NBER WORKING PAPER SERIES

CHARACTERIZING THE DRUG DEVELOPMENT PIPELINE FOR PRECISION MEDICINES

Amitabh Chandra
Craig Garthwaite
Ariel Dora Stern

Working Paper 24026
http://www.nber.org/papers/w24026

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2017

We are grateful to Brian Alexander, Can Huang, Jennifer Kao, Rebecca Sachs, Mark Trusheim, and participants in workshops for the NBER Project on Economic Dimensions of Personalized and Precision Medicine, the NBER-AIEA Conference, and the Bates White Life Sciences Symposium, for helpful comments. Ben Berger, Holly Breuer, Andrew Marder, Caroline Marra, and Alice Ndikumana provided excellent research and programming assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Characterizing the Drug Development Pipeline for Precision Medicines
Amitabh Chandra, Craig Garthwaite, and Ariel Dora Stern
NBER Working Paper No. 24026
November 2017
JEL No. I1,L5,O3

ABSTRACT

Precision medicines – therapies that rely on genetic, epigenetic, and protein biomarkers – create a better match between individuals with specific disease subtypes and medications that are more effective for those patients. These treatments are expected to be both more effective and more expensive than conventional therapies, implying that their introduction is likely to have a meaningful effect on health care spending patterns. In addition, precision medicines can change the expected profitability of therapies both by allowing more sophisticated pricing systems and potentially decreasing the costs of drug development through shorter and more focused trials. As a result, this could change the types of products that can be profitably brought to market. To better understand the landscape of precision medicines, we use a comprehensive database of over 130,000 global clinical trials, over the past two decades. We identify clinical trials for likely precision medicines (LPMs) as those that use one or more relevant biomarkers. We then further segment trials based on the nature of the biomarker(s) used and other trial features with economic implications. Given potential changes in the incentives for bringing products to market, we also examine the relative importance of public agencies such as the National Institutes of Health (NIH) and different types of private firms in developing precision medicines.

Amitabh Chandra
John F. Kennedy School of Government
Harvard University
79 JFK Street
Cambridge, MA 02138
and NBER
amitabh_chandra@harvard.edu

Craig Garthwaite
Department of Strategy
Kellogg School of Management
Northwestern University
2001 Sheridan Road
Evanston, IL 60208
and NBER
c-garthwaite@kellogg.northwestern.edu

Ariel Dora Stern
Harvard Business School
Morgan Hall 433
Boston, MA 02163
astern@hbs.edu

1. INTRODUCTION

While lacking a universally agreed upon definition, Precision Medicine is broadly known as an approach to disease treatment and prevention that takes into account variability in environment, lifestyle, and genes for each person.[1] The concept of targeted interventions has a long history across the practice of medicine, however, recent technological advancements have made it increasingly possible to tailor the development and utilization of medical technologies. This possibility has attracted interest from the medical and broader scientific communities. For example, in early 2015, the White House announced a "bold new research effort to revolutionize how we improve health and treat disease," and launched a Precision Medicine Initiative with a $215 million investment in 2016.[2] Other countries such as France and China have also announced major public investments ranging from the equivalent of several hundreds of millions of U.S dollars to several billion over coming years. Major investments to advance precision medicine have also been announced by a number of U.S. research institutions such as Harvard University and the University of California San Francisco.[3]

Below, we consider a subset of the broad set of practices encompassed by "precision medicine" and focus specifically on the clinical development of precision *medicines*, i.e. those new therapies focused on biomarker-defined patient subgroups. Precision medicines, and in particular, therapies that rely on genetic, epigenetic, and protein biomarkers, can help patients by using identifiable biological features (biomarkers) to define disease subtypes. The technology to rapidly and accurately sequence genes has increasingly facilitated understand the "-omic" (genomic and proteomic) characteristics of disease in recent years. This, in turn, has broadened the scope for drug development focusing on targeted therapies for newly-identifiable sub-groups of patients. Indeed, the public efforts noted above lag private endeavors in this area: the pharmaceutical industry[4] has already commercialized almost 150 drugs with pharmacogenomic information in

[1] https://www.nih.gov/research-training/allofus-research-program
[2] https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative
[3] http://solidarites-sante.gouv.fr/IMG/pdf/genomic_medicine_france_2025.pdf
https://www.genomeweb.com/clinical-translational/france-plans-invest-670m-genomics-personalized-medicine
https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative
http://www.nature.com/news/china-embraces-precision-medicine-on-a-massive-scale-1.19108
http://www.hbs.edu/news/releases/Pages/kraft-family-foundation-establishes-endowment.aspx
https://www.ucsf.edu/news/2015/08/131341/new-center-will-advance-life-saving-genome-based-diagnostic-tools
[4] Throughout the chapter, reference to the pharmaceutical industry and pharmaceutical manufacturers refers to all firms developing drugs to treat medical conditions, including pharmaceutical and biotechnology firms

their label, according to the U.S. Food and Drug Administration (FDA),[5] suggesting there are already substantial economic incentives for private firms to invest in the development of precision medicines.

We focus on precision medicines because, in theory, they allow for a more effective match between individuals with specific disease sub-types and medications that are more effective for those sub-types. While the science underlying these medicines is broadly interesting and is the subject of a growing body of research, this ability to more precisely match patients and medications based on likely efficacy also fundamentally changes many of the *economic* incentives that pharmaceutical manufacturers face in the drug development process. Given the growing importance of these medicines, these changing incentives could have far reaching implications on the entire pharmaceutical industry.

Perhaps most importantly, the ability to develop more targeted products may influence the decision process for which drugs to bring to market. These decisions will then subsequently be reflected in the equilibrium prices and availability of new pharmaceutical products. For example, almost by definition, precision medicines tend to target smaller patient populations than more traditional medicines. This may mean that manufacturers will shift their attention to the subset of products able to command high(er) prices – and thus are more likely to justify the fixed costs of developing the medication. These higher priced products are likely to include those products with large clinical benefits in relatively small patient populations. In addition, since these drugs are more efficacious within a smaller patient population, the marginal customer is expected to have a greater willingness to pay, allowing for higher profit maximizing prices on the part of manufacturers. These two factors together provide an economic rationale for the broadly higher prices observed for precision medicines.

Economic incentives could also, all else held equal, result in some products no longer being brought to market because manufacturers don't believe they can reasonably expect to recuperate their research and development (R&D) expenditures from the relatively small target patient populations. For example, with increasingly small patient populations we might expect a decrease in brand-brand competition for particular patients as new entrants find the potential for competing for a small market to be an unattractive economic opportunity. Perhaps more concerning, a similar dynamic could exist for the eventual generic

[5] http://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/pmc_personalized_medicine_by_the_numbers.pdf

markets for precision medicines, which would extend the period of pricing power far beyond the period of patent protection.

Potentially counteracting this effect is manufacturers' ability to create identifiable subgroups of patients based on their willingness to pay, so-called "indication-based pricing." Such an ability on the part of manufacturers increases the scope for future price-discrimination as manufacturers could, in theory, more easily charge higher prices for high-value indications and lower prices for indications or patients where therapies will work less well (Chandra and Garthwaite, 2017). This would (weakly) increase the profits from any particular product and would, in turn, (weakly) increase the subset of early-stage products that pharmaceutical manufacturers would consider as candidates for commercialization. In addition, greater potential therapeutic benefit may result in smaller, shorter clinical trials because fewer patients would be needed and shorter periods of time will be sufficient for demonstrating statistically significant improvements in outcomes. Smaller and/or faster trials would both decrease the costs of bringing a drug to market and could increase the drug's effective patent length,[6] increasing the set of pipeline drugs considered as potentially worthwhile R&D investments. These factors together could counteract some of the negative entry incentives that might be created by small patient populations.

Despite the potential for precision medicines to both reduce some of the costs of drug development and also increase the patient benefits created by new products, markets for some medicines may still be so small that private firms will lack the necessary incentives for bringing therapies to market. This would create a potential role for government funding of research in these areas from sources such as the National Institutes of Health (NIH).

Finally, the emergence of a new technology could create opportunities for additional specialization of firms into different stages of the development process and/or create new markets for mergers and acquisitions (M&A) among pharmaceutical companies. This could, for example, lead to early-stage drug discovery being increasingly pursued by a subset of highly specialized (e.g. small, research-focused) firms.

[6] Patent life for a drug in the U.S. is generally 20 years from the date the application is filed and manufacturers can file a patent application any time before or during a drug's development process. Therefore, the time that a drug spends in clinical trials (i.e. before the drug can be marketed) are typically counted against the 20-year patent life. Marketing exclusivity is different from patent life and is granted by the FDA upon drug approval. Exclusivity typically lasts for 5 years, though there are extensions to exclusivity for certain cases, such as orphan drugs and pediatric indications. (https://www.fda.gov/downloads/drugs/developmentapprovalprocess/smallbusinessassistance/ucm447307.pdf)

More broadly, it is possible that the emergence of precision medicines will shift the division of labor between small biotechnology companies and large pharmaceutical companies across different stages of the R&D process.

To help understand this collection of potential economic implications of precision medicines, we aim to provide a detailed characterization of the existing drug development efforts in this area. We begin at a broad level by examining the aggregate development of likely precision medicines (LPMs), those pipeline drugs whose clinical trials have signature features of precision medicine R&D. We identify and report on clinical trials for such medicines by therapeutic area and over time. Since cancers represent a set of diseases in which precision therapies are already successfully used, and since cancer applications of precision medicine are expected to grow rapidly over the coming years, we separately characterize cancer LPMs. Understanding the nature of these innovations provides first order information on the wide-ranging health care spending implications of these emerging medications.

We then examine other aspects of clinical trials that provide additional insight into the economic mechanisms of drug development that are shaping the nature of innovation in this area. We begin by considering the characteristics (e.g. geography, indication, sponsorship) of clinical research between LPM vs. non-LPM trials. We then summarize the role of the NIH in supporting the existing pipeline of precision medicines, by asking what share of LPM clinical trials cite the support of NIH grants. Finally, we consider the types of firms pursuing clinical trials in LPMs, considering how LPM R&D activities has evolved over recent years.

2. Precision Medicines and the Drug Development Process

As discussed above, we focus on the development of precision medicines – those products that use biomarkers to target particular subgroups of patients. To better understand how these products are defined and developed, we begin by providing some background information on the science of biomarkers and their use by various economic actors in the drug development process.

Precision medicines and biomarkers

The FDA defines a "biomarker" as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or biological responses to a therapeutic intervention."[7] A familiar example can be seen in the common medical practice of using glycated hemoglobin (HbA1c), an indicator of average blood glucose levels over time, as a measure of the effectiveness of a therapeutic agent in controlling diabetes. In this example, the biomarker (which indicates therapeutic efficacy) is HbA1c. However, biomarkers can also be used to carve out patient subtypes of diseases because a treatment may work differently in patients who vary in their biomarker subtypes. In this case, a biomarker can be used predictively to determine *ex ante* how likely a given patient is to benefit from a therapy. For example, among patients with non-small cell lung cancer, those with the ALK (anaplastic lymphoma kinase) gene mutation will benefit more from therapies like alectinib (Alecensa®) than patients without this mutation. Similarly, the CFTR (cystic fibrosis transmembrane conductance regulator) modulator ivacaftor (Kalydeco®) has been approved for people with cystic fibrosis (CF) who have at least one of thirty-eight CF mutations—out of more than 1700 mutations in the gene that causes the disease. This amounts to approximately 3,500 potential patients in the United States.[8]

Many biomarkers associated with precision medicines are genomic in nature. The FDA defines a genomic biomarker as "a measurable DNA and/or RNA characteristic that is an indicator of normal biologic processes, pathogenic processes, and/or response to therapeutic or other interventions" and can be a measurement of the expression, function, or regulation of a gene (FDA, 2008). In recent years, there have been large-scale public gene sequencing efforts – e.g. the NIH's funding of *The Cancer Genome* Atlas.[9] At the same time, a host of new genomics companies have sprung up, providing genetic sequencing technologies, including both software and hardware. An early 2017 report found that companies in genomics and sequencing raised more money in 2016 than any other category of digital health companies (Rock Health, 2017).

[7] https://www.fda.gov/Drugs/NewsEvents/ucm424545.htm

[8] Since Kalydeco® (ivacaftor) was initially approved in 2012 for patients with the G551D mutation, the FDA has subsequently approved its use for patients with any 1 of 38 mutations. According to the Cystic Fibrosis Foundation, recent approvals in May 2017 and August 2017 added an estimated 900 and 600 patients in the US to the estimated 2,000 who were already eligible for treatment with ivacaftor. (https://www.cff.org/News)

[9] https://cancergenome.nih.gov

In response to the growing market and the scientific and regulatory knowledge needed to commercialize such technologies, public funding organizations and regulators have joined forces to harmonize language around biomarkers: in 2015, the joint leadership council of the FDA and NIH identified "the harmonization of terms used in translational science and medical product development…with a focus on terms related to study endpoints and biomarkers" as a priority need. One product of this effort was the publication of the "BEST ("Biomarkers, EndpointS, and other Tools) Resource" in December of 2016 (FDA and NIH, 2016). Appendix A lists the biomarker definitions established to-date by the FDA-NIH Biomarker Working group.

Yet these broad discussions about biomarkers often fail to differentiate among a diverse set of biomarker applications, each of which has different economic implications. Biomarkers can reveal useful information about disease diagnosis and prognosis, predict the treatment efficacy or toxicity of a medication, serve as markers of disease progression, and often serve as auxiliary (or so-called "surrogate") endpoints in clinical trials. Complicating matters further, some biomarkers can be used in more than one way, while others have just one known role.[10] While all of these applications of biomarkers have the potential to shape the practice of (more) personalized medicine and improve drug development and clinical practice, only a small subset has the potential to assist in the development of precision *medicines*, those therapies targeted at specific patient populations who are more likely to benefit. It is the latter group of biomarkers – and the clinical trials driven by their use – that we specifically consider here.

A key opportunity in precision medicine is therapeutic innovation. As we improve our understanding of the genetic and cellular basis of disease, it will be possible to use genetic and protein biomarkers to classify patients into increasingly more specific subtypes where specific medicines will be more effective. In addition, biomarkers that can serve as surrogate endpoints can lead to faster clinical trials, which may influence decisions about whether to pursue treatments for specific diseases (Budish, Roin, and Williams 2015). However, the development of drugs that rely on biomarkers can also introduce challenges to the traditional clinical trial process, such as increased difficulty in trial recruitment due to smaller target patient populations. Additionally, trial design and execution can be significantly more complex when a companion diagnostic (used to identify the biomarker) needs to be approved alongside the drug (Fridlyand, et.al 2013).

[10] Biomarkers come in many types (genomic, proteomic, cellular, biochemical, structural, etc.) and can take on a range of roles (uses) in both drug development and clinical practice. These are explained below and listed in Tables 2 and 3.

Regardless of the specific application, an increase in the use of biomarkers has the potential to markedly change the development and approval process for pharmaceutical innovation.

The drug development pipeline

To describe the drug development pipeline for precision medicines, we characterize all phases of development-oriented clinical trials for new drug candidates over the past twenty-two years. Clinical trials oriented towards drug development typically consist of three main phases, which commence following a manufacturer's successful completion of preclinical studies and submission of an Investigational New Drug (IND) application. Phase I is primarily designed to assess product safety and appropriate dosage. Phase I trials run for several months and typically include 20-100 healthy volunteers or individuals with the target disease. Phase II trials are much larger, enrolling up to several hundred individuals with the target disease and typically lasting between several months to two years. Phase II trials are intended to study drug efficacy and side effects. Phase III trials – usually the final stage of pre-market clinical research – are the largest, enrolling anywhere from a few hundred to a few thousand individuals with the target disease. These trials are designed to study clinical efficacy and to monitor and collect data on adverse reactions to new drugs. Sometimes also referred to as "pivotal studies," Phase III trials typically take 1-4 years to run – but can take far longer (or shorter) depending on the normal progression of the disease studied.[11] Once Phase III results are available, manufacturers must submit a New Drug Application (NDA) or Biologics License Application (BLA) to the FDA that includes the full set of results from preclinical and clinical studies. The FDA then has up to 10 months to review the application and determine whether to grant marketing approval.[12]

[11] https://www.fda.gov/ForPatients/Approvals/Drugs/ucm405622.htm

[12] In recent decades, the FDA has introduced several expedited approval programs for drugs intended to treat serious conditions. "Fast Track" designation allows for frequent meetings with an FDA review team and is for drugs for which there is evidence of addressing an unmet medical need or treating an infectious disease. "Breakthrough Therapy" is for drugs that have preliminary clinical evidence indicating substantial improvement over available therapies and guarantees intensive guidance from the FDA as early as Phase I while also providing several opportunities for expedited and rolling review of results. The "Accelerated Approval" pathway is for drugs that demonstrate an effect on a surrogate endpoint that is reasonably likely to predict clinical benefit and provides the potential for approval based on that surrogate endpoint or an intermediate clinical endpoint. "Priority Review" requires the FDA to review marketing applications within 6 months rather than 10 and is available in a number of circumstances. https://www.fda.gov/downloads/Drugs/Guidances/UCM358301.pdf

The role of major pharmaceutical R&D actors

Clinical trials can be funded by private companies – both small privately-financed and large publicly-listed organizations – as well as by universities/academic medical centers, and by public actors such as the NIH. The latter has historically been more focused on early-stage research with a particular focus on basic science[13] (therefore, to the extent NIH-funded studies lead to drug development projects, one would expect NIH support to be more likely to appear in the context of earlier-stage clinical trials). This focus stems from the economic role of the NIH as not only the world's largest funder of biomedical research (with nearly $32.3 billion invested in 2016[14]), but also a provider of public goods in the form of investments in basic research.[15]

How might we expect patterns of NIH investment to differ among LPM trials? First, markets for precision medicines may be smaller (because the biomarker segments the patient population) and thus less attractive to private actors. At the same time, however, rare diseases are known to have strong lobbies: Hegde and Sampat (2015) find that approximately 3-15% of NIH grants for rare diseases are influenced by politics, suggesting that lobbying plays a role in the allocation of public resources. It is therefore possible that there could be relatively more NIH funding of later-stage precision medicines trials in response to disease group lobbies or in order to address shortfalls in private investment in these diseases. Second, LPM trials may be more innovative and closer to the frontier of biomedical research, a fact that should increase their likelihood of being supported by a competitive research grant. On the other hand, in many cases, these trials are sponsored by for-profit companies looking to commercialize targeted therapies, which can potentially be sold at higher prices, making even small markets more financially attractive (Stern, Alexander, and Chandra, 2017). In this case, private market interest in R&D projects for LPMs may amplify any additional propensity for such projects to receive NIH funding. The likelihood of observing public funding in LPM trials relative to other clinical trials (conditional on phase and drug indication) is therefore an empirical question. We study the role played by each of these actors in the development of LPMs and how these roles have changed over recent decades.

[13] https://nexus.od.nih.gov/all/2016/03/25/nihs-commitment-to-basic-science/
[14] https://www.hhs.gov/about/budget/budget-in-brief/nih/index.html
[15] The stated mission of the NIH is "to seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability."

3. THE ECONOMICS OF PRECISION MEDICINE

As noted above, not all biomarkers imply precision medicines. Here, we outline the economics of precision medicine to better understand how and why biomarkers are important for understanding the potential future of the pharmaceutical market. We argue that biomarkers that provide surrogate endpoints help manufacturers by speeding up clinical trials – e.g. through the use of the FDA's accelerated approval process.[16] This increase in the speed of clinical trials may provide the incentive for pharmaceutical manufacturers to target drugs for different conditions, thus potentially bringing new innovation to the market (Budish, Roin, and Williams, 2015). Conditional on approval however, such drugs may be priced lower because the evidence base for their approval was less certain.[17] At a broad level, the effect of the types of biomarkers that can be used as surrogate trial endpoints is limited to changes in the length of the drug development process (via the ability to run shorter clinical trials).[18]

In contrast, biomarkers that predict treatment benefit (by defining the subset of patients who are most appropriate for therapy) can have far reaching consequences. These include the ability to run faster trials because a therapeutic effect is easier to detect as a result of the greater putative efficacy in the indicated population, but also have a tendency to change expected market sizes. Further, as we have noted elsewhere, such biomarkers could facilitate indication-based pricing, which could expand access to patients, but also mean that higher prices will be charged for patients who have a biomarker that indicates the drug will be most effective (Chandra and Garthwaite, 2017). The broad contours of this type of price-discrimination are illustrated through a fictional example presented in Appendix B.[19]

Biomarkers can facilitate a drug market being segmented into identifiable groups based on the expected efficacy of the product – and thus the willingness to pay for the product. In a setting where pharmaceutical manufacturers are able to charge only a single price, these subgroups allow firms to effectively choose which patients to serve. For example, where the population receiving the least value is quite large, the

[16] https://www.fda.gov/drugs/resourcesforyou/healthprofessionals/ucm313768.htm

[17] This may be particularly true, for example, in cases where precision medicines are approved based on limited data and/or surrogate endpoints. Additional evidence substantiating their benefit on actual patient outcomes is likely to be required before clinicians and health organizations adopt these medications and reimbursement levels are determined (Dzau and Ginsburg 2016).

