is already approved. Finally, we explore general principles that arise from analyzing the moral efficiency of anti-cancer drug development.

The moral efficiency of drug development varies in ways that are not apparent at the individual trial level, and often in ways that may be counter-intuitive. Seemingly, over the course of a drug's development, there are cases where risk and benefit do not improve, but worsen. The early successes of drugs such as sunitinib and imatinib were followed by extensive testing that was mostly unsuccessful later on. The volume of post-approval trial activity testing imatinib was much greater than its pre-approval activity, and it was much less efficient by the measures we defined. The clinical impact of post-approval monotherapy exploration seen in chapter 3 was modest, with significant levels of patient involvement and burden. Combination therapy in chapter 4 recapitulated these dynamics, but on an even greater scale. Clinical agnosticism and other similar dynamics that result from extensive programmes of post-approval drug testing also threaten the scientific value of human research.

Moral efficiency measures have implications throughout drug development. I will outline four implications of the insights gleaned from this project. First, the analysis of moral efficiency of drug development may bear on decisions to launch or continue clinical trials. Second, moral efficiency, as well as the concept of clinical trial “portfolios” may guide the interpretation and evaluation of previously completed clinical trials. Third, patients making decisions regarding enrolment in a clinical trial—or doctors referring them—may wish to consider the profile of risk and benefit of the larger programme of research that a clinical trial is embedded in. Fourth, health research policy and institutional priority setting can also be informed by a broader view of clinical trial ethics.
This project also opens up new areas of inquiry into research ethics. I will describe five broad categories of new potential research, and briefly outline examples of each. Finally, I will outline several limitations of the methods used.

**Ethical review of clinical trials**

Ethical decisions regarding whether to launch or continue a clinical trial are conducted by committees tasked with reviewing the trial, and they are generally based on an analysis of the individual trial in question only. In light of this project, there are potential moral issues that may emerge from the consideration of the efficiency of drug development, and these may be relevant to this kind of decision-making. For example, the sunitinib monotherapy portfolio contains examples of potentially duplicative negative clinical trials testing the same indication. While each of the duplicative trials with negative results was duly considered and approved by a research ethics board and found to be acceptable in isolation, an examination of the larger portfolio might make future research ethics boards reconsider launching or continuing trials in similar cases. Research ethics boards may consider requiring a review of other completed or ongoing clinical trials to determine whether launching a new trial would contribute new information regarding the potential uses of a drug, or if it would potentially represent an inefficiency.

This thesis contained several examples where risk and benefit worsened over the course of a drug’s development overall. Chapters 3 and 4 found no examples of FDA approval for new monotherapy indications or new combinations that were launched into testing after a drug was already approved. Insofar as these trends are generalizable to newer generations of drugs, they should be brought to bear on decisions to launch or continue new clinical trials. Research ethics boards considering a clinical trial of an approved drug in a new indication may consider raising the standard for evidence that they require to justify launching a trial of a new trajectory when it comes after a series of failed attempts to extend a drug to new indications.

The launch of a new indication or combination should also be considered in light of the prospect of producing prolonged periods of clinical agnosticism. A clinical trial that is powered to generate a hypothesis rather than resolve it may leave clinicians with minimal guidance, and it may encourage the off-label use of an ineffective drug. Patients enrolled on such a trial will be exposed to experimental drugs that carry the risk of serious adverse events, and may undergo...
burdensome research procedures that are unredeemed by confirmatory testing. In light of the potential for extended periods of clinical agnosticism, institutional review should consider the likelihood of advancement to confirmatory testing when evaluating a prospective new hypothesis-generating clinical trial. There may be other reasons that outweigh concerns regarding advancement to confirmatory testing, such as the case of rare diseases, but the potential for unresolved uncertainty regarding the efficacy of a drug is an important consideration nonetheless. Policies of this type have also been suggested for underpowered trials.62

**Interpretation and evaluation of previously completed clinical trials**

The synthesis of evidence provided by multiple clinical trials of the same treatment is generally non-controversial. Systematic reviews and meta-analyses are already widely regarded as some of the highest-level and reliable medical evidence available.63 Less widely appreciated is the potential that a clinical trial’s inclusion in a large portfolio of related trials may confound proper statistical inference. Multiple testing of correlated hypotheses is common among development programmes for successful anti-cancer drugs, and as we have shown, it is common for drug development portfolios to include a large number of trials. Exploratory clinical trials in cancer are powered with a particular Type I error rate, often 5%. This means that unless there is correction for statistical multiplicity, a large portfolio of clinical trials is likely to contain a false positive result.

