Volume Title: Economic Dimensions of Personalized and Precision Medicine

Volume Authors/Editors: Ernst R. Berndt, Dana P. Goldman, and John W. Rowe, editor

Volume Publisher: University of Chicago Press

Volume ISBNs: 978-0-226-61106-8 (cloth); 978-0-226-61123-5 (electronic)

Volume URL: http://www.nber.org/books/bern-13

Conference Date: September 13–14, 2017

Publication Date: April 2019

Chapter Title: Characterizing the Drug Development Pipeline for Precision Medicines

Chapter Author(s): Amitabh Chandra, Craig Garthwaite, Ariel Dora Stern

Chapter URL: http://www.nber.org/chapters/c13994

Chapter pages in book: (p. 115 – 157)

5

Characterizing the Drug Development Pipeline for Precision Medicines

Amitabh Chandra, Craig Garthwaite, and Ariel Dora Stern

5.1 Introduction

While lacking a universally agreed upon definition, precision medicine is broadly understood to be an approach to disease treatment and prevention that takes into account variability in environment, lifestyle, and genes for each person.[1] Although the concept of targeted interventions for certain types of patients has a long history across the practice of medicine, recent technological advancements in genetic sequencing, large-scale data storage and analysis, and computing power have made it increasingly possible to tailor the development and utilization of medical technologies. This possibility has drawn attention and funding from beyond traditionally interested parties such as firms and investors in the biotech and pharmaceutical industries. For example, in early 2015, the White House announced a "bold new

Amitabh Chandra is the Malcolm Wiener Professor of Social Policy and director of health policy research at the Harvard Kennedy School of Government and a research associate of the National Bureau of Economic Research. Craig Garthwaite is associate professor of strategy and the director of the Health Enterprise Management Program (HEMA) at the Kellogg School of Management at Northwestern University and a research associate of the National Bureau of Economic Research. Ariel Dora Stern is assistant professor of business administration and Hellman Faculty Fellow in the technology and operations management unit at Harvard Business School.

We are grateful to Brian Alexander, Can Huang, Jennifer Kao, Rebecca Sachs, Mark Trusheim, a helpful anonymous reviewer, 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, Lila Kelso, Andrew Marder, Caroline Marra, and Alice Ndikumana provided excellent research and programming assistance. For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see http://www.nber.org/chapters/c13994.ack.

1. https://www.nih.gov/research-training/allofus-research-program.

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 to several billions of US dollars over the coming years. Major investments to advance precision medicine have also been announced by a number of US 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*, that is, those therapies focused on biomarker-defined patient subgroups. Precision medicines—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 an understanding of the "-omic" (i.e., 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 subgroups of patients. Notably, the public efforts noted above have lagged private endeavors in this area: the pharmaceutical industry[4] has already commercialized almost 150 drugs with pharmacogenomic information in their label, according to the US 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 better match between individuals with specific disease subtypes and medications that are more effective for those subtypes. While the science underlying these medicines is broadly interesting and is the subject of a growing body of research, the 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, changing

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.

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

incentives can have far-reaching implications on the entire pharmaceutical industry.

Perhaps most importantly, the ability to develop more targeted products may influence the decision process that determines which new drugs firms attempt 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, all else equal—and thus are more likely to justify the fixed costs of developing the medication. These higher-priced products are likely to include those with large clinical benefits, which may be more apparent in readily identifiable 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 do not 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 potential entrants find the prospects of competing for a small(er) market to be an unattractive economic opportunity. Perhaps more concerning for price competition, a similar dynamic could exist for the eventual generic and biosimilar 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, theoretically, 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 with an existing biomarker and increase the subset of early stage products that pharmaceutical manufacturers would consider as candidates for commercialization.

In addition, greater expected therapeutic benefit may result in smaller and/or shorter clinical trials because fewer patients would be needed and/or shorter periods of time would 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, along with innovations in clinical trial designs—such as adaptive platform trials, which are particularly well-suited to precision medicine approaches—could counteract some of the negative entry incentives that might be created by small patient populations, however the costs associated with trial recruiting are likely to be higher than traditional "all-comers" trials that include a broad set of patients with a given disease.

Despite the potential for precision medicines to both reduce some of the costs of drug development and also increase the patient value 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. More generally, 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 patterns in development efforts of likely precision medicines (LPMs), those pipeline drugs whose clinical trials have signature features of precision medicine R&D. We identify and report on the frequency of clinical trials for such products 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.

6. Patent life for a drug in the United States is generally twenty 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) is typically counted against the twenty-year patent life. Marketing exclusivity is different from patent life and is granted by the FDA upon drug approval. Exclusivity typically lasts for five 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).

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 consider the characteristics (e.g., geography, indication, sponsorship) of clinical research between LPM versus non-LPM trials over the years covered in our data. Finally, we consider the types of firms pursuing clinical trials in LPMs, considering how LPM R&D activities have evolved over recent years.

5.2 Precision Medicines and the Drug Development Process

As discussed above, we focus here 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.

5.2.1 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, as 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 cystic fibrosis transmembrane conductance regulator (CFTR) 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 1,700 mutations in the gene that causes the disease. This amounts to approximately 3,500 potential patients in the United States.[8]

Many of the biomarkers that are associated with the use of precision

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

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

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—for example, 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).

In response to the growing therapeutic 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 2016 (FDA and NIH 2016). Appendix A (http://www.nber.org/data-appendix/c13994/appendix.pdf) 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 may have different economic implications. Biomarkers can reveal useful information about disease diagnosis and prognosis. They can also be used to predict the treatment efficacy or toxicity of a therapy, serve as markers of disease progression, and can serve as auxiliary (or so-called surrogate) endpoints in clinical trials. Further, some biomarkers can be used in more than one way, while others have just one known role.[10]

Other work has discussed how different parties in the US health care system are (or are not) incentivized to develop biomarkers—including discovery and establishment (see, e.g., Stern, Alexander, and Chandra 2018). While all of these applications of biomarkers have the potential to shape the practice of personalized medicine and may help improve drug development and clinical practice, only a small subset will have the potential to assist in the development of precision *medicines*, those therapies targeted at specific patient populations who are more likely to benefit. For the development of

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

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 5.2 and 5.3.

precision medicines, it is useful to consider which subset of biomarkers are likely to be of value to pharmaceutical innovators in bringing new therapies to market—either because the use of biomarkers will lead to a higher probability of drug approval or higher expected profits, given approval. In both cases, these tend to be the types of biomarkers that can be used for diagnosis and prognosis as well as predictive biomarkers, which are those that can be "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" (FDA-NIH 2016). It is these latter groups of biomarkers—and the clinical trials driven by their use—that we specifically consider in the empirical analysis below.

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 completion of 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 and harder-to-find target patient populations. Additionally, trial design and execution can be significantly more complex when a companion diagnostic (used to measure and/or identify the biomarker *itself*) needs to be approved alongside the drug (Fridlyand et al. 2013). 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.

5.2.2 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 twenty-two recent years. Clinical trials oriented toward 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 premarket 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 one to four 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 submit a New Drug Application (NDA) or Biologics License Application (BLA) to the FDA that includes the full set of results from the product's preclinical and clinical studies. The FDA then has up to ten months to review the application and determine whether to grant marketing approval.[12]

5.2.3 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] 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 investment to differ among LPM trials? The 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

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. The "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. The "Breakthrough Therapy" designation is given 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 used 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. Finally, "Priority Review" requires the FDA to review marketing applications within six months rather than the standard ten, and is available in a number of circumstances. https://www.fda.gov/downloads/Drugs/Guidances/UCM358301.pdf.

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. 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."

other sources of funding. Further, as trials target increasingly specific subpopulations of patients, the operational costs and complexity of running clinical trials may increase, either of which could further shift small companies toward specializing in early stage R&D activities and letting larger, more established (e.g., publicly listed) companies focus on the final stages of regulatory approval and product commercialization. We study the role played by large firm actors in the development of LPMs and ask if these roles have changed over recent decades.

5.3 The Economics of Precision Medicine

As previously noted, not all biomarker uses are associated with precision medicines. Here, we outline some simple economics of precision medicine to better understand how and why biomarkers are important for understanding the potential future of the pharmaceutical market.

Biomarkers that constitute surrogate endpoints help manufacturers by speeding up clinical trials—for example, through the use of the FDA's accelerated approval process, whereby a product can be approved on the basis of intermediate patient outcomes that are a good proxy for a therapy's ultimate effectiveness.[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] For the most part, the effect of the types of biomarkers that can be used as surrogate trial endpoints has been and is likely to remain 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, since a therapeutic effect will be easier to detect as a result of the greater putative efficacy in the indicated population, as well as a tendency to change expected market sizes once the frequency of that biomarker in the broader disease population is known. Further, as we have noted elsewhere, such biomarkers could facilitate indication-based pricing, which could expand access to

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).

some patients, but would also mean that higher prices could be charged for patients with a biomarker that indicates the drug will be more effective (Chandra and Garthwaite 2017).

In this setting, biomarkers can facilitate a drug market being segmented into identifiable groups based on the expected efficacy of the product, and as a result a segmentation of patients by willingness to pay for the product. When pharmaceutical manufacturers are able to charge only a single price, the existence of known, distinct patient subgroups would effectively allow firms to choose which patients to serve. For example, where the population receiving lower (but positive) value is quite large, the manufacturer may choose to set a low price and sell to a larger market. However, when the lower-value population is quite small, the manufacturer may instead choose a higher price and forgo sales to those patients who derive the least value from the product. Economists will note that this represents the classic monopolist's dilemma, where pharmaceutical firms trade margins for quantity.