[18] For a detailed discussion of how the use of surrogate endpoints impacts drug development incentives, see Budish, Roin, and Williams (2015).

[19] This figure depicts the monetary value of a hypothetical product for three different indications (for example, patient populations defined by the presence of biomarkers), the size of the patient populations affected by each indication, and the prices charged for the product under different pricing regimes.

manufacturer can set a low price and sell to a larger market. However, when the low-value population is quite small, the manufacturer can choose a higher price and forgo sales to lower-value patients.

Economists will note that this represents the classic monopolist's dilemma, where pharmaceutical firms must trade margin for quantity. For this reason, firms attempt to find ways to sell the same product to different customers based on their valuation – a strategy known as price discrimination. If firms develop a mechanism for charging indication-based prices, the existence of well-established, readily identifiable biomarkers will become an important tool for facilitation price discrimination. When this is feasible, the most extreme outcome is that the manufacturer is able to capture all of the surplus as profits. Depending on the distribution of patients, this could (but need not) expand access to lower-value indications. However, an indication-based pricing strategy weakly increases the profits of firms developing precision medicines. As a result, the expanded use of biomarkers has the potential to provide additional incentives to develop products that otherwise would not be commercialized.

Pricing aside, biomarkers that predict treatment efficacy reduce market size, which in turn, could reduce some of the incentives for innovation. That said, some biomarkers could allow manufacturers to more easily qualify for "orphan drug" designation through the Orphan Drug Act of 1983 (ODA) by carving out an indication that affects fewer than 200,000 patients. If a medicine receives FDA approval for a new drug (a "new molecular entity") that treats an orphan condition, it receives tax credits equaling 50% of clinical trials expenses and seven years of marketing exclusivity (two years longer than non-orphan drugs). These incentives have been shown to be powerful: more than 516 medicines for over 450 different rare diseases have been approved through the ODA[20] and in 2015 alone, 47% of novel drugs approved were orphan drugs.[21] When such an approval happens, it will also raise prices because of the (extended) protections from generic competition offered by the Orphan Drug Act, and the fact that smaller markets will attract less follow-on competition. Further, in small markets, brand-brand competition will likely be far less robust than in large markets as potential entrants see little expected benefit in competing for a smaller market. To some extent this phenomenon has already been observed in early biosimilar competition in the European Union (Scott Morton, et al., 2017; Berndt and Trusheim, 2015). Thus, even after

[20] https://www.accessdata.fda.gov/scripts/opdlisting/oopd/index.cfm
[21] http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/ucm474696.htm

exclusivity periods end, there may not be a substantial enough market to stimulate price competition through generic entry.[22] As a result, a major shift in the innovative market towards precision medicine could result in less price competition through a meaningful decline in the attractiveness of the generic/biosimilar drug market and, as a result, a meaningful increase in total drug spending.

Finally, the complexity of developing products in this space combined with the use of new and emerging technologies may result in greater specialization for the drug development process. This could involve a greater share of products beginning their lifecycle at small research-focused firms than would be true in more traditional segments of the pharmaceutical industry.

4. DATA

We use data from the Cortellis Competitive Intelligence Clinical Trials Database (Cortellis), which is compiled by Clarivate (and formerly by Thompson Reuters). The database includes over 270,000 global and US-based clinical trials. Cortellis includes full coverage of 24 clinical trial registries from around the world, including clinicaltrials.gov, which is maintained by the National Institutes of Health (NIH), and European Clinical Trials Database (EudraCT), which is maintained by the European Medicines Agency (EMA). Biomedical researchers are strongly encouraged to register trials for publication in medical journals and, as of 2005, trials must be registered to an approved public clinical trial registry prior to patient enrollment in order to be considered for publication in any International Committee of Medical Journal Editors (ICMJE) member journals (De Angelis, Catherine, et al., 2004).

Because both publication and registration are integral parts of the new drug development process, the set of registered trials included in Cortellis should capture all relevant development efforts – in particular, in the years since 2005, after which time the International Committee of Medical Journal Editors required trial registration in order to publish the results of clinical trials in member journals.[23] Cortellis has full

[22] Competition in follow-on drug markets has been discussed by a number of researchers (e.g. Scott Morton, 1999) and in recent years by Berndt, Conti, and Murphy (2017), Scott Morton, Stern, and Stern (2017), and others. More generally, larger markets attract more entrants while smaller markets have been shown to attract less competition, all else equal (DuBois et al., 2015; Acemoglu and Linn, 2004).

[23] We believe that coverage of registered trials is comprehensive and we further expect a high share of trials to be registered in the post-2004 period (De Angelis, Catherine, et al., 2004). However, we note that certain types of trials – e.g. smaller trials without regulatory oversight may still be missing in our data. Kao (2017), describes these types of trials and how they may be designed to signal "off-label usability" to physicians. While an understanding of these types of unregistered trials

coverage of all ICMJE approved trial registries (Clinical Trial Registration, 2016) and Cortellis data have been used in several published studies in peer reviewed biomedical journals such as Lancet Infectious Disease (Phyo, Aung Pyae, et al., 2016) and Nature Reviews Drug Discovery (Bespalov, Anton, et al., 2016). Appendix C includes a detailed timeline of important dates related to the registration of clinical trials and the establishment of the U.S. registry *clinicaltrials.gov*.

Data composition and summary statistics

We queried the Cortellis database for all clinical trials with a launch date between January 1, 1995 and December 31, 2016 for a total of 22 calendar years of clinical trial starts. We identify the full set of phase I, II, and III[24] clinical trials, along with detailed clinical trial information associated with each trial. A few facts are notable: first, the total number of registered trials worldwide has grown over time for each phase of clinical research (Figure 1) and in particular for phase II trials.[25] In 2016, roughly six thousand phase II trials were launched globally, nearly double the number of registered trials launched a decade earlier, in 2006.

For each trial, the Cortellis database also provides information on the trial's relevant drug indication(s), any biomarkers used in the trial, and the trial's sponsor(s). In addition, we are able to classify trials according to broad set of descriptive categories – in particular, the presence (or absence) of one or more biomarker(s) used in the trial. For each biomarker, we are separately able to consider data on its type and use (role). A complete list of the descriptive variables we consider and their frequencies in the clinical trials data set are provided in Table 1.

is important for understanding pharmaceutical firm strategy, we do not believe they are likely to be the types of trials that we attempt to identify in this study, which are those specifically intended to commercialize targeted therapies.

[24] For the simplest classification of trails into phases, we assign combined trials (e.g. combined Phase II & Phase III) to the lower of the two phases involved. For example, a combined Phase II/ Phase III trial would be classified as having started Phase II in the year that the trial launched. In robustness tests, we create separate sub-categories for combined Phase I/II and II/III trials and include controls for these combined trials in regression analyses. Subsequent regression results are not sensitive to this distinction, so we use the simplified 3-phase classification in tables and figures for simplicity.

[25] The recent spike in the number of global clinical trials (and Phase II trials in particular) is driven by growth in non-US trails (see Appendix tables for a version of Figure 1 that presents only U.S. trials).

To aggregate the detailed indications reported in the Cortellis database to more usable categories, we used a dataset[26] of indications matched to ICD-9 codes to link each trial in our dataset to a 3-digit ICD-9 code. The matched indication-ICD-9 dataset was independently checked for accuracy by three research assistants using an online ICD-9 medical coding reference manual,[27] and any discrepancies between their matches were resolved by a fourth research assistant. Each indication was ultimately assigned to one ICD-9 code, corresponding to a total of 65 ICD-9 sub-chapters (listed in Appendix D). Trials with any indications matching ICD-9 codes 140-239 were classified as cancer trials.

We also capture key information about the clinical trial's sponsor(s). Trial sponsors are identified by name and type, including academic investigators, government, non-government, company, and other sponsors. Importantly, the database also lists associated clinical trial registry identifiers such as unique trial registration numbers from clinicaltrials.gov as well as NIH grant numbers that supported the research. We use these to identify whether a trial benefited from *any* acknowledged NIH funding (regardless of the sponsor's identity). NIH funded trials can be segmented by type of NIH funding using the activity code embedded in the NIH project number(s) listed. Appendix E describes how NIH project types are identified.

The categorical variable "biomarker type" indicates the biological feature that a given biomarker identifies. Biomarker types include genomic biomarkers, proteomic biomarkers, biochemical biomarkers, cellular biomarkers, physiological biomarkers, structural biomarkers, and anthropomorphic biomarkers. Definitions of biomarker types and their frequencies of use in clinical trials both a) overall and b) over time are reported in Table 2. Importantly, these types are not mutually exclusive, since a given biomarker – e.g. a receptor such as EGFR (epidermal growth factor receptor) – can be both a genomic and proteomic biomarker. This is because genomic characteristics will lead to differential expression of EGFR – making it a biomarker of particular genomic features –but EGFR is *itself* a protein and therefore a proteomic biomarker as well. For this reason, there can be correlation in the frequencies of biomarkers types across trials.

[26] We are grateful to Manuel Hermosilla, Craig Garthwaite, and David Dranove, who generously shared their version of a 3-digit ICD-9 crosswalk dataset with us. This dataset was assembled through the use of two independent medical coders separately constructing a crosswalk. Discrepancies were adjudicated by a third expert and additional outside research.
[27] ICD9Data.com

Biomarker data and defining pipeline precision medicines

Cortellis includes fairly broad categories of biomarker uses as they may relate to clinical trials. These include disease markers, toxic effect markers, and therapeutic effect markers. Disease-related biomarkers indicate if a disease already exists (diagnostic biomarker), or how such a disease may develop in an individual case regardless of the type of treatment (prognostic biomarker). Therapeutic effect-related biomarkers provide an indication of the progress of a product on the patient during treatment. Toxic effect-related biomarkers indicate a treatment-related adverse reaction. Other biomarker roles are "not determined" because they do not have any of the roles described above in a particular trial. In practice, we are interested in a *subset* of the trials that use disease-related biomarkers – namely, those in which we observe the unambiguous features of products that would likely come to market as targeted therapeutics upon successful progression through the R&D process. This is because this subset of biomarkers facilitates *ad hoc* patient selection for therapy.

Our working definition of likely precision medicines (LPMs) is that they encompass the set of pipeline products that are being developed using the types of diseases-related biomarkers that are relevant for identifying subpopulations that are likely to be more (or less) responsive to medications. We therefore employ a second, biomarker-specific database from Clarivate in order to link biomarkers to their *detailed* roles in clinical trials. The detailed biomarkers data (DBD) from Clarivate include additional detail (in the form of "detailed biomarker roles") on all known clinical biomarkers and their paired uses and indications in clinical research. For example, human epidermal growth factor receptor 2 (HER2) is a (genomic) biomarker that can be used for (a) selection for therapy and (b) predicting treatment efficacy – both of which are detailed biomarker roles – among patients with breast cancer (the indication). Based on using a trial's "breast cancer" indication and knowing that the HER2 biomarker was used in that clinical trial, one can probabilistically assign both a biomarker type and a detailed biomarker role (or, in some cases, more than one) to that trial. Generally, in order to assign biomarker types and biomarker roles to our full set of clinical trials, we match the named biomarker(s) associated with each trial (when there are any) and the indication(s) of that trial to the DBD.

Definitions of detailed biomarker roles and the frequencies of their use in clinical trials are reported in Table 3. A biomarker may have multiple associated uses, making it important to correctly link a biomarker

associated with a given clinical trial and indication to its use *in that setting*. Therefore, the process of matching a biomarker-indication pair from the Cortellis clinical trials data with a biomarker-indication pair from the DBD is a crucial step in correctly assigning biomarker roles to individual clinical trials. We define LPMs in two ways using these detailed biomarker roles. These classifications are consistent with the NIH-FDA definitions of biomarkers and how they are employed and were separately discussed with an oncologist, who is the principal investigator on a biomarker-driven clinical trial.

In the first, "generous" definition of LPMs, we identify trials using biomarkers whose roles include diagnosis, differential diagnosis, predicting drug resistance, predicting treatment efficacy, predicting treatment toxicity, screening, and selection for therapy. The rationale for the generous definition is that all of these biomarkers can be used in the development of targeted therapeutics and are likely to be associated with the development of precision medicines.

In the second, "restrictive" definition of LPMs, we identify the subset of trials from the "generously" defined group that specifically employ biomarkers for prediction (predicting drug resistance, predicting treatment efficacy, and predicting treatment toxicity), with the vast majority of these trials identified as LPM trials due to the use of biomarkers that can help predict treatment efficacy (Table 3). We consider each in turn and further consider the interaction of these roles with specific biomarker *types* (genomic and proteomic) that are more likely to be used in trials for precision medicines.

5. CHARACTERIZING THE LPM DEVELOPMENT PIPELINE

We characterize the number and type of drugs using biomarkers in their clinical trials as well as those that can be considered LPMs by therapeutic area and over time. Since cancers represent a set of diseases in which precision therapies are already successfully used, and since cancer applications of precision medicine are expected to grow in coming years, we separately characterize the cancer applications of pipeline precision medicines in detail.

Biomarkers and LPMs in clinical trials

We begin at perhaps the broadest point, by first identifying all trials that use one or more biomarker(s) of any kind (Figure 2). Notably, both the share and total number of clinical trials employing biomarkers

has increased markedly over recent decades. We next focus only on the subset of biomarker uses that are associated with LPMs, by both the generous and restrictive definitions (Figure 3). Both the number and percentage of LPM trials increased over our period of observation, as seen in Figure 3. We further note that the two definitions of LPMs track each other closely over time – both in Figure 3 as well as in the subsequent sub-sample analyses described below. Table 4 presents the count (column 1) and percentage (column 2) of LPMs in clinical trials in each year of our data. Columns 3-8 present the same results by clinical trial phase. Even by the most restrictive definition of LPM trials, by 2016, approximately 7.5% of trials were for LPMs, roughly double the percentage observed a decade earlier (3.8%).

LPM trials are associated with the use of different types of biomarkers and the relative and absolute frequencies of these types have evolved over time. Biomarker types are not mutually exclusive; for example, there is extremely high overlap between proteomic and genomic biomarkers, since the vast majority of genomic mutations (e.g. in cancer) manifest themselves through differences in protein expression. Figure 4 shows how these types were represented in each phase (by both generous and conservative definitions of LPMs), over our years of observation. Genomic/proteomic biomarkers were the most commonly used in recent years, featured in the vast majority of LPM trials, a statistic that is consistent with LPMs being driven primarily by understanding gene and protein expression and how these factors predict the likely success of medications.

Pipeline precision cancer therapies

Figure 5 and Table 5 present data on the frequency of LPMs in cancer trials only. Several features of these trials are notable – especially in comparison. First, LPM trials are more than an order of magnitude more common in cancer indications: in 2015 and 2016, roughly 25% (or more) of all cancer drug trials were LPM trials, but only 1-2% of trials for non-cancer indications were LPM trials. In regression analysis (Table 8), we also see that a cancer indication is a strong statistical predictor of a LPM trial and the growth of LPMs among cancer drugs explains the lion's share of growth in LPM trials over the past two decades. These results are completely consistent with the clinical belief that the majority of applications of precision medicines in coming years will be in the context of targeted therapies for cancer.

Institutional factors

Next, we consider the LPM development pipeline in light of a number of specific institutional factors. We consider US-based vs. non-US-based trials. The United States is, by far, the world's largest pharmaceutical consumer (International Trade Administration, 2016) and it would therefore be reasonable to expect trials for LPMs to be driven to by both local demand (Costinot et al., 2016) as well as local regulations (FDA, 2004). Figure 6 shows the number and share of U.S. LPM trials. The total number of LPM trials conducted within the United States is comparable to the total number conducted abroad, but the share of LPM trials among U.S. trials is roughly double that of international trials. This finding is consistent with the fact that U.S. drug prices are typically higher than those of other countries (Kanavos et al., 2013), making it more appealing for pharmaceutical manufacturers to bring drugs to market in the United States as soon as possible. These facts are also reflected in our regression analysis (Table 8) which indicates that U.S. trials are, on average, roughly 1 percentage point more likely to involve LPMs at any point in time than their non-U.S. counterparts in the same year, all else equal.

We next consider LPM trials with vs. without NIH funding. Since NIH grants are concentrated in U.S. research institutions, we focus our analysis of NIH funding on U.S. trials only. The first panel of Figure 7 shows the share of clinical trials by phase that received NIH funding in each year. Although the total value of NIH funding available has grown over time, the total number of registered clinical trials has grown more rapidly, leading to a declining *share* of total U.S. trials with NIH funding. Among LPM trials (both definitions), the share of trials with NIH funding has been relatively constant, albeit somewhat noisily measured. On average, roughly 5-6% of Phase I and Phase II trials (but a lower share of Phase III trials) have received NIH support in recent years (with higher averages, but also higher variances in earlier years of observation; Figure 7). Table 6 presents the absolute shares of all trials – not restricting to LPM trials – receiving NIH funding over our sample. Overall, NIH support is *less* common among LPM trials (of both definitions) relative to overall rates of NIH support of clinical trials in the United States.

In addition, we briefly consider whether LPM trials appear to be related to disease severity.[28] We use the Institute for Health Metrics and Evaluation's *Global Health Data Exchange* to collect data on "global burden of disease" for all cancers.[29] For both the United States (alone) as well as globally, we assemble data on years of life lost (YLL) due to each cancer.[30] For all cancers, we identify the relevant ICD-9 code

[28] We are grateful to NBER conference participants for this suggestion.
[29] These Data are publicly available at http://www.healthdata.org/gbd
[30] We use this measure because it is one of the only metrics that has yearly data dating back to the 1990s.

and can then match YLL to the cancer trials in our data (as described above and in the 11 cancer ICD-9 subchapters presented in Appendix D). Table 7 presents results from two sets of t-tests of differences in means with unequal variances. We find evidence that among cancer trials, LPM trials are associated with significantly more U.S. and global YLL for the product's intended indication than non-LPM trials on average.

Finally, we consider the types of firms – namely publicly listed companies vs. (typically smaller) privately held firms – engaging in the development of LPMs (Figure 8 and Table 8). The correct assignment of individual trials to their sponsor firms (and according firm types) is both difficult and fundamental for our analysis. Because acquisitions are common and firm ownership may change over time, we probabilistically assign each trail in our dataset to the firm that sponsored the trial and its type (e.g. publicly listed vs. privately held) *at the time the trial was launched*. Although we are not able to assign these types with complete accuracy, we are mathematically able to construct upper and lower bounds for whether each sponsor firm was publicly listed at the time of a trial. Aggregating our data, we are able to construct upper and lower bounds for the *share* of publicly listed firms over time and across phases (Figure 8). Appendix F presents details how these bounds were calculated and a short proof of the bounding exercise. Overall, we find that publicly listed firms are significantly more likely to pursue LPM trials, regardless of whether we use the upper or lower bound for the measure for whether or not a firm was public at the time of a given trial.

We conclude with regression analysis (Tables 9-10).[31] We are circumspect in interpreting our regression results: the coefficients calculated are not *causally* estimated; rather they represent differences between categories in our sample, controlling for other factors. However, the coefficients are useful in that they allow for interpretation of multivariate associations. Table 9 presents linear probability models using facets of trials to predict the likelihood that any given trial is an LPM. Table 9a presents these models using the generous definition of LPMs as the binary outcome and Table 9b presents the same set of regression models using the conservative definition as the binary outcome.

Through both panels, a set of statistical relationships emerge. For example, the linear probability models presented in Tables 9a-9b indicate that the total share of LPM trials has been increasing over time by

[31] As noted above, in our analyses we assign combined trials (e.g. combined Phase II & Phase III) to the lower of the two phases involved. In robustness tests, we create separate sub-categories for combined Phase I/II and II/III trials and include controls for these combined trials in regression analyses. Results are not sensitive to this distinction, so we use the more parsimonious 3-phase classification in tables and figures for simplicity.

between 0.3 and 0.5 percentage points per year. Other relationships seen in earlier tables and figures are also apparent. Most prominent among these is the overwhelming importance of cancer trials: trials for cancer indications are 13-15 percentage points more likely to be LPM trials than those for non-cancer indications. Indeed, the coefficient on the binary indicator for whether a trial is a cancer trial is an order of magnitude larger than the association between time, location, trial phase or firm type. Trials with U.S. sites are more likely than non-U.S. trials to be LPM trials, but only by about one percentage point – in other words, comparing this result to the overall time trend in the data, U.S. trials seem to be about 2-3 years "ahead" of non-U.S. trials in their inclusion of LPMs. We also find that the role of publicly listed firms is similar in magnitude and direction: trials pursued by publicly listed firms are 1-2 percentage points more likely to be LPM trials than those of privately held firms, all else equal. With respect to public funding, we find that, on average, NIH-supported trials are more likely to be LPMs. In all specifications, the coefficient on NIH-support is associated with a roughly a 1 percentage point higher probability of being an LPM, however this relationship is only statistically significant in some specifications. This indicates that the (albeit noisy) relationships seen in Figure 7 are not fully robust to controlling for other trial features.