Beyond multiplicity, there is also Stein’s paradox, which applies to portfolios of drug trials.64 There are two sources of variability within a drug’s portfolio: the underlying heterogeneity due to differences in the trials (populations, indications, etc.) as well as the random variation in treatment effects. Paradoxically, unless treatment effects are “shrunk” toward the treatment mean of the entire portfolio, observations of large effects are likely to represent over-estimates, and exaggerate the effect size.14

Investigators and research ethics boards should consider at the outset of a trial the possibility that their trial may produce a false positive because of Type I error or Stein’s paradox. This may inform decisions regarding the design and statistical powering of clinical trials, or whether they are launched at all. Funders and institutional decision-makers may choose to prioritize more
basic science to develop novel treatments rather than extend the use of older ones. They may also
decide to prioritize confirmatory testing of hypotheses generated by trials that may represent
potential over-estimates of treatment effect.

**Patient decision-making**

The risk and benefit to patients considering enrolling in clinical trials should be informed by
reference to the properties of the larger portfolio of research in which a proposed clinical trial is
embedded. For example, while imatinib rightly has a reputation for being safe and effective, there
was an expansive programme of testing following its licensure that led to several new licensed
indications, but also numerous failed attempts to extend its utility. The end of the imatinib
portfolio was similar in some ways to the drug development programmes of multi-tyrosine
kinases, including broad indication exploration that was redeemed with few regulatory successes.
The development of sunitinib also provided an example of worsening risk and benefit over the
course of a drug’s development. Patient decision-making should be informed that contrary to the
tacit portrayal of clinical translation as an orderly pipeline of increasingly safe and effective
clinical trials, there may be a worsening profile of risk and benefit later in a drug’s development.

Patients are often cognizant that they may not personally benefit from enrolment on a clinical
trial, and rather base their decision to enrol in a clinical trial on the expectation that they will be
helping to advance medical practice. In many cases, they are willing to take on risk and
potentially burdensome research procedures because they believe that doing so will provide
better data to guide future medical practice. Our data suggest that few cases of trajectories
launched after a drug’s first approval lead to new FDA approvals or clinical practice guideline
recommendations. Further, a trial that generates clinical agnosticism may result in prolonged
periods of uncertainty in which ineffective and costly drugs are used off-label. As previously
discussed, Type I errors and Stein’s paradox may also confound correct statistical inference in
large programmes of research attempting to extend a drug’s utility. Thus, a patient’s expectation
that they are helping to advance medical practice may be less well-founded in cases where there
is little expectation that a clinical trial will produce clinically actionable information, and it may
be subverted entirely in cases where a trial represents a false positive result or produces
prolonged periods of clinical agnosticism.
Health research policy and institutional priority setting

Health research policy and priority setting should be influenced by an accurate assessment of the current state of research that includes an assessment of drug development at the portfolio level. The generally accepted portrayal of clinical trialling is one of increasingly safe and informative studies progressing from phase 1 to phase 3 and culminating in licensure.\textsuperscript{12}

Our analysis shows that risk and benefit to patients does not necessarily improve over the course of a portfolio, and extensive programmes of clinical trials do not necessarily yield a proportionate amount of clinical impact, measured in terms of FDA approvals, clinical guideline recommendations or randomized testing. Clinical trials represent opportunity costs, and this project may suggest that extensive attempts to extend a drug’s utility to new indications or combinations may provide less clinical impact than investing such resources into developing new unlicensed treatments.

Regulators and funders may also consider statistically shrinking estimates of effect sizes within portfolios when making regulatory decisions. This will produce more accurate estimates, and it will also discourage some investigators from running hypothesis-generating clinical trials that may bring about clinical agnosticism.