For this reason, firms often 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 facilitating price discrimination. When such price discrimination is feasible, the most extreme outcome is that a manufacturer would be 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.

In a world where a product with a biomarker exists, an indication-based pricing strategy weakly increases the profits of firms. As a result, the expanded use of biomarkers has the potential to provide additional incentives to develop products that would otherwise not be commercialized. The broad contours of this type of price discrimination are illustrated through a fictional example presented in appendix B (http://www.nber.org/data-appendix/c13994/appendix.pdf).[19]

However, the profit implications of the decision to pursue a biomarker-based product in the first place is more complicated. First, it is important to note that the assumption that price discrimination is weakly profit maximizing for an innovator firm is based on the fact that this firm always has the ability to abandon a price discrimination scheme if it proves to be unprofitable. However, a firm cannot as easily discard information about the efficacy of a product that is commercialized through a biomarker-driven trial. Thus,

19. This figure depicts the monetary value of a hypothetical product for three different indications (e.g., 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.

depending on the distribution of customers and the relative efficacy of the drug across subpopulations, it could be that ex post a firm finds that the use of a biomarker as part of its R&D strategy decreases the profits from a drug compared to a counterfactual where the drug was only priced based on the average efficacy of the product in the broad patient population (i.e., one where the knowledge from the biomarker did not exist).

Yet firms must choose whether or not to pursue use of a biomarker for patient selection under some degree of uncertainty about its likely consequences. As a result, the set of trials that use biomarkers in pursuit of new therapeutic approvals is likely to be a systemically selected subset of the possible trials that *could* have used biomarkers for patient selection, as firms make educated guesses about how the use of a biomarker will affect the size of their eligible patient population and the probability of new product approval—where the two factors are likely to work in opposite directions.

This type of trade-off is not simply a theoretic point; a situation with many of these features has played out in recent development projects for PD-1 inhibitors pembrolizumab (Keytruda) and nivolumab (Opdivo), anticancer immunotherapies from Merck and Bristol-Myers Squibb (BMS), respectively.[20] Bristol-Myers Squibb pursued "all-comers" trials for nivolumab—that is, the firm decided to forgo biomarker selection in favor of broader indications. Merck, however, opted to commercialize pembrolizumab through a series of biomarker-enriched trials (the trials selected for patients with positive PD-L1 expression). Ultimately, Merck's drug has been more successful—in part due to a series of blanket approvals by the FDA in mid-2017 for patients with *any* cancer with a specific molecular signature, the FDA's "first tissue/site agnostic indication" approval.[21] Thus, although the use of biomarkers for selection led to a more profitable drug for Merck, it was not clear, ex ante that this would be the case. More generally, personalization may, in certain cases, ultimately reduce profitability—but not necessarily costs—of some therapies.

Pricing aside, biomarkers that predict treatment efficacy reduce market size, which in turn may reduce some of the incentives for innovation. At the same time, some biomarkers could allow manufacturers to more easily qualify for an "orphan drug" designation through the Orphan Drug Act of 1983 (ODA) by focusing on developing a therapy for a disease subpopula-

20. PD-1 is a checkpoint protein, which prevents a patient's T-cells from attacking cancer and other cells in the body. PD-1 is described as being something like an "off switch," which binds to PD-L1, a protein that is on both normal and cancer cells. Monoclonal antibodies can be used therapeutically to bind to either the PD-1 checkpoint protein or PD-L1, preventing binding and allowing the immune system to target cancer cells. https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types/immunotherapy/immune-checkpoint-inhibitors.html).

21. https://www.fda.gov/drugs/informationondrugs/approveddrugs/ucm560040.htm. For a detailed account of the pembrolizumab development process, see: https://www.forbes.com/sites/davidshaywitz/2017/07/26/the-startling-history-behind-mercks-new-cancer-blockbuster/#7ef5cbb8948d.

tion of 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 percent of clinical trial expenses and seven years of marketing exclusivity (two years longer than the standard five years granted for nonorphan 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,[22] and in 2015 alone 47 percent of novel drugs approved were orphan drugs.[23] When an approval happens, it will also raise spending and reduce price competition over the medium term due to the (extended) protections from generic competition offered by the ODA, and the fact that smaller markets will attract less follow-on competition.[24] In particular, 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. To some extent this phenomenon has already been observed in early biosimilar competition in the European Union (Scott Morton, Stern, and Stern 2018; Berndt and Trusheim 2015). Thus, even after exclusivity periods end, there may not be a large enough market to stimulate price competition through follow-on (i.e., generic or biosimilar) entry.[25] As a result, a major shift in new therapeutics toward precision medicines could result in less price competition through a meaningful decline in the attractiveness of follow-on competition and, as a result, a meaningful increase in total drug spending.

While conventional wisdom suggests that the use of biomarkers will lead to smaller target markets, we note that there are settings in which the use of biomarkers might serve to expand the size of the potential patient market. This would happen if a drug has side effects that discourage physicians from using it, but a biomarker identifies patients who do not suffer these side effects or in whom the drug is particularly effective. Such approvals are likely to become more common through the growth and adoption of new clinical trial designs such as "basket trials." In a basket trial, patients are enrolled based on a shared *mutation*, regardless of their disease type—that is, patients with colon and lung cancer, as well as the same mutation in a particular gene, would be included in the same trial (West 2017). When a certain mutation or protein expression is relatively common across cancers or autoimmune diseases, biomarker-enriched trials may ultimately serve to drive additional indications, or faster approvals. Both would increase

22. https://www.accessdata.fda.gov/scripts/opdlisting/oopd/index.cfm.
23. http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/ucm474696.htm.
24. For additional discussion of the implications of the ODA, see Bagley et al. (forthcoming).
25. 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 (2018), 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).

market size because manufacturers will not have to seek indications on an individual basis. However, such "market expanding" precision therapies are still relatively rare (notably, the FDA's first tissue/site agnostic indication only occurred in 2017). While biomarker-driven market expansion may be important looking ahead, we expect the LPM development pipeline in the years we study here to mostly be characterized by the types of products that are likely to shrink rather than expand target patient populations.

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

5.4 Data

We use data from the Cortellis Competitive Intelligence Clinical Trials Database (Cortellis), which is compiled by Clarivate (and formerly by Thomson Reuters). The database includes over 270,000 global and US-based clinical trials. Cortellis includes full coverage of twenty-four clinical trial registries from around the world, including ClinicalTrials.gov, which is maintained by the National Institutes of Health (NIH) and the 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 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 relevant development efforts—in particular, in the years since 2005, after which time the ICMJE required trial registration in order to publish the results of clinical trials in member journals.[26] Cortellis has full coverage of all ICMJE-approved trial registries (Clinical Trial Registration 2016) and Cortellis data have been used in several published studies in peer-

26. 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 et al. 2004). However, we note that certain types of trials—for example, 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 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.

reviewed biomedical journals such as *Lancet Infectious Disease* (Phyo et al. 2016) and *Nature Reviews Drug Discovery* (Bespalov et al. 2016). Appendix C (http://www.nber.org/data-appendix/c13994/appendix.pdf) includes a detailed timeline of important dates related to the registration of clinical trials and the establishment of the US clinical trial registry (clinicaltrials.gov).

5.4.1 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 twenty-two calendar years of clinical trial starts. We identify the full set of known Phase I, II, and III[27] clinical trials, along with other detailed 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 5.1), and in particular for Phase II trials.[28] In 2016, roughly 6,000 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 intended disease indication(s), any biomarkers used in the trial, and the trial's sponsors. In addition, we are able to classify trials according to a 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 categorize its type and use (role). We also capture key information about the clinical trial's sponsors. Trial sponsors are identified by name and type, including academic investigators, government, nongovernment, company, and other sponsors. A complete list of the descriptive variables we consider and their frequencies in the clinical trials data set are provided in table 5.1.

To aggregate the detailed indications reported in the Cortellis database into more usable categories, we used a data set[29] of standardized indications matched to International Classification of Diseases (ICD)-9 codes to link each trial in our data set to a three-digit ICD-9 code. The matched indication ICD-9 data set was independently checked for accuracy by three research

27. For the simplest classification of trials 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 as Phase II in the year that the trial launched. In robustness tests, we create separate subcategories 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 three-phase classification in tables and figures for simplicity.

28. The recent spike in the number of global clinical trials (and Phase II trials in particular) is driven by growth in non-US trials (see appendix tables for a version of figure 5.1 that presents only US trials; http://www.nber.org/data-appendix/c13994/appendix.pdf).

29. We are grateful to Manuel Hermosilla, Craig Garthwaite, and David Dranove, who generously shared their version of a three-digit ICD-9 crosswalk data set with us. This data set 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.

