Published in final edited form as:

J Health Econ. 2011 July ; 30(4): 603–615. doi:10.1016/j.jhealeco.2011.05.007.

Opportunities and Benefits as Determinants of the Direction of Scientific Research[*]

Jay Bhattacharya[†] and **Mikko Packale**[‡]

[†]Stanford University School of Medicine, CHP/PCOR, 117 Encina Commons, Stanford, CA 94305-6019

[‡]University of Waterloo, Department of Economics, 200 University Avenue West, Waterloo, ON N2L 3G1

Abstract

Scientific research and private-sector technological innovation differ in objectives, constraints, and organizational forms. Scientific research may thus not be driven by the direct practical benefit to others in the way that private-sector innovation is. Alternatively, some–yet largely unexplored-mechanisms drive the direction of scientific research to respond to the expected public benefit. We test these two competing hypotheses of scientific research. This is important because any coherent specification of what constitutes the socially optimal allocation of research requires that scientists take the public practical benefit of their work into account in setting their agenda. We examine whether the composition of medical research responds to changes in disease prevalence, while accounting for the quality of available research opportunities. We match biomedical publications data with disease prevalence data and develop new methods for estimating the quality of research opportunities from textual information and structural productivity parameters.

Keywords

Scientific Research; Induced Innovation; Research Opportunity; Technological Opportunity; Non-Profit Incentives; Medicine; Textual Econometrics

1 Introduction

Scientific research and private-sector technological innovation differ in objectives, constraints, and organizational forms. For example, the for-profit objective that drives private-sector innovation is muted in much scientific research.[1] This particular difference is

[*]We thank Ian Cockburn, David Cutler (Co-Editor), Amy Finkelstein, Raphael Godefrey, Darius Lak-dawalla, Neeraj Sood, Scott Stern, Bruce Weinberg, anonymous referees, and participants at the NBER Summer Institute 2008 Productivity meeting for helpful comments. Previous versions of this paper are titled “Is Medicine an Ivory Tower? Induced Innovation, Technological Opportunity, and For-Profit vs. Non-Profit Innovation.”

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

JEL Classification: O31, O33, I12, L65.

important in part because other differences are likely linked to it. For example, Aghion et al. (2008) view the fact that individual researchers have more control over their agenda in scientific research than in private-sector innovation as the defining characteristic of academia. They conjecture that this difference is due to the non-profit nature of scientific research.

A key virtue of for-profit allocation is that decisions made by for-profit firms must necessarily respond to changes in the market, or else risk failure. There is abundant evidence that for-profit producers innovate according to market demand. Non-profit allocation, on the other hand, imposes looser budget constraints (Lakdawalla and Philipson, 2006). In principle, looser constraints could divorce production decisions from demand. For example, the choice of topics could be driven by the prospect of influencing other scientists (e.g. Dasgupta and David, 1994, and Saha and Weinberg, 2008) rather than the expected social benefit.

These considerations raise the possibility that the direction of scientific research does not respond to market demand in the way that private-sector technological innovation does. Alternatively, some-yet largely unexplored–mechanisms drive the direction of scientific research to respond to the expected public benefit, as has been argued by Rosenberg (1982). In this paper we test these two competing hypotheses. This question is important because any coherent specification of what constitutes the socially optimal allocation of research would require that scientists take the public benefit of their work into account in setting their agenda.

To test these two competing hypotheses of scientific research, we examine whether the composition of medical research responds to changes in disease prevalence. For drug-related medical research we also condition on the quality of available research opportunities. We focus on disease-driven medical research examined here because it represents the majority of research in medicine at least in terms of publication output.[2]

Our focus on medicine is appropriate because, while there may be good reasons to insulate some research activities from the vagaries of the market, academic medicine is not such a market. There is little extant evidence that academic medicine actually does so respond to the market (that is, to the epidemiology of patient health) and the view of academic medicine as an “ivory tower” persists. The role of technological progress in producing gains against diseases implies that the study of factors that determine the direction and magnitude of that progress–including the study of what determines the direction of academic medical research-are especially important from health and health economic perspectives.

[1] In the Background Appendix we examine the connections between industrial R&D and academic research in the biomedical sector and how pharmaceutical innovation reflects largely the functioning of for-profit incentives and academic medicine and biomedical publications reflect largely non-profit incentives.

[2] Throughout our sample period over 60% of publications in medicine are linked to a disease (this can be seen from Figure 2.1 in Section 7.1). Our match of publication data and disease prevalence data captures roughly 50% of all disease-linked research (see Section 7.1). The three measures of drug-related research that we employ (see Section 4.1) represent between 20% and 45% of all disease-matched research (see Section 7.1).

Our analysis is agnostic about why medical research would respond to changes in disease prevalence and research opportunities. The available data do not enable us to differentiate between theories of scientific research such as altruism, prestige maximization (see e.g. Merton 1973 [1942], Glaeser, 2003, and Stern 2004), and the availability of government funding.[3] The specific mechanisms are important but so is understanding the relationship between the direction of scientific research and characteristics that determine the socially optimal allocation.

Only a handful of studies have examined the determinants of scientific research and nonprofit innovation in general. Rosenberg (1982) emphasizes that private-sector technological innovation yields important inputs to scientific research. He conjectures that the direction of scientific research is in part driven by the quality of research opportunities and the expected rewards from research. Lichtenberg (1999) and Lichtenberg (2006) find a positive correlation between public biomedical funding and both disease prevalence and disease severity and between cancer prevalence and the number of biomedical publications. In contrast with these two analyses, we use exogenous variation in disease prevalence to identify the induced innovation effect. Finkelstein (2004) finds that the impact of vaccine policies on the number of new patent applications is small and statistically insignificant for both non-profit and for-profit entities. Unlike all three analyses, we condition on available research opportunities.[4]

The literature on the determinants of the direction of private-sector technological innovation is more extensive. The induced innovation hypothesis originated in Hicks (1932) and Schmookler (1966). Recent empirical studies of the induced innovation hypothesis in the pharmaceutical industry include Acemoglu and Linn (2004), Finkelstein (2004), Lichtenberg and Waldfogel (2003) and Yin (2008).[5] Our research opportunity concept corresponds to the technological opportunity concept examined by Scherer (1965) and Schmookler (1966) as well as by Popp (2002).[6]

Our methodology to estimate the quality of research opportunities builds on the methodology of Caballero and Jaffe (1993), Jaffe and Trajtenberg (1996) and Popp (2002). Our method extends this by permitting the probability that a given knowledge cohort is used in research to depend not only on the quality of a given knowledge cohort but also on the

[3] Throughout our sample period-based on medical researchers' self-reports-only 11% (13%) of disease- linked research (all research) is supported by National Institutes of Health (NIH) or other U.S. government sources (authors' calculations from the publications data). These low numbers suggest that our estimates are not necessarily driven by responses to changes in government funding priorities.

[4] Examination of the research opportunity effect and development of associated methods is important for three reasons. First, while it is implausible that researchers *within* a research field would not redirect research effort in response to changes in research opportunities, it is not nearly as evident that scientists would very often switch fields to take advantage of greater research opportunities. Second, it is obviously important to condition on research opportunities in estimating the induced innovation effect if the two variables are correlated. Third, from an allocative efficiency perspective it is important to understand how the research opportunity effect varies across organizational forms and across types of individuals.

[5] Newell, Jaffee and Stavins (1999) and Popp (2002) examine induced innovation in the energy sector. In addition to Acemoglu and Linn (2004), also DellaVigna and Pollet (2007) exploit changes in the age demographics of the population for identification.

[6] The previous version of this paper (Bhattacharya and Packalen, 2008a) included estimates of the induced innovation effect in pharmaceutical innovation, which we omit here for presentational clarity. The analyses of aging and obesity induced innovation are related to the empirical studies on preference externalities by Waldfogel (2003) and George and Waldfogel (2003). In a companion paper (Bhattacharya and Packalen, 2008b) we calculate the welfare effect of the induced innovation externality of obesity. The reader is also referred to this companion paper for references to the medical and economic literatures on obesity.

quality of other existing knowledge cohorts. Also, we construct the scientific opportunity variable from textual information, rather than citation information. This considerably expands the information base from which research opportunities can be measured. For example, citations in scientific publications seldom capture research opportunities generated by private-sector technological innovation, whose role Rosenberg (1982) emphasized.[7]

2 Theory

We present a model in which the socially optimal allocation of research across diseases is influenced by disease prevalence and quality of research opportunities, implying that any good allocation mechanism would induce research to respond to these characteristics. The analysis has also implications for how to measure quality of research opportunities.

2.1 A Model of the Social Benefit from Medical Research

We assume that each unit of research is identified by three characteristics: the disease i which the research examines, the year t in which the research is conducted, and the cohort f of the research opportunities that are pursued in the research.[8] The benefit from research depends on three factors: 1) the extent of research effort (N_{itf}), 2) the number of people who benefit from the research (M_{it}), and 3) the quality of the research opportunities.[9]

The third factor, quality of research opportunities, captures the idea that the benefit from an inframarginal unit of research is higher when inputs to the research process provide researchers fertile applications compared to when inputs to the research process hold only potential for average or below average applications. Inputs to the research process can be tangible or intangible. Tangible research inputs in medicine include approved drugs developed by pharmaceutical companies. Ideas are an example of intangible research inputs, some of which are recorded as citations in publications. In existing literature research inputs are measured from citations. In the complementary approach developed here research inputs and the associated opportunities and cohorts are measured from textual content of research publications and the year in which a given concept first appears in the publications.[10]

The quality of research opportunities depends on two factors: 1) the baseline productivity of research inputs in the opportunity cohort f in research on the disease i, which we denote by a_{if}, and 2) the elapsed time $t - f$ since the initial discovery of the research inputs in cohort f. The first factor reflects the fact that the productivity of research inputs in cohort f in research

[7] Related work includes Azoulay et al. (2007, 2009) who determine patentability research from the textual content of publications, and the graphical analysis of topic bursts by Mane and Börner (2004).

[8] These assumptions are, of course, simplifications as a research project in medicine does not necessarily examine only one disease and may rely on opportunities that do not all belong to the same opportunity cohort f. We address these issues in our empirical analysis (see Sections4.1–4.2).

[9] We thus assume that the benefit from research does not depend on the severity of the disease. This is in part due to lack of exogenous variation in severity over time, but is also consistent with the findings of Acemoglu and Linn (2004). From a theoretical perspective, for an increase in the severity of a disease to increase the expected benefit from research on the disease the increase in severity should be accompanied with an increase in the expected progress that could be made against the disease. The validity of latter condition is not evident to us given the incremental nature of technological progress.

[10] As an example of the research inputs that are captured using this approach, consider the active ingredient “cyclosporine”, which was the first effective immunosuppressant and thereby enabled transplants. Cyclosporine was first mentioned in the title/abstract of a medical research publication in 1981. We set the preceding year as the cohort. By 2005 it was mentioned in tite title/abstract of 17,682 publications.

on disease *i* depends on how suitable the research inputs are in research on the disease. The second factor reflects the fact that diffusion and exhaustion of knowledge is generally gradual. A new research input contributes little when it is first discovered and only a few researchers know about it. It will also contribute little on the margin after everyone knows about it and most of the potential of the knowledge cohort has been exhausted.