We conclude our regression analysis by briefly considering predictors of trial duration. One implication of precision medicine is that trials themselves can be conducted more efficiently, if effect sizes are expected to be large. Efficiency improvements could occur on the dimension of enrollment (fewer patients required) or on the dimension of trial duration (less time needed to draw statistically sound evidence); we consider only the latter possibility here. Table 10 presents results from a set of linear regression models predicting trial duration. These models include a number of trial features as regressors and present multivariate associations in our dataset. As above, these coefficients cannot be interpreted causally; rather, they represent average associations between salient features of trials and the amount of time required for trial completion.

The first three columns of Table 10 present models predicting trial duration in LPM trials, while the last three columns present identical models in non-LPM trials. A number of interesting relationships emerge. First, we note the difference in the coefficient on the intercept in the two sets of linear models: LPM trials take roughly 20 months longer to complete relative to non-precision trials, all else equal. This may be due to the challenges of enrolling patients with less common sub-types of a disease as well as the fact that non-precision trials include a number of studies (e.g. for antibiotics) that can be run extremely quickly, thereby lowering the average time to completion in the second group of trials. With respect to trial

phases, phase I trials are the omitted category in all models. For LPM trials, phase II trials are only about two months longer than phase I trials, on average, and this differences is only statistically significant in one out of the three specifications. This is quite different than what is observed in non-precision trials, where phase II studies take 5-6 months longer than phase I studies to complete. Among LPM trials, phase III studies have durations over a year longer than their phase I and phase II counterparts, a bigger difference than among non-precision trials, where phase III studies are only 7-9 months longer on average. This suggests, that while LPM trials may be longer on average, precision medicine biomarkers may be able to close the gap between Phase 1 and later phases.

Interestingly, although cancer trials appear to always take longer to complete, on average than non-cancer trials, the additional trial length associated with LPM cancer trials is 6-7 months *less* than the additional trial length associated with non-precision cancer trials in these models. One interpretation of this is that precision medicines speed up cancer trials perhaps because of surrogate endpoints or enrichment. We caution the over-interpretation of this relationship because it does not hold up when examined in further detail: in Appendix Table III we consider the same sets of models for cancer trials alone and show similar patterns across many coefficients in the regression models, but differences in the estimated constants between LPM trials vs. non-precision trails in cancer. Finally, we note that as economic incentives would predict, trials supported by the NIH have longer durations on average (e.g. longer studies may require public support to run) and trials sponsored by publicly listed firms have shorter durations, on average (e.g. such firms are likely under pressure from investors to bring products to market). While none of these facts provide conclusive evidence on the causes of differences in trial length, the associations are intriguing and suggest the value of future research into the determinants of clinical trial length – especially since clinical trials represent a significant component of both the time and financial cost associated with new drug development.

6. CONCLUSION

By taking a detailed view of the global clinical trial pipeline over recent decades, we are able to describe a number of trends and industry-level changes. Beyond growth in the number of registered clinical trials,

we document a number of patterns that that have implications for cost-growth in health care and pharmaceutical pricing. First, we document that the use of biomarkers in clinical trials has grown significantly, with an important subset of those representing the types of biomarkers that have the potential to be used in the development of targeted therapies. Such therapies are likely to be more effective, but will also likely come with higher prices. Although the raw numbers of trials using biomarkers in the development of precision medicines is still dwarfed by the total number of clinical trials, the growth in such trials has been large in percentage terms, approximately doubling every decade over the past 22 years.

Our results should be interpreted with a number of caveats. Firstly, the findings presented here are only as representative as the global registries on which our primary clinical trial dataset is based. While we have noted above that there are good reasons to believe that these registries are highly representative of the set of pipeline drugs pursuing regulatory approvals in the dozen most recent years of our data, some trials may not have been reported in earlier years. In particular, we believe that the data in the years after 2004 are more likely to capture clinical trials, due to changes in trial registration requirements for academic journal publication (discussed above). Unfortunately, we do not have a way of estimating the type and direction of selection into trial registries that may have occurred.

Secondly, we note that our characterization of trials as either LPM or non-LPM trials is, by nature, probabilistic, based on observable features of these trials and the drugs in them. While the categories we use are unambiguously more conservative than simply considering any use of biomarkers in clinical trials, they may still capture some trials and pipeline products that do not, in fact, represent true precision medicines.

Finally, and perhaps most importantly, we have characterized the drug development *pipeline*, which is not necessarily synonymous with characterizing the *actual set of products* that are commercialized. If failure rates in clinical research are endogenously determined with other characteristics related to commercialization strategies (e.g. single-product vs. multi-product firms, as seen in Guedj and Scharfstein, 2004), characterizing trials may not accurately reflect future products. To the extent that there is selection in R&D project discontinuations based on features not included in our analysis, the set of products that ultimately comes to market may look different than the late-stage clinical trial pipeline would suggest.[32]

[32] On average, success rate for a drug entering clinical trials is approximately 10%. This rate is even lower for oncology therapeutics at roughly 5%. (https://www.bio.org/sites/default/files/Clinical%20Development%20Success%20Rates%202006-2015%20-%20BIO,%20Biomedtracker,%20Amplion%202016.pdf)

Yet we believe that we have also made progress in characterizing recent trends and developments in clinical research related to precision medicines. By taking a big-picture view of global clinical trials, we can observe how LPMs have grown in number and share of trials over recent decades. We can also bring empirical data to bear on predictions from medicine and economics, which would suggest that certain types of drugs (e.g. for cancers) and certain markets (e.g. in the United Sates) are likely to have a greater share of LPMs. Within LPMs, we see a large and growing share of products that incorporate genomic and proteomic biomarkers in their development, suggesting the growing importance of sequencing technologies for both R&D and patient care. Further, recent trajectories have implications for health care spending: to the extent that LPMs grow in market share, they will drive up costs for drugs that target specific groups of patients and also open up opportunities for indication-based pricing.

References:

Acemoglu, Daron, and Joshua Linn. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." Quarterly journal of economics 3 (2004): 1049-1090.

Berndt, E.R., Conti, R.M. and Murphy, S.J., 2017. *The Landscape of US Generic Prescription Drug Markets, 2004-2016* (No. w23640). National Bureau of Economic Research.

Bespalov, Anton, et al. "Failed trials for central nervous system disorders do not necessarily invalidate preclinical models and drug targets." *Nature Reviews Drug Discovery* (2016).

Berndt, Ernst R., and Mark R. Trusheim. "Biosimilar and Biobetter Scenarios for the US and Europe: What Should We Expect?." In *Biobetters*, pp. 315-360. Springer New York, 2015.

Budish, Eric, Benjamin N. Roin, and Heidi Williams. "Do firms underinvest in long-term research? Evidence from cancer clinical trials." The American economic review 105.7 (2015): 2044-2085.

Chandra, A. and Garthwaite, C., 2017. The Economics of Indication-Based Drug Pricing. New England Journal of Medicine, 377(2), pp.103-106.

Costinot, A., Donaldson, D., Kyle, M. and Williams, H., 2016. The more we die, the more we sell? a simple test of the home-market effect (No. w22538). National Bureau of Economic Research.

Cohen, Joshua P., and Abigail E. Felix. "Personalized medicine's bottleneck: diagnostic test evidence and reimbursement." Journal of personalized medicine 4.2 (2014): 163-175.

De Angelis, Catherine, et al. "Clinical trial registration: a statement from the International Committee of Medical Journal Editors." *New England Journal of Medicine* 351.12 (2004): 1250-1251.

Dubois, Pierre, Olivier de Mouzon, Fiona Scot Morton, and Paul Seabright. "Market size and pharmaceutical innovation." The RAND Journal of Economics 46, no. 4 (2015): 844-871.

Dzau, V.J., Ginsburg, G.S. Realizing the Full Potential of Precision Medicine in Health and Health Care. JAMA. 2016; 316(16):1659–1660.

FDA, Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research. "Guidance for Industry: Information Program on Clinical Trials for Serious or Life-Threatening Diseases and Conditions." (Draft Guidance) *U.S. Department of Health and Human Services, Food and Drug Administration* (2004).

FDA, Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research. "Guidance for Industry E15 Definitions for Genomic Biomarkers, Pharmacogenomics, Pharmacogenetics,

Genomic Data and Sample Coding Categories." *U.S. Department of Health and Human Services, Food and Drug Administration* (2008).

Fridlyand, J., Simon, R., Walrath, J., Roach, N., Buller, R., Schenkein, D., Flaherty, K., Allen, J., Sigal, E., & Scher, H. 2013, Considerations for the successful co-development of targeted cancer therapies and companion diagnostics, Nature Reviews Drug Discovery, 12, 10, pp. 743-755

Guedj, I., & Scharfstein, D. (2004). *Organizational scope and investment: Evidence from the drug development strategies and performance of biopharmaceutical firms* (No. w10933). National Bureau of Economic Research.

Hegde, Deepak, and Bhaven Sampat. "Can private money buy public science? Disease group lobbying and federal funding for biomedical research." *Management Science* 61, no. 10 (2015): 2281-2298.

International Trade Administration, Department of Commerce. 2016 ITA Pharmaceuticals Top Markets Report." (2016).

Kanavos, Panos, Alessandra Ferrario, Sotiris Vandoros, and Gerard F. Anderson. "Higher US branded drug prices and spending compared to other countries may stem partly from quick uptake of new drugs." Health affairs 32, no. 4 (2013): 753-761.

Kao, Jennifer L. "R&D Decisions for New Medical Technologies: Evidence from New Use Approvals and Off-Label Uses." *Working Paper* (2017).

Phyo, Aung Pyae, et al. "Antimalarial activity of artefenomel (OZ439), a novel synthetic antimalarial endoperoxide, in patients with Plasmodium falciparum and Plasmodium vivax malaria: an open-label phase 2 trial." *The Lancet Infectious Diseases* 16.1 (2016): 61-69.

Rock Health "2016 Year End Funding Report" Available at https://rockhealth.com/reports/2016-year-end-funding-report-a-reality-check-for-digital-health (2017).

Scott Morton, Fiona M. "Entry decisions in the generic pharmaceutical industry." The Rand journal of economics (1999): 421-440.

Scott Morton, Fiona M., Ariel Dora Stern, and Scott Stern. "The impact of the entry of biosimilars: evidence from Europe." Harvard Business School Working Paper No. 16-141, 2016.

Stern, Ariel D., Brian M. Alexander, and Amitabh Chandra, 2017. How economics can shape precision medicines. Science, 355(6330), pp.1131-1133.

Stern, Ariel D., Brian M. Alexander, and Amitabh Chandra. "Innovation Incentives and Biomarkers." Clinical Pharmacology & Therapeutics (2017).