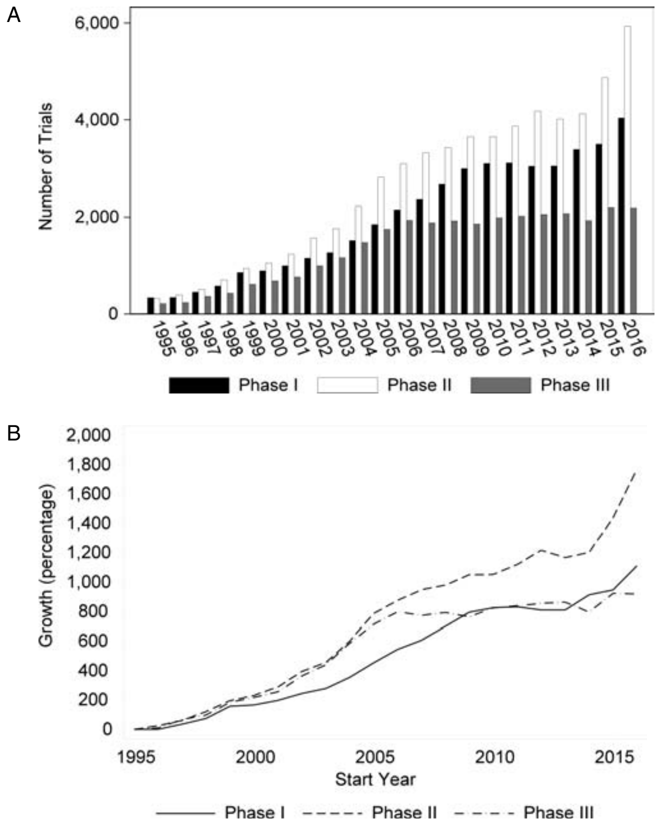

Fig. 5.1 Clinical trials over time. *A*, number of registered Phase I–III trials (1995–2016); *B*, growth in number of registered Phase I–III trials since 1995.

assistants using an online ICD-9 medical coding reference manual,[30] 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 sixty-five ICD-9 subchapters (listed in appendix D; http://www.nber.org/data-appendix/c13994/appendix.pdf). Trials with any indications matching ICD-9 codes 140–239 were classified as cancer trials.

The categorical variable "biomarker type" indicates the biological feature that a given biomarker identifies. Biomarker types include genomic, proteomic, biochemical, cellular, physiological, structural, 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 5.2. Importantly, these types are not mutually exclusive since a given

30. ICD9Data.com.

Table 5.1 **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 I clinical (includes Phase I/Phase II trials)	0.3305	131,971	0.3653	49,540
Phase II clinical (includes Phase II/Phase III trials)	0.4367	131,971	0.4263	49,540
Phase III clinical	0.2328	131,971	0.2083	49,540
Trial site in United States	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 5.2 **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

Notes: Biomarker types are defined as follows. 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.

biomarker—for example, 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 expressions 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 presence of biomarkers types across trials.

5.4.2 Biomarker Data and Defining Pipeline Precision Medicines

The Cortellis data include fairly broad categories of biomarker uses as they 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 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 LPMs is that they encompass the set of pipeline products that are being developed using the types of disease-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 database (DBD) from Clarivate includes 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. In the database, both of these are included as detailed biomarker roles for using the biomarker HER2 in studies of breast cancer (the indication). In the Cortellis database, we can then link DBD data as follows: given a trial's breast cancer indication and knowing that the HER2 biomarker was used in that clinical trial, one can assign both a biomarker type and a detailed biomarker role (or, in some cases, more than one) to that trial. In this way, assignment of biomarker roles (from the DBD) to trials (in Cortellis) is achieved via a matching algorithm that requires a precise match between both the indication and biomarker. For example, in order to link trial x to biomarker role y, we would match as follows:

$$\text{trial}_x + \underline{\text{biomarker}_{BIO} + \text{indication}_{IND}} \leftrightarrow \underline{\text{biomarker}_{BIO} + \text{indication}_{IND}} + \text{biomarker role}_y$$

where the terms on the left represent variables in the Cortellis database, the terms on the right represent variables in the DBD, and the underlined terms (indication_{IND} and biomarker_{BIO}) are used as exact matching criteria.

Definitions of detailed biomarker roles and the frequencies of their use in clinical trials are reported in table 5.3. 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 FDA-NIH definitions of biomarkers and how they are employed.[31]

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 "generous" trials that *specifically employ biomarkers for prediction* (these include biomarkers whose roles include 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 5.3). To be clear, both definitions measure *likely* precision-medicine trials, but the former, more inclusive definition includes trials in which biomarkers were used to define and select patient populations for the trial and this role may not fit everyone's perception of a precision medicine trial (e.g., selecting patients with a certain disease type or patients whose presentation of a biomarker, such as a protein, is measured as above/below a cutoff relevant to disease diagnosis).

5.5 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

31. The classification definitions were also separately discussed with an oncologist, who at the time of writing was serving as the principal investigator on a biomarker-driven clinical trial.

Table 5.3 **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

Notes: 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.

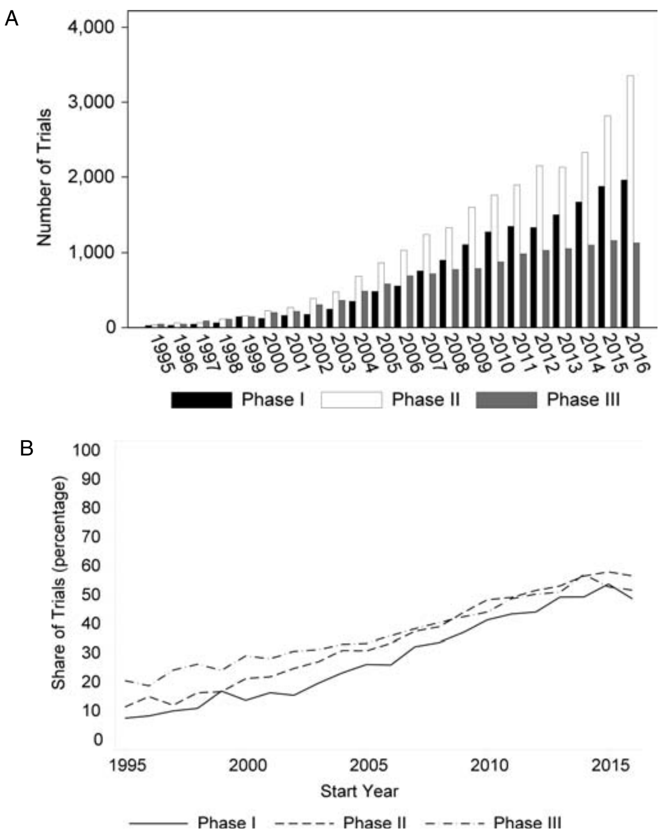

Fig. 5.2 Clinical trials employing biomarkers. *A*, number of registered Phase I–III trials using at least one biomarker; *B*, share of trials using at least one biomarker.

of precision medicine are expected to grow in coming years, we separately characterize the cancer applications of pipeline precision medicines in detail.

5.5.1 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 5.2). 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 trials with biomarker uses that are associated with LPMs, by both the generous and restrictive definitions (figure 5.3). Both the number and percentage of LPM trials increased over our period of observation, as seen in figure 5.3. We further note that the two definitions of LPMs track each other closely over time—both in figure 5.3 as well as in the subsequent subsample analyses described below. Table 5.4 presents the count (column [1]) and percentage

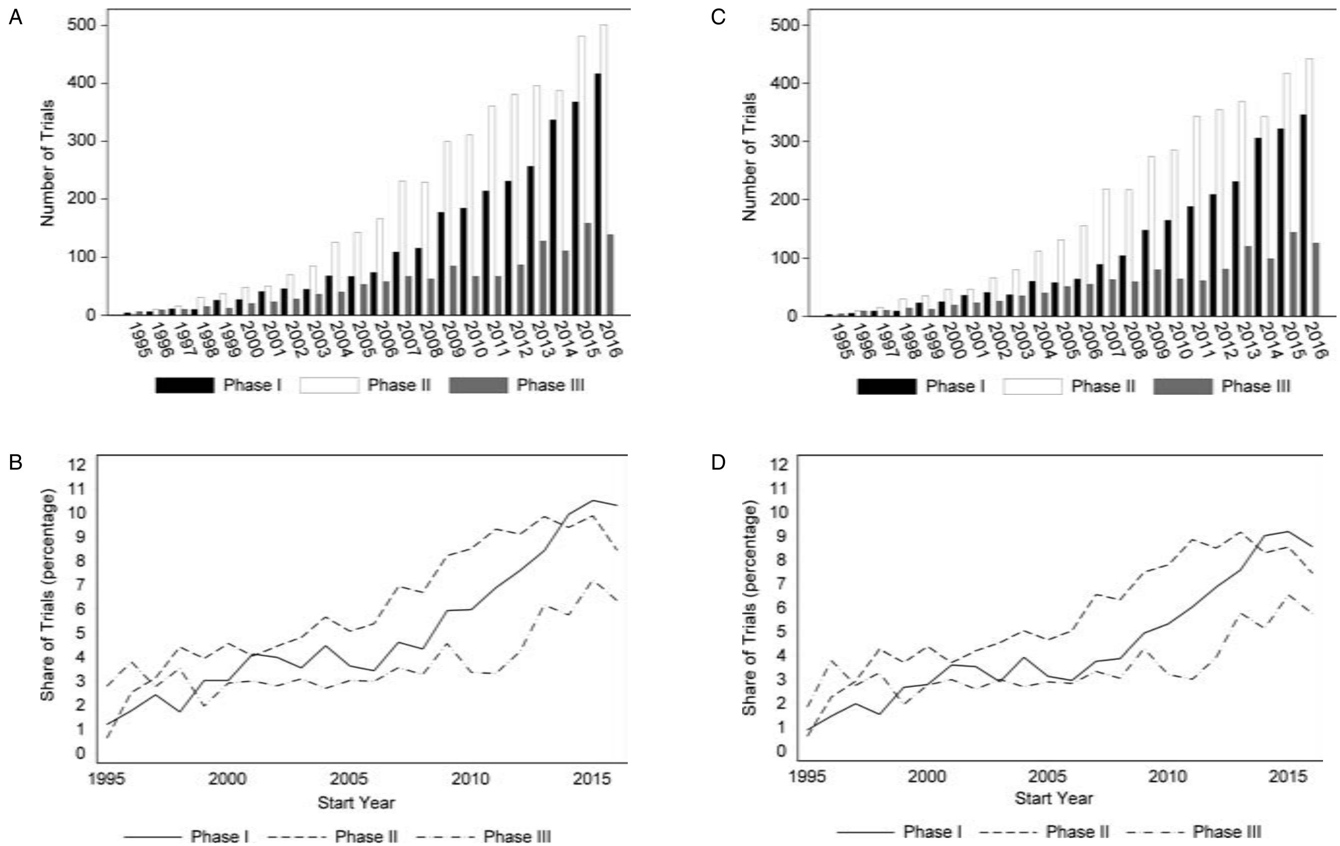

Fig. 5.3 Clinical trials for LPMs. ***A*****, number of registered LPM (generous definition) trials by phase;** ***B*****, share of trials with LPM biomarkers (generous definition);** ***C*****, number of registered LPM (restrictive definition) trials by phase;** ***D*****, share of trials with LPM biomarkers (restrictive definition).**