We assume a specific functional form for the total benefit from research across diseases:

$$\sum_i M_{it} \sum_{f=f_0}^{t} \left\{\alpha_{if} \times e^{-\beta_1(t-f)} \times [1-e^{-\beta_2(t-f)}]+\varepsilon_{itf}\right\} \times ln\,(N_{itf}), \qquad (1)$$

where ε_{itf} denote factors that are observable to medical researchers but not the econometrician, satisfy $E[\varepsilon_{itf}] = 0$, and are independently distributed. Factors $[1-e^{-\beta_2(t-f)}]$ and $e^{-\beta_1(t-f)}$, respectively, represent diffusion and exhaustion of opportunities in each cohort *f*.[11]

2.2 Socially Optimal Allocation

Let N^*_{itf} denote the optimal allocation of research effort when allocation across input cohorts *f within* each disease *i* is optimal. First-order conditions for the optimum imply that

$$p^*_{itf}=\frac{\alpha_{if} \times e^{-\beta_1(t-f)} \times [1-e^{-\beta_2(t-f)}]+\varepsilon_{itf}}{\sum_{f'=f_0}^{t}\left\{\alpha_{if'} \times e^{-\beta_1(t-f')} \times [1-e^{-\beta_2(t-f')}]+\varepsilon_{itf'}\right\}}, \qquad (2)$$

where $p^*_{itf} \equiv \frac{N^*_{itf}}{\sum_{f'=f_0}^{t} N^*_{itf'}}$.[12] Equation (2) states that the share of research on disease *i* that relies on opportunity cohort *f* is equal to the ratio of the quality of the opportunity cohort *f* in research and the sum of the qualities of all available research opportunity cohorts.

Let $N_{it} \equiv \sum_{f=f_0}^{t} N^*_{itf}$. When allocation of research effort across opportunity cohorts *f* within a disease is optimal, expression (1) can be rewritten as

[11] We do not model explicitly the effect that the amount of research in the preceding years may have on the benefit from research in a given year. This assumption is innocuous if marginal research in each year does not influence the quality of research opportunities in future years.

[12] The first-order condition for optimum is $M_{it}\times\{\alpha_{if}\times e^{-\beta_1(t-f)}\times[1-e^{-\beta_2(t-f)}]+\varepsilon_{itf}\}/N^*_{itf}=M_{it}\times\{\alpha_{if'}\times e^{-\beta_1(t-f')}\times[1-e^{-\beta_2(t-f')}]+\varepsilon_{itf'}\}/N^*_{itf'}$, for all (*i, t, f, f′*). Denoting $c_{itf} \equiv \alpha_{if} \times e^{-\beta_1(t-f)} \times [1-e^{-\beta_2(t-f)}]+\varepsilon_{itf}$ this condition can be rewritten as $c_{itf} \times N^*_{itf'}=N^*_{itf} \times c_{itf'}$, for all (*i,t,f,f′*). Summing both sides of the equation $c_{itf} \times N^*_{itf'}=N^*_{itf} \times c_{itf'}$, over all $f' \in \{f_0,\ldots,t\}$ gives $c_{itf} \times \sum_{f'=f_0}^{t} N^*_{itf'}=N^*_{itf} \times \sum_{f'=f_0}^{t} c_{itf'}$ for all (*i, t, f*). Rearranging and using the definitions of c_{itf} and p^*_{itf} gives the relationship (2).

$\sum_i M_{it} \sum_{f=f_0}^{t} \left\{ \alpha_{if} \times e^{-\beta_1(t-f)} \times [1-e^{-\beta_2(t-f)}] + \varepsilon_{itf} \right\} \times ln\,(N_{it} \times p^*_{itf})$. First-order conditions for the optimal allocation of research effort *across* diseases imply that

$$N_{it} = \left(\sum_i N_{it} \right) \times \frac{M_{it} \sum_{f=f_0}^{t} \alpha_{if} \times e^{-\beta_1(t-f)} \times [1-e^{-\beta_2(t-f)}] + \varepsilon_{itf}}{\sum_i M_{it} \sum_{f=f_0}^{t} \left\{ \alpha_{if} \times e^{-\beta_1(t-f)} \times [1-e^{-\beta_2(t-f)}] + \varepsilon_{itf} \right\}} \quad (3)$$

holds for all (*t*, *i*). The factor $\sum_{f=f_0}^{t} \left\{ \alpha_{if} \times e^{-\beta_1(t-f)} \times [1-e^{-\beta_2(t-f)}] + \varepsilon_{itf} \right\}$ in equation (3) is the sum of the qualities of available research opportunity cohorts in the disease *i* in year *t*, and we use this sum as a measure of the quality of research opportunities in research on the disease *i* in year *t*. We denote this measure of the quality of research opportunities by

$$K_{it} \equiv \sum_{f=f_0}^{t} \left\{ \alpha_{if} \times e^{-\beta_1(t-f)} \times [1-e^{-\beta_2(t-f)}] + \varepsilon_{itf} \right\}. \quad (4)$$

We can rewrite equation (3) as $N_{it} = \sum_i N_{it} / \left(\sum_i M_{it} \sum_{f=f_0}^{t} K_{it} \right) \times M_{it} \times K_{it}$. Assuming that $N_{it} > 0$ and $K_{it} > 0$ for all (*i*, *t*) this can be rewritten as

$$ln\, N_{it} = ln\, M_{it} + ln\, K_{it} + u_t, \quad (5)$$

where $u_t \equiv ln \left[\sum_i N_{it} / \left(\sum_i M_{it} \sum_{f=f_0}^{t} K_{it} \right) \right]$.

2.3 Implications for Empirical Analysis of Research Allocation

Equation (5) describes a proportional relationship between research effort in a disease and the quality of research opportunities, and a proportional relationship between the research effort in a disease and disease prevalence. With a different functional form for the overall benefit from research both relationships would still be positive but non-proportional. We allow for this possibility in our empirical framework.

So far, we have ignored the fact that a factor that changes disease prevalence may also change research within a disease. For example, population aging and rising obesity may shift research effort away from drug-related medical research on a disease and toward research that is still applied and disease-driven but more focused on the physiology of the disease in the old-age and obese populations. The estimated disease prevalence effect may then even be negative for drug-related research.[13] We allow for this possibility in our empirical framework.

[13] Even when the total research effort on the disease changes, our measure of that effort (publications) may not change if one type of research (say, research that examines the physiology of the disease in a subpopulation) requires more inputs than another type of research (say, drug-related research).

3 Estimation of the Quality of Research Opportunities

Our definition of the quality of research opportunities follows from our theoretical model and is given by (4).[14] Provided that we can obtain estimates $\hat{\alpha}_{if}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ of the parameters α_{if}, β_1 and β_2, we can estimate the quality of research opportunities using the formula

$$\hat{K}_{it} \equiv E[K_{it}|\hat{\alpha}_{if}, \hat{\beta}_1, \hat{\beta}_2] \approx \sum_{f=f_0}^{t} \hat{\alpha}_{if} \times e^{-\hat{\beta}_1(t-f)} \times \left[1 - e^{-\hat{\beta}_2(t-f)}\right]. \quad (6)$$

The econometric challenge is to estimate the parameters α_{if}, β_1 and β_2. We start with equation (2). We measure p^*_{itf} from textual information in publications (see Section 4.2). Denoting $\alpha_{it} \equiv 1/\sum_{f=f_0}^{t} \left\{\alpha_{if} \times e^{-\beta_1(t-f)} \times [1 - e^{-\beta_2(t-f)}] + \varepsilon_{itf}\right\}$ equation (2) may be rewritten as

$$p^*_{itf}/\alpha_{it} = \alpha_{if} \times e^{-\beta_1(t-f)} \times [1 - e^{-\beta_2(t-f)}] + \varepsilon_{itf}. \quad (7)$$

When $t - f_0$ is large, we have that $\sum_{f=f_0}^{t} \varepsilon_{itf} \approx 0$, which modifies the definition of α_{it} to

$$\alpha_{it} \equiv 1/\sum_{f=f_0}^{t} \left\{\alpha_{if} \times e^{-\beta_1(t-f)} \times [1 - e^{-\beta_2(t-f)}]\right\} \quad (8)$$

and the relationship (2) to

$$p^*_{itf} = \frac{\alpha_{if} \times e^{-\beta_1(t-f)} \times [1 - e^{-\beta_2(t-f)}] + \varepsilon_{itf}}{\sum_{f=f_0}^{t} \alpha_{if} \times e^{-\beta_1(t-f)} \times [1 - e^{-\beta_2(t-f)}]}. \quad (9)$$

This equation forms our estimating equation for the parameters α_{if}, β_1,and β_2.[15]

In our empirical analysis the number of parameters α_{if} is over 5, 000. Estimating the parameters α_{if}, β_1,and β_2 using non-linear least squares and equation (9) is therefore computationally quite demanding. Instead, we first estimate the parameters β_1 and β_2 using non-linear least squares applied to the equation (7) while assuming fixed values for the parameters α_{if} and α_{it}.[16] We then estimate the parameters α_{if} using the following three-step

[14] Popp (2002), Caballero and Jaffe (1993) and Jaffe and Trajtenberg (1996) each use either the formula (6) or a close equivalent. One advantage of our structural approach is that potential limitations in empirical work become more transparent. For example, the analysis reveals that if expression (1) is not a good representation of the benefit from research, the measure of the quality of research opportunities variable will likely be influenced by the extent of research on the disease (see Section 6.2). It is therefore important to examine whether such potential reverse causality influences estimates of the "research opportunity effect". We are not aware of this issue having been considered in existing work.

[15] The innovation relative to the prior literature is that the probabilities [Inline] are allowed to depend not only on the quality of opportunity cohort f but also on the quality of all available opportunity cohorts through the denominator in equation (9).

iterative procedure: Step 1. Calculate initial estimates of a_{it} by plugging in the estimates of β_1 and β_2 as well as arbitrary starting values of a_{if}= 1 for all i, f into (8); Step 2. Using estimates of a_{it}, β_1,and β_2 estimate parameters a_{if} by least squares applied to the equation (7) while holding a_{it}, β_1,and β_2 fixed; Step 3. Recompute a_{it} by plugging in estimates of a_{if}, β_1,and β_2 into the expression (8). If the new value of a_{it} is sufficiently close to the old value, declare convergence. If not, iterate the previous step until convergence. This iterative procedure yields estimates of the parameters a_{if}, which we then combine with the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ to construct a measure of the quality of research opportunities using formula (6).

4 Data and Measurement of Variables

4.1 Publications Data and Match to Disease Prevalence Data

We measure research effort in medicine from the indexed MEDLINE database on approximately 16 million biomedical publications, from 1950 to the present. The indexing is done by professionals with biomedical training using the hierarchical Medical Subject Headings (MESH) vocabulary of over 20,000 terms. We use the MESH (2007 version) codes in the MEDLINE data to identify the disease(s) examined in each publication.

We construct a match between the publications data and the disease prevalence data, which is indexed by the ICD-9 classification system. We limit this matching effort in several ways that are detailed in the Data Appendix. The match, included in the previous version of this paper (Bhattacharya and Packalen, 2008a), yields 127 separate matches between a disease or a group of diseases and a MESH entry/entries. The 127 diseases belong to 12 disease classes. We limit the empirical analysis to these 127 diseases.

We measure the research effort related to a disease by the number of publications that are matched to the disease.[17] A publication may be indexed to multiple diseases. We allow for this possibility by counting publications that are matched to more than one disease the same way we would count the matches if each match were from a separate publication.