Figures

Figure 1: Clinical trials over time

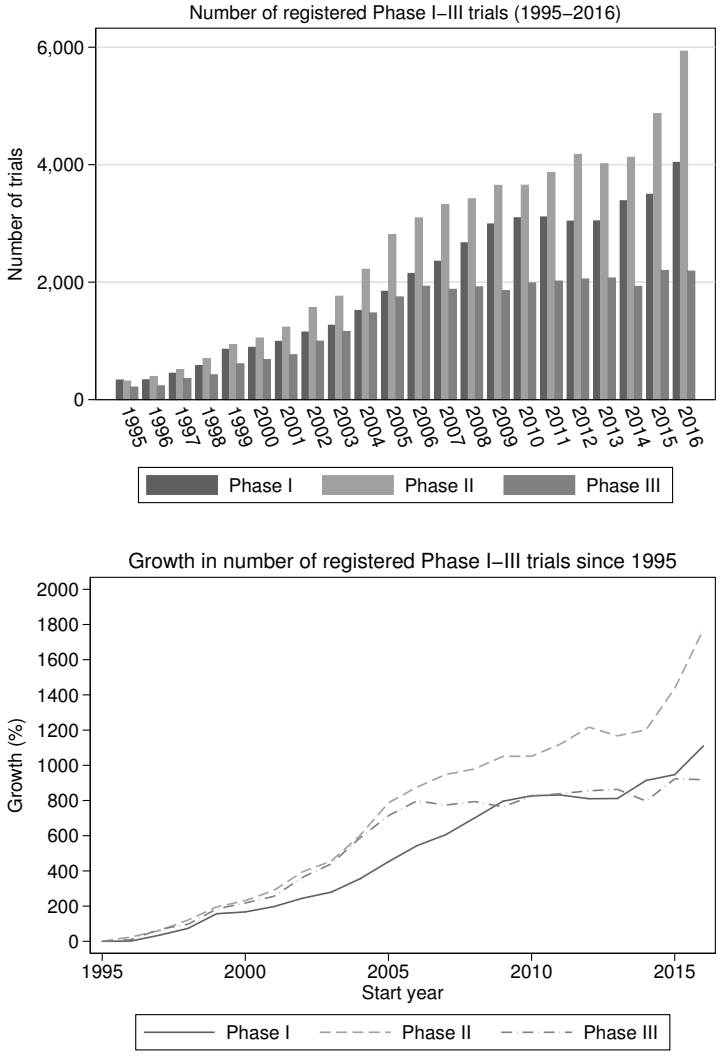


Figure 2: Clinical trials using biomarkers

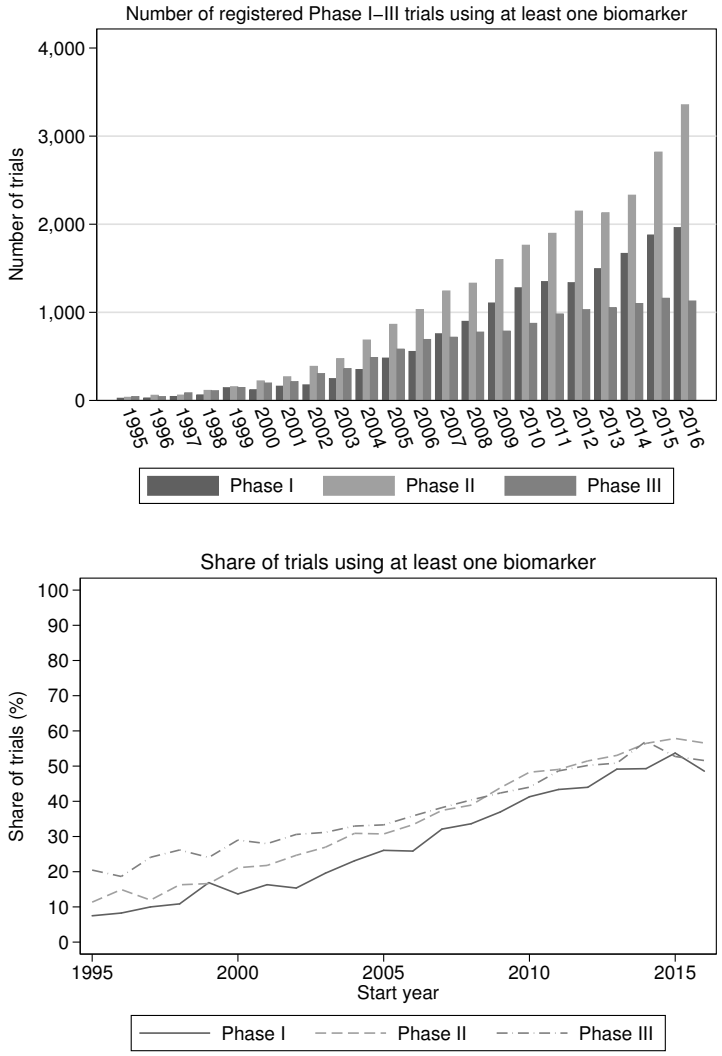


Figure 3: Clinical trials for LPMs

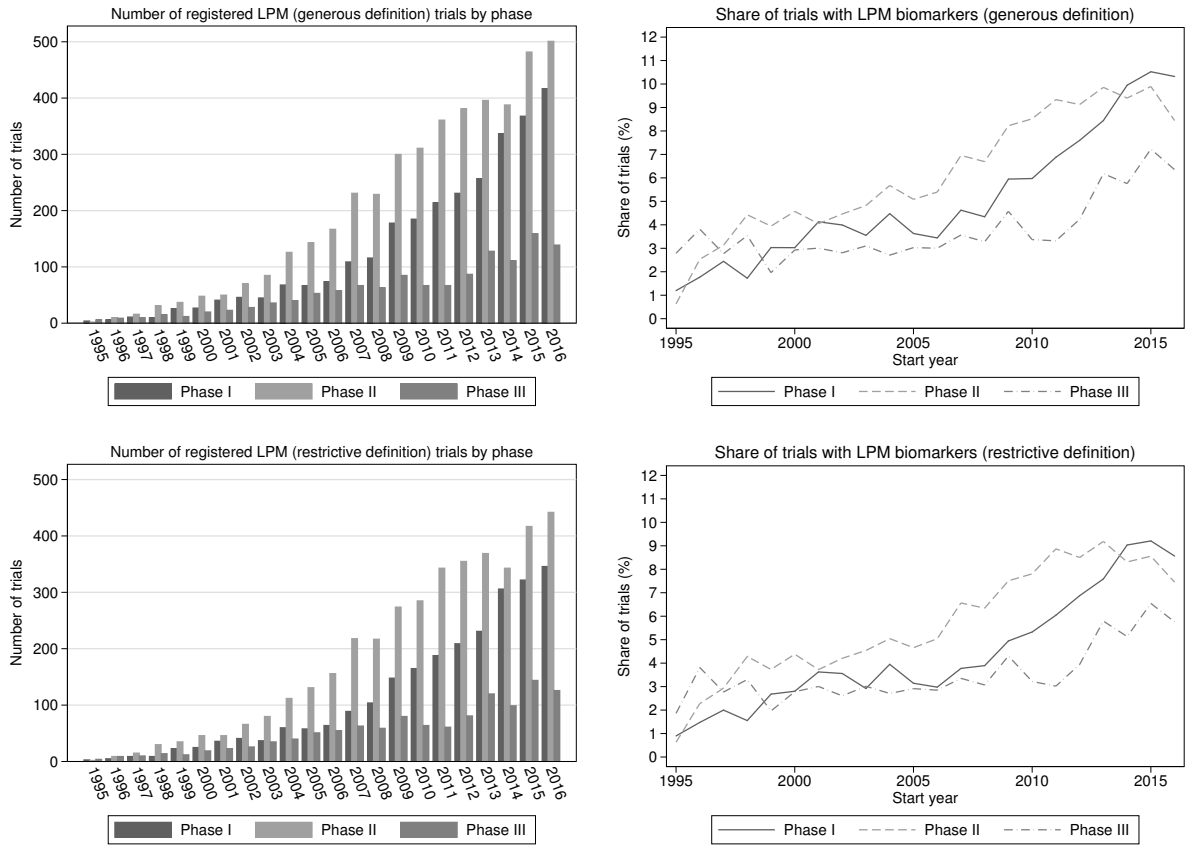


Figure 4: Types of biomarkers used in LPM trials

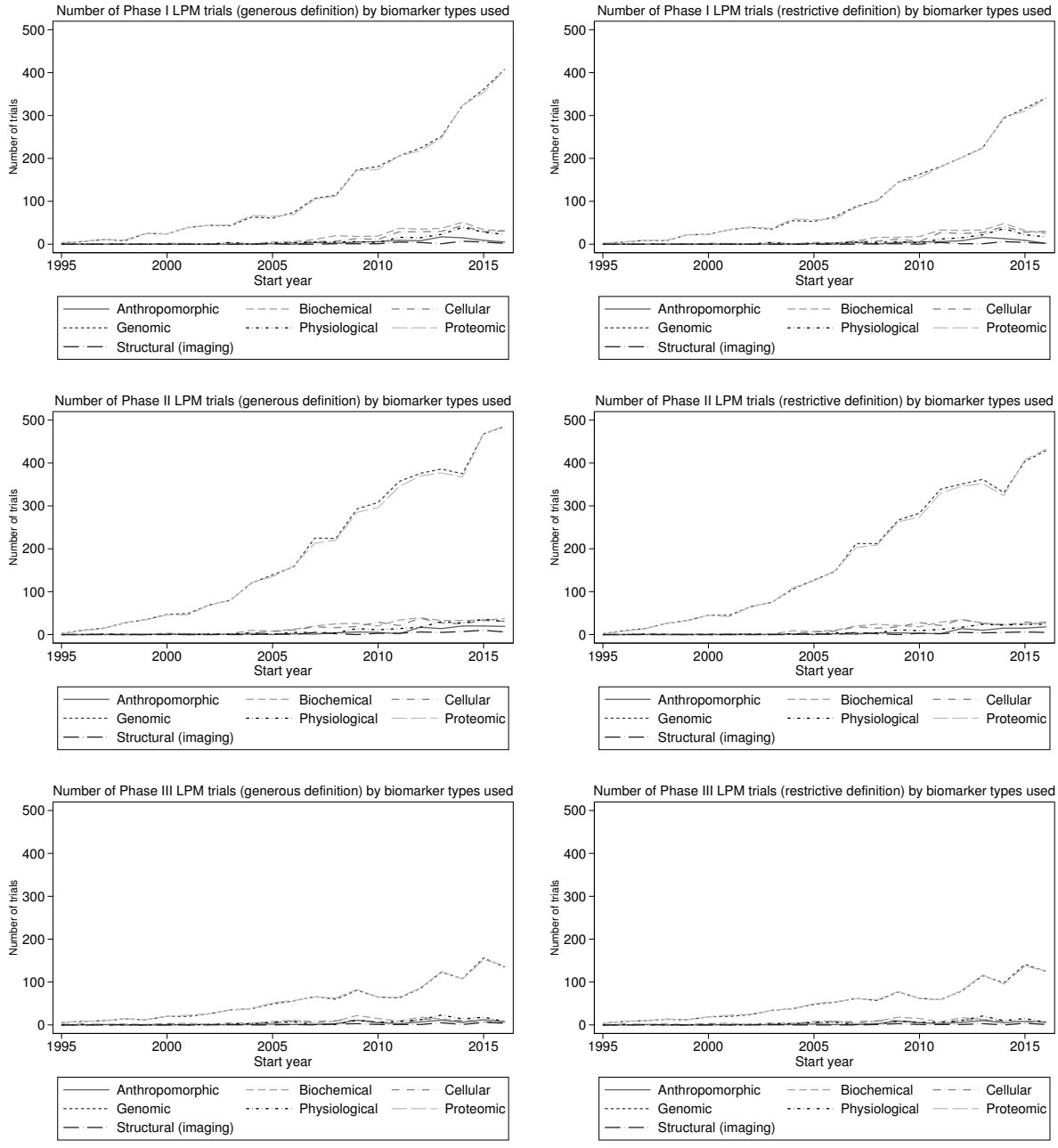


Figure 5: Clinical trials for LPMs, cancer indications only

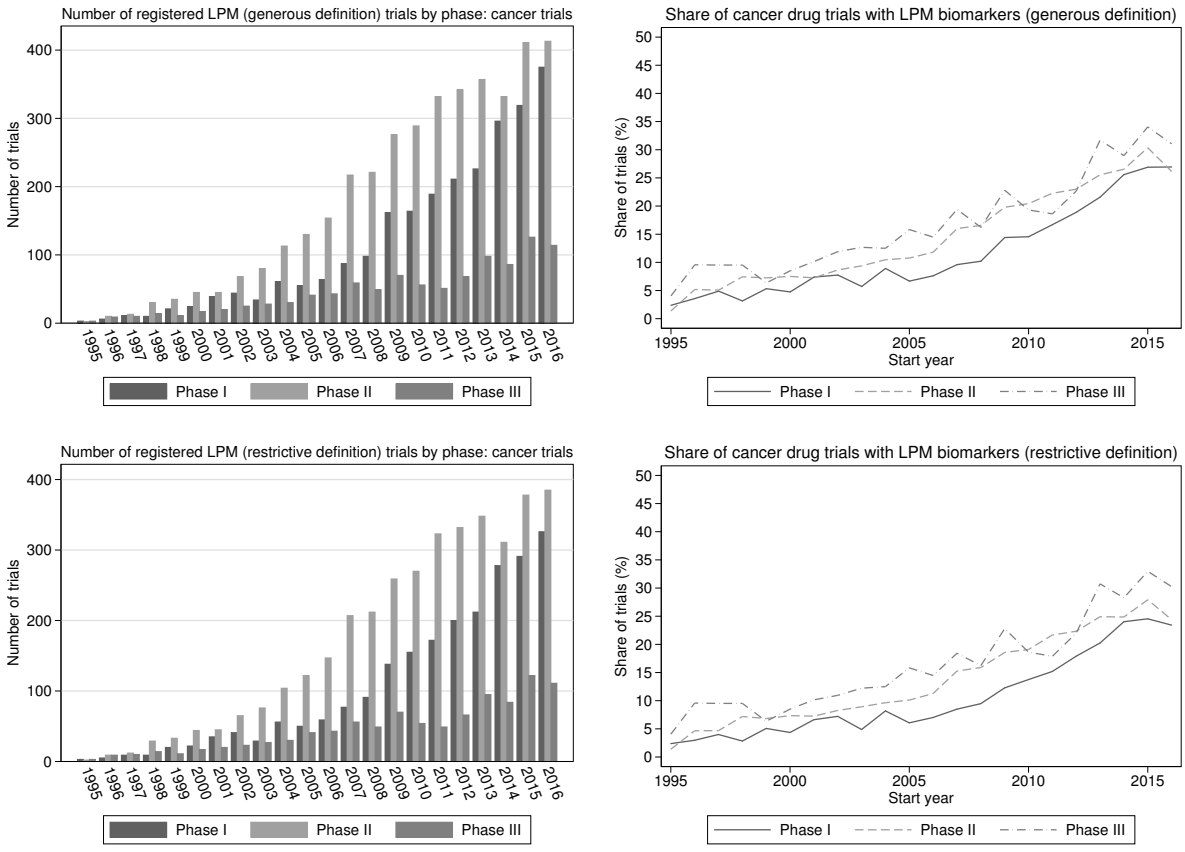


Figure 6: Clinical trials for LPMs, U.S. trials only

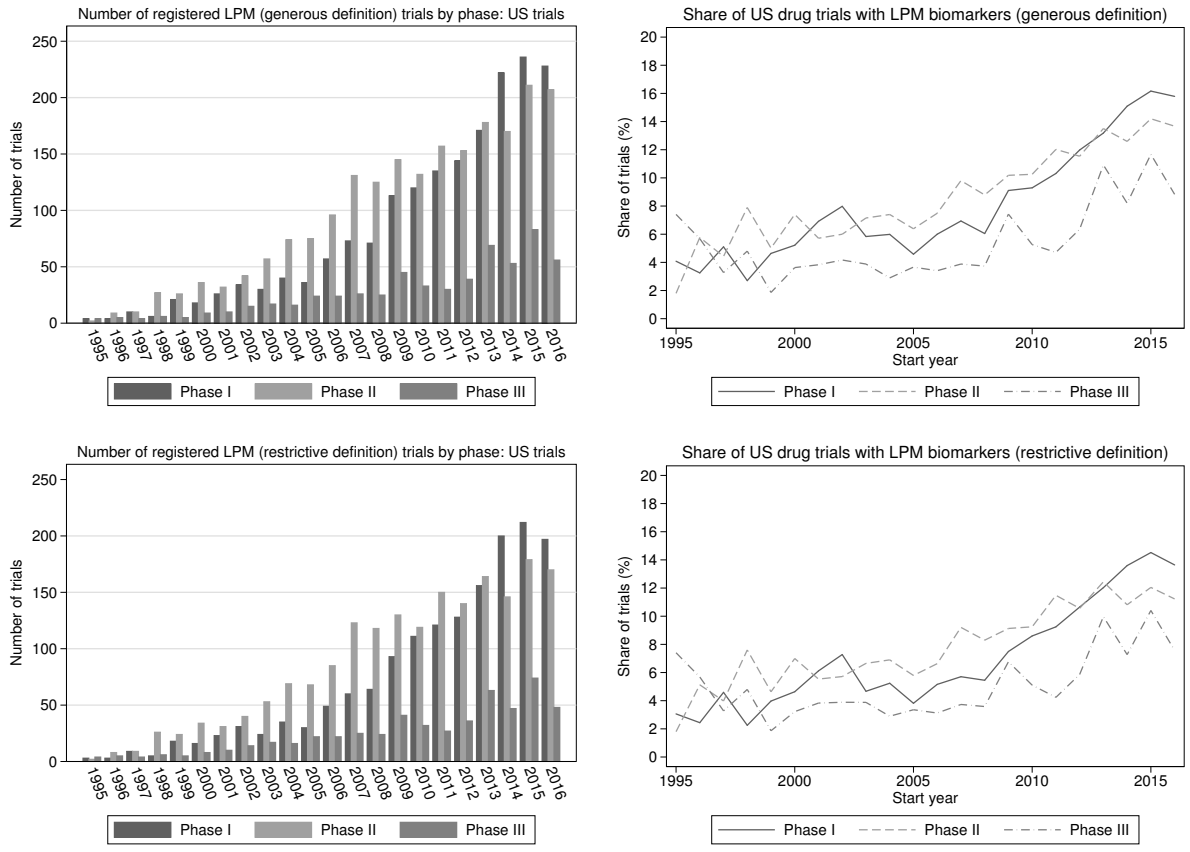


Figure 7: Trials and LPM trials with NIH funding (U.S. trials only)

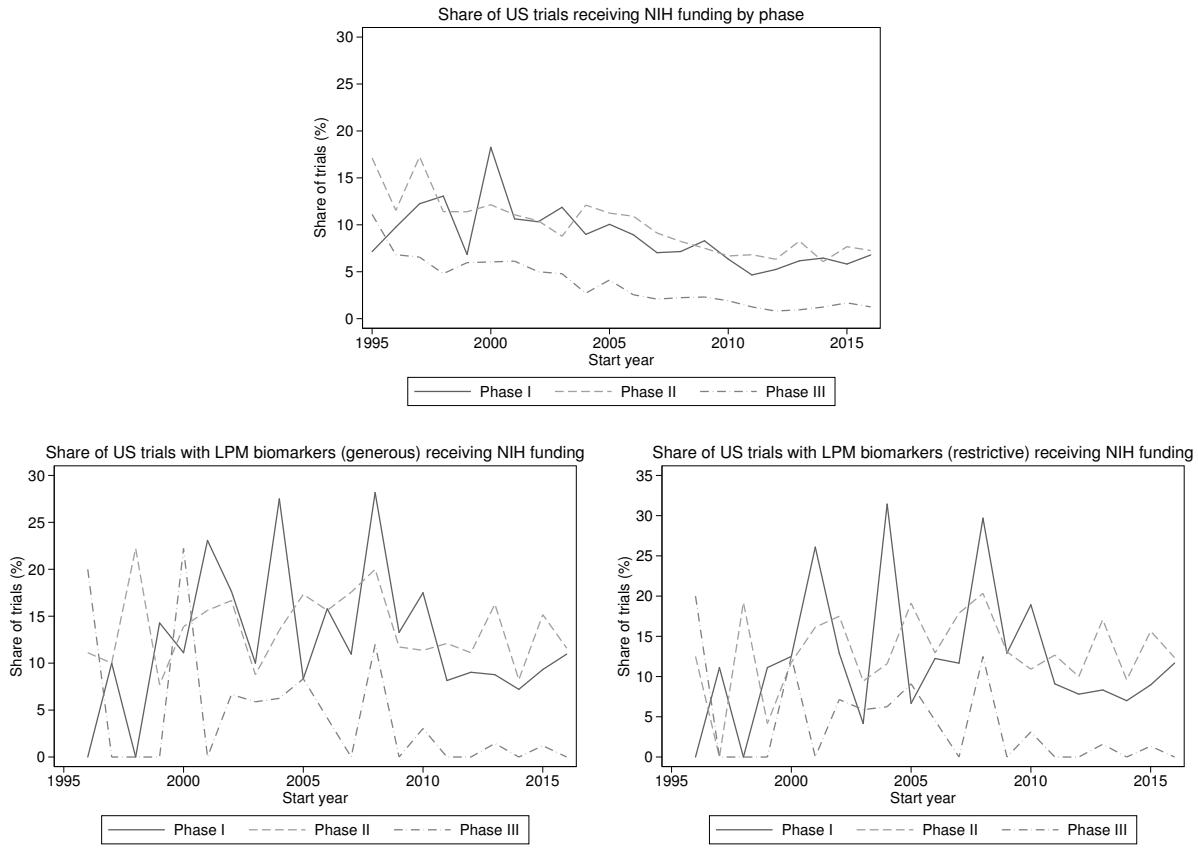
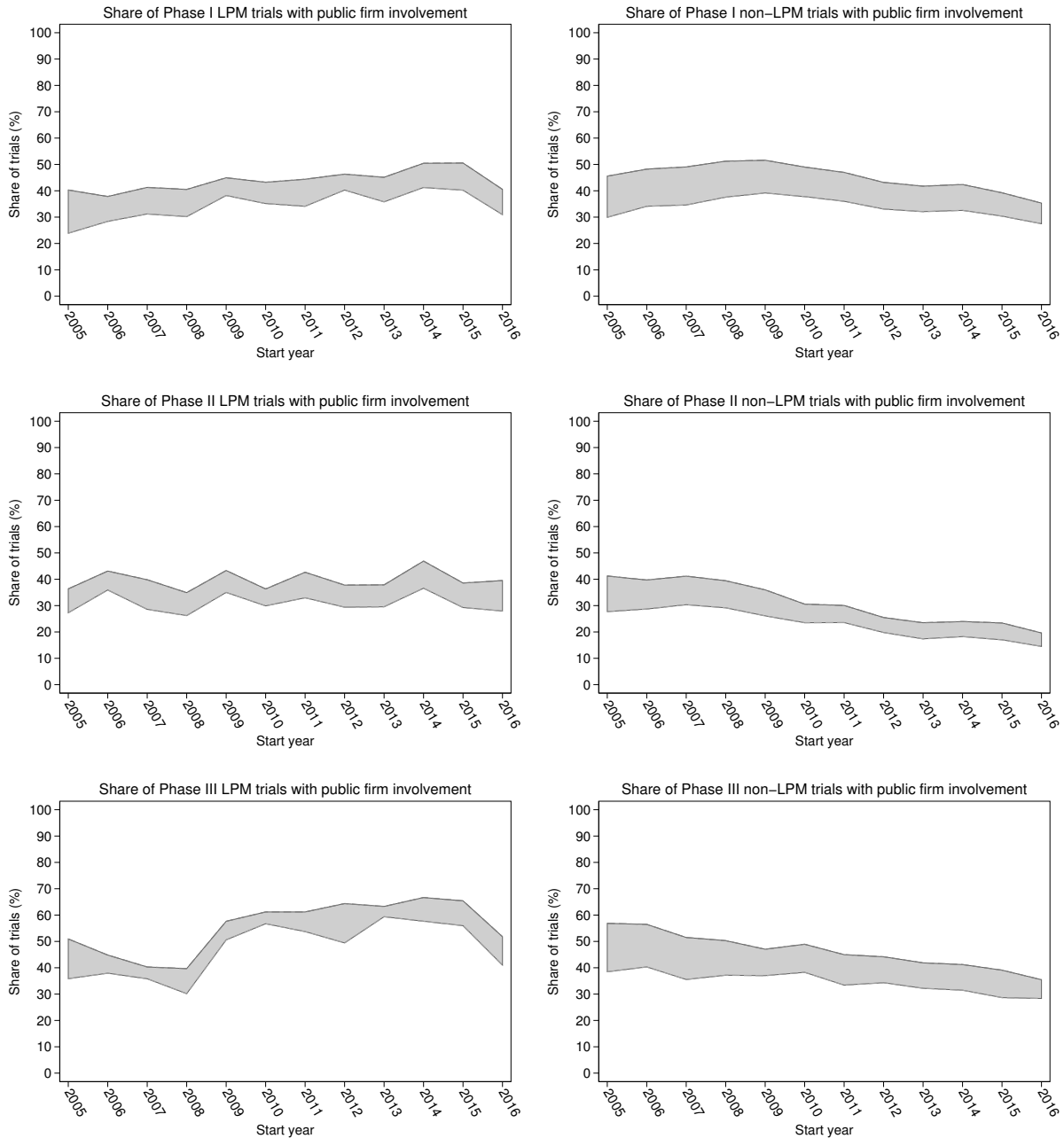


Figure 8: Public vs. privately-held firms (representation in LPM trials)



Tables

Table 1: Summary statistics for selected variables

	All trials		US trials	
	Mean	Observations	Mean	Observations
Uses biomarker	0.4092	131,971	0.4619	49,540
Generous LPM	0.0643	131,971	0.0907	49,540
Restrictive LPM	0.0581	131,971	0.0813	49,540
Phase 1 Clinical (includes Phase 1/Phase 2 trials)	0.3305	131,971	0.3653	49,540
Phase 2 Clinical (includes Phase 2/Phase 3 trials)	0.4367	131,971	0.4263	49,540
Phase 3 Clinical	0.2328	131,971	0.2083	49,540
Received NIH funding	0.0282	131,971	0.0703	49,540
Trial site in US	0.4368	113,410	1.0000	49,540
Publicly-listed firm (lower bound)	0.2903	131,971	0.3436	49,540
Publicly-listed firm (upper bound)	0.3977	131,971	0.4588	49,540
Drug indication for neoplasm (cancer)	0.3352	131,971	0.4002	49,540
Biomarker role: disease	0.0842	131,971	0.1145	49,540
Biomarker role: toxic effect	0.0496	131,971	0.0699	49,540
Biomarker role: therapeutic effect	0.3371	131,971	0.3758	49,540
Biomarker role: not determined	0.0023	131,971	0.0024	49,540
Biomarker type: anthropomorphic	0.0350	131,971	0.0400	49,540
Biomarker type: biochemical	0.1248	131,971	0.1300	49,540
Biomarker type: cellular	0.0308	131,971	0.0424	49,540
Biomarker type: genomic	0.2321	131,971	0.2845	49,540
Biomarker type: physiological	0.0849	131,971	0.0865	49,540
Biomarker type: proteomic	0.2426	131,971	0.2942	49,540
Biomarker type: structural (imaging)	0.0177	131,971	0.0200	49,540
Biomarker role (detailed): diagnosis	0.2948	117,180	0.3448	43,777
Biomarker role (detailed): differential diagnosis	0.1829	117,180	0.2041	43,777
Biomarker role (detailed): predicting drug resistance	0.0624	117,180	0.0778	43,777
Biomarker role (detailed): predicting treatment efficacy	0.2568	117,180	0.3060	43,777
Biomarker role (detailed): predicting treatment toxicity	0.0474	117,180	0.0493	43,777
Biomarker role (detailed): screening	0.0523	117,180	0.0547	43,777
Biomarker role (detailed): selection for therapy	0.0938	117,180	0.1111	43,777
Biomarker role (detailed): disease profiling	0.1909	117,180	0.2269	43,777
Biomarker role (detailed): monitoring disease progression	0.1293	117,180	0.1394	43,777
Biomarker role (detailed): monitoring treatment efficacy	0.2998	117,180	0.3481	43,777
Biomarker role (detailed): monitoring treatment toxicity	0.0464	117,180	0.0469	43,777
Biomarker role (detailed): not determined	0.0090	117,180	0.0114	43,777
Biomarker role (detailed): prognosis	0.2375	117,180	0.2797	43,777
Biomarker role (detailed): prognosis - risk stratification	0.0564	117,180	0.0660	43,777
Biomarker role (detailed): risk factor	0.2407	117,180	0.2770	43,777
Biomarker role (detailed): staging	0.1103	117,180	0.1280	43,777
Biomarker role (detailed): toxicity profiling	0.0085	117,180	0.0082	43,777
N		131,971		49,540

Table 2: Number of trials employing biomarkers by type

	Any biomarker	Anthropomorphic	Biochemical	Cellular	Genomic	Physiological	Proteomic	Structural
Overall	53,998	4,620	16,472	4,070	30,634	11,205	32,011	2,340
1995	105	4	29	1	59	22	60	4
1996	131	5	34	6	77	16	84	4
1997	193	10	62	8	119	24	125	2
1998	288	12	74	6	165	58	182	5
1999	448	16	119	22	292	68	307	8
2000	542	33	149	28	349	83	360	9
2001	645	36	190	38	406	94	426	9
2002	869	53	263	36	558	135	579	21
2003	1,085	80	358	51	698	156	732	28
2004	1,524	126	469	68	950	216	997	34
2005	1,928	135	580	118	1,157	314	1,218	58
2006	2,280	178	737	138	1,379	377	1,462	73
2007	2,718	220	831	207	1,687	437	1,751	98
2008	3,005	252	970	245	1,813	548	1,900	101
2009	3,492	288	1,137	251	2,157	627	2,248	114
2010	3,916	334	1,239	304	2,333	740	2,418	134
2011	4,228	366	1,353	357	2,525	828	2,638	164
2012	4,517	408	1,463	406	2,566	994	2,661	206
2013	4,681	439	1,446	382	2,544	1,104	2,666	241
2014	5,099	518	1,576	434	2,647	1,310	2,762	270
2015	5,857	546	1,610	438	2,944	1,499	3,086	374
2016	6,447	561	1,783	526	3,209	1,555	3,349	383

Biomarker types:
Anthropomorphic biomarkers are markers of the body shape/form
Biochemical biomarkers are substrates or products of chemical reactions in the body
Cellular biomarkers are whole cells
Genomic biomarkers are variants in the DNA sequence or in the transcription level;
Physiological biomarkers are body processes
Proteomic biomarkers are variants in protein sequence, protein levels in a given tissue, protein interactions and enzyme activities
Structural biomarkers are anatomical structures

Table 3: Number of trials employing biomarkers by detailed role

	Any biomarker	Diagnosis	Differential Diagnosis	Predicting drug resistance	Predicting treatment efficacy	Predicting treatment toxicity	Screening	Selection for therapy
Overall	39,207	34,545	21,429	7,312	30,091	5,556	6,133	10,988
1995	105	68	45	7	62	8	8	13
1996	131	88	49	22	81	14	14	19
1997	193	130	83	38	122	31	39	38
1998	288	210	137	66	199	53	42	68
1999	448	341	201	76	310	85	57	97
2000	542	369	233	88	343	78	59	118
2001	645	458	275	121	421	85	81	138
2002	869	624	395	151	578	122	109	203
2003	1,085	764	487	174	691	157	132	263
2004	1,524	1,051	675	240	954	224	190	332
2005	1,928	1,306	799	286	1,189	263	239	408
2006	2,280	1,575	1,004	370	1,396	308	291	510
2007	2,718	1,882	1,215	444	1,693	369	332	617
2008	3,005	2,046	1,360	496	1,832	430	362	661
2009	3,492	2,352	1,578	649	2,145	504	482	842
2010	3,916	2,539	1,540	581	2,210	343	444	768
2011	4,228	2,738	1,698	582	2,379	376	502	890
2012	4,517	2,909	1,780	574	2,494	376	462	906
2013	4,681	2,932	1,778	609	2,530	396	500	964
2014	5,099	3,071	1,809	548	2,552	409	519	934
2015	5,857	3,355	2,005	574	2,816	427	589	1,070
2016	6,447	3,737	2,283	616	3,094	498	680	1,129