Future research

This project raises a number of possible avenues for future research. First, I will propose several areas of research that will close gaps in knowledge left by my thesis that can be addressed by the methods that my thesis develops. Second, I will consider how my thesis sets a benchmark for comparing future analyses of drug development programmes. New developments in cancer therapy, such as the advent of precision medicine and immunotherapy provide opportunities to measure changes in risk, benefit and clinical impact across entire drug development paradigms. Third, some of the work in my thesis suggests the hypothesis that the prevalence of a cancer indication may correlate with the amount of clinical trial activity. The order with which indications are explored also appears to be related to successful clinical translation. I will explore how these hypotheses may guide future analyses of moral efficiency. Fourth, new developments in the practice of clinical trials such as prospective registration, adaptive trial designs and “mega
phase 1” trials may provide new areas for future research into the moral efficiency of clinical trials and guide future trial activity. Finally, I will propose a project that attempts to correct for potential over-estimation of effect sizes within drug development portfolios, which has implications for the interpretation and conduct of clinical trials.

The focus of my thesis was translational drug development—clinical trials that were aimed at achieving a drug’s first FDA approval, or extending it to new indications or combinations. Hence, any trial activity that occurred in a licensed indication after a drug’s approval (for example, any trial testing sunitinib in renal cell carcinoma after 2006) was excluded from our analysis. The risk and potential harms to patient-subjects in post-approval clinical trials in approved indications is much lower, but still nonzero, as new schedules and doses are tested, and patients may need to undergo burdensome research procedures that they might not have encountered while receiving care outside the research setting. In some cases post-marketing clinical trials are thinly-veiled marketing activities referred to as “seeding studies” which pose threats to the integrity of the human research enterprise. A future programme of research could quantify threats to the moral efficiency of drug development by measuring the extent of post-marketing clinical trial activities in approved indications, and the moral efficiency in terms of human enrolment, burdensome research procedures, and threats to scientific value such as inappropriate trial design and statistical underpowering.

My thesis addresses the moral efficiency of drug development for a set of drugs that was approved by the FDA. The projects presented here do not consider clinical trial activity that occurs in the testing of drugs that never achieve licensure. Future researchers could study a matched sample drugs that did not receive FDA approval. This cohort of drugs could be constructed by querying clinicaltrials.gov for drugs that were tested in clinical trials during the same timeframe as the drugs studied in my thesis, and including drugs where active development has ceased without achieving regulatory approval within a defined period of follow-up. The systematic review methods developed in my thesis could be applied to this sample of drugs, and the efficiency of developing successful drugs that is reported here could be used as a benchmark. This would allow us to determine the cost of ruling out the use of a clinically promising treatment, in terms of clinical trial activity, volume of patient-subject enrolment and direct patient risk and benefit.
A matched cohort of unsuccessful drugs as described above would quantify the cost of ruling out the use of a promising new drug, but it would not tell us how much unsuccessful clinical trial activity there is in total. My thesis quantifies the extent of successful clinical trial activity within a two year timeframe. To make a measurement of the moral efficiency of clinical trial activity that includes both approved and non-approved drugs, and to establish the proportion of clinical trial activity leading to an FDA approval, a cross-section of all clinical trials within a two-year timeframe could be constructed, based on registry entries in clinicaltrials.gov. An estimate of the amount of clinical trial activity in drugs that never receive licensure and the burden to patient-subjects associated with it would quantify one of the residuals of the moral efficiency of human research in anti-cancer drug development.

The results reported here provide a benchmark against which to measure the efficiency of future research programmes. There have been many broad paradigms of cancer drug development including cytotoxic drugs, targeted therapies, precision medicine and immunotherapy. The drugs studied in my sample are mostly examples of targeted therapies, and so it remains to be seen whether the patterns observed in my thesis will be repeated in future paradigms. Precision medicine and immunotherapy as paradigms for drug development may change the balance of risk and benefit to the human subjects of anti-cancer research, as well as the output in terms of clinically actionable information. When there is sufficient time for follow-up, a sample of drugs could be constructed from more recent paradigms of drug development, and the moral efficiency with which they are translated to useful treatments could be measured using the template provided by my thesis, and benchmarking results against what is already known about drugs developed between 2005-2007.