We use several alternative strategies to measure drug-related medical research. First, we classify all publications that are matched to an active ingredient as being drug-related medical research and count a publication that is matched to n different cohorts of ingredients as n units of research. Second, we classify all publications that are matched to an ingredient as being drug-related medical research and count each such publication as one unit of research. Third we classify all publications that are indexed with the “major topic” flag and the qualifier term “drug therapy”, “drug effects” or “pharmacology” as being drug-related research and count each such publication as one unit of research. We refer to these three constructed measures as *DRUG 1, DRUG 2*, and *DRUG 3*, respectively.

[16] We assume that a_{if}= 1 and a_{it} = 1 for all i,t,f. The estimating equation therefore becomes $p^*_{itf}{=}e^{-\beta_1(t-f)}\times\left[1-e^{-\beta_2(t-f)}\right]{+}\varepsilon_{itf}$. Omitting a multiplicative constant in this specification is both innocuous and necessary because the true value of the parameter β_2 is typically very small and the variation in $t-f$ is limited which make the factor $\left[1-e^{-\beta_2(t-f)}\right]$ approximately equal to $\beta_2 \times (t-f)$ in the sample.

[17] This could be generalized by employing article-specific citation or journal impact factor based quality- adjusted measures of research effort. We leave this generalization for future work.

We also use several strategies to measure other medical research. First, we classify all publications that are not matched to an ingredient as other medical research and count each such publication as one unit of research. Second, we classify all publications that are 1) not matched to an ingredient, 2) not indexed are indexed with any of the qualifier terms “drug therapy”, “drug effects” or “pharmacology”, and 3) not indexed with the term “Chemicals and Drugs”, as other medical research and count each such publication as one unit of research (this method excludes most of research that is conducted using unapproved drugs that do not appear in our list of FDA approved ingredients). Third, we classify all research that is indexed with the qualifier term “surgery” or “transplantation” as other medical research and count each such publication as one unit of research. We refer to these three constructed measures as *OTHER 1, OTHER 2*, and *OTHER 3*, respectively.

4.2 Measurement of Research Inputs and their Cohort

We rely on textual information in publications to measure research opportunities, rather than citations as is standard in the patent literature. In principle, also an analysis of research inputs in scientific research could be based on citations. However, citation data are not always widely available (even only post-1947 U.S. patents display citation data) and while an important input in scientific research is private-sector innovation (Rosenberg, 1982) citations in scientific publications seldom reveal their presence.

To limit the scope of the data extraction exercise, we construct the measure of the quality of research opportunities only for drug-related medical research, for which an important (and easily available) subset of the research inputs is the set of approved active ingredients.[18] To determine which ingredients are used as research inputs in each medical research publication, we search through the titles and abstracts of all publications for all approved active ingredients.[19] A match between ingredient name and publication indicates that the ingredient was an input to the research process that led to the publication.

Our measure of drug-related research captures applied drug-related medical research.[20] Such research is primarily conducted by academic researchers and published in academic

[18] Sufficient computational resources would enable one to index all words in all research publications and then determine the cohort of each concept represented by each new word, allowing one to identify the research opportunities associated with each publication without a predetermined list of research inputs. The Google Books Ngram project (Michel et al., 2011) applied a related approach to examine the popularity of different cultural interests over time.
By focusing on active ingredients, we ignore new concepts and technologies such as genomics and recombinant DNA which undoubtedly have opened new research opportunities even for drug-related research. This does not pose a problem for our analyses to the extent that such general technologies affect research on all diseases equally (or at least their introduction is not correlated with our independent variables). More importantly, despite this caveat the textual analysis developed here captures a much richer set of research inputs than can be captured by citation data alone.

[19] We identify active ingredients from the Federal Drug Administration (FDA) data on drug approvals during 1939–2006. As we cannot distinguish between active ingredients and their derivatives in the publications data, we consider the first word of each entry in the list of approved ingredients to be the ingredient name. This yields a list of 1,448 ingredients. This list includes drugs that were in use before the FDA was established. We do not use approval year information in the FDA data because medical research that pursues opportunities created by a new drug often starts before FDA approval (e.g. cyclosporine was first approved in 1983 but was mentioned in publications published two years earlier). Also, FDA administrative data is not always complete (e.g. the data report the approval year(s) of subsequent formulations, but not the base drug).

[20] Pharmaceutical research that leads to the discovery of new active ingredients precedes the drug-related research we examine. When active ingredient name is used in publications, the intended therapeutic use of the drug is already known and the associated patent application already filed. Pharmaceutical manufacturers apply for an active ingredient name for a drug during phase I or phase II clinical trials which happen after the pre-clinical testing has been completed and the drug has received an investigational drug application.

journals; basic drug development research is conducted by either pharmaceutical firms or academics. We set the cohort f of each measured research input (active ingredient) to the year prior to the year when the input is first mentioned in the a publication. We lump together research inputs with the same cohort f. A publication may mention research inputs from multiple cohorts. We count such multiple matches from one publication the same way we would count the matches if each match was from a separate publication.

4.3 Data on Disease Prevalence

To construct population aging and obesity epidemic related measures of disease prevalence over time we combine cross-sectional disease prevalence data with panel data on population characteristics. We estimate cross-sectional disease prevalence for each age and BMI group from Medical Expenditure Panel Survey data for years 1996–2005. We use Surveillance Epidemiology and End Results data for years 1975–2004 to estimate the share of people in each age group in each year. For each age group we use National Health Interview Survey data for years 1976–2005 to estimate the share of people in each BMI group in each year.

4.4 Sample Period

We set 1975–2005 as the sample period in the regression analyses, and in estimating the quality of research opportunities we limit the range of research input cohorts f to 1960–2001 and the range of publication years t to 1970–2002. Please see the data appendix for details.

5 The Empirical Models

We rely on population aging and obesity epidemic induced exogenous changes in disease prevalence to identify the "induced innovation effect" (see Section 6.1). Construction of the corresponding disease prevalence variables M_{it}^{AGING} and $M_{it}^{OBESITY}$ is discussed below.[21] The employed regression model, which is based on equation (5) and the discussion in Section 2.3, is:

$$ln\ N_{it}=\beta_A\ ln\ M_{it}^{AGING}+\beta_O\ ln\ M_{it}^{OBESITY}+\beta_K\ ln\ \hat{K}_{it}+\alpha_i+\alpha_t+u_{it}. \quad (10)$$

The variable N_{it} is a measure of medical research effort (publications) on the disease i in year t. We construct a measure of the quality of research opportunities, K_{it}, only for drug-related medical research but include this variable also in our analyses of other medical research to test whether our estimate of the "research opportunity effect" in drug-related research is biased by reverse causality (see Section 6.2). The variable u_t is the error term.

Parameters α_i and α_t represent disease and year fixed effects. Year fixed effects capture changes in publication norms and effort associated with each publication over time. We also employ an alternative specification with disease fixed effects α_i and disease-class specific year fixed effects $\alpha_{d,t}$. Parameters of interest are different in these specifications if the elasticity of substitution of research effort between diseases depends on whether the diseases

[21] See the Data Appendix to Section 4.3 for the derivation and rationale for this decomposition.

are in the same disease class. In the former (latter) specification the identifying variation for each parameter is the variation in the regressor within each disease relative to the corresponding variation for all other diseases (for other diseases in the same disease class).

The variable M_{it}^{AGING} is constructed as $M_{it}^{AGING} \equiv \sum_{j=1}^{5} \sum_{k=1}^{3} \mu_{i,j,k} \times s_{j,t}^{AGE} \times s_{j,k,t_0}^{BMI}$, where $\mu_{i,j,k}$ is the prevalence of disease *i* among people in the age group *j* who are in the BMI group *k*, $s_{j,t}^{AGE}$ is the share of people in the age group *j* in year *t*, and s_{j,k,t_0}^{BMI} is the share people in the age group *j* who are in the BMI group *k* in the initial year t_0.[22] Effectively, M_{it}^{AGING} reflects the prevalence of disease *i* in year *t* when age distribution varies over time but body weight distribution is fixed to the level seen in the initial year t_0 in the sample. The variable $M_{it}^{OBESITY}$ is constructed analogously as $M_{it}^{OBESITY} \equiv \sum_{j=1}^{5} \sum_{k=1}^{3} \mu_{i,j,k} \times s_{j,t_0}^{AGE} \times s_{j,k,t}^{BMI}$.

6 Identification Strategy

6.1 Identification of Induced Innovation Effects

The causal effect of potential market size (the number of people with a disease) on innovation cannot be inferred from the correlation between observed innovation and observed market size because causation runs in both directions. Acemoglu and Linn (2004) circumvent this problem by examining the relationship between changes in pharmaceutical innovation and changes in potential market size that are caused by population aging induced changes in disease prevalence. This identification strategy works when the effect of aging on disease prevalence varies across diseases, the age demographics of the population have changed over time, and the changes in the age demographics are mostly caused by changes in fertility.

We follow this general identification strategy in our analysis. However, we also take into account changes in the body weight distribution of the population. Using obesity as an exogenous source of variation is appropriate if the effect of obesity on disease prevalence varies across diseases, the body weight distribution of the population has changed over time, and the obesity epidemic is mostly exogenous to the rate of medical innovation.[23] Our descriptive statistics show that the first two conditions hold (see Section 7.1). It is also reasonable to expect that the third condition holds, since both the theoretical and empirical literature on the obesity epidemic point to factors other than developments in medical technology.[24]

[22] Parameters $\mu_{i,j,k}$, $s_{j,t}^{AGE}$ and $s_{j,k,t}^{BMI}$ are estimated from data on disease prevalence and demographics (see Section 4.3). The age groups are 0–18, 18–35, 35–50, 50–65 and 65+. The BMI groups are 18.5–25, 25–30 and 30–50. As we use disease and year fixed effects we can ignore population size and population growth in estimating disease prevalence. Please see the Data Appendix to Section 4.3 for further details.

[23] Of course, the changes in obesity (and in age demographics) must also be uncorrelated with factors captured by the error term.

[24] Our focus on aging and obesity is not meant to dispute that other factors also influence the extent of medical research. This focus is dictated by data availability and by the fact that both factors are known to have had a large impact on the prevalence of many diseases. Potentially important omitted factors include changes in disease severity (seefootnote 9), and changes in insurance coverage (see below). Also, as a population becomes wealthier its willingness to invest in developing treatments to diseases that affect mainly

6.2 Identification of the Research Opportunity Effect

With a different functional form for the benefit from medical research the optimal allocation of research effort across opportunity cohorts within a disease would depend on the extent of research on the disease. Consequently, changes in the level of research on a disease would impact estimates of the quality of research opportunities. There might thus be a positive empirical relationship between these two variables even if there was no causal effect from the quality of research opportunities on the extent of research effort.

To address this potential concern we take advantage of the fact that we examine two categories of medical research, namely drug-related medical research and other medical research. Unobserved effects that influence these two types of research are likely correlated. Consequently, if there is (is no) reverse causality from the level of drug-related research effort to the measure of quality of research opportunities in drug-related research, this measure of the quality of research opportunities will likely also be correlated (will likely be uncorrelated) with the level of other medical research. We can therefore test for the presence of reverse causality in our estimates of the research opportunity effect by including the measure of the quality of research opportunities in drug-related medical research also as a regressor in the analyses of the determinants of other medical research.[25]

7 Results

7.1 Descriptive Statistics

Changes in the age and body weight distributions in the U.S. population are well known, so we omit displaying them here. While the change has been gradual for both distributions, the change in the body weight distribution began more recently. Figure 1 shows the effects of these changes on the prevalence of each disease during the sample period (1975 vs. 2005). For both variables there is considerable variation in the effect (from −10% to +20%). These identifying variations are also not so correlated that the effects cannot be separately identified.