Table 5.4 **Likely precision medicine (LPM) trials (1995–2016)**

	All count	All (%)	P1 count	P1 (%)	P2 count	P2 (%)	P3 count	P3 (%)
	Generous definition							
1995	12	1.39	4	1.20	2	.63	6	2.79
1996	25	2.58	6	1.77	10	2.53	9	3.81
1997	37	2.80	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	3.10
2004	234	4.49	68	4.48	126	5.68	40	2.71
2005	263	4.10	67	3.63	143	5.09	53	3.03
2006	299	4.17	74	3.44	167	5.40	58	3.00
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	7.60	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	9.40	111	5.76
2015	1,009	9.55	368	10.50	482	9.89	159	7.23
2016	1,057	8.69	417	10.30	501	8.44	139	6.35
	Restrictive definition							
1995	9	1.04	3	.89	2	.63	4	1.86
1996	23	2.37	5	1.47	9	2.28	9	3.81
1997	34	2.57	9	2.00	15	2.94	10	2.77
1998	53	3.11	9	1.55	30	4.29	14	3.30
1999	70	2.91	23	2.68	35	3.73	12	1.96
2000	90	3.43	25	2.80	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	4.90	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	8.50	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

(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 percent of trials were for LPMs, roughly double the percentage observed a decade earlier (3.8 percent).

The 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 5.4 shows how these types were represented in each phase over our years of observation. Genomic/proteomic biomarkers were the most commonly used in recent years and featured in the vast majority of LPM trials, a fact that is consistent with LPMs being driven primarily by understanding gene and protein expression and how these factors predict the likely success of medications.

5.5.2 Pipeline Precision Cancer Therapies

Figure 5.5 and table 5.5 present data on the frequency of LPMs in cancer trials only. Several features of these trials are notable, especially when compared with one another. First, LPM trials are more than an order of magnitude more common in cancer indications; in 2015 and 2016, roughly 25 percent (or more) of all cancer drug trials were LPM trials, but only 1–2 percent of trials for noncancer indications were LPM trials. In regression analysis (table 5.8), we also see that a cancer indication is a strong statistical predictor of an 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 accord with the commonly held belief that the majority of applications of precision medicines in coming years will be in the context of targeted therapies for cancer.

5.5.3 Institutional Factors

Next, we consider the LPM development pipeline in light of a number of other institutional factors. We consider US-based versus 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 by both local demand (Costinot et al. 2016) and local regulations (FDA 2004). Figure 5.6 shows the number and share of US 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 US trials is roughly double that of international trials. This finding is consistent with the fact that US drug prices are typically higher than those of other countries (Kanavos et al. 2013), making it more appealing for pharmaceutical manufacturers

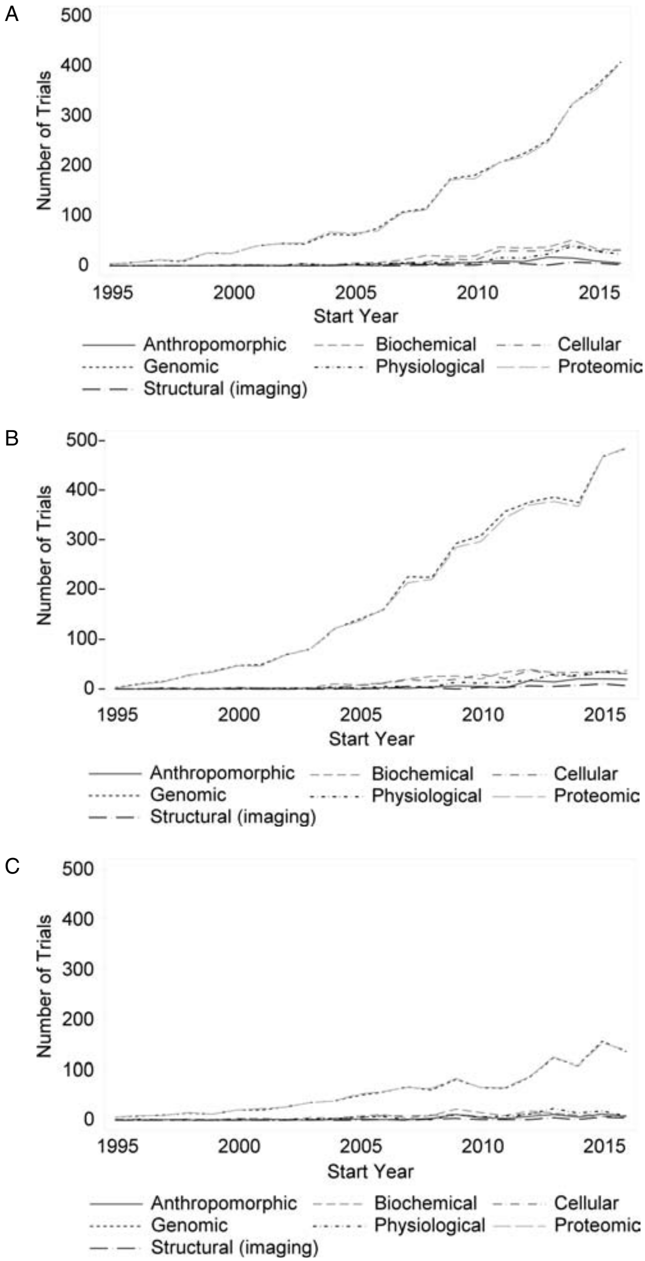

Fig. 5.4 Types of biomarkers used in LPM trials. *A*, number of Phase I LPM trials (generous definition) by biomarker types used; *B*, number of Phase II LPM trials (generous definition) by biomarker types used; *C*, number of Phase III LPM trials (generous definition) by biomarker types used; *D*, number of Phase I LPM trials (restrictive definition) by biomarker types used; *E*, number of Phase II LPM trials (restrictive definition) by biomarker types used; *F*, number of Phase III LPM trials (restrictive definition) by biomarker types used.

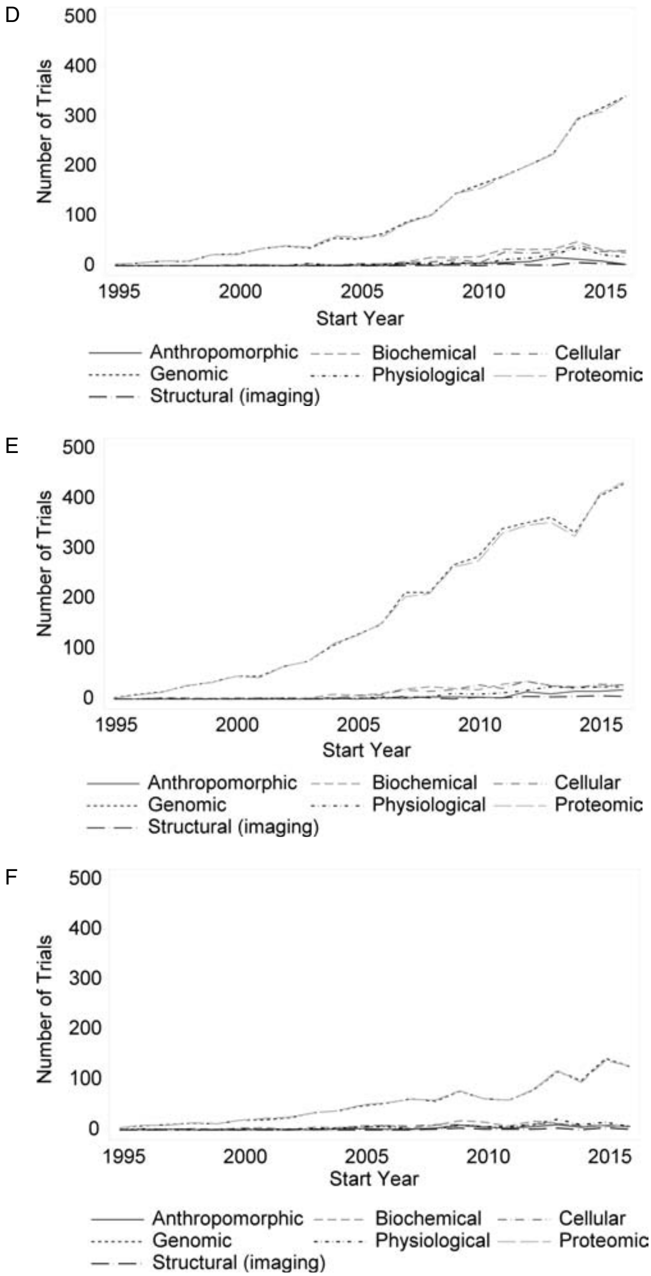
D
500
400
300
200
100
0
Number of Trials
1995
2000
2005
2010
2015
Start Year
Anthropomorphic
Biochemical
Cellular
Genomic
Physiological
Proteomic
Structural (imaging)
E
500
400
300
200
100
0
Number of Trials
1995
2000
2005
2010
2015
Start Year
Anthropomorphic
Biochemical
Cellular
Genomic
Physiological
Proteomic
Structural (imaging)
F
500
400
300
200
100
0
Number of Trials
1995
2000
2005
2010
2015
Start Year
Anthropomorphic
Biochemical
Cellular
Genomic
Physiological
Proteomic
Structural (imaging)

Fig. 5.4 (cont.)