Biomarker roles (uses) that are related to the development of LPMs, generously defined, are included above
The restrictive definition of LPMs limits the definition to those related only to prediction: predicting drug resistance, treatment efficacy, and treatment toxicity and is driven by "predicting treatment efficacy."
Biomarker roles (uses) that are unrelated to developing LPMs, but included in the data are: disease profiling, monitoring disease progression, monitoring treatment efficacy, monitoring treatment toxicity, prognosis, prognosis - risk stratification, risk factor, staging, and toxicity profiling

Table 4: Likely precision medicine (LPM) trials (1995-2016):

	Generous definition							
	All Count	All %	P1 Count	P1 %	P2 Count	P2 %	P3 Count	P3 %
1995	12	1.39	4	1.2	2	.631	6	2.79
1996	25	2.58	6	1.77	10	2.53	9	3.81
1997	37	2.8	11	2.44	16	3.13	10	2.77
1998	56	3.29	10	1.72	31	4.43	15	3.54
1999	75	3.12	26	3.03	37	3.94	12	1.96
2000	95	3.62	27	3.03	48	4.57	20	2.93
2001	114	3.81	41	4.13	50	4.05	23	3.01
2002	144	3.87	46	3.99	70	4.46	28	2.81
2003	166	3.96	45	3.55	85	4.82	36	3.1
2004	234	4.49	68	4.48	126	5.68	40	2.71
2005	263	4.1	67	3.63	143	5.09	53	3.03
2006	299	4.17	74	3.44	167	5.4	58	3
2007	407	5.39	109	4.62	231	6.96	67	3.57
2008	408	5.09	116	4.34	229	6.69	63	3.28
2009	563	6.63	178	5.95	300	8.22	85	4.57
2010	563	6.44	185	5.97	311	8.52	67	3.37
2011	642	7.14	214	6.88	361	9.34	67	3.32
2012	699	7.54	231	7.6	381	9.13	87	4.24
2013	781	8.55	257	8.44	396	9.86	128	6.18
2014	836	8.85	337	9.95	388	9.4	111	5.76
2015	1,009	9.55	368	10.5	482	9.89	159	7.23
2016	1,057	8.69	417	10.3	501	8.44	139	6.35
	Restrictive definition							
	All Count	All %	P1 Count	P1 %	P2 Count	P2 %	P3 Count	P3 %
1995	9	1.04	3	.898	2	.631	4	1.86
1996	23	2.37	5	1.47	9	2.28	9	3.81
1997	34	2.57	9	2	15	2.94	10	2.77
1998	53	3.11	9	1.55	30	4.29	14	3.3
1999	70	2.91	23	2.68	35	3.73	12	1.96
2000	90	3.43	25	2.8	46	4.38	19	2.78
2001	105	3.51	36	3.63	46	3.72	23	3.01
2002	133	3.58	41	3.56	66	4.21	26	2.61
2003	152	3.63	37	2.92	80	4.54	35	3.01
2004	212	4.06	60	3.95	112	5.05	40	2.71
2005	240	3.75	58	3.14	131	4.66	51	2.91
2006	275	3.83	64	2.98	156	5.04	55	2.85
2007	370	4.9	89	3.78	218	6.56	63	3.35
2008	380	4.74	104	3.89	217	6.34	59	3.07
2009	502	5.91	148	4.95	274	7.51	80	4.31
2010	514	5.88	165	5.33	285	7.81	64	3.22
2011	592	6.58	188	6.04	343	8.87	61	3.02
2012	645	6.96	209	6.88	355	8.5	81	3.94
2013	720	7.88	231	7.59	369	9.19	120	5.79
2014	748	7.92	306	9.03	343	8.31	99	5.13
2015	883	8.35	322	9.21	417	8.56	144	6.55
2016	914	7.52	346	8.56	442	7.45	126	5.76

Table 5: Likely precision medicine (LPM) trials: cancer only (1995-2016):

Generous definition								
	All Count	All %	P1 Count	P1 %	P2 Count	P2 %	P3 Count	P3 %
1995	8	2.33	3	2.38	2	1.39	3	4.05
1996	24	5.3	6	3.57	10	5.21	8	8.6
1997	34	5.81	11	4.89	13	5.1	10	9.52
1998	54	6.26	10	3.15	30	7.52	14	9.52
1999	67	6.41	21	5.34	35	7.26	11	6.47
2000	86	6.62	24	4.75	45	7.56	17	8.5
2001	104	7.74	39	7.39	45	7.28	20	10.2
2002	137	8.77	44	7.76	68	8.66	25	11.9
2003	142	8.53	34	5.72	80	9.41	28	12.7
2004	204	10.2	61	8.93	113	10.5	30	12.6
2005	226	9.89	55	6.67	130	10.8	41	15.8
2006	261	10.7	64	7.62	154	11.8	43	14.5
2007	363	14.2	87	9.6	217	16.1	59	19.4
2008	368	14.2	98	10.2	221	16.6	49	16.3
2009	507	18	162	14.4	275	19.8	70	22.9
2010	509	18	164	14.6	289	20.5	56	19.4
2011	572	19.8	189	16.7	332	22.4	51	18.8
2012	620	21.4	211	18.9	341	23	68	22.7
2013	680	24.8	226	21.6	356	25.6	98	31.8
2014	713	26.5	296	25.6	332	26.7	85	28.9
2015	855	29.5	319	27	410	30.5	126	34.3
2016	899	27.1	375	27	410	26.1	114	31.4
Restrictive definition								
	All Count	All %	P1 Count	P1 %	P2 Count	P2 %	P3 Count	P3 %
1995	8	2.33	3	2.38	2	1.39	3	4.05
1996	22	4.86	5	2.98	9	4.69	8	8.6
1997	31	5.3	9	4	12	4.71	10	9.52
1998	52	6.03	9	2.84	29	7.27	14	9.52
1999	64	6.12	20	5.09	33	6.85	11	6.47
2000	83	6.38	22	4.36	44	7.39	17	8.5
2001	100	7.45	35	6.63	45	7.28	20	10.2
2002	129	8.26	41	7.23	65	8.28	23	11
2003	132	7.93	29	4.88	76	8.94	27	12.2
2004	190	9.51	56	8.2	104	9.67	30	12.6
2005	213	9.32	50	6.07	122	10.1	41	15.8
2006	249	10.2	59	7.02	147	11.3	43	14.5
2007	340	13.3	77	8.5	207	15.3	56	18.4
2008	352	13.6	91	9.48	212	15.9	49	16.3
2009	467	16.6	138	12.3	259	18.6	70	22.9
2010	479	17	155	13.8	270	19.2	54	18.7
2011	544	18.9	172	15.2	323	21.8	49	18
2012	598	20.7	200	18	332	22.4	66	22.1
2013	654	23.8	212	20.3	347	25	95	30.8
2014	673	25	278	24	311	25	84	28.6
2015	791	27.3	291	24.6	378	28.1	122	33.2
2016	820	24.7	326	23.5	383	24.4	111	30.6

Table 6: Share of trials receiving NIH funding

	P1 All Trials	P1 Gen. LPM	P1 Rest. LPM	P2 All Trials	P2 Gen. LPM	P2 Rest. LPM	P3 All Trials	P3 Gen. LPM	P3 Rest. LPM
1995	2.10	5.99	3.26	0.00	100.00	0.00	0.00	100.00	0.00
1996	3.54	4.81	3.39	0.00	10.00	11.11	0.00	11.11	11.11
1997	5.56	8.02	3.05	9.09	6.25	0.00	11.11	0.00	0.00
1998	5.34	6.00	1.89	0.00	19.35	6.67	0.00	16.67	0.00
1999	4.08	6.93	2.95	15.38	8.11	0.00	13.04	5.71	0.00
2000	7.29	6.10	2.64	7.41	10.42	10.00	8.00	8.70	5.26
2001	4.73	5.67	3.27	14.63	12.00	0.00	16.67	13.04	0.00
2002	3.91	4.97	2.51	15.22	12.86	3.57	9.76	13.64	3.85
2003	5.13	4.14	2.07	6.67	5.88	2.78	2.70	6.25	2.86
2004	4.08	5.81	1.15	16.18	7.94	2.50	18.33	7.14	2.50
2005	4.34	4.94	1.60	4.48	9.79	3.77	3.45	10.69	3.92
2006	4.05	4.75	0.98	12.16	9.58	1.72	9.38	7.69	1.82
2007	3.22	3.73	0.75	7.34	9.96	0.00	7.87	10.09	0.00
2008	3.29	3.62	0.88	17.24	10.92	4.76	18.27	11.06	5.08
2009	3.54	3.18	0.81	8.43	6.00	0.00	8.11	6.57	0.00
2010	2.74	2.47	0.75	11.35	4.82	1.49	12.73	4.56	1.56
2011	1.96	2.43	0.69	5.14	5.26	0.00	5.85	5.54	0.00
2012	2.07	2.04	0.24	5.63	4.46	0.00	4.78	3.94	0.00
2013	2.66	2.81	0.34	5.84	7.32	0.78	5.63	7.59	0.83
2014	3.13	2.25	0.52	5.04	3.87	0.00	4.90	4.37	0.00
2015	2.63	2.48	0.82	6.52	6.64	0.63	6.52	6.71	0.69
2016	2.67	1.85	0.64	6.47	4.79	0.00	7.23	4.75	0.00
N	43615	57636	30720	2837	4365	1283	2478	3991	1195

Table 7: Burden of disease: Millions of years of life lost for associated diseases (average)

	U.S. only	Global
non-LPM	11.66	188.20
LPM	14.65	202.03
t-statistic	19.30	5.57

Table 8a: Likely precision medicine LPM trials: publicly listed firm (upper bound) involvement (1995-2016):

Generous definition								
	All Count	All %	P1 Count	P1 %	P2 Count	P2 %	P3 Count	P3 %
1995	3	25	1	25	0	0	2	33.3
1996	1	4	1	16.7	0	0	0	0
1997	6	16.2	1	9.09	5	31.3	0	0
1998	8	14.3	3	30	4	12.9	1	6.67
1999	8	10.7	2	7.69	5	13.5	1	8.33
2000	11	11.6	3	11.1	4	8.33	4	20
2001	32	28.1	12	29.3	14	28	6	26.1
2002	40	27.8	12	26.1	20	28.6	8	28.6
2003	56	33.7	7	15.6	35	41.2	14	38.9
2004	65	27.8	15	22.1	39	31	11	27.5
2005	106	40.3	27	40.3	52	36.4	27	50.9
2006	126	42.1	28	37.8	72	43.1	26	44.8
2007	164	40.3	45	41.3	92	39.8	27	40.3
2008	152	37.3	47	40.5	80	34.9	25	39.7
2009	259	46	80	44.9	130	43.3	49	57.6
2010	234	41.6	80	43.2	113	36.3	41	61.2
2011	290	45.2	95	44.4	154	42.7	41	61.2
2012	307	43.9	107	46.3	144	37.8	56	64.4
2013	347	44.4	116	45.1	150	37.9	81	63.3
2014	426	51	170	50.4	182	46.9	74	66.7
2015	476	47.2	186	50.5	186	38.6	104	65.4
2016	439	41.5	169	40.5	198	39.5	72	51.8
Restrictive definition								
	All Count	All %	P1 Count	P1 %	P2 Count	P2 %	P3 Count	P3 %
1995	2	22.2	1	33.3	0	0	1	25
1996	1	4.35	1	20	0	0	0	0
1997	5	14.7	0	0	5	33.3	0	0
1998	8	15.1	3	33.3	4	13.3	1	7.14
1999	8	11.4	2	8.7	5	14.3	1	8.33
2000	11	12.2	3	12	4	8.7	4	21.1
2001	30	28.6	11	30.6	13	28.3	6	26.1
2002	37	27.8	11	26.8	18	27.3	8	30.8
2003	55	36.2	7	18.9	34	42.5	14	40
2004	65	30.7	15	25	39	34.8	11	27.5
2005	97	40.4	25	43.1	47	35.9	25	49
2006	122	44.4	27	42.2	70	44.9	25	45.5
2007	150	40.5	39	43.8	86	39.4	25	39.7
2008	144	37.9	44	42.3	77	35.5	23	39
2009	241	48	73	49.3	123	44.9	45	56.3
2010	218	42.4	74	44.8	104	36.5	40	62.5
2011	273	46.1	88	46.8	146	42.6	39	63.9
2012	289	44.8	101	48.3	135	38	53	65.4
2013	331	46	113	48.9	141	38.2	77	64.2
2014	388	51.9	163	53.3	159	46.4	66	66.7
2015	435	49.3	172	53.4	168	40.3	95	66
2016	395	43.2	150	43.4	179	40.5	66	52.4

Table 8b: Likely precision medicine LPM trials: publicly listed firm (lower bound) involvement (1995-2016):

Generous definition								
	All Count	All %	P1 Count	P1 %	P2 Count	P2 %	P3 Count	P3 %
1995	1	8.33	0	0	0	0	1	16.7
1996	1	4	1	16.7	0	0	0	0
1997	3	8.11	1	9.09	2	12.5	0	0
1998	3	5.36	2	20	0	0	1	6.67
1999	3	4	0	0	2	5.41	1	8.33
2000	5	5.26	0	0	2	4.17	3	15
2001	23	20.2	10	24.4	8	16	5	21.7
2002	27	18.8	6	13	15	21.4	6	21.4
2003	40	24.1	4	8.89	24	28.2	12	33.3
2004	51	21.8	10	14.7	31	24.6	10	25
2005	74	28.1	16	23.9	39	27.3	19	35.8
2006	103	34.4	21	28.4	60	35.9	22	37.9
2007	124	30.5	34	31.2	66	28.6	24	35.8
2008	114	27.9	35	30.2	60	26.2	19	30.2
2009	216	38.4	68	38.2	105	35	43	50.6
2010	196	34.8	65	35.1	93	29.9	38	56.7
2011	228	35.5	73	34.1	119	33	36	53.7
2012	248	35.5	93	40.3	112	29.4	43	49.4
2013	285	36.5	92	35.8	117	29.5	76	59.4
2014	345	41.3	139	41.2	142	36.6	64	57.7
2015	378	37.5	148	40.2	141	29.3	89	56
2016	326	30.8	129	30.9	140	27.9	57	41
Restrictive definition								
	All Count	All %	P1 Count	P1 %	P2 Count	P2 %	P3 Count	P3 %
1995	0	0	0	0	0	0	0	0
1996	1	4.35	1	20	0	0	0	0
1997	2	5.88	0	0	2	13.3	0	0
1998	3	5.66	2	22.2	0	0	1	7.14
1999	3	4.29	0	0	2	5.71	1	8.33
2000	5	5.56	0	0	2	4.35	3	15.8
2001	21	20	9	25	7	15.2	5	21.7
2002	24	18	5	12.2	13	19.7	6	23.1
2003	40	26.3	4	10.8	24	30	12	34.3
2004	51	24.1	10	16.7	31	27.7	10	25
2005	69	28.8	16	27.6	34	26	19	37.3
2006	101	36.7	21	32.8	59	37.8	21	38.2
2007	112	30.3	29	32.6	61	28	22	34.9
2008	111	29.2	34	32.7	59	27.2	18	30.5
2009	201	40	62	41.9	100	36.5	39	48.8
2010	182	35.4	59	35.8	86	30.2	37	57.8
2011	213	36	68	36.2	111	32.4	34	55.7
2012	232	36	87	41.6	104	29.3	41	50.6
2013	271	37.6	89	38.5	110	29.8	72	60
2014	319	42.6	137	44.8	123	35.9	59	59.6
2015	345	39.1	137	42.5	128	30.7	80	55.6
2016	296	32.4	118	34.1	125	28.3	53	42.1

Table 9a: Predicting LPM trials (linear probability models)

Outcome = LPM trial, generous definition								
	All Years				2005-2016 Only			
Trial start year	0.0038*	0.0038*	0.0038*	0.0038*	0.0050	0.0050	0.0050	0.0050
	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
Phase 2 Clinical (includes phase 2/3 trials)	0.0097	0.0100	0.0097	0.0100	0.0124	0.0127	0.0129	0.0131
	(0.0095)	(0.0096)	(0.0093)	(0.0094)	(0.0109)	(0.0110)	(0.0107)	(0.0107)
Phase 3 Clinical	0.0168	0.0166	0.0169	0.0167	0.0192	0.0189	0.0196	0.0193
	(0.0145)	(0.0145)	(0.0143)	(0.0143)	(0.0160)	(0.0159)	(0.0157)	(0.0157)
Trial site in US=1	0.0126***	0.0118***	0.0120**	0.0112**	0.0132***	0.0122***	0.0087*	0.0078*
	(0.0020)	(0.0021)	(0.0032)	(0.0031)	(0.0026)	(0.0028)	(0.0035)	(0.0035)
Cancer trial =1	0.1372***	0.1373***	0.1364***	0.1365***	0.1502***	0.1502***	0.1444***	0.1444***
	(0.0147)	(0.0147)	(0.0120)	(0.0120)	(0.0184)	(0.0184)	(0.0134)	(0.0133)
NIH funding	0.0113	0.0128*	0.0111	0.0126*	0.0100	0.0114	0.0080	0.0095
	(0.0064)	(0.0060)	(0.0060)	(0.0056)	(0.0078)	(0.0071)	(0.0064)	(0.0058)
Biomarker type: genomic=1	0.2427*	0.2427*	0.2426*	0.2426*	0.2401*	0.2401*	0.2397*	0.2397*
	(0.1103)	(0.1102)	(0.1105)	(0.1104)	(0.1129)	(0.1128)	(0.1132)	(0.1131)
Public firm (lower bound)	0.0109*		0.0109*		0.0133*		0.0132*	
	(0.0043)		(0.0043)		(0.0059)		(0.0060)	
Public firm (upper bound)		0.0124*		0.0124*		0.0141*		0.0140*
		(0.0048)		(0.0048)		(0.0062)		(0.0062)
Trial site in US=1 × Cancer trial=1			0.0017	0.0018			0.0137	0.0136
			(0.0074)	(0.0074)			(0.0129)	(0.0130)
N	108749	108749	108749	108749	92568	92568	92568	92568
R^2	0.271	0.271	0.271	0.271	0.279	0.279	0.279	0.279

* p<0.05, ** p<0.01, *** p<0.001

All models include a constant; robust standard errors clustered at the level of the ICD-9 chapter

Table 9b: Predicting LPM trials (linear probability models)

Outcome = LPM trial, restrictive definition								
	All Years				2005-2016 Only			
Trial start year	0.0034*	0.0034*	0.0034*	0.0034*	0.0044	0.0044	0.0044	0.0044
	(0.0014)	(0.0013)	(0.0014)	(0.0013)	(0.0023)	(0.0023)	(0.0023)	(0.0023)
Phase 2 Clinical (includes phase 2/3 trials)	0.0127	0.0131	0.0129	0.0132	0.0153	0.0156	0.0158	0.0162
	(0.0101)	(0.0101)	(0.0099)	(0.0100)	(0.0117)	(0.0117)	(0.0115)	(0.0115)
Phase 3 Clinical	0.0215	0.0212	0.0217	0.0214	0.0236	0.0233	0.0241	0.0238
	(0.0153)	(0.0153)	(0.0152)	(0.0151)	(0.0170)	(0.0169)	(0.0168)	(0.0167)
Trial site in US=1	0.0091***	0.0081***	0.0075**	0.0065*	0.0090***	0.0078***	0.0035	0.0025
	(0.0014)	(0.0015)	(0.0024)	(0.0022)	(0.0017)	(0.0019)	(0.0024)	(0.0025)
Cancer trial =1	0.1360***	0.1362***	0.1338***	0.1340***	0.1488***	0.1488***	0.1418***	0.1418***
	(0.0139)	(0.0139)	(0.0111)	(0.0111)	(0.0175)	(0.0175)	(0.0124)	(0.0124)
NIH funding	0.0093	0.0111*	0.0088	0.0106	0.0100	0.0117*	0.0076	0.0094*
	(0.0056)	(0.0052)	(0.0054)	(0.0051)	(0.0059)	(0.0053)	(0.0050)	(0.0044)
Biomarker type: genomic=1	0.2174	0.2174	0.2173	0.2172	0.2142	0.2143	0.2137	0.2138
	(0.1086)	(0.1085)	(0.1087)	(0.1087)	(0.1110)	(0.1109)	(0.1112)	(0.1112)
Public firm (lower bound)	0.0138*		0.0138*		0.0163*		0.0162*	
	(0.0054)		(0.0054)		(0.0071)		(0.0072)	
Public firm (upper bound)		0.0154*		0.0154*		0.0171*		0.0170*
		(0.0057)		(0.0057)		(0.0073)		(0.0073)
Trial site in US=1 × Cancer trial =1			0.0047	0.0048			0.0165	0.0164 ¿
			(0.0069)	(0.0069)			(0.0123)	(0.0123)
Constant	-6.8957*	-6.9257*	-6.9057*	-6.9359*	-8.8324	-8.9462	-8.8072	-8.9201
	(2.7655)	(2.7107)	(2.7552)	(2.7004)	(4.6835)	(4.6680)	(4.6941)	(4.6796)
N	108749	108749	108749	108749	92568	92568	92568	92568
R^2	0.254	0.254	0.254	0.254	0.261	0.262	0.262	0.262