Figure 2.1 depicts the count of all publications and the count of publications with an abstract by the year of publication. The graph also shows the count of publications that are indexed with a disease (*Publications Indexed with a Disease*) and the count of publications that are indexed with a disease that is matched to an ICD-9 disease by our match (*Publications Matched*). Their comparison reveals that our match captures roughly 50% of all disease-linked medical research. The count of matches of publications to a disease (*Publication-Disease Matches*) is higher than the number of publications matched to at least one disease (*Publications Matched*) as a publication may be indexed to multiple diseases.

financially vulnerable populations may change. It is also possible that the attitudes toward these (or other populations such as children) change over time for not yet understood reasons, and that these changes influence the allocation of research.

Using data on insurance coverage from the NHIS, we conducted additional analyses to examine whether taking public and private insurance coverage into account in constructing the disease prevalence variables changes the results. This approach did not qualitatively change the estimates or their statistical significance.

[25] If the estimate of the research opportunity parameter β_K is close to zero when the dependent variable is a measure of other medical research, it is an indication that a positive estimate of the coefficient β_K when the dependent variable is a measure of drug-related medical research is not a result of reverse causality.

The two panels of Figures 2.2 depict the counts of publication-disease matches for each measure of drug-related medical research and other medical research in each year.[26] The count of publications for each measure is an important determinant of the precision of our estimates because the variance of the share of publications that are matched to a disease is expected to be inversely related to the count of publications and the estimated effects are identified from the effects on the share of publications that are matched to each disease. This is also the reason why we report weighted regression results.[27]

7.2 Estimates of the Quality of Research Opportunities

We estimate the quality of research opportunities, $\hat{K}_{it}$ using formula (6) and the iterative procedure described in Section 3. The estimates of the diffusion and exhaustion parameters are $\hat{\beta}_1$ = 0.0628 (s.e. 0.0045) and $\hat{\beta}_2$ = 0.0027 (s.e. 0.0004). Figure 2.3 shows that the predicted probability that is calculated based on these estimates tracks the observed probability closely except for when the ingredient age is 35 and over.[28]

7.3 Induced Innovation and Research Opportunity Effects

7.3.1 All Medical Research—Results for all medical research are shown in Table 1. The count of all publications N_{it}^{ALL} corresponds to the measure *Publication-Disease Matches* in Figure 2.1.[29] Columns 1 and 2 show that aging-induced increases in disease prevalence have increased research effort. By contrast, there is no evidence of a corresponding effect for obesity-induced changes in disease prevalence. Columns 3 and 4 show a positive relationship between the quality of research opportunities variable and the amount of total research.

Columns 3 and 4 also show that inclusion of the research opportunity variable renders the effect of aging-induced changes in disease prevalence statistically insignificant. However, a careful examination of residual plots from these regression (see Bhattacharya and Packalen, 2008a) suggests that, with the exception of the outlier disease 299 (Child Development Diseases in the Mental Disorders class) there is a robust positive relationship between aging-induced changes in disease prevalence and the changes in the overall research effort in the disease. Columns 5 and 6 show that when the disease 299 and the two other children’s mental health diseases (314 and 315) are excluded, the relationship between aging-induced changes in disease prevalence and the overall research effort is again statistically significant. Because the change in the age distribution has had such an unusual effect on the predicted disease prevalence for the disease 299 (see Figure 1) and because the dramatic increases in

[26] Their comparison with the count of matches of publications to a disease (*Publication-Disease Matches*) in Figure 2.1 shows that throughout the sample period the broadest (narrowest) measures of drug-related research and other research, respectively, represent roughly 45% and 55% (20% and 10%) of total matches.

[27] While we do not report the unweighted regressions, the residual graphs shown in Bhattacharya and Packalen (2008a) serve the same purpose and also show that the results are not the product of outliers.

[28] The share of publications that use ingredients aged 35 and over is artificially inflated by the fact that the publications data consists mostly of publications published after 1950. Our methodology of assigning the discovery year thus assigns cohorts between 1950 and 1965 for a disproportionate number of ingredients, as can be seen from Figure 2e in Bhattacharya and Packalen (2008a).

[29] The observations are weighted by the total count of publications matched to the disease during the sample period. That is, each observation is weighted by $\sum_{t=1975}^{2005} N_{it}^{ALL}$. The number of observations varies across columns because an observation is omitted if either $\hat{K}_{it}=0$ or $N_{it}^{ALL}=0$.

the number of diagnoses and research interest in the children's mental health diseases have been well recognized but without agreement over causes, in the subsequent analyses we exclude the three children's mental health diseases.[30]

The magnitude of our estimates of the impact of aging-induced changes in disease prevalence on medical research in columns 5 and 6 is similar to the magnitude of the corresponding estimate in Acemoglu and Linn (2004) for aging-induced pharmaceutical innovation. These estimates in our study and in Acemoglu and Linn (2004) are larger than the estimates in Finkelstein (2004) for the impact of vaccine policies on the number of new clinical trials.[31]

7.3.2 Drug-Related Medical Research—Results for the three measures of drug-related medical research are shown in Table 2.[32] There is robust evidence for aging-induced changes in the composition of drug-related medical research across diseases. By contrast, there is no evidence for a positive relationship between obesity-induced changes in disease prevalence and the amount of drug-related research on the disease. If anything, results suggest a negative relationship. Further below we examine a potential explanation for this (see Section 7.3.4).

There is also robust evidence for the hypothesis that the quality of opportunities influences the allocation of drug-related medical research effort across diseases.[33] As expected, estimates of this effect in Table 2 are larger compared to the case when the dependent variable is constructed from all medical research (see columns 3–6 in Table 1). Our estimates of this effect demonstrate-for the first time in the literature-that the direction of scientific research responds to changes in the quality of opportunities. The magnitude of this response relative to aging-induced changes in research effort can reflect either the possibility that the true response is indeed relatively weak, as changes in the quality of opportunities may be hard to identify at the time, or measurement error in the opportunity variable.

One potential problem is that mismeasurement and exclusion of the opportunity variable will lead to upward bias in estimates of the induced innovation effect if the opportunity variable and the variable associated with the induced innovation effect are correlated. To address this concern we estimated a constrained regression corresponding to the specification in column

[30] Research on children's mental health diseases has increased dramatically since the early 1990s and this increase is undoubtedly tied with the increase in the number of diagnoses for these diseases during the same period. While the unusual increase in the interest in these diseases is well known there is no agreement on why the increase has occurred. One explanation is that the increase in the diagnoses and the increase in research to the children's mental health diseases are consequences of the availability of dramatically better treatment options for these diseases, especially in the form of better knowledge of the effects of several drugs such as methylphenidate (ritalin). Methylphenidate was discovered in the 1950s and our measure of the quality of research opportunity is unable to predict the increase in research to these diseases because the increase happens 40 years after the discovery of the drug. An alternative explanation for why children's mental health diseases are outliers is that during the sample period there may have been a general disproportionate increase in research in diseases that primarily affect the children. We plan to explore this possibility in future research.

[31] Finkelstein's (2004) estimates too are derived from a difference-in-difference methodology. They reflect the impact on the composition of innovation rather than impact on the total extent of innovation. Acemoglu and Linn (2004) note this and emphasize that the responses to changes in relative market sizes across diseases and changes in total market size can be quite different. For this reason we do not use our estimates to construct an estimate of the dynamic welfare impact of induced medical research. Another interesting question left for future research is whether the responses are closer to optimal for private or public sector actors.

[32] For the measure *DRUG k*, where $k \in \{1,2,3\}$, each observation is weighted by $\sum_{t=1975}^{2005} N_{it}^{DRUG\ k}$.

[33] Figure 3d in Bhattacharya and Packalen (2008a) shows that this result is not driven by outliers.

6 of Table 2 with the coefficient on the opportunity variable fixed at 1.48 (two times its estimated value). In this constrained regression the coefficient on the aging variable decreases to 2.30 (from 4.06) but remains statistically significant at the 5% level. While this speaks to the robustness of the results to some measurement error in the opportunity variable, it must be interpreted with caution: if the true value of the coefficient on the opportunity variable is lower than the constrained value and the opportunity and disease prevalence variables are correlated, then the constrained estimates of the induced innovation effect will likely be biased downward. Extending methodologies relating to the measurement of the opportunity variable, as we have done in this paper, is therefore important also from the perspective of estimating induced innovation effects.

7.3.3 Other Medical Research—Results for the three measures of other medical research are shown in Table 3.[34] The coefficients on the quality of research opportunities in drug-related research are much smaller in Table 3 than the estimates of the coefficient on the same variable are in the analyses of drug-related research (see Table 2). These estimates are now also statistically insignificant, except in column 1 in which the dependent variable is the most inclusive measure of other medical research and which is thus the most likely of the three measures of other medical research to include some drug-related publications. As we argued in Section 6.2, such a null finding is evidence against the concern that reverse causality is the reason for the observed positive research opportunity effect for drug-related research.[35]

Columns 1–4 provide evidence of aging-induced changes in the composition of other medical research across diseases but again show no evidence of similar obesity-induced changes. Results for surgery-related research (the measure *OTHER 3*) in columns 5 and 6 show that the relationship between aging-induced changes in disease prevalence and the extent of surgery-related research on the disease is positive but not statistically significant. Results in columns 5–6 also suggest a possible negative relationship between obesity-induced changes in disease prevalence and the extent of surgery-related research on the disease. We examine a potential explanation for this next.

7.3.4 Effects on the Composition of Medical Research Within Diseases— Changes in disease prevalence may also have effects on the composition of research within diseases and such changes may influence estimates of the determinants of the extent of research effort across diseases (see Section 2.3). Accordingly, in Table 4 we report estimates of the determinants of the composition of research within diseases.[36]

[34] For the measure *OTHER k*, where $k \in \{1,2,3\}$, each observation is weighted by $\sum_{t=1975}^{2005} N_{it}^{OTHERS\ k}$.

[35] The power of this reverse causality test of course depends on the correlation between the unobserved factors that influence the extent of drug-related medical research and the unobserved factors that influence other medical research. The correlation between the residuals from the regression of the extent of drug-related medical research on disease prevalence and the opportunity variable and the residuals from the regression of other medical research on the same variables (or on only the disease prevalence variable) is between .26 and .44 depending on the fixed effects specification and the chosen measures of drug-related and other medical research. The relationship is highly ($p < 0.001$) statistically significant in all cases.

[36] In columns 1–2 (columns 3–4) each observation is weighted by $\sum_{t=1975}^{2005} N_{it}^{DRUG\ 3}$ (by $\sum_{t=1975}^{2005} N_{it}^{OTHER\ 3}$).

In columns 1 and 2, the dependent variable is the logarithm of the ratio of the most restrictive measure of drug-related research to all research. As expected, the results indicate a positive relationship between the opportunity variable and the share of research that is drug-related. The positive but statistically insignificant point estimate on the aging variable leaves open the possibility that drug-related research reacts to aging-induced changes in disease prevalence more strongly than all medical research. The negative but statistically insignificant coefficient on obesity-induced changes in disease prevalence implies that an obesity-induced increase in the prevalence of a disease may decrease the share of research on the disease that is drug-related and increase the share of research on the disease that is still applied but focused on the physiology of disease in the obese.