The work represented by my thesis suggests that there are patterns to the order and persistence with which indications are explored in post-approval testing. More prevalent cancer indications seem to be pursued more persistently within a drug’s development. For example, within the sunitinib monotherapy portfolio, there were four separate phase 2 clinical trials of breast cancer, none of which produced a positive result on its primary endpoint. The hypothesis that more prevalent indications are tested more heavily could be tested by constructing a sample of drug-indication trajectories using the methods defined in my thesis, and measuring the prevalence of...
the indication as well as measures of persistence, such as the number of clinical trials launched, and the volume of patients enrolled.

My thesis also suggests that over the course of a successful drug’s development, the benefit and risk for patients worsens. The order in which drug-indication trajectories are pursued within a drug’s development may correlate with measures of that trajectory’s success, such as FDA licensure, uptake into clinical practice guidelines, advancement to randomized testing or even direct measures of patient benefit and burden. Indeed, among our sample of 12 drugs, none of the trajectories launched into testing after a drug’s first approval received licensure. This suggests that later stages of development consist of clinical trials that are testing hypotheses that are more of a “long shot” than ones pursued early on. This hypothesis could be tested by constructing a sample of drug-indication trajectories based on a random sample of trials from clinicaltrials.gov from a timeframe that allows sufficient follow-up to determine whether it is ultimately vindicated with FDA approval. These trials could be characterized in terms of their inclusion in a drug-indication trajectory, and that trajectory could be numbered according to the order in which it is pursued within the drug’s development. This would allow us to measure correlations between the order of trajectories within a drug’s development and whether it is met with regulatory approval, uptake into clinical practice guidelines or advancement to randomized testing. The relationship between the order of indication and measures of direct patient benefit and burden could also be measured. This would have implications for the review of future trials, research priority-setting, and for patient decision-making when considering enrolment in a clinical trial.

New developments in the practice of clinical trials also open up avenues for applying and extending the methods and tools used to pursue the projects that comprise this thesis. Prospective registration of clinical trials has been legally mandated in certain cases, and required by many medical journals. This was not the case when many of the drugs in my sample first entered human testing, and so we were unable to rely on clinical trial registries in order to populate our sample. The advent of clinicaltrials.gov and other international clinical trial registries allows for many of the techniques described in this thesis to be automated, although many data points will still require human curation. Converting these techniques from traditional systematic reviews of clinical trial journal publications to tools for the analysis of clinical trial
registry entries will allow for rapid development of hypotheses (c.f. The Drug Trials Visualiser, Appendix G), reproducibility of methods and reduced non-publication bias.

The development of cancer drugs that were approved between 2005-2007 was, in some ways, very straightforward. Phase 1 trials in my sample were generally small one-armed trials designed to measure the maximum tolerated dose of a drug and characterize its pharmacokinetics and pharmacodynamics. Most phase 2 trials in my sample were small, often Simon Two-Stage trials designed to produce testable hypotheses about potential clinical efficacy. Phase 3 trials studied in my thesis were usually large, multi-arm trials designed to confirm a clinical hypothesis. The design of clinical trials has since changed, and the advent of “mega phase 1 trials” and new adaptive trial designs also provide opportunities for new research and to extend and adapt the techniques developed in my thesis, to see if the dynamics observed continue to exist, and to use the work reported here as a benchmark for other research programmes. Adaptive trial designs are meant to make oncology drug development “more informative, more accurate and shorter.”

A sample of oncology drug development strategies that employ adaptive trial designs could be constructed, and compared using the methods defined in my thesis to determine whether these new trial designs also provide efficiency when evaluated according to measures of clinical impact such as FDA approvals or recommendations by clinical practice guidelines, and measures of patient risk and benefit, such as survival, serious adverse event rates, and volume of enrolment.

Machine learning techniques have also been applied in attempts to extend drugs to new indications and combinations. A future programme of research could review the clinical successes and patient burden associated with this new technique for drug exploration, and use traditional clinical trialling as a benchmark for evaluating their moral efficiency.