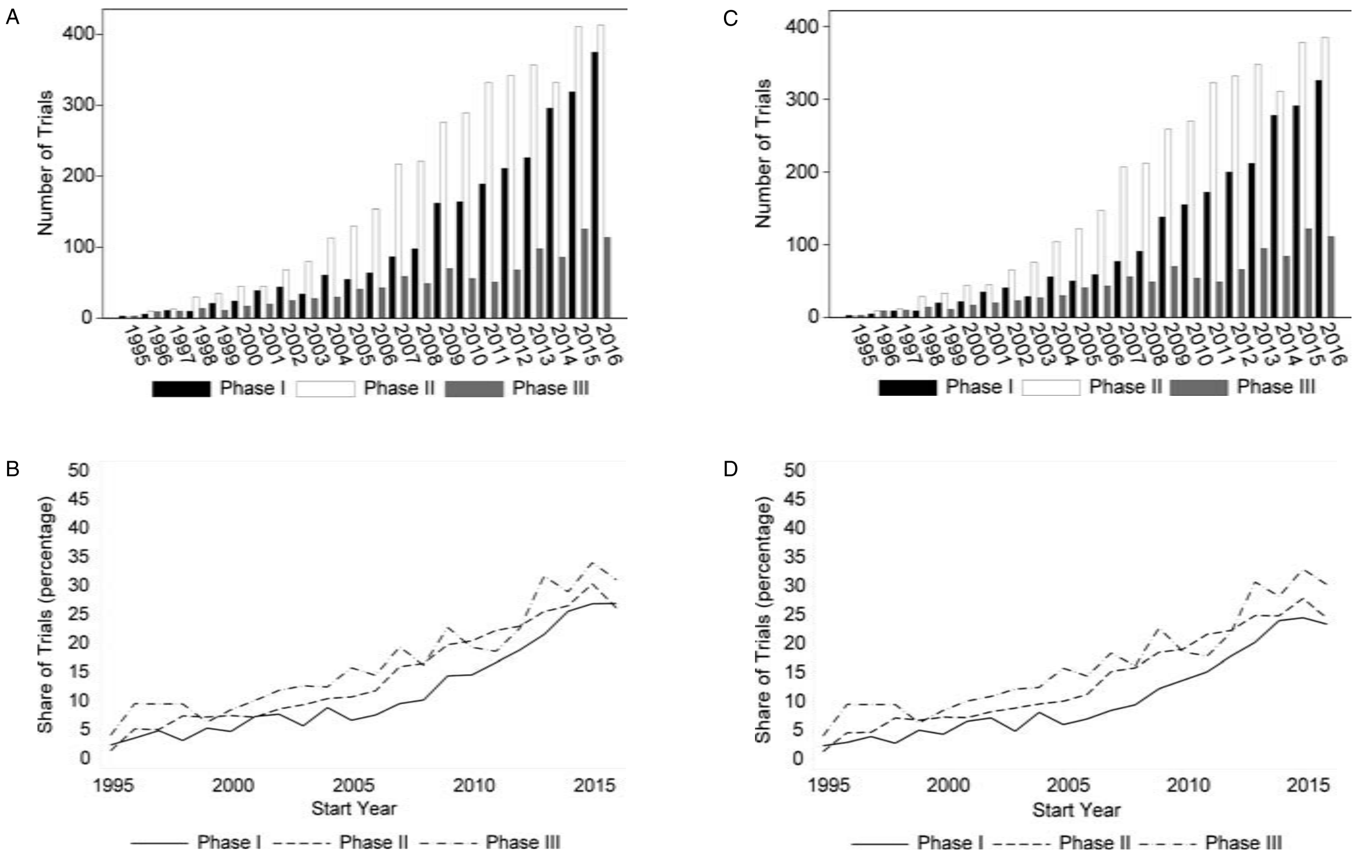

Fig. 5.5 Clinical trials for LPMs, cancer indications only. ***A*****, number of registered LPM (generous definition) trials by phase: cancer trials;** ***B*****, share of cancer drug trials with LPM biomarkers (generous definition);** ***C*****, number of registered LPM (restrictive definition) trials by phase: cancer trials;** ***D*****, share of cancer drug trials with LPM biomarkers (restrictive definition).**

Table 5.5 **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	5.30	6	3.57	10	5.21	8	8.60
1997	34	5.81	11	4.89	13	5.10	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	8.50
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	9.60	217	16.1	59	19.4
2008	368	14.2	98	10.2	221	16.6	49	16.3
2009	507	18.0	162	14.4	275	19.8	70	22.9
2010	509	18.0	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	23.0	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	27.0	410	30.5	126	34.3
2016	899	27.1	375	27.0	410	26.1	114	31.4
	Restrictive definition							
1995	8	2.33	3	2.38	2	1.39	3	4.05
1996	22	4.86	5	2.98	9	4.69	8	8.60
1997	31	5.30	9	4.00	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	8.50
2001	100	7.45	35	6.63	45	7.28	20	10.2
2002	129	8.26	41	7.23	65	8.28	23	11.0
2003	132	7.93	29	4.88	76	8.94	27	12.2
2004	190	9.51	56	8.20	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	8.50	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	17.0	155	13.8	270	19.2	54	18.7
2011	544	18.9	172	15.2	323	21.8	49	18.0
2012	598	20.7	200	18.0	332	22.4	66	22.1
2013	654	23.8	212	20.3	347	25.0	95	30.8
2014	673	25.0	278	24.0	311	25.0	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

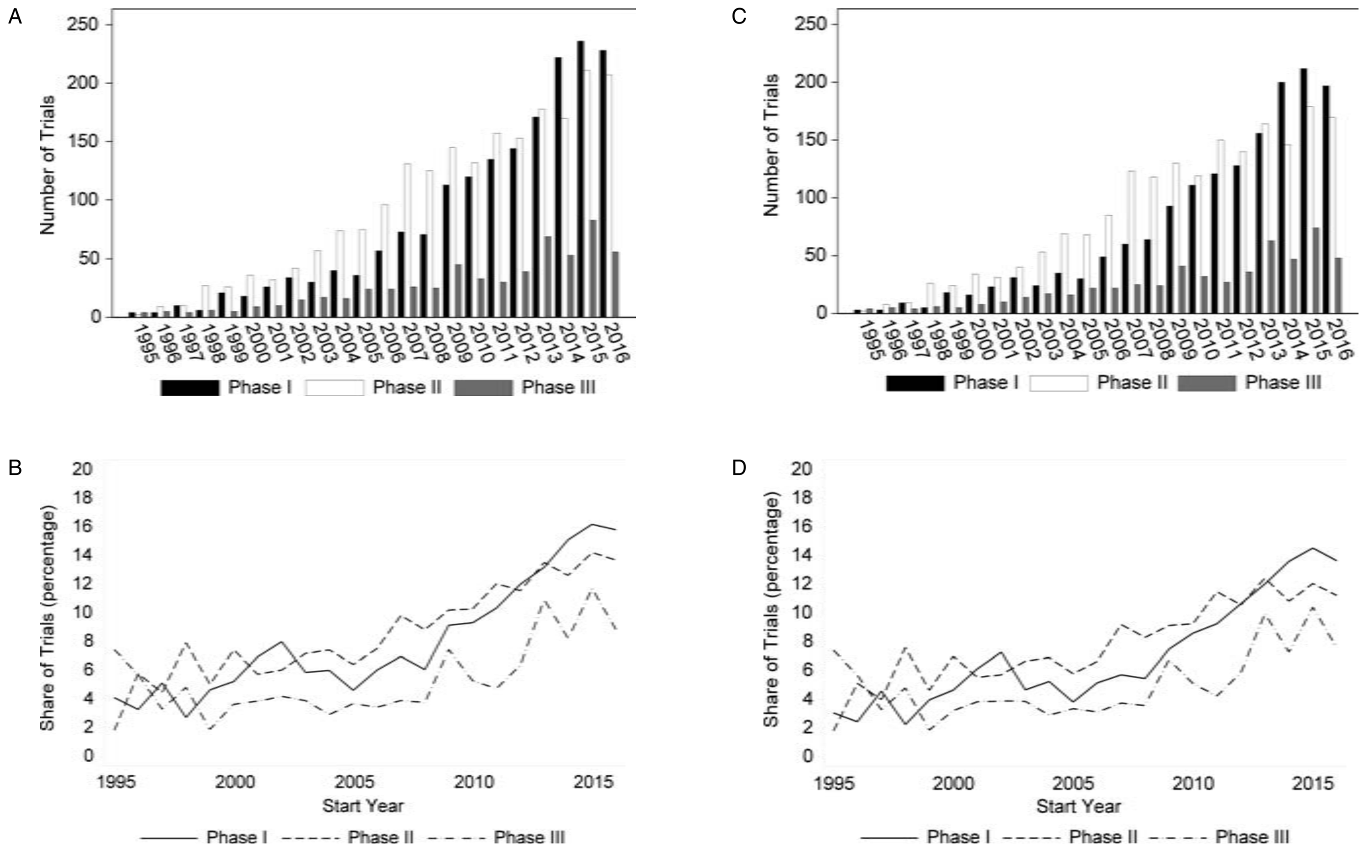

Fig. 5.6 Clinical trials for LPMs, US trials only. *A*, number of registered LPM (generous definition) trials by phase: US trials; *B*, share of US drug trials with LPM biomarkers (generous definition); *C*, number of registered LPM (restrictive definition) trials by phase: US trials; *D*, share of US drug trials with LPM biomarkers (restrictive definition).