In columns 3 and 4, the dependent variable is the logarithm of the ratio of surgery-related research to all research. The negative estimate of the opportunity effect is additional evidence against the aforementioned reverse causality concern. Further, it is evidence that an increase in the quality of research opportunities in drug-related research shifts research effort away from other types research to drug-related research. We find no relationship between aging-induced changes in disease prevalence and the ratio of research that is surgery-related. The negative coefficient on obesity-induced changes in disease prevalence again suggests the possibility that an obesity-induced increase in the prevalence of a disease shifts resources away from general research to obesity-specific research on the disease.

8 Conclusion

Our results show that the composition of medical research across diseases responds to population aging induced changes in disease prevalence. This result is robust to the inclusion of a carefully constructed control measuring the quality of research opportunities in drug-related medical research. The results also suggest that obesity induced increase in the prevalence of a disease shifted research away from more general drug-related medical research and from more general surgery-related research on the disease, toward obesity-specific research on the disease.

These results provide support for the hypothesis that the direction of scientific research in medicine (conducted primarily in academic medical centers) responds to changes in the downstream market for that research, namely patients, just as private-sector technological innovation does. The distinction between scientific research and private-sector technological innovation is important because the two can differ in many ways, including for-profit vs. non-profit status and the level of control that individual researchers have.

While we do not examine the mechanisms that induce the direction of scientific research to respond to these factors, our analysis shows that these-yet largely unexplored–mechanisms have a desirable property in the sense that they induce scientific research to respond to factors that in part determine the socially optimal allocation of research resources. Our analysis is an important input into analyses of these mechanisms as it refutes the view of scientific medical research as an ivory tower in which scientists' desire to influence other scientists is the primary determinant of the direction of research. We expect that future research on these mechanisms, on how far or close the allocation of scientific research is from the socially optimal allocation, and on what factors besides opportunities and societal benefits are important drivers of the direction of scientific research, will be both fertile and worthy.

Acknowledgments

Bhattacharya thanks the National Institute on Aging for funding his work on this paper.

Data Appendix

Data Appendix to Section 4.1

We limit the match effort to diseases for which the MEPS disease incidence data includes at least 100 observations.[37] We do not match ICD-9 codes that include either the word “Other” or the word ’Unspecified” in the title because these ICD-9 codes typically include a variety of different diseases and are therefore difficult to match to the MESH vocabulary. Neither do we match diseases in the pregnancy category (class 11), in the congenital category (class 14), in the perinatal category (class 15), in the symptoms category (class 16), in the injuries category (class 17) or in the services category (class V). These classes are excluded from the match effort both in order to limit the scope of our match effort and because of the difficulty of matching diseases in these categories. If a match from an individual disease to a MESH entry/entries is not possible we try to match a group of ICD-9 codes to a MESH entry/entries. The 127 matched diseases account for 377,482 of the 745,355 disease mentions in the MEPS disease incidence data.

Because MESH is a hierarchical vocabulary, we also count all research that is indexed to any subnode of a matched MESH term as research that is related to the matched disease or group of diseases.[38] As the MESH vocabulary has changed over the years we make an effort to check that the MESH terms for the matched diseases have not changed in a way that would influence the research effort estimate. For the diseases for which the related publications from a year during the sample period are likely to have been indexed by terms other than the matched MESH entry/entries we exclude the observations from such years and from any of the preceding years. In Bhattacharya and Packalen (2008a) the match for such diseases is marked with an asterisk and the year prior to which any observations are excluded.

Data Appendix to Section 4.3

To estimate disease incidence for each age and BMI group we use the Medical Expenditure Panel Survey (MEPS) data from years 1996–2005.[39] Each subject is followed in MEPS for two years. For each subject we aggregate the observations in each year into one observation.

[37] We exclude HIV/AIDS because the disease does not appear in the publications database until the early 1980s and because the variations in the incidence of HIV/AIDS are obviously not mainly driven by aging or the obesity epidemic.

[38] We manually remove several matches of ICD-9 diseases to terms for neoplasms in MESH when the same neoplasm term is also mapped to a disease in the ICD-9 disease class 2 (neoplasms). MESH has 4982 disease terms. The match maps 1338 terms in MESH to the 127 diseases. 51 of the matched terms are mapped to 2 diseases and one term in MESH is mapped to 3 diseases. All other terms are mapped to only 1 disease.

[39] Because the trends in the changes in the age and body weight distributions have been similar across the developed nations we do not believe that using data on disease incidence, age demographics and obesity for the United States but data on world-wide publications is a significant concern.

MEPS includes a list of self-reported diseases that are coded by the International Classification of Diseases, Ninth Revision (ICD-9). MEPS does not include BMI information for years 1996–2000. We therefore use the National Health Interview Survey (NHIS) data from years 1996–2000 and the match between NHIS and MEPS to obtain BMI information for the observations in those years. Except for subjects in the age group 0–18 we exclude subjects without either age or BMI information.[40] The resulting MEPS data includes 262,958 observations on 149,737 subjects.

We use the Surveillance Epidemiology and End Results (SEER) data from years 1975–2004 to estimate the share of people in each age group in each year.[41] For each age group we use the NHIS data from years 1976–2005 to estimate the share of people in each BMI group in each year.[42]

In estimating the disease incidence parameters $\mu_{i,j,k}$ (see Section 5) we allow these parameters to vary by sex, race (black/non-black), insurance status (private/not private) and year but for expositional simplicity we omit these issues in the main text. As we don't measure changes in insurance coverage across time we do not examine the effect that changes in the insurance coverage across time may have on the benefit from medical research and on the extent of research.

The decomposition of changes in disease incidence to population aging and obesity epidemic induced changes (see Section 5) arises as follows. Let M_{it_0} denote the incidence of disease *i* in the initial year *t*$_0$. Let R_{it}^{AGING} denote the effect of aging alone on the incidence of disease *i* so that if only population aging affected the incidence of disease *i* the incidence of disease *i* would be $M_{it_0}R_{it}^{AGING}$ in year *t*. Let $\tilde{R}_{it}^{OBESITY}$ denote the additional effect of the obesity epidemic on the incidence of disease *i* so that if only aging and obesity affected the incidence of disease *i* the incidence of disease *i* would be $M_{it}{=}M_{it_0}R_{it}^{AGING}\tilde{R}_{it}^{OBESITY}$ in year *t*. Let $R_{it}^{OBESITY}$ denote the effect of obesity alone on the incidence of disease *i* so that if only obesity affected the incidence of disease *i* the incidence of disease *i* would be $M_{it_0}R_{it}^{AGING}$ in year *t*. Because R_{it}^{AGING} is small, $R_{it}^{OBESITY} \approx \tilde{R}_{it}^{OBESITY}$. Therefore, $ln\left(M_{it_0}R_{it}^{AGING}\tilde{R}_{it}^{OBESITY}\right) \approx ln\left(M_{it_0}R_{it}^{AGING}R_{it}^{OBESITY}\right)$. We can therefore decompose the total effect $ln\left(M_{it_0}R_{it}^{AGING}\tilde{R}_{it}^{OBESITY}\right)$ into an aging effect $ln\left(R_{it}^{AGING}\right)$ and an obesity effect $ln\left(R_{it}^{OBESITY}\right)$. Because the empirical specifications include either disease fixed effects, we can use the variables $ln\left(M_{it_0}R_{it}^{AGING}\right)$ and $ln\left(M_{it_0}R_{it}^{OBESITY}\right)$–

[40] Interpreting BMI of children is not as straightforward as interpreting BMI of adults. Hence, we do not distinguish the disease incidence by body weight for the age group 0–18. Consequently, we set $s_{1,1,t}^{BMI}{=}1$, $s_{1,2,t}^{BMI}{=}0$ and $s_{1,3,t}^{BMI}{=}0$ for all *t*. Because people the age group 0–18 have small average expenditures and also the effect of the obesity epidemic on disease incidence is small for this age group, ignoring the effect of the obesity epidemic on the disease incidence of this age group has a negligible influence on the potential market size variable M_{it}^{TOTAL}.

[41] We impute the values for 2005 by assuming that the change in the population in each age group from 2004 to 2005 was the same as it was from 2003 to 2004.

[42] We impute the values for 1975 by assuming the the body weight distribution was the same in 1975 as it was in 1976.

instead of the variables $ln\left(R_{it}^{AGING}\right)$ and $ln\left(R_{it}^{OBESITY}\right)$–as regressors. In the text these variables $ln\left(M_{it_0}R_{it}^{AGING}\right)$ and $ln\left(M_{it_0}R_{it}^{OBESITY}\right)$ are denoted by $ln\left(M_{it}^{AGING}\right)$ and $ln\left(M_{it}^{OBESITY}\right)$, respectively.

This decomposition reflects the fact that the effect that an obesity-induced change in disease incidence has had on the extent of research may be different than the effect that a corresponding aging-induced change in disease incidence has had on the extent of research. These two effects would be different, for example, if the implications of aging on disease incidence have been better understood than the implications of obesity on disease incidence, or if the change in age demographics was more expected than the obesity epidemic.

Data Appendix to Section 4.4

As our discussion of the descriptive statistics in Section 7.1 shows, there is a discontinuous jump in the share of publications with abstracts in the database from 1974 to 1975. Moreover, a number of diseases are indexed with different MESH terms before 1975 and especially before 1970 than they are after 1975. For these reasons we choose 1975–2005 as our sample period.

We determine the cohort of an ingredient (the year before the first mention of the ingredient-see Section 4.3) from the publications in years 1906–2005. In estimating the parameters that govern the quality of research opportunities (see Section 4) we limit the range of cohorts f to years 1960–2001 because there is a discontinuous jump in 1950 in the number of publications that are indexed in MEDLINE and because there is a discontinuous fall in the number of ingredients in a cohort from 2001 to 2002 due to the lag between the year in which an ingredient is first mentioned in the publications database and the year of FDA approval of the ingredient.[43] Because of this lag, because many of the diseases are indexed with different terms before 1970, and because in the subsequent analysis our focus is on the sample period 1975–2005, in estimating the quality of research opportunities (see Section 4) we limit the range of the years t to 1970–2002.

Background Appendix: Non-Profit Nature of Publications in Medicine

The intertwined nature of industrial R&D activity and academic research activity is well established. The two are connected in many ways, especially in the biomedical sector. In this appendix we discuss each of these connections. And in each case we argue that despite the connection, pharmaceutical innovation reflects largely the functioning of for-profit incentives and biomedical publications reflect largely the functioning of non-profit incentives.

[43] We multiply the initially estimated research opportunity by a factor that compensates for truncation. We assume that the average baseline productivity is the same before and after any truncation point. That is, the estimates are multiplied by $\{\sum_{t-f=1}^{\infty}e^{-\hat{\beta}_1(t-f)}\times[1-e^{-\hat{\beta}_2(t-f)}]\}/\{\sum_{t-f=1}^{t-1960}e^{-\hat{\beta}_1(t-f)}\times[1-e^{-\hat{\beta}_2(t-f)}]\}$ for all years $t\leq 2001$. For $t> 2001$ we also compensate for truncation due to the upper bound.