* p<0.05, ** p<0.01, *** p<0.001

All models include a constant; robust standard errors clustered at the level of the ICD-9 chapter

Table 10: Dependent variable: Trial duration in months

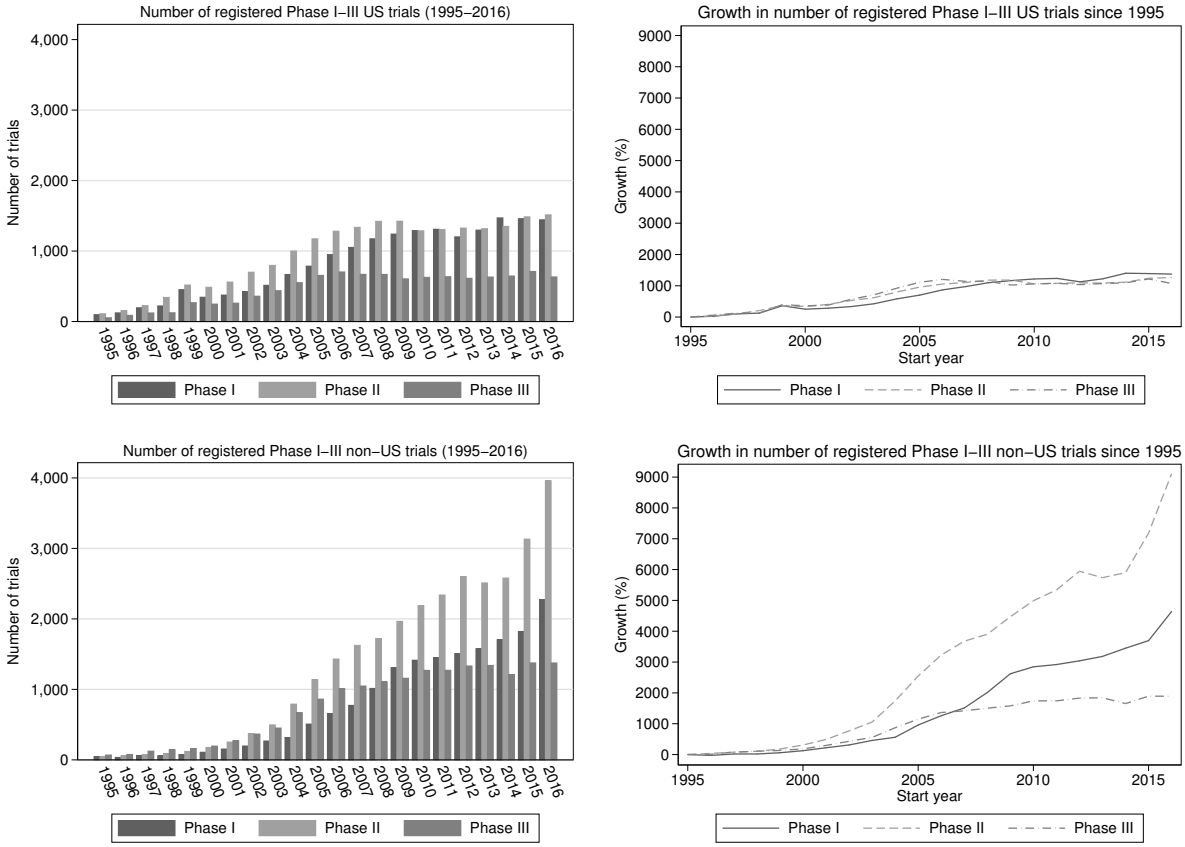
	LPM Trials			Non-LPM Trials		
	All Trials	All Trials	All U.S.	All Trials	All Trials	All U.S.
Phase 2 Clinical (inc. Phase 2/3 trials)	2.028	2.063*	2.557	6.441***	6.184***	5.284***
	(1.045)	(1.043)	(1.358)	(0.222)	(0.222)	(0.337)
Phase 3 Clinical	14.164***	14.367***	13.400***	9.262***	9.374***	7.095***
	(1.564)	(1.554)	(2.063)	(0.253)	(0.252)	(0.397)
Trial site in US	2.998**	3.425***		3.975***	4.355***	
	(0.981)	(0.989)		(0.189)	(0.189)	
Cancer trial	12.930***	13.038***	12.391***	19.387***	19.214***	18.171***
	(1.282)	(1.275)	(1.755)	(0.269)	(0.268)	(0.374)
Received NIH funding	8.026***	7.033**	6.680**	12.839***	11.871***	11.714***
	(2.243)	(2.271)	(2.355)	(0.730)	(0.732)	(0.758)
Public firm (lower bound)	-2.608**			-6.111***		
	(0.973)			(0.187)		
Public firm (upper bound)		-4.308***	-4.866***		-7.238***	-7.210***
		(1.006)	(1.394)		(0.192)	(0.298)
Constant	60.199***	59.999***	63.247***	38.158***	38.697***	44.402***
	(6.136)	(6.113)	(6.710)	(1.276)	(1.275)	(1.472)
N	2743	2743	1760	50186	50186	26101
R^2	0.334	0.337	0.324	0.311	0.317	0.286

* p<0.05, ** p<0.01, *** p<0.001

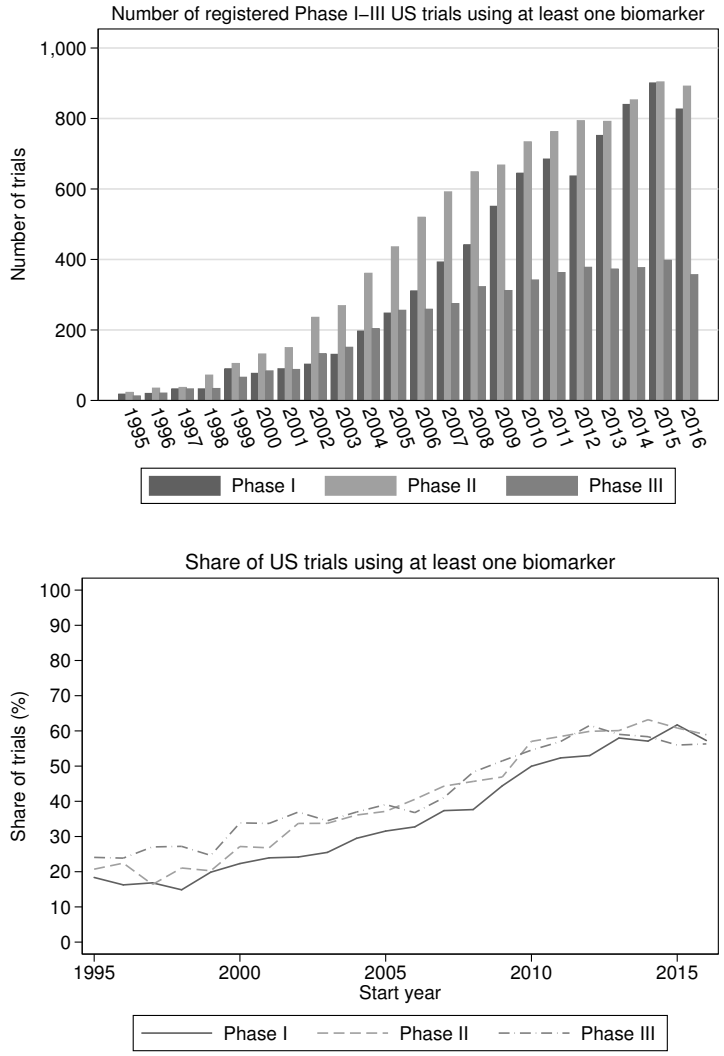
Sample includes all trials launched after 2000 with known end dates. Duration is winsorized to remove extreme outliers. All OLS models include a constant, year fixed effects, and robust standard errors. All models in this table use the "generous" definition of LPM trials.

Appendices

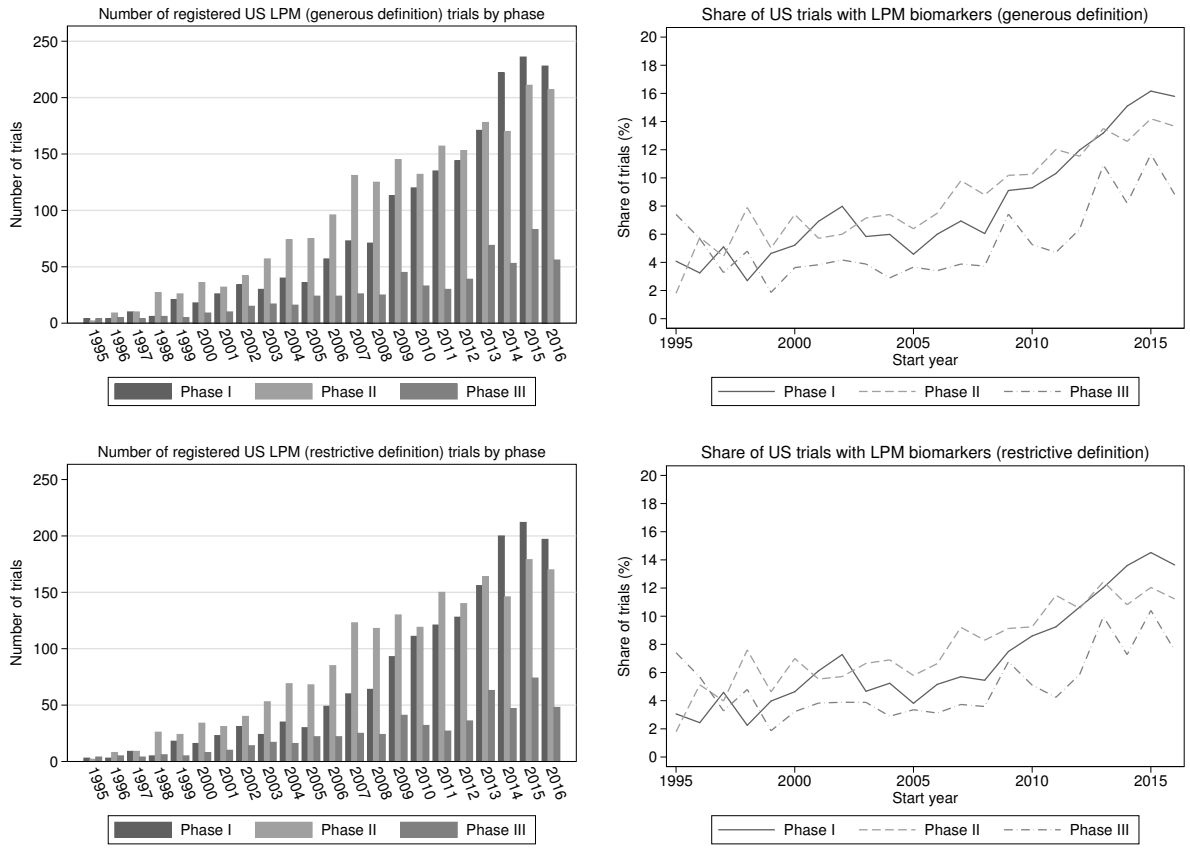
Appendix Figure A: U.S. Clinical trials over time



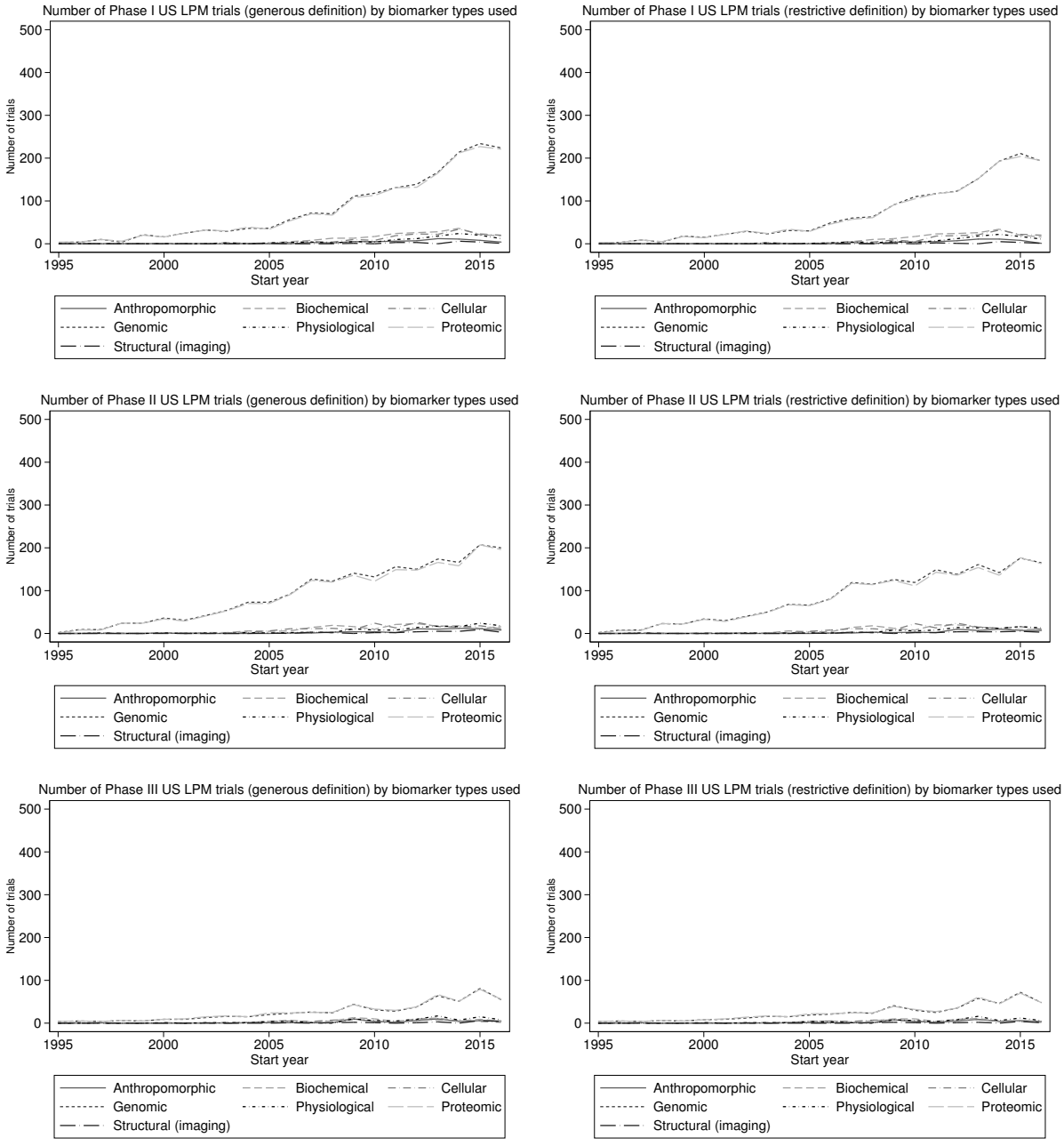
Appendix Figure B: U.S. Clinical trials using biomarkers



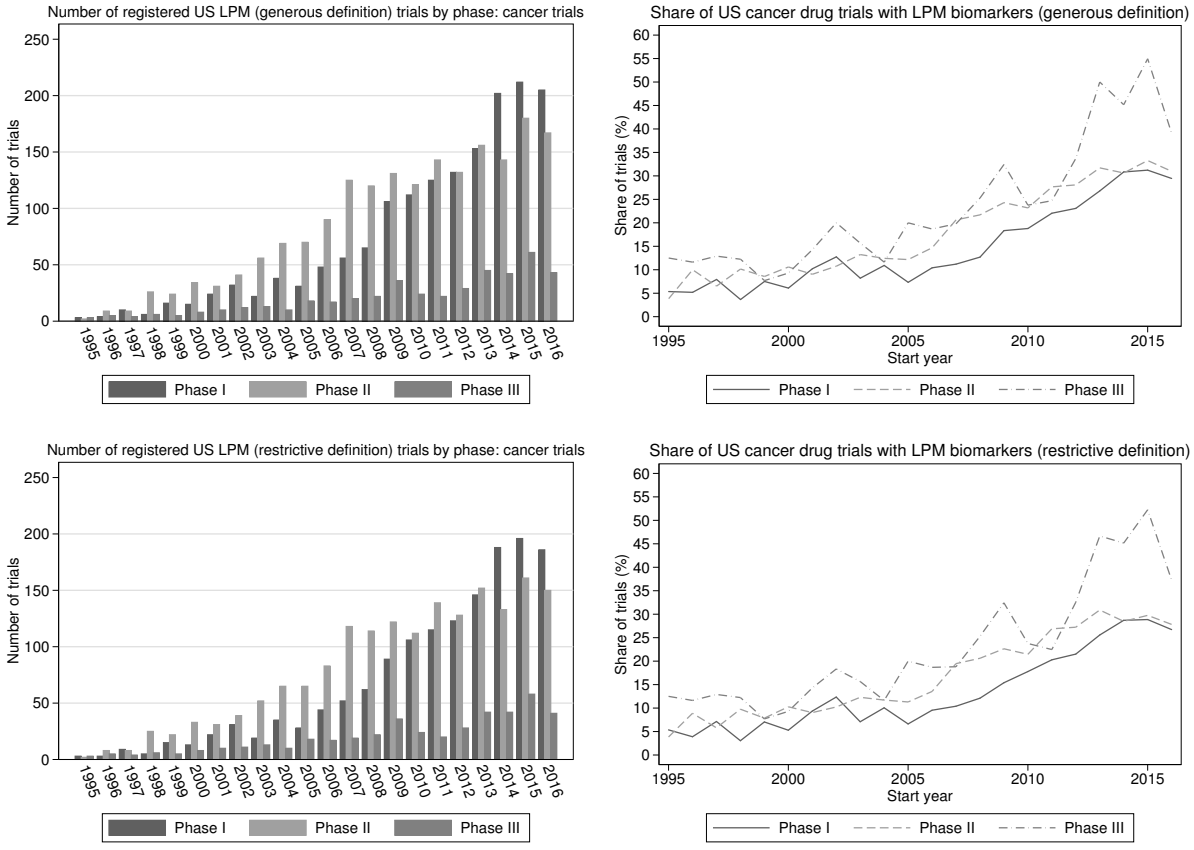
Appendix Figure C: U.S. Clinical trials for LPMs



Appendix Figure D: Types of biomarkers used in U.S. LPM trials



Appendix Figure E: U.S. clinical trials for LPMs, cancer indications only



Appendix Table I: U.S. likely precision medicine (LPM) trials (1995-2016):

Generous definition								
	All Count	All %	P1 Count	P1 %	P2 Count	P2 %	P3 Count	P3 %
1995	8	6.15	3	5.36	2	4	3	12.5
1996	18	8.57	4	5.19	9	10	5	11.6
1997	23	7.82	10	7.94	9	6.57	4	12.9
1998	38	8.15	6	3.66	26	10.3	6	12.2
1999	45	8.09	16	7.55	24	8.57	5	7.81
2000	57	8.78	15	6.1	34	10.7	8	9.3
2001	65	10.1	24	10.2	31	9.09	10	14.3
2002	85	12.3	32	12.8	41	10.8	12	20
2003	91	11.8	22	8.21	56	13.2	13	15.7
2004	117	11.8	38	10.9	69	12.5	10	11.6
2005	119	11	31	7.33	70	12.2	18	20
2006	155	13.3	48	10.4	90	14.7	17	18.7
2007	201	16.7	56	11.2	125	20.7	20	19.8
2008	207	18	65	12.7	120	21.7	22	25.3
2009	271	22.2	106	18.3	129	24.1	36	32.7
2010	257	21.1	112	18.8	121	23.3	24	24
2011	290	24.8	125	22	143	27.8	22	24.7
2012	293	26.1	132	23.1	132	28.3	29	34.1
2013	354	30.8	153	26.8	156	31.9	45	50
2014	387	31.9	202	30.8	143	30.8	42	45.2
2015	453	34.2	212	31.3	180	33.5	61	55
2016	413	30.9	205	29.5	165	31	43	39.4
Restrictive definition								
	All Count	All %	P1 Count	P1 %	P2 Count	P2 %	P3 Count	P3 %
1995	8	6.15	3	5.36	2	4	3	12.5
1996	16	7.62	3	3.9	8	8.89	5	11.6
1997	21	7.14	9	7.14	8	5.84	4	12.9
1998	36	7.73	5	3.05	25	9.88	6	12.2
1999	42	7.55	15	7.08	22	7.86	5	7.81
2000	54	8.32	13	5.28	33	10.4	8	9.3
2001	63	9.75	22	9.36	31	9.09	10	14.3
2002	81	11.7	31	12.4	39	10.3	11	18.3
2003	84	10.9	19	7.09	52	12.3	13	15.7
2004	110	11.1	35	10.1	65	11.7	10	11.6
2005	111	10.2	28	6.62	65	11.3	18	20
2006	144	12.4	44	9.54	83	13.5	17	18.7
2007	189	15.7	52	10.4	118	19.6	19	18.8
2008	198	17.2	62	12.1	114	20.7	22	25.3
2009	245	20	89	15.4	120	22.4	36	32.7
2010	242	19.9	106	17.8	112	21.5	24	24
2011	274	23.4	115	20.3	139	27	20	22.5
2012	279	24.8	123	21.5	128	27.4	28	32.9
2013	340	29.6	146	25.6	152	31.1	42	46.7
2014	363	30	188	28.7	133	28.7	42	45.2
2015	415	31.3	196	29	161	30	58	52.3
2016	376	28.1	186	26.8	149	28	41	37.6