Table 5.6 **Burden of disease: millions of years of life lost for associated diseases (average)**

	United States only	Global
Non-LPM	11.66	188.20
LPM	14.65	202.03
t-statistic	19.30	5.57

to conduct clinical research in order to bring drugs to market in the United States as soon as possible. These facts are also reflected in our regression analysis (table 5.8), which indicates that US trials are, on average, roughly 1 percentage point more likely to involve LPMs at any point in time relative to their non-US counterparts in the same year, all else equal. Off of a relatively low overall share of LPM trials, this difference represents a double-digit percentage increase in the likelihood of observing an LPM trial.

In addition, we briefly consider whether LPM trials appear to be related to disease severity.[32] 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.[33] For both the United States (alone) as well as globally, we assemble data on years of life lost (YLL) due to each cancer.[34] For all cancers, we identify the relevant ICD-9 code and match YLL to the cancer trials in our data (as described above and in the eleven cancer ICD-9 subchapters presented in appendix D; http://www.nber.org/data-appendix/c13994/appendix.pdf). Table 5.6 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 US 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 versus (typically smaller) privately held firms—engaging in the development of LPMs (figure 5.7 and tables 5.7A and 5.7B). The correct assignment of individual trials to their sponsor firms (and according to firm types) is both difficult and fundamental for our analysis. Because acquisitions are common and firm ownership may change over time, we use imputation to assign each trial in our data set 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

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

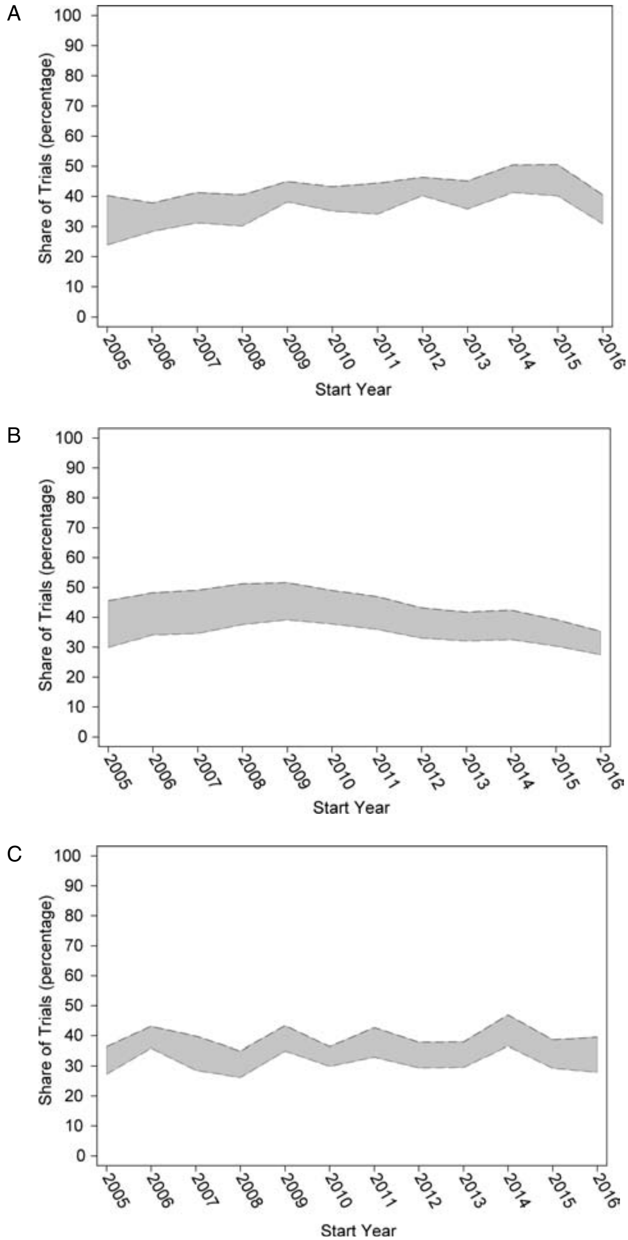

Fig. 5.7 Public versus privately held firms (representation in LPM trials). ***A*, share of Phase I LPM trials with public firm involvement; *B*, share of Phase I non-LPM trials with public firm involvement; *C*, share of Phase II LPM trials with public firm involvement; *D*, share of Phase II non-LPM trials with public firm involvement; *E*, share of Phase III LPM trials with public firm involvement; *F*, share of Phase III non-LPM trials with public firm involvement.**

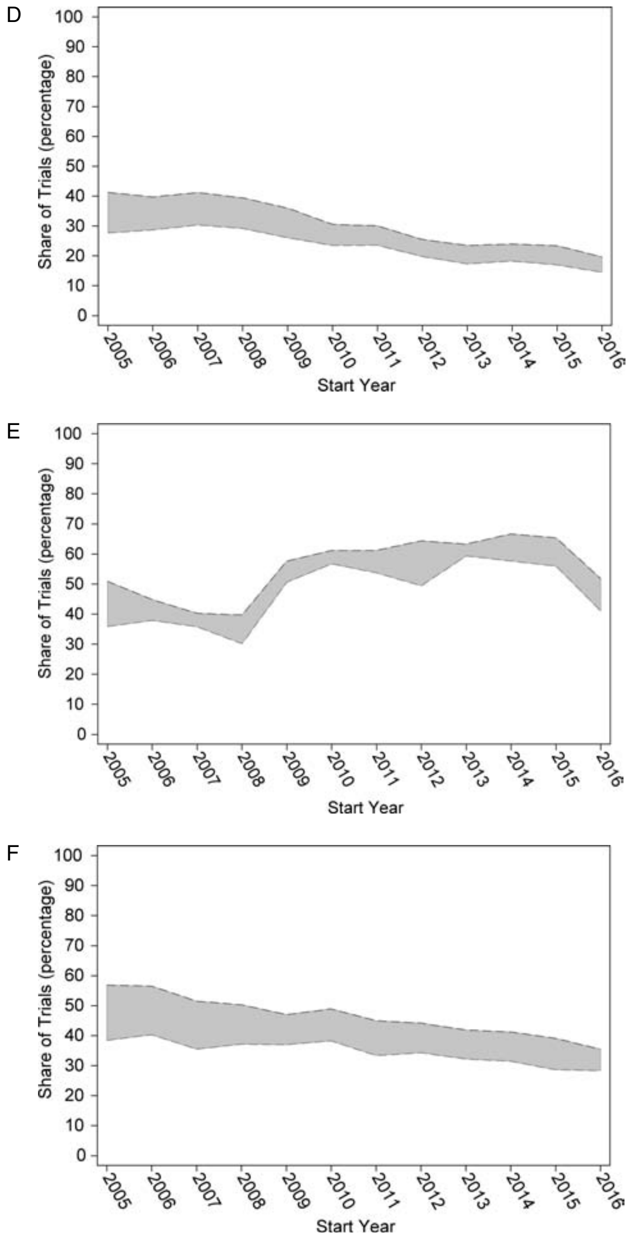
D
E
F
Share of Trials (percentage)
0
10
20
30
40
50
60
70
80
90
100
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Start Year

Fig. 5.7 (cont.)

Table 5.7A **Likely precision medicine LPM trials: publicly listed firm (upper bound) involvement (1995–2016)**

	All count	All (%)	P1 count	P1 (%)	P2 count	P2 (%)	P3 count	P3 (%)
	Generous definition							
1995	3	25.0	1	25.0	0	0.00	2	33.3
1996	1	4.00	1	16.7	0	0.00	0	0.00
1997	6	16.2	1	9.09	5	31.3	0	0.00
1998	8	14.3	3	30.0	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	20.0
2001	32	28.1	12	29.3	14	28.0	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	31.0	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	46.0	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	51.0	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							
1995	2	22.2	1	33.3	0	0.00	1	25.0
1996	1	4.35	1	20.0	0	0.00	0	0.00
1997	5	14.7	0	0.00	5	33.3	0	0.00
1998	8	15.1	3	33.3	4	13.3	1	7.14
1999	8	11.4	2	8.70	5	14.3	1	8.33
2000	11	12.2	3	12.0	4	8.70	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	40.0
2004	65	30.7	15	25.0	39	34.8	11	27.5
2005	97	40.4	25	43.1	47	35.9	25	49.0
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	39.0
2009	241	48.0	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	38.0	53	65.4
2013	331	46.0	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	66.0
2016	395	43.2	150	43.4	179	40.5	66	52.4