1. Pharmaceutical Innovation Reflects Mostly For-Profit Incentives

First, many of the innovations that are introduced by pharmaceutical companies are based on knowledge generated in the public sector (see e.g. Cockburn and Henderson, 1998 and Ward and Dranove, 1995). This fact is at odds with the interpretation in the related literature on the determinants of pharmaceutical innovation (e.g. Acemoglu and Linn, 2004), where research in the pharmaceutical industry is often taken as an example of innovation in the for-profit sector. However, we believe that, to a first approximation, it is reasonable to consider the majority of pharmaceutical innovation as reflective of the functioning of for-profit incentives. The alternative-that pharmaceutical firms do not make substantial choices about which lines of research to pursue but decide their research agenda mainly on the basis of prior basic research done in the public sector–seems to us less reasonable.

2. Research Publications Are Mostly from Academic Research Institutions

Second, individuals employed in pharmaceutical companies also publish in academic journals and co-author research papers with university researchers (see e.g. Cockburn and Henderson, 1998 and Adams and Clemons, 2008). Biomedical publishing will therefore reflect, in part, how for-profit incentives respond to determinants of innovation. Unfortunately no comprehensive study exists on what part of biomedical publishing can be attributed to industry. However, Adams and Clemons (2006) present summary statistics on the origin of scientific publications in a database of over 5000 journals across the sciences during the time period 1980–1999. Their analysis shows that during this time period the top 110 U.S. universities published 800,000 papers in medicine and the top 200 U.S. R&D firms published less than 30,000 papers in medicine. While this comparison is not comprehensive because it does not compare the biomedical publications of all U.S. universities with the biomedical publications of all pharmaceutical and biotechnology firms, the comparison suggests that the contribution of industry to academic publishing during this time period is substantially less voluminous than the contribution of research universities to academic publishing.

This conclusion is also supported by a National Science Foundation study which mentions that in 2003 in the science and engineering sector the academic sector accounted for almost three quarters of the publications originating in the U.S.[44] The remaining one quarter of the publications is attributed to industry, government and non-profits. The study also finds that only 6.0% of the publications that have at least one academic author have an industry coauthor. The available evidence thus suggests that the results in our analysis and the results in any other analysis that examines the determinants of publications in medicine in a comprehensive manner mainly reflect the publishing behavior of academic research institutions and the associated non-profit incentives as opposed to the publishing behavior of industry and the associated for-profit incentives.

[44] National Science Foundation, Division of Science Resources Statistics (NSF/SRS) 2006, “Industrial Funding of Academic R&D Continues to Decline in FY 2004,” NSF 06–315.

3. Academic Researchers in Medicine Mostly Do Not Patent

Third, in addition to publishing their work in academic publications university researchers also apply for patents. If patents and the associated for-profit incentives are a significant driving force behind academic biomedical research then biomedical publishing as the other product of biomedical research would also reflect the functioning of for-profit incentives. However, the analysis of patenting in medicine by Azoulay et al. (2007b) shows that during the period from 1981 to 2000 only 5% of faculty members in medical schools applied for a patent that was successfully granted.[45] Especially when this already low percentage figure is combined with the fact that most patented innovations bring no revenue to the patentee, we conclude that patenting and the associated for-profit incentives are likely not a significant determinant of biomedical research and publishing. Moreover, this conclusion is even stronger for our analysis as we only consider biomedical publications that are applied biomedical research in the sense that the publication is related to a specific disease and Azoulay et al. (2007b) find that biomedical patenting is much more common for basic research than for applied research.

Because patenting is not very common in medicine and particularly in the applied research that is our focus, patenting by other researchers is unlikely to influence the direction of research much in medicine. This is somewhat in contrast with the analysis of the anticommons hypothesis in biotechnology by Murray and Stern (2007) who find a modest anticommons effect. However, their analysis was based on selecting the biomedical publication (Nature Biotechnology) in which, *ex ante*, patenting by other researchers was the likeliest to have an effect on publication behavior. We also note that, to the best of our knowledge, in-force patents on approved drugs are not known to hinder the type of disease-driven medical research that we examine here.

4. Academic R&D is Mostly Funded by Non-Industry Sources

Fourth, some of the research activities of academic institutions are financed by industry. If industry funding is a major source of funding for academic R&D, the direction of academic R&D might simply reflect the direction of industry R&D and the associated for-profit incentives. However, a National Science Foundation study shows that in science and engineering during 1993–2004 academic R&D funds provided by industry have been less than 8% of all R&D funding.[46] The industry funding of academic R&D was $2.1 billion in 2004. This figure is substantially less than total university research expenditures ($43.0 billion in 2004) and federal support for university research expenditures ($27.4 billion in 2004). That figure is also substantially less than university research expenditures in medical sciences alone ($14 billion in 2004) and federal support for medical sciences ($9.4 billion in 2004). Considering the balance of the evidence, we conclude that research in academic medicine largely (though not exclusively) exemplifies the products of non-profit incentives.

[45] The analysis also shows that academic biomedical patents accounted for only 25% of total biomedical patenting even during the peak period of academic biomedical patenting (the late 1990s).

[46] National Science Foundation, Division of Science Resources Statistics (NSF/SRS) 2006, “Where Has All the Money Gone? Declining Industrial Support of Academic R&D,” NSF 06–328.

References

Acemoglu D, Linn J. Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry. Quarterly Journal of Economics. 2004; 119:1049–90.

Adams JD, Clemons JR. The NBER-Rensselaer Polytechnic Institute Scientific Papers Database: Characteristics and Purpose. Mimeo. 2006

Adams JD, Clemons JR. The Origins of Industrial Scientific Discoveries. NBER working paper. 2008; (13823)

Aghion P, Dewatripont M, Stein JC. Academic Freedom, Private-Sector Focus, and the Process of Innovation. RAND Journal of Economics. 2008; 39:617–35.

Azoulay P, Ding W, Stuart T. The Determinants of Faculty Patenting Behavior: Demographics or Opportunities? Journal of Economic Behavior & Organization. 2007; 63:599–623.

Azoulay P, Ding W, Stuart T. The Effect of Academic Patenting on the Rate, Quality, and Direction of (Public) Research Output. Journal of Industrial Economics. 2009; 57:637–76.

Azoulay P, Michigan R, Sampat BN. The Anatomy of Medical School Patenting. The New England Journal of Medicine. 2007b; 357:2049–56. [PubMed: 18003961]

Bhattacharya J, Packalen M. Is Medicine an Ivory Tower? Induced Innovation, Technological Opportunity, and For-Profit vs. Non-Profit Innovation. NBER working paper No 13862. 2008a

Bhattacharya J, Packalen M. The Other Ex-Ante Moral Hazard in Health. NBER working paper No 13863. 2008b

Caballero, RJ., Jaffe, AB. How High are The Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth. In: Blanchard, OJ., Fisher, S., editors. NBER Macroeconomics Annual 1993. Cambridge: MIT Press; 1993. p. 15-74.

Cameron AC, Gelbach JB, Miller DL. Bootstrap-Based Improvements for Inference with Clustered Errors. NBER technical working paper No 344. 2007

Cockburn IM, Henderson RM. Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery. The Journal of Industrial Economics. 1998; 46:157–82.

Dasgupta P, David PA. Towards a new Economics of Science. Research Policy. 1994; 23:487–521.

DellaVigna S, Pollet JM. Demographics and Industry Returns. American Economic Review. 2007; 97:1667–702.

Finkelstein A. Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry. Quarterly Journal of Economics. 2004; 119:527–64.

George L, Waldfogel J. Who Affects Whom in Daily Newspaper Markets? Journal of Political Economy. 2003; 111:765–84.

Glaeser, EL., editor. The Governance of Not-for-Profit Organizations. Chicago: The University of Chicago Press; 2003.

Jaffe AB, Trajtenberg M. Flows of Knowledge from Universities and Federal Labs: Modeling the Flow of Patent Citations Over Time and Across Institutional and Geographic Boundaries. Proceedings of the National Academy of Sciences. 1996; 93:12671–7.

Hicks, JR. Theory of Wages. London: Macmillan; 1932.

Lakdawalla D, Philipson T. The Nonprofit Sector and Industry Performance. Journal of Public Economics. 2006; 90:1681–98.

Lichtenberg, FR. The Allocation of Publicly Funded Biomedical Research. In: Berndt, E., Cutler, D., editors. Medical Care Output and Productivity: Studies in Income and Wealth. Vol. LXIII. Chicago: University of Chicago Press; 1999.

Lichtenberg FR. Importation and Innovation. NBER working paper No 12539. 2006

Lichtenberg FR, Waldfogel J. Does Misery Love Company? Evidence from Pharmaceutical Markets Before and After the Orphan Drug Act. NBER working paper No 9750. 2003

Mane KK, Börner K. Mapping topics and topic bursts in PNAS. Proceedings of the National Academy of Sciences. 2004; 101:5287–90.

Merton, RK. The Normative Structure of Science. In: Merton, RK., editor. The Sociology of Science: Theoretical and Empirical Investigations. Chicago: The University of Chicago Press; 1973. [1942]

Michel J-B, Shen Y-K, Aiden AP, Veres A, Gray MK, The Google Books Team. Pickett JP, Hoiberg D, Clancy D, Norvig P, Orwant J, Pinker S, Nowak MA, Aiden EL. Quantitative Analysis of Culture Using Millions of Digitized Books. Science. 2011 Jan 14.:176–82.

Murray F, Stern S. Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the Anti-Commons Hypothesis. Journal of Economic Behavior and Organizations. 2007; 63:648–87.

Newell RA, Jaffee A, Stavins R. The Induced Innovation Hypothesis and Energy-Saving Technological Change. Quarterly Journal of Economics. 1999; 114:907–40.

Popp D. Induced Innovation and Energy Prices. American Economics Review. 2002; 92:160–80.

Rosenberg, N. Inside the Black Box: Technology and Economics. Cambridge: Cambridge University Press; 1982.

Saha, SB., Weinberg, B. The Economics of Ivory Towers. Mimeo; 2008.

Scherer FM. Firm Size, Market Structure, Opportunity, and the Output of Patented Inventions. American Economic Review. 1965; 55:1097–125.

Schmookler, J. Invention and Economic Growth. Cambridge: Harvard University Press; 1966.

Stern S. Do Scientists Pay to Be Scientists? Management Science. 2004; 50:835–53.

Waldfogel J. Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated Product Markets. RAND Journal of Economics. 2003; 34:557–68.

Ward MR, Dranove D. The Vertical Chain of Research and Development in the Pharmaceutical Industry. Economic Inquiry. 1995; 33:70–87.