Appendix Table II: U.S. likely precision medicine (LPM) trials: cancer only (1995-2016):

Generous definition								
	All Count	All %	P1 Count	P1 %	P2 Count	P2 %	P3 Count	P3 %
1995	8	6.15	3	5.36	2	4	3	12.5
1996	18	8.57	4	5.19	9	10	5	11.6
1997	23	7.82	10	7.94	9	6.57	4	12.9
1998	38	8.15	6	3.66	26	10.3	6	12.2
1999	45	8.09	16	7.55	24	8.57	5	7.81
2000	57	8.78	15	6.1	34	10.7	8	9.3
2001	65	10.1	24	10.2	31	9.09	10	14.3
2002	85	12.3	32	12.8	41	10.8	12	20
2003	91	11.8	22	8.21	56	13.2	13	15.7
2004	117	11.8	38	10.9	69	12.5	10	11.6
2005	119	11	31	7.33	70	12.2	18	20
2006	155	13.3	48	10.4	90	14.7	17	18.7
2007	201	16.7	56	11.2	125	20.7	20	19.8
2008	207	18	65	12.7	120	21.7	22	25.3
2009	271	22.2	106	18.3	129	24.1	36	32.7
2010	257	21.1	112	18.8	121	23.3	24	24
2011	290	24.8	125	22	143	27.8	22	24.7
2012	293	26.1	132	23.1	132	28.3	29	34.1
2013	354	30.8	153	26.8	156	31.9	45	50
2014	387	31.9	202	30.8	143	30.8	42	45.2
2015	453	34.2	212	31.3	180	33.5	61	55
2016	413	30.9	205	29.5	165	31	43	39.4
Restrictive definition								
	All Count	All %	P1 Count	P1 %	P2 Count	P2 %	P3 Count	P3 %
1995	8	6.15	3	5.36	2	4	3	12.5
1996	16	7.62	3	3.9	8	8.89	5	11.6
1997	21	7.14	9	7.14	8	5.84	4	12.9
1998	36	7.73	5	3.05	25	9.88	6	12.2
1999	42	7.55	15	7.08	22	7.86	5	7.81
2000	54	8.32	13	5.28	33	10.4	8	9.3
2001	63	9.75	22	9.36	31	9.09	10	14.3
2002	81	11.7	31	12.4	39	10.3	11	18.3
2003	84	10.9	19	7.09	52	12.3	13	15.7
2004	110	11.1	35	10.1	65	11.7	10	11.6
2005	111	10.2	28	6.62	65	11.3	18	20
2006	144	12.4	44	9.54	83	13.5	17	18.7
2007	189	15.7	52	10.4	118	19.6	19	18.8
2008	198	17.2	62	12.1	114	20.7	22	25.3
2009	245	20	89	15.4	120	22.4	36	32.7
2010	242	19.9	106	17.8	112	21.5	24	24
2011	274	23.4	115	20.3	139	27	20	22.5
2012	279	24.8	123	21.5	128	27.4	28	32.9
2013	340	29.6	146	25.6	152	31.1	42	46.7
2014	363	30	188	28.7	133	28.7	42	45.2
2015	415	31.3	196	29	161	30	58	52.3
2016	376	28.1	186	26.8	149	28	41	37.6

Appendix Table III: Dependent variable: Trial duration in months (cancer trials only)

	LPM Trials		Non-LPM Trials	
Phase 2 Clinical (inc. Phase 2/3 trials)	1.619	1.626	3.106***	3.045***
	(1.126)	(1.124)	(0.483)	(0.482)
Phase 3 Clinical	16.378***	16.478***	13.128***	13.043***
	(1.884)	(1.875)	(0.932)	(0.928)
Trial site in US	3.062**	3.404**	3.207***	3.497***
	(1.095)	(1.106)	(0.481)	(0.482)
Received NIH funding	8.011***	7.172**	8.347***	7.721***
	(2.342)	(2.376)	(1.113)	(1.126)
Public firm (lower bound)	-2.162*		-5.506***	
	(1.081)		(0.473)	
Public firm (upper bound)		-3.551**		-6.032***
		(1.117)		(0.470)
Constant	70.225***	70.159***	58.019***	58.086***
	(6.410)	(6.384)	(1.785)	(1.786)
N	2289	2289	12423	12423
R^2	0.308	0.310	0.197	0.199

* p<0.05, ** p<0.01, *** p<0.001

Sample includes all trials launched after 2000 with known end dates. Duration is winsorized to remove extreme outliers. All OLS models include year fixed effects, and robust standard errors.

APPENDIX A

This table lists the formal definition of different biomarker types as defined by the FDA-NIH Biomarker Working group (2016)

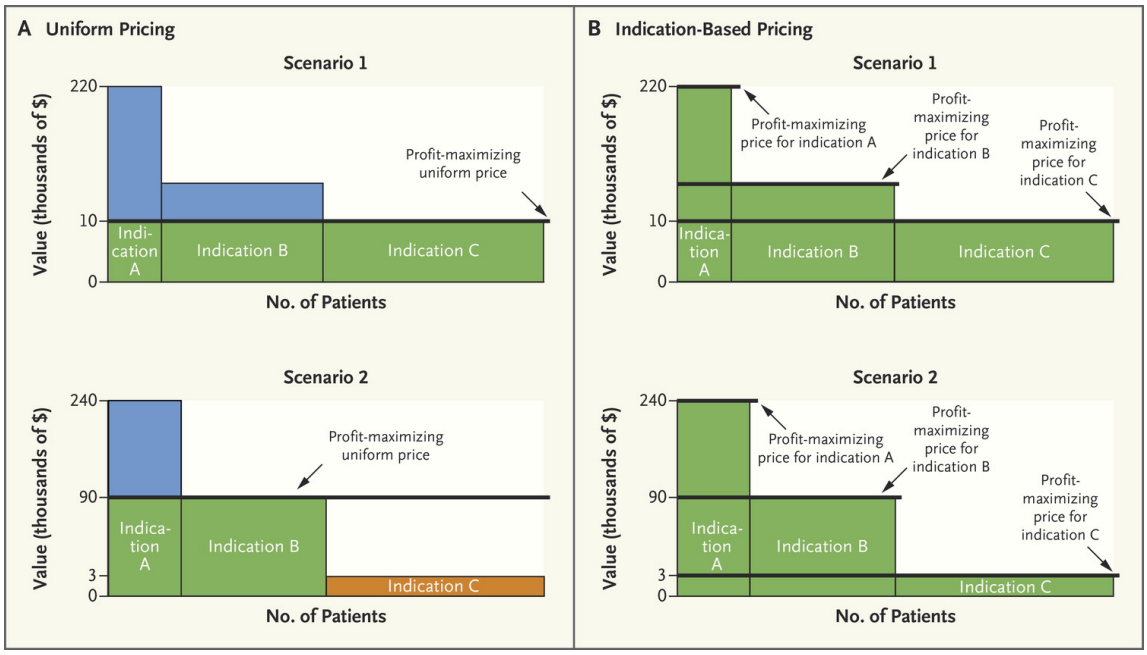
Biomarker type	Official definition	Examples
Diagnostic Biomarker	A biomarker used to detect or confirm presence of a disease or condition of interest or to identify individuals with a subtype of the disease.	1) Sweat chloride may be used as a diagnostic biomarker to confirm cystic fibrosis (Farrell et al. 2008). 2) Glomerular filtration rate (GFR) may be used as a diagnostic biomarker to identify patients with chronic kidney disease (National Kidney Foundation 2002).
Monitoring Biomarker	A biomarker measured serially for assessing status of a disease or medical condition or for evidence of exposure to (or effect of) a medical product or an environmental agent.	1) HIV-RNA may be used as a monitoring biomarker to measure and guide treatment with antiretroviral therapy (ART) (AIDSinfo 2007). 2) Serial measurements of symphysis-fundal height during pregnancy can be used during antenatal screening to detect fetal growth disturbances (Papageorghiou et al. 2016).
Pharmacodynamic / Response Biomarker	A biomarker used to show that a biological response has occurred in an individual who has been exposed to a medical product or an environmental agent.	1) Circulating B lymphocytes may be used as a pharmacodynamic/response biomarker when evaluating patients with systemic lupus erythematosus to assess response to a B-lymphocyte stimulator inhibitor (Stohl and Hilbert 2012). 2) Urinary level of glycosaminoglycans may be used as a pharmacodynamic/response biomarker when evaluating the effect of enzyme replacement therapy for patients with mucopolysaccharidosis type 1 (Jameson et al. 2016).
Predictive Biomarker	A biomarker used to identify individuals who are more likely than similar individuals without the biomarker to experience a favorable or unfavorable effect from exposure to a medical product or an environmental agent.	1) Certain cystic fibrosis transmembrane conductance regulator (CFTR) mutations may be used as predictive biomarkers in clinical trials evaluating treatment for cystic fibrosis, to select patients more likely to respond to particular treatments (Davies et al. 2013). 2) Human leukocyte antigen allele (HLA)–B*5701 genotype may be used as a predictive biomarker to evaluate human immunodeficiency virus (HIV) patients before abacavir treatment, to identify patients at risk for severe skin reactions (AIDSinfo 2007).
Prognostic Biomarker	A biomarker used to identify likelihood of a clinical event, disease recurrence or progression in patients who have the disease or medical condition of interest.	1) Breast Cancer genes 1 and 2 (BRCA1/2) mutations may be used as prognostic biomarkers when evaluating women with breast cancer, to assess the likelihood of a second breast cancer (Basu et al. 2015). 2) Gleason score may be used as a prognostic biomarker when evaluating patients with prostate cancer to assess the likelihood of cancer progression (Epstein et al. 2016; Gordetsky and Epstein 2016).
Safety Biomarker	A biomarker measured before or after an exposure to a medical product or an environmental agent to indicate the likelihood,	1) Hepatic aminotransferases and bilirubin may be used as safety biomarkers when evaluating potential hepatotoxicity (Senior 2014).

	presence, or extent of toxicity as an adverse effect.	2) Serum creatinine may be used as a safety biomarker when evaluating patients on drugs that affect kidney function to monitor for nephrotoxicity (Wasung et al. 2015).
Susceptibility / Risk Biomarker:	A biomarker that indicates the potential for developing a disease or medical condition in an individual who does not currently have clinically apparent disease or the medical condition.	1) Factor V Leiden may be used as a susceptibility/risk biomarker to identify individuals with a predisposition to develop deep vein thrombosis (DVT) (Kujovich 2011). 2) Infection with certain human papillomavirus (HPV) subtypes may be used as a susceptibility/risk biomarker to identify individuals with a predisposition to develop cervical cancer (Khan et al. 2005; Schiffman et al. 2011).

Note: Some examples of biomarkers cited in this appendix may be applicable for more than one type of biomarker. For example, in some cases predictive biomarkers used to identify individuals who are more likely to experience a favorable effect from a drug can also be used as diagnostic biomarkers in the initial detection or confirmation of the disease (e.g. CFTR mutations in Cystic Fibrosis).

APPENDIX B

Effects of uniform pricing versus indication-based pricing.



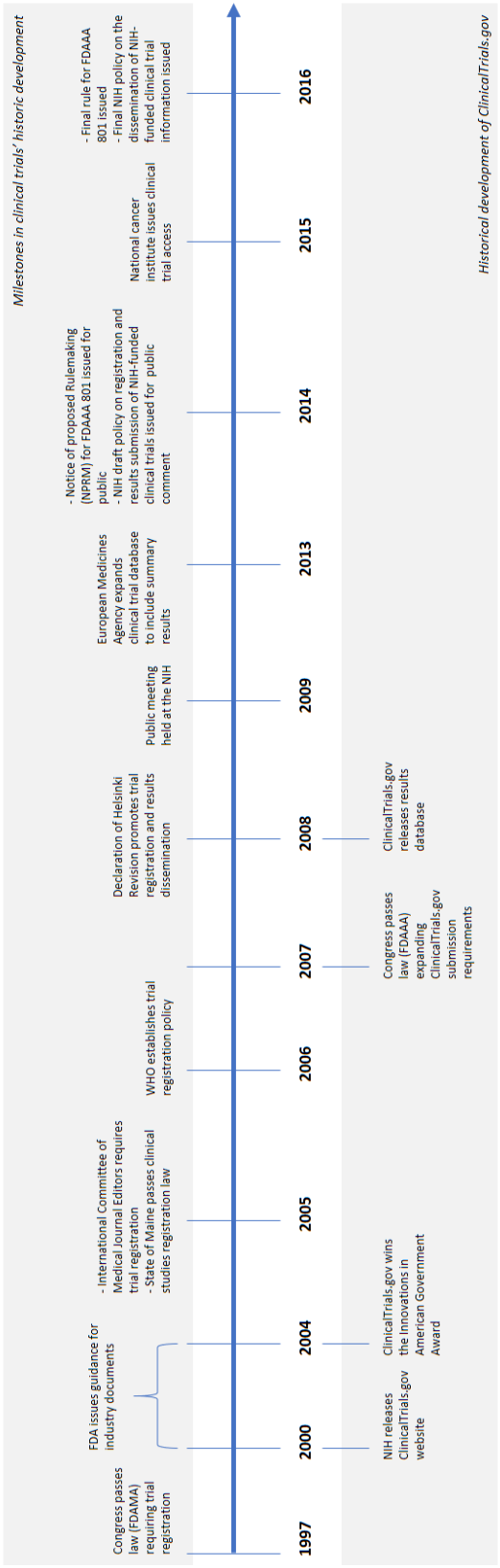
From Chandra, A. and Garthwaite, C. "The Economics of Indication-Based Drug Pricing." *New England Journal of Medicine*, 377(2), pp.103-106. Copyright © (2017) Massachusetts Medical Society. Reprinted with permission. http://www.nejm.org/doi/full/10.1056/NEJMp1705035

In Panel A, the upper graph represents a uniform-pricing context in which patients with indication A receive the most benefit and those with indication C receive the least. The population with indication C is large, and the value of treatment to this group is close to the value for indication B. As a result, the manufacturer's profit-maximizing price allows all patients to obtain the drug. At this price, the manufacturer earns profits represented by the green area. But the firm faces a trade-off. By setting the price in this way, the manufacturer forgoes profits that could be earned by charging higher prices to patients with indications A and B. These forgone profits, represented by the blue areas, are captured by these patients as consumer surplus — the value difference between the most consumers are willing to pay and what they actually pay. The lower graph in Panel A shows a different scenario, in which the product's valuation for patients with indication C is very low. In this case, it's a better trade-off for the manufacturer to set a high price, at which it knows the payer will allow only patients with indications A and B to obtain the drug. The manufacturer accepts the loss of sales to patients with indication C in exchange for higher profits earned from patients with indications A and B. Comparing these graphs, we see that when the valuation of the product for indication C is relatively low, manufacturers set a higher uniform price, the payer curtails sales to patients with indication C (orange area), and patients with indications A and B obtain less consumer surplus than they did in the first scenario.

Panel B of the graph represents the same set scenarios with respect to the distribution of patients and valuations but allows for indication based pricing by the manufacturer. The scenario presented is an extreme example where a monopoly provider is able to set the price exactly at the willingness to pay of the

consumer population and thus capture all of the surplus. For scenario 1, the same sets of patients are served by the manufacturer is now able to capture all of the surplus. Scenario 2 represents an output expanding scenario where the manufacturer now finds it profitable to sell to patients with indication C, while also raising the price on the indication A patients that receive the most value from the drug. In total, the introduction of indication based pricing is shown to weakly increase prices, profits, and the quantity sold.

APPENDIX C


Milestones in clinical trials' historic development
Congress passes law (FDAMA) requiring trial registration
FDA issues guidance for industry documents
- International Committee of Medical Journal Editors requires trial registration
- State of Maine passes clinical studies registration law
WHO establishes trial registration policy
Declaration of Helsinki Revision promotes trial registration and results dissemination
Public meeting held at the NIH
European Medicines Agency expands clinical trial database to include summary results
- Notice of proposed Rulemaking (NPRM) for FDAAA 801 issued for public
- NIH draft policy on registration and results submission of NIH-funded clinical trials issued for public comment
National cancer institute issues clinical trial access
- Final rule for FDAAA 801 issued
- Final NIH policy on the dissemination of NIH-funded clinical trial information issued
1997
2000
2004
2005
2006
2007
2008
2009
2013
2014
2015
2016
NIH releases ClinicalTrials.gov website
ClinicalTrials.gov wins the Innovations in American Government Award
Congress passes law (FDAAA) expanding ClinicalTrials.gov submission requirements
ClinicalTrials.gov releases results database
Historical development of ClinicalTrials.gov

Selected Explanation as provided by the Website of ClinicalTrials.gov (2017):

1997: Congress Passes Law (FDAMA) Requiring Trial Registration
The first U.S. Federal law to require trial registration was the Food and Drug Administration Modernization Act of 1997 (FDAMA) (PDF). Section 113 of FDAMA required the National Institutes of Health (NIH) to create a public information resource on certain clinical trials regulated by the Food and Drug Administration (FDA)

2000: NIH Releases ClinicalTrials.gov Web Site
The first version of ClinicalTrials.gov was made available to the public on February 29, 2000. At the time, ClinicalTrials.gov primarily included NIH-funded studies.

2000–2004: FDA Issues Guidance for Industry Documents
In 2000 FDA issued a draft Guidance for Industry document, which provided recommendations for researchers submitting information to ClinicalTrials.gov. A final guidance document that incorporated comments from the public was issued in 2002.

2004: ClinicalTrials.gov Wins the Innovations in American Government Award
The Innovations in American Government Awards program highlights exemplary models of government innovation and advances efforts to address the Nation's most pressing public concerns.

2005: International Committee of Medical Journal Editors Requires Trial Registration
In 2005 the International Committee of Medical Journal Editors (ICMJE) began requiring trial registration as a condition of publication.

2005: State of Maine Passes Clinical Studies Registration Law (Repealed in 2011)
In 2005 the State of Maine passed a law requiring prescription drug manufacturers or labelers to submit clinical study registration and results information to ClinicalTrials.gov. In 2011 the law was repealed; it is no longer in effect.

2006: World Health Organization Establishes Trial Registration Policy
In 2006 the World Health Organization (WHO) stated that all clinical trials should be registered, and it identified a minimum trial registration dataset of 20 items and in 2007 launched the International Clinical Trials Registry Platform (ICTRP).

2007: Congress Passes Law (FDAAA) Expanding ClinicalTrials.gov Submission Requirements
In 2007 the requirements for submission to ClinicalTrials.gov were expanded after Congress passed the Food and Drug Administration Amendments Act of 2007 (FDAAA). Section 801 of FDAAA (FDAAA 801) required more types of trials to be registered; additional trial registration information; and the submission of summary results, including adverse events, for certain trials. The law also included penalties for noncompliance, such as the withholding of NIH grant funding and civil monetary penalties of up to $10,000 a day.

2008: ClinicalTrials.gov Releases Results Database
In September 2008, as required by FDAAA 801, ClinicalTrials.gov began allowing sponsors and principal investigators to submit the results of clinical studies.[33]

[33] The submission of adverse event information was optional when the results database was released but was required beginning in September 2009.

2008: Declaration of Helsinki Revision Promotes Trial Registration and Results Dissemination
In October, 2008 the 59th World Medical Association (WMA) General Assembly amended the Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. Two newly added principles (paragraphs 19 and 30) considered the prospective registration and the public disclosure of study results to be ethical obligations.

2009: Public Meeting Held at the National Institutes of Health
In accordance with FDAAA 801, NIH held a public meeting in April 2009 to solicit input from interested individuals about future regulations that will expand the information on ClinicalTrials.gov.

2013: European Medicines Agency Expands Clinical Trial Database to Include Summary Results
In October 2013 the European Medicines Agency (EMA) released a new version of the European Clinical Trials Database (EudraCT). Notably, the EudraCT summary results data requirements are "substantially aligned" with those of the ClinicalTrials.gov results database.

2014: Notice of Proposed Rulemaking (NPRM) for FDAAA 801 Issued for Public Comment
In November 2014 the U.S. Department of Health and Human Services issued a notice of proposed rulemaking (NPRM) describing the proposed requirements and procedures for registering and submitting the results, including adverse events, of clinical trials on ClinicalTrials.gov, in accordance with FDAAA 801.