Table 5.7B 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	0.00	0	0.00	1	16.7
1996	1	4.00	1	16.7	0	0.00	0	0.00
1997	3	8.11	1	9.09	2	12.5	0	0.00
1998	3	5.36	2	20.0	0	0.00	1	6.67
1999	3	4.00	0	0.00	2	5.41	1	8.33
2000	5	5.26	0	0.00	2	4.17	3	15.0
2001	23	20.2	10	24.4	8	16.0	5	21.7
2002	27	18.8	6	13.0	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	25.0
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	35.0	43	50.6
2010	196	34.8	65	35.1	93	29.9	38	56.7
2011	228	35.5	73	34.1	119	33.0	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	56.0
2016	326	30.8	129	30.9	140	27.9	57	41.0
	Restrictive definition							
1995	0	0.00	0	0.00	0	0.00	0	0.00
1996	1	4.35	1	20.0	0	0.00	0	0.00
1997	2	5.88	0	0.00	2	13.3	0	0.00
1998	3	5.66	2	22.2	0	0.00	1	7.14
1999	3	4.29	0	0.00	2	5.71	1	8.33
2000	5	5.56	0	0.00	2	4.35	3	15.8
2001	21	20.0	9	25.0	7	15.2	5	21.7
2002	24	18.0	5	12.2	13	19.7	6	23.1
2003	40	26.3	4	10.8	24	30.0	12	34.3
2004	51	24.1	10	16.7	31	27.7	10	25.0
2005	69	28.8	16	27.6	34	26.0	19	37.3
2006	101	36.7	21	32.8	59	37.8	21	38.2
2007	112	30.3	29	32.6	61	28.0	22	34.9
2008	111	29.2	34	32.7	59	27.2	18	30.5
2009	201	40.0	62	41.9	100	36.5	39	48.8
2010	182	35.4	59	35.8	86	30.2	37	57.8
2011	213	36.0	68	36.2	111	32.4	34	55.7
2012	232	36.0	87	41.6	104	29.3	41	50.6
2013	271	37.6	89	38.5	110	29.8	72	60.0
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

our data, we are able to construct upper and lower bounds for the *share* of publicly listed firms over time and across phases (figure 5.7). Appendix E (http://www.nber.org/data-appendix/c13994/appendix.pdf) presents details about 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 (already) publicly listed at the time of a given trial's launch.

We conclude with regression analyses (tables 5.8A, 5.8B, and 5.9).[35] 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. Tables 5.8A and 5.8B present linear probability models using facets of trials to predict the likelihood that any given trial is an LPM using the generous and restrictive definitions as binary outcomes, respectively.

Through both panels, a set of statistical relationships emerge. For example, the linear probability models presented in tables 5.8A and 5.8B indicate that the total share of LPM trials has been increasing over time by between 0.3 and 0.5 percentage points per year, with slightly higher point estimates from models focusing on the most recent twelve years of data only. We also find that precision trials are likely to be spread across all phases of trial development (it is not the case that they are significantly less common in later phases relative to their frequency in Phase I trials). Other relationships seen in the trends presented in earlier tables and figures are also apparent. Most prominent among these is the importance of cancer trials; trials for cancer indications are 13–15 percentage points more likely to be LPM trials than those for noncancer indications. Off of an average share of LPM trials that only first reached 10 percent in recent years, this means that conditional on having a cancer trial, the probability that the trial is for an LPM more than doubles. 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 US sites are more likely than non-US trials to be LPM trials, but only by about 1 percentage point—in other words, comparing this result to the overall time trend in the data, US trials seem to be about two to three years ahead of non-US 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

35. 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 subcategories 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 three-phase classification in tables and figures for simplicity.

Table 5.8A **Predicting likely precision medicine 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	0.0099	0.0097	0.0100	0.0124	0.0127	0.0129	0.0131
	(0.0095)	(0.0095)	(0.0093)	(0.0094)	(0.0109)	(0.0110)	(0.0107)	(0.0107)
Phase 3 clinical	0.0167	0.0164	0.0168	0.0165	0.0190	0.0188	0.0195	0.0193
	(0.0146)	(0.0145)	(0.0143)	(0.0143)	(0.0160)	(0.0159)	(0.0158)	(0.0157)
Trial site in United States = 1	0.0134***	0.0127***	0.0125**	0.0118**	0.0138***	0.0130***	0.0090*	0.0082*
	(0.0021)	(0.0023)	(0.0032)	(0.0032)	(0.0027)	(0.0029)	(0.0036)	(0.0036)
Cancer trial = 1	0.1374***	0.1376***	0.1363***	0.1364***	0.1504***	0.1504***	0.1443***	0.1443***
	(0.0148)	(0.0148)	(0.0120)	(0.0120)	(0.0186)	(0.0186)	(0.0133)	(0.0133)
Biomarker type: genomic = 1	0.2429*	0.2429*	0.2428*	0.2428*	0.2403*	0.2403*	0.2398*	0.2398*
	(0.1103)	(0.1103)	(0.1105)	(0.1104)	(0.1128)	(0.1128)	(0.1131)	(0.1131)
Public firm (lower bound)	0.0104*		0.0104*		0.0129*		0.0129*	
	(0.0044)		(0.0044)		(0.0061)		(0.0061)	
Public firm (upper bound)		0.0117*		0.0117*		0.0136*		0.0135*
		(0.0049)		(0.0049)		(0.0063)		(0.0063)
Trial site in United States = 1 × cancer trial = 1			0.0024	0.0025			0.0142	0.0142
			(0.0077)	(0.0076)			(0.0133)	(0.0133)
Constant	–7.6404*	–7.6590*	–7.6465*	–7.6654*	–10.0729	–10.1599	–10.0572	–10.1441
	(2.8601)	(2.8172)	(2.8436)	(2.8008)	(4.8924)	(4.8829)	(4.9004)	(4.8911)
N	108,749	108,749	108,749	108,749	92,568	92,568	92,568	92,568
R^2	0.271	0.271	0.271	0.271	0.279	0.279	0.279	0.279

Notes: All models include a constant; robust standard errors clustered at the level of the ICD-9 chapter.
***Significant at the 0.1 percent level.
**Significant at the 1 percent level.
*Significant at the 5 percent level.

Table 5.8B **Predicting likely precision medicine 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*	0.0043	0.0044	0.0043	0.0044
	(0.0014)	(0.0013)	(0.0014)	(0.0013)	(0.0023)	(0.0023)	(0.0023)	(0.0023)
Phase 2 clinical (includes	0.0127	0.0130	0.0129	0.0132	0.0153	0.0156	0.0159	0.0161
Phase 2/3 trials)	(0.0101)	(0.0101)	(0.0099)	(0.0100)	(0.0117)	(0.0117)	(0.0115)	(0.0115)
Phase 3 clinical	0.0214	0.0211	0.0216	0.0213	0.0235	0.0232	0.0241	0.0237
	(0.0154)	(0.0153)	(0.0152)	(0.0151)	(0.0170)	(0.0169)	(0.0169)	(0.0168)
Trial site in United States = 1	0.0097***	0.0089***	0.0079**	0.0070**	0.0096***	0.0086***	0.0039	0.0029
	(0.0016)	(0.0016)	(0.0025)	(0.0024)	(0.0018)	(0.0021)	(0.0026)	(0.0027)
Cancer trial = 1	0.1362***	0.1364***	0.1338***	0.1339***	0.1490***	0.1491***	0.1417***	0.1418***
	(0.0140)	(0.0140)	(0.0111)	(0.0111)	(0.0176)	(0.0176)	(0.0124)	(0.0124)
Biomarker type: genomic = 1	0.2175	0.2176	0.2174	0.2174	0.2144	0.2145	0.2138	0.2139
	(0.1086)	(0.1086)	(0.1087)	(0.1087)	(0.1110)	(0.1109)	(0.1112)	(0.1112)
Public firm (lower bound)	0.0134*		0.0134*		0.0159*		0.0159*	
	(0.0054)		(0.0054)		(0.0072)		(0.0072)	
Public firm (upper bound)		0.0148*		0.0149*		0.0166*		0.0166*
		(0.0058)		(0.0057)		(0.0073)		(0.0073)
Trial site in United States = 1			0.0052	0.0054			0.0170	0.0170
× cancer trial = 1			(0.0071)	(0.0071)			(0.0125)	(0.0125)
Constant	–6.8624*	–6.8852*	–6.8755*	–6.8987*	–8.8031	–8.9076	–8.7843	–8.8886
	(2.7860)	(2.7331)	(2.7727)	(2.7198)	(4.6997)	(4.6847)	(4.7055)	(4.6907)
N	108,749	108,749	108,749	108,749	92,568	92,568	92,568	92,568
R^2	0.254	0.254	0.254	0.254	0.261	0.262	0.262	0.262

Notes: All models include a constant; robust standard errors clustered at the level of the ICD-9 chapter.
***Significant at the 0.1 percent level.
**Significant at the 1 percent level.
*Significant at the 5 percent level.

1–2 percentage points more likely to be LPM trials than those of privately held firms, all else equal.

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 5.9 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 data set. As above, these coefficients cannot be interpreted causally; rather, they represent average associations between features of clinical trials and the amount of time required for trial completion.

The first three columns of table 5.9 present models predicting trial duration in LPM trials, while the last three columns present identical models in non-LPM trials. A number of interesting differences emerge. First, we note the difference in the coefficient on the intercept in the two sets of linear models; LPM trials take roughly twenty months longer to complete relative to nonprecision trials, all else equal. This may be due to the challenges of enrolling patients with less common subtypes of a disease, as well as the fact that nonprecision trials include a number of shorter studies (e.g., for antibiotics) that can be run extremely quickly, thereby lowering the average time to completion in the second group of trials. As in tables 8A and 8B, Phase I trials are the omitted category in all models. For LPM trials, Phase II trials are, all else equal, only about two months longer than Phase I trials, on average, and this difference is not statistically significant in any of the three specifications. This is quite different than what is observed in nonprecision trials, where Phase II studies take five to six 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 nonprecision trials, where Phase III studies are only seven to nine months longer, on average. This suggests that although LPM trials may be longer on average, the use of biomarkers may be able to close the gap between Phase I and later phases to some extent.