Yin W. Market Incentives and Pharmaceutical Innovation. Journal of Health Economics. 2008; 27:1060–1077. [PubMed: 18395277]

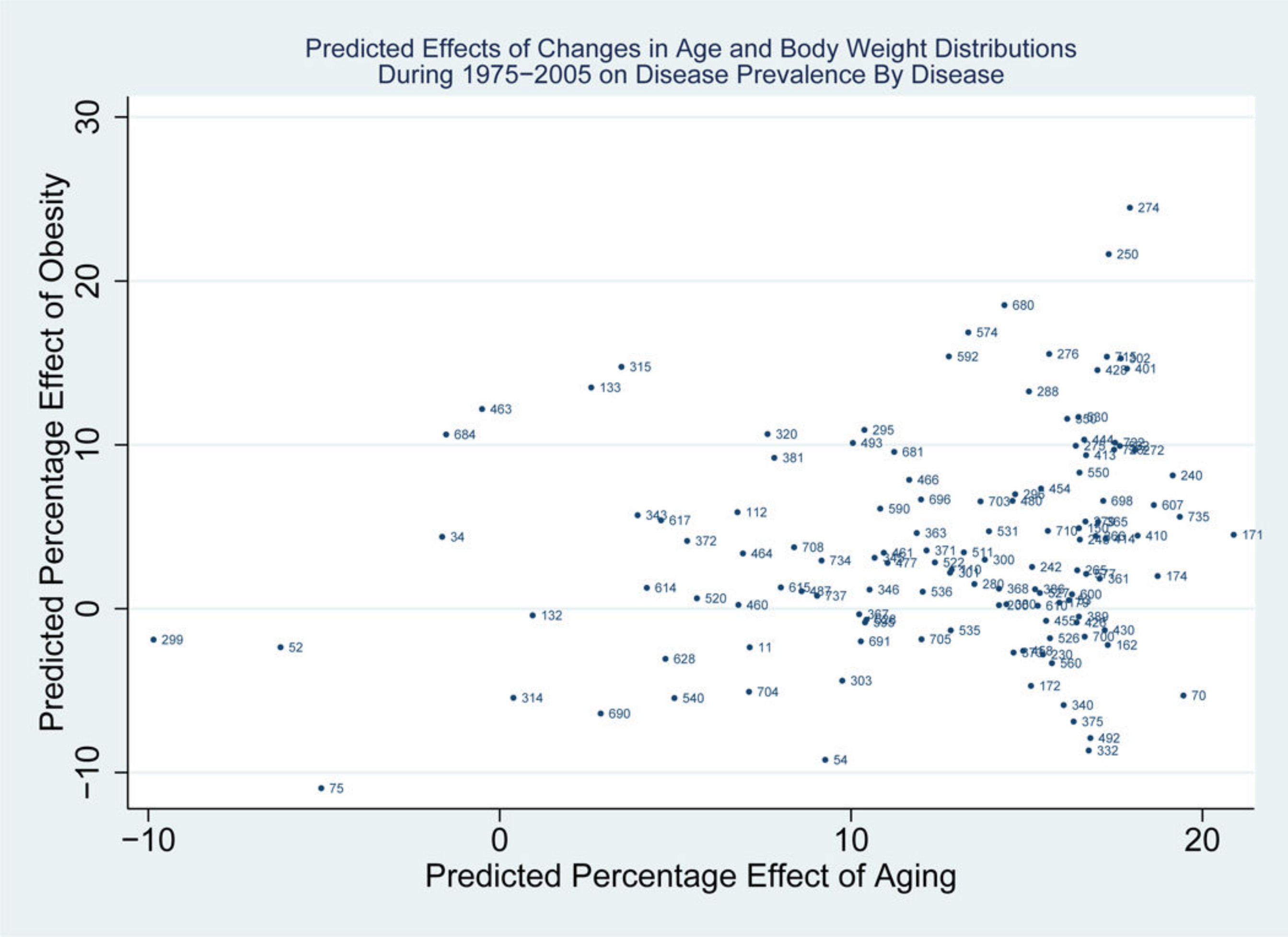

Figure 1.

Predicted Effects of Changes in Age and Body Weight Distributions on Disease Prevalence.

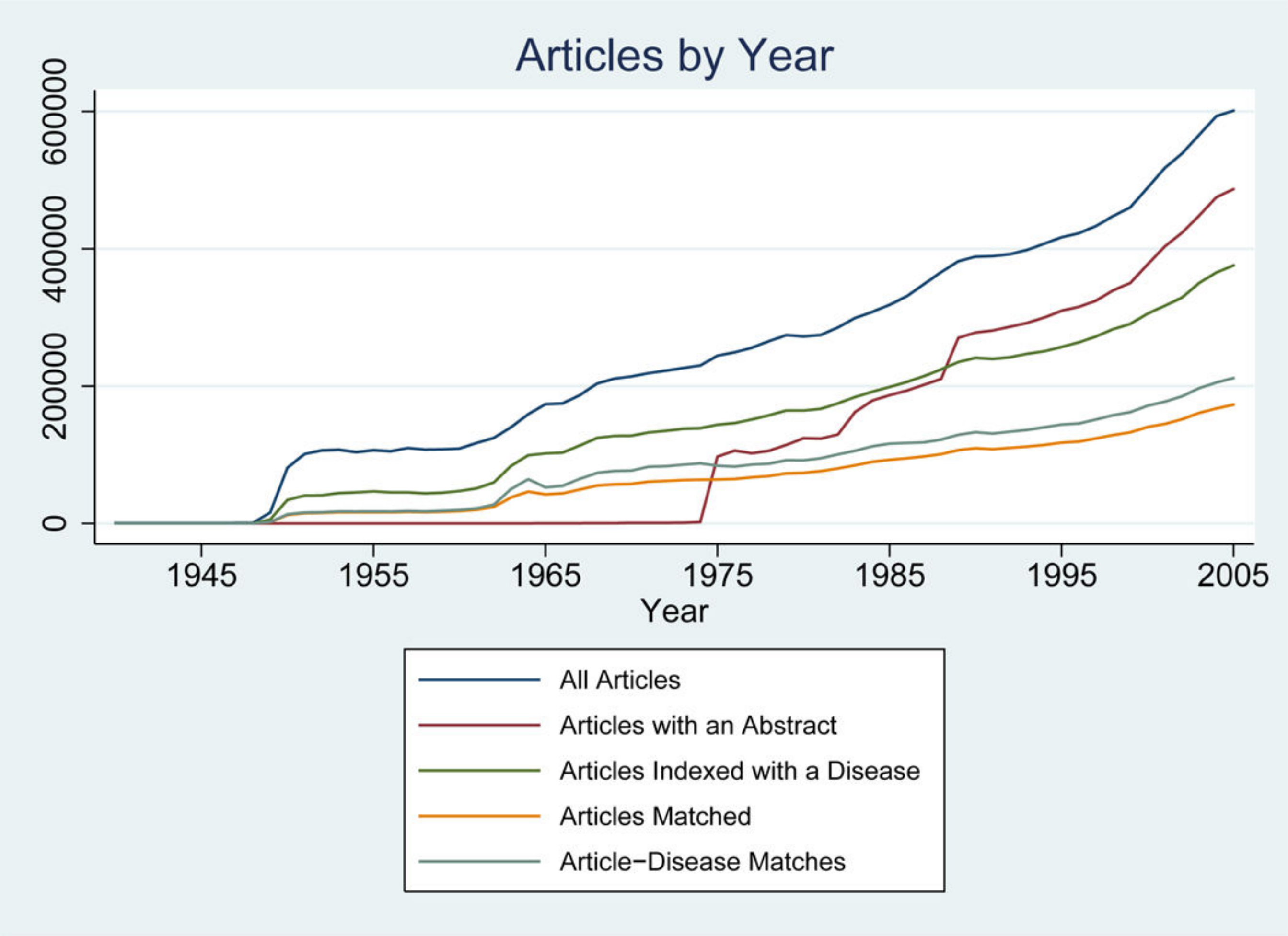

Figure 2.1.

Publications in the MEDLINE Database by Year.

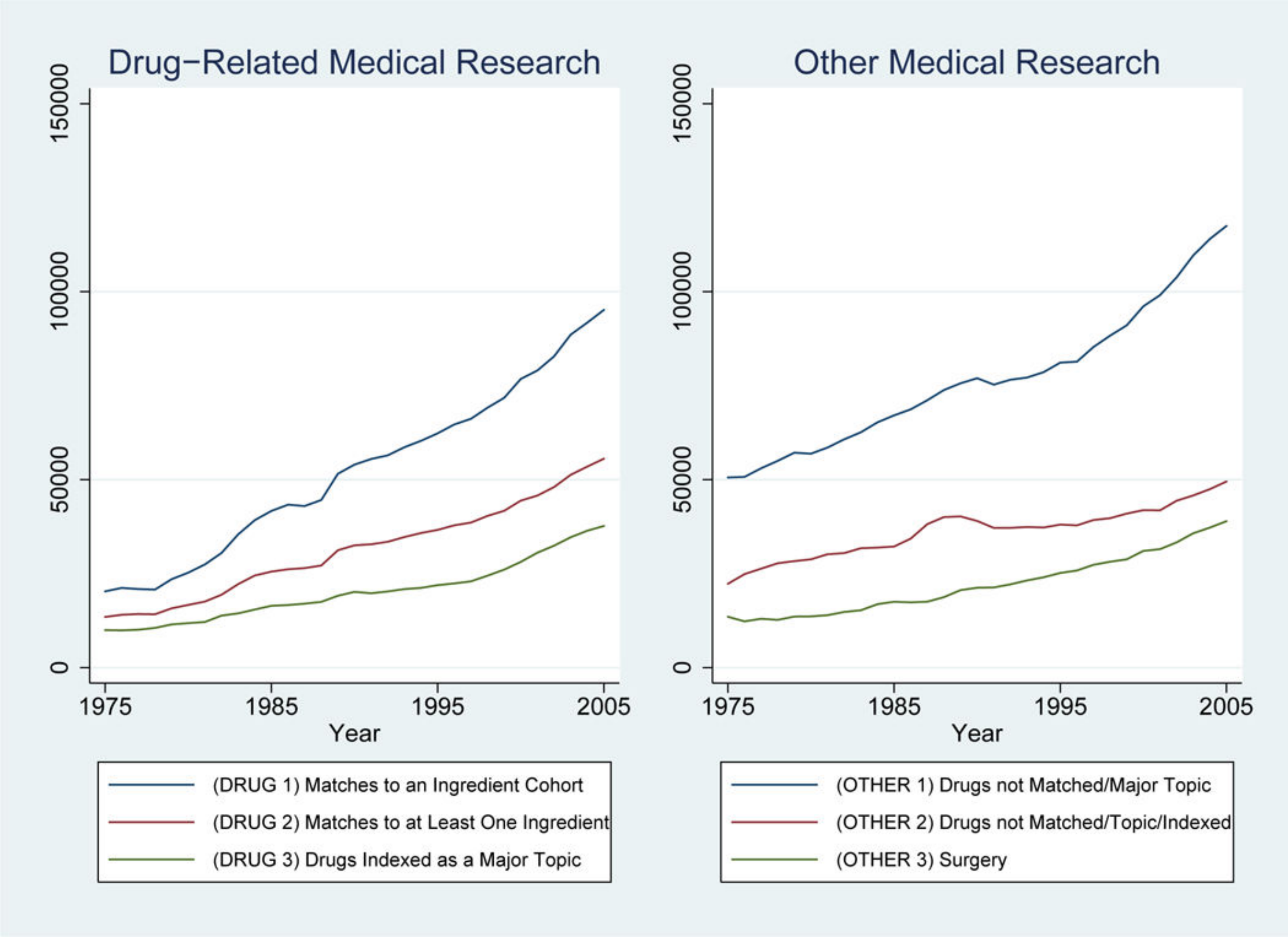

Figure 2.2.

Matches to Drug-Related and Other Publications in the MEDLINE Database by Year.