2014: NIH Draft Policy on Registration and Results Submission of NIH-Funded Clinical Trials Issued for Public Comment.
In November 2014 NIH proposed a policy to ensure that every clinical trial (see the Revised NIH Definition of "Clinical Trial") that receives NIH funding is registered on ClinicalTrials.gov and has summary results submitted and posted in a timely manner, whether subject to FDAAA 801 or not.

2015: National Cancer Institute Issues Clinical Trial Access Policy
In January, 2015 the NIH National Cancer Institute (NCI) issued its Policy Ensuring Public Availability of Results from NCI-supported Clinical Trials. The policy states, "Final Trial Results are expected to be reported in a publicly accessible manner within twelve (12) months of the Trial's Primary Completion Date regardless of whether the clinical trial was completed as planned or terminated earlier."

2016: Final Rule for FDAAA 801 Issued
In September 2016, the U.S. Department of Health and Human Services issued a Final Rule for Clinical Trials Registration and Results Information Submission (42 CFR Part 11) that clarifies and expands the regulatory requirements and procedures for submitting registration and summary results information of clinical trials on ClinicalTrials.gov, in accordance with FDAAA 801. The final rule is intended to make it clear to sponsors, investigators, and the public which trials must be submitted, when they must be submitted, and whether compliance has been achieved.

2016: Final NIH Policy on the Dissemination of NIH-funded Clinical Trial Information Issued
In September 2016, NIH issued a final policy to promote broad and responsible dissemination of information from NIH-funded clinical trials through ClinicalTrials.gov. Under this policy, every clinical trial funded in whole or in part by NIH is expected to be registered on ClinicalTrials.gov and have summary results information submitted and posted in a timely manner, whether subject to FDAAA 801 or not.

APPENDIX D

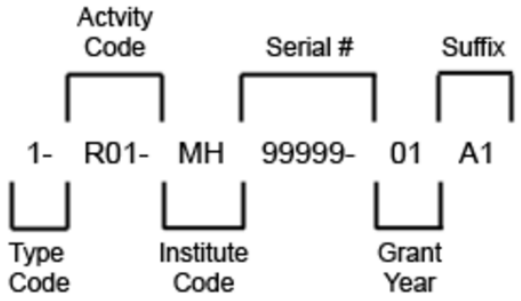
ICD-9 Sub-chapter	Number of trials	Neoplasm (cancer) Sub-chapter
Intestinal Infectious Diseases	402	No
Tuberculosis	414	No
Zoonotic Bacterial Diseases	80	No
Other Bacterial Diseases	1749	No
Human Immunodeficiency Virus (HIV) Infection	2909	No
Poliomyelitis And Other Non-Arthropod-Borne Viral Diseases And Prion Diseases Of Central Nervous System	232	No
Viral Diseases Generally Accompanied By Exanthem	627	No
Arthropod-Borne Viral Diseases	210	No
Other Diseases Due To Viruses And Chlamydiae	3344	No
Rickettsioses And Other Arthropod-Borne Diseases	174	No
Syphilis And Other Venereal Diseases	74	No
Other Spirochetal Diseases	14	No
Mycoses	663	No
Helminthiases	86	No
Other Infectious And Parasitic Diseases	532	No
Late Effects Of Infectious And Parasitic Diseases	3	No
Malignant Neoplasm Of Lip, Oral Cavity, And Pharynx	468	Yes
Malignant Neoplasm Of Digestive Organs And Peritoneum	8793	Yes
Malignant Neoplasm Of Respiratory And Intrathoracic Organs	5891	Yes
Malignant Neoplasm Of Bone, Connective Tissue, Skin, And Breast	9034	Yes
Malignant Neoplasm Of Genitourinary Organs	7110	Yes
Malignant Neoplasm Of Other And Unspecified Sites	9340	Yes
Malignant Neoplasm Of Lymphatic And Hematopoietic Tissue	8981	Yes
Neuroendocrine Tumors	382	Yes
Benign Neoplasms	440	Yes
Carcinoma In Situ	1	Yes
Neoplasms Of Uncertain Behavior	2377	Yes
Neoplasms Of Unspecified Nature	2312	Yes
Disorders Of Thyroid Gland	135	No
Diseases Of Other Endocrine Glands	6639	No
Nutritional Deficiencies	526	No
Other Metabolic And Immunity Disorders	5532	No
Diseases Of The Blood And Blood-Forming Organs	3392	No
Psychoses	2855	No
Neurotic Disorders, Personality Disorders, And Other Nonpsychotic Mental Disorders	4348	No
Intellectual Disabilities	5	No
Inflammatory Diseases Of The Central Nervous System	150	No
Organic Sleep Disorders	257	No

Hereditary And Degenerative Diseases Of The Central Nervous System	3541	No
Pain	228	No
Other Headache Syndromes	33	No
Other Disorders Of The Central Nervous System	2466	No
Disorders Of The Peripheral Nervous System	1024	No
Disorders Of The Eye And Adnexa	2440	No
Diseases Of The Ear And Mastoid Process	393	No
Acute Rheumatic Fever	1	No
Chronic Rheumatic Heart Disease	110	No
Hypertensive Disease	1378	No
Ischemic Heart Disease	1933	No
Diseases Of Pulmonary Circulation	613	No
Other Forms Of Heart Disease	2515	No
Cerebrovascular Disease	1285	No
Diseases Of Arteries, Arterioles, And Capillaries	1179	No
Diseases Of Veins And Lymphatics, And Other Diseases Of Circulatory System	1605	No
Acute Respiratory Infections	455	No
Other Diseases Of The Upper Respiratory Tract	1047	No
Pneumonia And Influenza	1794	No
Chronic Obstructive Pulmonary Disease And Allied Conditions	3159	No
Pneumoconioses And Other Lung Diseases Due To External Agents	18	No
Other Diseases Of Respiratory System	914	No
Diseases Of Oral Cavity, Salivary Glands, And Jaws	841	No
Diseases Of Esophagus, Stomach, And Duodenum	1040	No
Appendicitis	20	No
Hernia Of Abdominal Cavity	20	No
Noninfectious Enteritis And Colitis	1213	No
Other Diseases Of Intestines And Peritoneum	993	No
Other Diseases Of Digestive System	1576	No
Nephritis, Nephrotic Syndrome, And Nephrosis	1508	No
Other Diseases Of Urinary System	1207	No
Diseases Of Male Genital Organs	793	No
Disorders Of Breast	37	No
Inflammatory Disease Of Female Pelvic Organs	816	No
Other Disorders Of Female Genital Tract	1454	No
Ectopic And Molar Pregnancy	12	No
Other Pregnancy With Abortive Outcome	91	No
Complications Mainly Related To Pregnancy	396	No
Normal Delivery, And Other Indications For Care In Pregnancy, Labor, And Delivery	130	No
Complications Occurring Mainly In The Course Of Labor And Delivery	20	No
Complications Of The Puerperium	84	No
Infections Of Skin And Subcutaneous Tissue	205	No

Other Inflammatory Conditions Of Skin And Subcutaneous Tissue	2100	No
Other Diseases Of Skin And Subcutaneous Tissue	1536	No
Arthropathies And Related Disorders	3237	No
Dorsopathies	545	No
Rheumatism, Excluding The Back	1220	No
Osteopathies, Chondropathies, And Acquired Musculoskeletal Deformities	982	No
Congenital Anomalies	789	No
Maternal Causes Of Perinatal Morbidity And Mortality	4	No
Other Conditions Originating In The Perinatal Period	155	No
Symptoms	6901	No
Nonspecific Abnormal Findings	402	No
Ill-Defined And Unknown Causes Of Morbidity And Mortality	195	No
Fractures	134	No
Sprains And Strains Of Joints And Adjacent Muscles	22	No
Intracranial Injury, Excluding Those With Skull Fracture	226	No
Internal Injury Of Thorax, Abdomen, And Pelvis	83	No
Open Wounds	252	No
Injury To Blood Vessels	7	No
Late Effects Of Injuries, Poisonings, Toxic Effects, And Other External Causes	3	No
Superficial Injury	28	No
Contusion With Intact Skin Surface	15	No
Burns	119	No
Injury To Nerves And Spinal Cord	204	No
Certain Traumatic Complications And Unspecified Injuries	138	No
Poisoning By Drugs, Medicinal And Biological Substances	60	No
Toxic Effects Of Substances Chiefly Nonmedicinal As To Source	78	No
Other And Unspecified Effects Of External Causes	2264	No
Complications Of Surgical And Medical Care, Not Elsewhere Classified	515	No
Persons With Potential Healthhazards Related To Communicable Diseases	54	No
Persons With Need For Isolation, Other Potential Health Hazards And Prophylactic Measures	41	No
Persons With Potential Health Hazards Related To Personal And Family History	16	No
Persons Encountering Health Services In Circumstances Related To Reproduction And Development	233	No
Persons With A Condition Influencing Their Health Status	835	No
Persons Encountering Health Services For Specific Procedures And Aftercare	31	No
Persons Without Reported Diagnosis Encountered During Examination And Investigation Of Individuals And Populations	214	No

APPENDIX E

The following explanation of NIH grant numbers are provided by the NIMH website (2017):
The parts of a complete NIH grant number indicate the following: type, activity code, Institute, serial #, grant year, and (possibly) a suffix. For example, the grant number: 1-R01-MH99999-01A1 indicates:



1- This is the **Type Code**. The most common types are:

- 1- never previously funded grants — that is, a new/first time grant application.
- 2- competing continuations — that is, a grant application that was previously funded for a period of time. This new continuing period of support requires peer review.
- 5- non-competing continuations — that is, a grant application that has been funded and is in the midst of its support period. For each year of the support period awarded, there is an administrative review of progress before the next annual installment of support is issued (no peer review is needed). The application that the PI submits as part of this process is called a "non-competing continuation application," and it contains a "progress report" for the period of support just completed.

R01- **Activity Code** indicates the type of grant mechanism. Examples include R01s (investigator initiated research grant), R03s (small grants), R13s (conference support grants), "K"s (career awards), "T"s (institutional training awards), etc.

MH- **Institute Code** identifies the NIH Institute with primary responsibility for payment of this application. For example, MH = National Institute of Mental Health (NIMH) and DA=National Institute on Drug Abuse (NIDA). Each NIH Institute has a two-letter code associated with it.

99999- **Serial Number** provides a unique identification to the project and is assigned sequentially for newly submitted applications. The Serial Number remains the same for as long as a project is active, even when the PI submits a competing continuation for a new period of support.

01- **Grant Year**. "01" indicates the first year of a grant application or funded grant.

A1- **Suffix**. "A1" indicates that the application was submitted once previously but did not receive a sufficiently strong priority score to merit funding. This application is an amended version of the original one also called a "resubmission." At NIH, an R01 may be submitted up to three separate times for review (i.e., an A2 application is the last amended version permitted). Other suffix terms are also used. For example, "S1" refers to a competing supplement request for a currently funded project.

APPENDIX F

Identifying publicly listed firms

In order to understand the "lineage" (ownership histories) of firms, we take advantage of data on a firm's "Ancestor" as provided by the Thompson Reuters Permanent Identifier ("PermID") database. Thompson Reuters describes the database as "a machine-readable identifier developed to create a unique reference for any data item" noting that a "PermID provides comprehensive identification across a wide variety of entity types including organizations, instruments, funds, issuers and people."[34] We match firms in the Cortellis data to the firms' PermIDs: 90.0% of the companies in the Cortellis database have PermID information (137,160 out of 152,357). Of the137,160 companies with PermIDs we matched 99.2% of them with the PermID data. This results in firm-specific data on whether or not a firm is publicly listed. The same database also allows us to observe if a firm has been acquired by a publicly listed firm ("ancestor"). Based on a combination of trial date (from Cortellis) and acquisition data (from the PermID database), we can understand whether a trial was sponsored by a publicly listed firm (*and/) or* whether or not the sponsor was a subsidiary of a publicly listed firm.

As a result of the data considerations described below, we assign upper and lower-bound measures of whether or not a firm was publicly listed at the time of an observed clinical trial as follows.

Firms

	Firm	Ancestor	$Public_0$	$Public_1$	$Public_2$	$Public_3$
		Ancestor (AKA parent) firm observed at time = T	Firm or its ancestor is publicly traded on trial date (unobserved *true* status)	Firm is publicly traded (observed at time = T)	Ancestor is publicly traded (observed at time = T)	Either $Public_1$ or $Public_2$ is TRUE
1.	Pfizer Inc	Pfizer Inc	TRUE	TRUE	TRUE	TRUE
2.	Pfizer Inc (India)	Pfizer Inc	TRUE	FALSE	TRUE	TRUE
3.	Small Bio Corp.	GSK	FALSE	FALSE	TRUE	TRUE
4.	Genentech	Roche	TRUE	FALSE	TRUE	TRUE
5.	Xenoport	Arbor Pharmaceuticals	TRUE	FALSE	FALSE	FALSE

[34] More detail can be found at https://financial.thomsonreuters.com/en/products/data-analytics/market-data/reference-data/permid-data-management.html

6.	ALK-Abello	Lundbeck Founda-tion	TRUE	TRUE	FALSE	TRUE

We use ancestor firms' public status instead of firms' (own) public status assigns legitimate subsidiaries to their parent company's status as wanted (Row 2); however, this method also assigns some acquired firms to an incorrect status.

In Row 3 above, Small Bio Corp. conducts a trial as a privately owned firm at time 0 and is acquired by GSK at time t > 0. Due to data limitations we observe only the most recent firm ancestor (GSK) at time of data collection T > t > 0, and thus the ancestor's public status at time T (TRUE) misrepresents Small Bio's status on the trial date. This is not an issue for firms that were publicly traded before being acquired as long as the acquiring firm is public as well (as in the example in Row 4). This is, however, a complication for firms that were publicly traded and then "delisted" after being purchased by a private firm (as in the example in Row 5).

Rarely, firms are listed as public with non-publicly traded ancestors. This generally indicates partial private ownership of a public firm (as in the example in Row 6).

None of the measures of $Public_j|_{j\in 1,2,3}$ match the unobserved true public status ($Public_0$) for each case, but they can still be useful in a bounding exercise. Because $Public_1$is never TRUE in any case that $Public_0$ is FALSE, it can be used as a lower bound for $Public_0$.

Measure 3 is NOT an upper bound on Measure 0 because, as is the case with Xenoport, $Public_0 = TRUE$ does not imply $Public_3 = TRUE$. However, the true share of trials run by public firms will be bounded above by Measure 3 share as long as there are more trials misclassified as public (due to a later acquisition) than misclassified as private. This is proven below:

$$Share\ Public_3$$

$$= \frac{\#Public\ Trials + \#Misclassified_{Private\rightarrow Public} - \#Misclassified_{Public\rightarrow Private}}{\#Trials}$$

$$If\ \#Misclassified_{Private\rightarrow Public} > \#Misclassified_{Public\rightarrow Private} \Rightarrow$$

$$Share\ Public_3 > \frac{\#Public\ Trials}{\#Trials} = Share\ Public_0$$

So in this case, $Share\ Public_3$ is an upper bound on the true share of trials funded by public firms.

We cannot directly measure the number of misclassified trials to test whether this assumption holds, but because these misclassifications result from mergers and acquisitions, public firms acquiring private firms will likely make up the lion's share of such activity and the bound will hold.

The process by which we calculate dummy variables indicating whether a trial is public by the different measures is outlined below:

1. For each firm
    a. $Public_1 = \mathbf{1}(firm\ is\ public\ in\ 2017);$
    b. $Public_2 = \mathbf{1}(firm's\ ancestor\ is\ public\ in\ 2017)$
2. For each trial and firm recode
    a. $Public_1 = 0\ if\ IPO\ Date > Trial\ Date.$
    b. $Public_2 = 0\ if\ Ancestor\ IPO\ Date > Trial\ Date.$
3. For each firm-ancestor pair calculate $Public_3 = \max\{Public_1, Public_2\}$.
4. For each trial, calculate whether any public firms were involved with the trial:
    a. $Public\ Trial_1 = \max\left\{\{Public_{1j}: j \in J\}\right\}$
    b. $Public\ Trial_3 = \max\left\{\{Public_{3j}: j \in J\}\right\}$ $for\ the\ set\ J\ of\ firm-ancestor\ pairs\ involved\ with\ the\ trial$

APPENDIX SOURCES:

AIDSinfo. Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents. December 1, 2007. Accessed October 2016. Available at: https://aidsinfo.nih.gov/guidelines/html/1/adult-and-adolescent-arv-guidelines/7/hla-b--5701-screening

Basu NN, Ingham S, Hodson J, Lalloo F, Bulman M, Howell A, Evans DG. Risk of contralateral breast cancer in BRCA1 and BRCA2 mutation carriers: a 30-year semi-prospective analysis. Fam Cancer. 2015 Dec;14(4):531–8. doi: 10.1007/s10689-015-9825-9. PubMed PMID: 26239694.

ClinicalTrials.gov. History, Policies, and Laws. July, 2017. Accessed August 2017. Available at: https://clinicaltrials.gov/ct2/about-site/history

Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. Grading Committee. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. Am J Surg Pathol. 2016 Feb;40(2):244–52. doi: 10.1097/PAS.0000000000000530. PubMed PMID: 26492179.

Farrell PM, Rosenstein BJ, White TB, Accurso FJ, Castellani C, Cutting GR, Durie PR, Legrys VA, Massie J, Parad RB, Rock MJ, Campbell PW 3rd. Cystic Fibrosis Foundation. Guidelines for diagnosis of cystic fibrosis in newborns through older adults: Cystic Fibrosis Foundation consensus report. J Pediatr. 2008 Aug;153(2):S4–S14. doi: 10.1016/j.jpeds.2008.05.005. PubMed PMID: 18639722.

FDA-NIH Biomarker Working Group. "BEST (Biomarkers, EndpointS, and other Tools) Resource." (2016).

Gordetsky J, Epstein J. Grading of prostatic adenocarcinoma: current state and prognostic implications. Diagn Pathol. 2016 Mar 9;11:25. doi: 10.1186/s13000-016-0478-2. PubMed PMID: 26956509.

Jameson E, Jones S, Remmington T. Enzyme replacement therapy with laronidase (Aldurazyme®) for treating mucopolysaccharidosis type I. Cochrane Database Syst Rev. 2016 Apr 1;4:CD009354. doi: 10.1002/14651858.CD009354.pub4. PubMed PMID: 27033167.

Khan MJ, Castle PE, Lorincz AT, Wacholder S, Sherman M, Scott DR, Rush BB, Glass AG, Schiffman M. The elevated 10-year risk of cervical precancer and cancer in women with human papillomavirus (HPV) type 16 or 18 and the possible utility of type-specific HPV testing in clinical practice. J Natl Cancer Inst. 2005 Jul 20;97(14):1072–9. doi: 10.1093/jnci/dji187. PubMed PMID: 16030305.

Kujovich JL, Factor V. Leiden thrombophilia. Genet Med. 2011 Jan;13(1):1–16. doi: 10.1097/GIM.0b013e3181faa0f2. PubMed PMID: 21116184.

National Kidney Foundation. K/DOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification and Stratification. Am J Kidney Dis 39:S1-S266, 2002 (suppl 1). Accessed December 2016. Available at: https://www.kidney.org/sites/default/files/docs/ckd_evaluation_classification_stratification.pdf

"NIMH » Research Funding Frequently Asked Questions (FAQs).". https://www.nimh.nih.gov/funding/grant-writing-and-application-process/research-funding-frequently-asked-questions-faqs.shtml (accessed August 8, 2017).

Papageorghiou A, Ohuma E, Gravett M, Hirst J, Silveira M, Lambert A, Carvalho M, Jaffer Y, Altman D, Noble J, Bertino E, Purwar M, Pang R, Ismail L, Victora C, Bhutta Z, Kennedy S, Villar J. International standards for symphysis-fundal height based on serial measurements from the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project: prospective cohort study in eight countries. BMJ. 2016 Oct;355:i5662. doi: 10.1136/bmj.i5662. PubMed PMID: 27821614.

Schiffman M, Glass AG, Wentzensen N, Rush BB, Castle PE, Scott DR, Buckland J, Sherman ME, Rydzak G, Kirk P, Lorincz AT, Wacholder S, Burk RD. A long-term prospective study of type-specific human papillomavirus infection and risk of cervical neoplasia among 20,000 women in the Portland Kaiser Cohort Study. Cancer Epidemiol Biomarkers Prev. 2011 Jul;20(7):1398–409. doi: 10.1158/1055-9965.EPI-11-0206. PubMed PMID: 21602310.

Senior JR. Evolution of the Food and Drug Administration approach to liver safety assessment for new drugs: current status and challenges. Drug Saf. 2014 Nov;37 Suppl 1:S9–17. doi: 10.1007/s40264-014-0182-7. PubMed PMID: 25352324.

Stohl W, Hilbert DM. The discovery and development of belimumab: the anti-BLyS-lupus connection. Nat Biotechnol. 2012 Jan 9;30(1):69–77. doi: 10.1038/nbt.2076. PubMed PMID: 22231104.

Wasung ME, Chawla LS, Madero M. Biomarkers of renal function, which and when? Clin Chim Acta. 2015 Jan 1;438:350–7. doi: 10.1016/j.cca.2014.08.039. PubMed PMID: 25195004.