A future programme of research may attempt to mathematically correct potential over-estimation of effect sizes within large portfolios that represent correlated hypotheses. As I described earlier, large portfolios of drug development may, in certain circumstances, contain examples of over-estimates of treatment effect, due to an under-appreciated statistical paradox. Large programmes of research could be evaluated in light of this, and effect sizes duly adjusted. A sample of successful drugs could be constructed, and their development portfolios systematically reviewed, capturing all clinical trials aimed at bringing the drug to market or extending its activity to new indications. The treatment effects of the clinical trials in this sample could be statistically

Carlisle 118
corrected, to “shrink” them toward the treatment mean of the entire portfolio, producing more accurate estimates. The more extreme positive effects within a drug’s portfolio will be reduced, which may result in more modest anti-cancer activity when portfolio effects are taken into account; or possibly even reversals in cases of comparisons of effective treatments. This will have implications for clinical trial interpretation and clinical practice.

I have mostly described portfolios in this thesis in terms of a single drug’s development. This is not, however, the only way that a portfolio could be defined. For example, a portfolio could be defined by a cross-section of drug development that captures all trials of a cancer indication regardless of the drug; it could also be defined in terms of a mechanism of action being explored. This suggests many ways that the principles explored by this thesis could be extended even further.

**Limitations**

These conclusions should be considered in light of the limitations of the projects that comprise this thesis. An analysis of the moral efficiency of clinical trials necessitates a long period of follow-up in order to evaluate the clinical impact of the hypotheses it represents. For example, a claim that a hypothesis-testing trial with a positive result was not vindicated with confirmatory testing would only hold weight if there was an adequate period of follow-up. Because of the speed with which clinical trials are designed, funded, launched and completed, projects of this nature therefore require a timescale on the order of years, or even decades. This means that, for many of the results found using these methods, there will always be doubt over whether they are applicable to the present human research landscape. There will also be doubt over whether there has been a sufficiently long period of follow-up.

Due to a need for adequate follow-up, the drugs we examined were first approved between 2005-2007. At this time, clinical trial registry was not yet legally mandated, and so we rely heavily on the published record, which may be a source of bias. A more recent sample that uses clinical trial registry entries may find different dynamics.

It remains to be seen whether the dynamics we have observed here will be repeated in newer generations of anti-cancer drugs, or whether they can be generalized to other domains of drug development. After all, even within our sample, only the most successful drugs had very extensive

Carlisle 119
development portfolios. But there may be reason to dampen enthusiasm for contemporary oncology research—the development of molecularly targeted agents was also, at the time, considered to be fundamentally different from previous anti-cancer drug development, and there was an expectation that this would lower levels of attrition, increasing the efficiency of drug development.\textsuperscript{79} Clinical agnosticism, at least, appears to generalize to drug development outside the context of anti-cancer drug development. In 2004, pregabalin was approved by the FDA for treatment of partial seizures, diabetic peripheral neuropathy, and post-herpetic neuralgia, however it has been prescribed extensively for chronic non-cancer pain on the basis of exploratory evidence only.\textsuperscript{80} Regardless of whether or not these patterns will continue, or if they apply outside the domain of cancer drug development, this work represents a baseline against which to compare future drug development programmes.

Finally, because of the meta-level nature of this project, simplifying assumptions needed to be made in order to make data from different trials commensurable. Not all clinical trials report survival endpoints, for example, and so for some analyses, we must restrict ourselves to only considering subsets where those data are reported. In many cases, we must resort to using rough proxy measures for patient burden, such as the number of patients enrolled, or the number of patients who experienced a grade 5 serious adverse event, and measures such as the eventual FDA approval of an indication as a measure of benefit. These measures will blur many of the finer details, and provide a very high-level view of the drug development landscape.

**Conclusion**

The moral efficiency of drug development can be expressed in terms of the relationship between the useful scientific value accrued from clinical trials and the human welfare expended to attain it. Examining the amount of clinical impact a clinical trial has in light of the risk and benefit experienced by patient-subjects is relevant when designing, conducting and reviewing a clinical trial. This has implications for research policy, priority setting, patient protections, and for guiding future research programmes.
References for introduction, summaries and discussion


19. Knowledge of the Tuskegee study and its impact on the willingness to participate in medical research studies. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2568333/. (Accessed: 10th December 2018)


Carlisle 123


40. Biomedical research: increasing value, reducing waste - The Lancet. Available at: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(13)62329-6/fulltext. (Accessed: 10th December 2018)