Interestingly, although cancer trials appear to always take longer to complete, on average, than noncancer trials, the additional trial length associated with LPM cancer trials is six to seven months *less* than the additional trial length associated with nonprecision cancer trials in these models. One interpretation of this is that precision medicines perhaps speed up cancer trials because of surrogate endpoints or enrichment. We are cautious not to overinterpret this relationship, because it does not hold up when examined in further detail; in appendix table III (http://www.nber.org/data-appendix/c13994/appendix.pdf) we consider the same sets of models for cancer trials

Table 5.9 **Dependent variable: trial duration in months**

	LPM trials			Non-LPM trials		
	All trials	All trials	All United States	All trials	All trials	All United States
Phase 2 clinical (inc. Phase 2/3 trials)	1.976	2.018	2.565	6.472***	6.192***	5.374***
	(1.048)	(1.045)	(1.359)	(0.224)	(0.223)	(0.339)
Phase 3 clinical	13.889***	14.136***	13.171***	9.054***	9.193***	6.859***
	(1.566)	(1.555)	(2.063)	(0.254)	(0.252)	(0.400)
Trial site in United States	3.767***	4.143***		4.610***	4.968***	
	(0.958)	(0.962)		(0.188)	(0.188)	
Cancer trial	13.258***	13.335***	12.781***	19.608***	19.403***	18.488***
	(1.283)	(1.275)	(1.750)	(0.269)	(0.268)	(0.375)
Public firm (lower bound)	–3.348***			–6.620***		
	(0.949)			(0.186)		
Public firm (upper bound)		–5.078***	–6.021***		–7.805***	–8.227***
		(0.973)	(1.317)		(0.190)	(0.292)
Constant	60.043***	59.840***	63.616***	38.916***	39.438***	45.857***
	(6.184)	(6.153)	(6.764)	(1.275)	(1.274)	(1.469)
N	2,743	2,743	1,760	50,186	50,186	26,101
R^2	0.330	0.334	0.320	0.304	0.311	0.277

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

***Significant at the 0.1 percent level.

**Significant at the 1 percent level.

*Significant at the 5 percent level.

alone and show similar patterns across many coefficients in the regression models, but differences in the estimated constants between LPM trials versus nonprecision trials in cancer. Finally, we note that as economic incentives would predict, 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.

5.6 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 have implications for cost growth in health care and pharmaceutical pricing. First, we show that the use of biomarkers in clinical trials has grown significantly, with an important subset 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 frequently 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 for new therapeutics, the growth in likely precision medicine trials has been large in percentage terms, approximately doubling every decade over the past twenty-two years.

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

Second, 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 biomarkers employed 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 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 subsequently commercialized. If failure rates in clinical research are endogenously determined with other characteristics related to commercialization strategies (e.g., single-product vs. multiproduct firms, as seen in Guedj and Scharfstein [2004]), characterizing *trials* may not accurately reflect the future *products* that emerge from those trials. To the extent that there is selection in R&D project discontinuations based on features that are 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.[36]

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 trials for LPMs have grown in number and share 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 States) 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 precision medicines grow in market share, they will drive up costs for many of the drugs that target specific groups of patients and also open up opportunities for previously difficult-to-implement firm strategies such as indication-based pricing. Such developments will also underscore the impact of strategic decisions regarding when and how to run biomarker-driven clinical trials during the therapeutic development process.

Appendix

Appears online at http://www.nber.org/data-appendix/c13994/appendix.pdf.

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

References

Acemoglu, Daron, and Joshua Linn. 2004. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *Quarterly Journal of Economics* 3:1049–90.
Bagley, Nicholas, Benjamin Berger, Amitabh Chandra, Craig Garthwaite, and Ariel D. Stern. Forthcoming. "The Orphan Drug Act at 35: Challenges and Opportunities for the 21st Century." *Innovation Policy and the Economy*.
Berndt, Ernst R., Rena M. Conti, and Stephen J. Murphy. 2017. "The Landscape of US Generic Prescription Drug Markets, 2004–2016." NBER Working Paper no. 23640, Cambridge, MA.
Berndt, Ernst R., and Mark R. Trusheim. 2015. "Biosimilar and Biobetter Scenarios for the US and Europe: What Should We Expect?" In *Biobetters: Protein Engineering to Approach the Curative*, edited by Amy Rosenberg and Barthélemy Demeule, 315–60. New York: Springer.
Bespalov, Anton, Thomas Steckler, Bruce Altevogt, Elena Koustova, Phil Skolnick, Daniel Deaver, Mark M. Millan, et al. 2016. "Failed Trials for Central Nervous System Disorders Do Not Necessarily Invalidate Preclinical Models and Drug Targets." *Nature Reviews Drug Discovery* 15 (7): 516.
Budish, Eric, Benjamin N. Roin, and Heidi Williams. 2015. "Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials." *American Economic Review* 105 (7): 2044–85.
Chandra, Amitabh, and Craig Garthwaite. 2017. "The Economics of Indication-Based Drug Pricing." *New England Journal of Medicine* 377 (2):103–06.
Costinot, A., Dave Donaldson, Margaret Kyle, and Heidi Williams. 2016. "The More We Die, the More We Sell? A Simple Test of the Home-Market Effect." NBER Working Paper no. 22538, Cambridge, MA.
De Angelis, Catherine, Jeffrey Drazen, Frank A. Frizelle, Charlotte Haug, John Hoey, Richard Horton, Sheldon Kotzin, et al. 2004. "Clinical Trial Registration: A Statement from the International Committee of Medical Journal Editors." *New England Journal of Medicine* 351 (12):1250–1251.
Dubois, Pierre, Olivier de Mouzon, Fiona Scott Morton, and Paul Seabright. 2015. "Market Size and Pharmaceutical Innovation." *RAND Journal of Economics* 46 (4):844–71.
Dzau, Victor J., and Geoffrey S. Ginsburg. 2016. "Realizing the Full Potential of Precision Medicine in Health and Health Care." *Journal of the American Medical Association* 316 (16): 1659–60.
FDA, Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research. 2004. "Guidance for Industry: Information Program on Clinical Trials for Serious or Life-Threatening Diseases and Conditions." Draft guidance, US Department of Health and Human Services, Food and Drug Administration.
———. 2008. "Guidance for Industry E15 Definitions for Genomic Biomarkers, Pharmacogenomics, Pharmacogenetics, Genomic Data and Sample Coding Categories." US Department of Health and Human Services, Food and Drug Administration.
FDA-NIH. 2016. "BEST (Biomarkers, EndpointS, and other Tools) Resource." Biomarker Working Group, US Food and Drug Administration and US National Institutes of Health. http://www.biostatsolutions.com/wp-content/uploads/2016/11/Bookshelf_NBK326791.pdf.
Fridlyand, Jane, Richard Simon, Jessica Walrath, Nancy Roach, Richard Buller, David Schenkein, Keith Flaherty, Jeff Allen, Ellen Sigal, and Howard Scher. 2013.

"Considerations for the Successful Co-development of Targeted Cancer Therapies and Companion Diagnostics." *Nature Reviews Drug Discovery* 12 (10):743–55.
Guedj, Ilan, and David Scharfstein. 2004. "Organizational Scope and Investment: Evidence from the Drug Development Strategies and Performance of Biopharmaceutical Firms." NBER Working Paper no. 10933, Cambridge, MA.
International Trade Administration, Department of Commerce. 2016. "2016 ITA Pharmaceuticals Top Markets Report." https://www.trade.gov/topmarkets/pdf/Pharmaceuticals_Executive_Summary.pdf.
Kanavos, Panos, Alessandra Ferrario, Sotiris Vandoros, and Gerard F. Anderson. 2013. "Higher US Branded Drug Prices and Spending Compared to Other Countries May Stem Partly from Quick Uptake of New Drugs." *Health Affairs* 32 (4):753–61.
Kao, Jennifer L. 2017. "R&D Decisions for New Medical Technologies: Evidence from New Use Approvals and Off-Label Uses." Working paper, Harvard University.
Phyo, Aung Pyae, Podjanee Jittamala, François H. Nosten, Sasithon Pukrittayakamee, Millika Imwong, Nicholas J. White, Stephan Duparc, et al. 2016. "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." *Lancet Infectious Diseases* 16 (1):61–69.
Rock Health. 2017. "2016 Year End Funding Report." https://rockhealth.com/reports/2016-year-end-funding-report-a-reality-check-for-digital-health.
Scott Morton, Fiona M. 1999. "Entry Decisions in the Generic Pharmaceutical Industry." *RAND Journal of Economics* 30 (3): 421–40.
Scott Morton, Fiona M., Ariel Dora Stern, and Scott Stern. 2018. "The Impact of the Entry of Biosimilars: Evidence from Europe." *Review of Industrial Organization* 53 (1): 173–210.
Stern, Ariel D., Brian M. Alexander, and Amitabh Chandra. 2017. "How Economics Can Shape Precision Medicines." *Science* 355 (6330): 1131–33.
———. 2018. "Innovation Incentives and Biomarkers." *Clinical Pharmacology & Therapeutics* 103 (1). https://www.doi.org/10.1002/cpt.876.
West, Howard Jack. 2017. "Novel Precision Medicine Trial Designs: Umbrellas and Baskets." *Journal of the American Medical Association Oncology* 3 (3): 423.