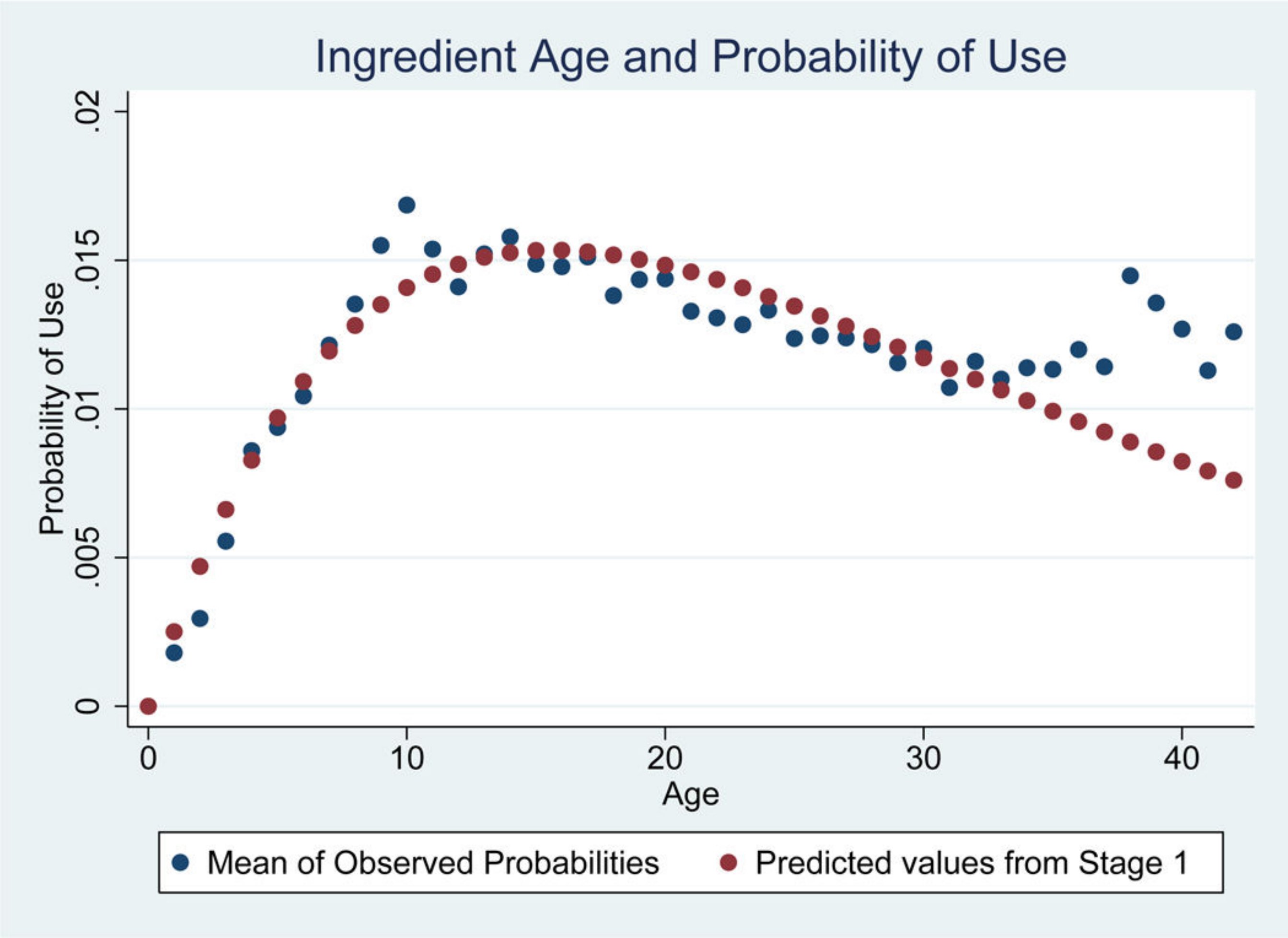

Figure 2.3.
Predicted and Actual Probability of Opportunity Cohort Use as a Function of Cohort Age.

Table 1

Determinants of the Allocation of the All Medical Research Across Diseases.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dependent variable:	$ln\left(N_{it}^{ALL}\right)$	$ln\left(N_{it}^{ALL}\right)$	$ln\left(N_{it}^{ALL}\right)$	$ln\left(N_{it}^{ALL}\right)$	$ln\left(N_{it}^{ALL}\right)$	$ln\left(N_{it}^{ALL}\right)$
$ln\left(\hat{K}_{it}\right)$			0.32 [0.16] p_{wild} = 0.027	0.22 [0.15] p_{wild} = 0.137	0.29 [0.14] p_{wild} = 0.014	0.22 [0.15] p_{wild} = 0.149
$ln\left(M_{it}^{AGING}\right)$	2.74 [1.43] p_{wild} = 0.037	2.40 [1.29] p_{wild} = 0.083	1.76 [1.16] p_{wild} = 0.125	1.84 1.18] p_{wild} = 0.168	2.66 [1.47] p_{wild} = 0.062	2.76 [1.28] p_{wild} = 0.023
$ln\left(M_{it}^{OBESITY}\right)$	0.43 [1.18] p_{wild} = 0.773	−0.004 [0.71] p_{wild} = 0.996	−0.25 [0.98] p_{wild} = 0.800	−0.13 [0.68] p_{wild} = 0.858	−0.07[1.05] p_{wild} = 0.947	−0.13 [0.68] p_{wild} = 0.857
Fixed effects	Disease, Class × Year	Disease, Year	Disease, Class × Year	Disease, Year	Disease, Class × Year	Disease, Year
Number of observations	3884	3884	3883	3883	3796	3796

Our statistical inference is based on p_{wild} which is calculated using the cluster-robust standard error (clustered at the class level) and the wild cluster bootstrapped distribution of the *t*-statistic (1000 iterations). Monte Carlo evidence favors this approach when the number of clusters is small and the clusters are unbalanced (Cameron et al., 2007). The wild cluster bootstrapped standard error (1000 iterations) is presented in brackets. In columns 5 and 6 children's mental health diseases (299, 314, 315) are omitted.

Table 2

Determinants of the Allocation of Drug-Related Medical Research Across Diseases.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	$ln\left(N_{it}^{DRUG1}\right)$	$ln\left(N_{it}^{DRUG1}\right)$	$ln\left(N_{it}^{DRUG2}\right)$	$ln\left(N_{it}^{DRUG2}\right)$	$ln\left(N_{it}^{DRUG3}\right)$	$ln\left(N_{it}^{DRUG3}\right)$
$ln\left(\hat{K}_{it}\right)$	0.64 [0.36] $p_{wild} = 0.021$	0.58 [0.31] $p_{wild} = 0.037$	0.59 [0.30] $p_{wild} = 0.022$	0.48 [0.28] $p_{wild} = 0.057$	0.85 [0.35] $p_{wild} = 0.006$	0.74 [0.31] $p_{wild} = 0.010$
$ln\left(M_{it}^{AGING}\right)$	2.51 [1.81] $p_{wild} = 0.200$	2.32 [1.09] $p_{wild} = 0.008$	2.46 [1.17] $p_{wild} = 0.185$	2.73 [1.22] $p_{wild} = 0.015$	3.85 [1.92] $p_{wild} = 0.030$	4.06 [1.93] $p_{wild} = 0.019$
$ln\left(M_{it}^{OBESITY}\right)$	−1.75 [2.02] $p_{wild} = 0.525$	−1.79 [1.29] $p_{wild} = 0.163$	−1.77 [1.91] $p_{wild} = 0.489$	−1.57 [1.15] $p_{wild} = 0.223$	−2.08 [2.11] $p_{wild} = 0.451$	−1.87 [1.49] $p_{wild} = 0.208$
Fixed effects	Disease, Class × Year	Disease, Year	Disease, Class × Year	Disease,Year	Disease, Class × Year	Disease,Year
Number of observations	3730	3730	3730	3730	3697	3697

Children's mental health diseases (299, 314, 315) are omitted. See the footnote to Table 1 for an explanation of the standard errors and *p*-values.

Table 3

Determinants of the Allocation of Other Medical Research Across Diseases.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	$ln\left(N_{it}^{OTHER1}\right)$	$ln\left(N_{it}^{OTHER1}\right)$	$ln\left(N_{it}^{OTHER2}\right)$	$ln\left(N_{it}^{OTHER2}\right)$	$ln\left(N_{it}^{OTHER3}\right)$	$ln\left(N_{it}^{OTHER3}\right)$
$ln\left(\hat{K}_{it}\right)$	0.20 [0.12] p_{wild} = 0.086	0.15 [0.13] p_{wild} = 0.339	0.06 [0.09] p_{wild} = 0.566	−0.007 [0.09] p_{wild} = 0.950	−0.11 [0.18] p$wild$ = 0.507	−0.06 [0.16] p$wild$ = 0.726
$ln\left(M_{it}^{AGING}\right)$	2.84 [1.53] p_{wild} = 0.056	2.82 [1.35] p_{wild} = 0.018	2.98 [1.57] p_{wild} = 0.062	2.79 [1.36] p_{wild} = 0.025	1.64 [1.24] p_{wild} = 0.181	2.59 [1.55] p_{wild} = 0. 114
$ln\left(M_{it}^{OBESITY}\right)$	0.23 [0.92] p_{wild} = 0.798	−0.11 [0.52] p_{wild} = 0.830	0.13 [0.89] p_{wild} = 0.897	−0.19 [0.69] p_{wild} = 0.787	−0.51 [1.20] p_{wild} = 0.684	−1.49 [0.81] p_{wild} = 0.066
Fixed effects	Disease, Class × Year	Disease, Year	Disease, Class × Year	Disease, Year	Disease, Class ×Year	Disease, Year
Number of observations	3796	3796	3796	3796	3723	3723

Children’s mental health diseases (299, 314, 315) are omitted. See the footnote to Table 1 for an explanation of the standard errors and p-values.

Table 4

Determinants of the Allocation of Medical Research Across Research Types Within Diseases.

	(1)	**(2)**	**(3)**	**(4)**
Dependent variable:	$ln(\frac{N_{it}^{DRUG3}}{N_{it}^{ALL}})$	$ln(\frac{N_{it}^{DRUG3}}{N_{it}^{ALL}})$	$ln(\frac{N_{it}^{OTHER3}}{N_{it}^{ALL}})$	$ln(\frac{N_{it}^{OTHER3}}{N_{it}^{ALL}})$
$ln\left(\hat{K}_{it}\right)$	0.53 [0.21] $p_{wild} = 0.005$	0.54 [0.21] $p_{wild} = 0.001$	−0.25 [0.14] $p_{wild} = 0.043$	−0.20 [0.12] $p_{wild} = 0.089$
$ln\left(M_{it}^{AGING}\right)$	0.79 [1.06] $p_{wild} = 0.494$	1.58 [1.05] $p_{wild} = 0.196$	−0.62 [1.04] $p_{wild} = 0.563$	−0.77 [0.92] $p_{wild} = 0.459$
$ln\left(M_{it}^{OBESITY}\right)$	−1.69 [1.28] $p_{wild} = 0.242$	−1.76 [1.07] $p_{wild} = 0.105$	−0.65 [0.46] $p_{wild} = 0.248$	−0.89 [0.48] $p_{wild} = 0.073$
Fixed effects	Disease, Class × Year	Disease, Year	Disease, Class × Year	Disease, Year
Number of observations	3697	3697	3723	3723

Children’s mental health diseases (299, 314, 315) are omitted. See the footnote to Table 1 for an explanation of the standard errors and *p*-values.