IS MEDICINE AN IVORY TOWER? INDUCED INNOVATION, TECHNOLOGICAL
OPPORTUNITY, AND FOR-PROFIT VS. NON-PROFIT INNOVATION

Jay Bhattacharya
Mikko Packalen

Is Medicine an Ivory Tower? Induced Innovation, Technological Opportunity, and For-Profit
vs. Non-Profit Innovation
Jay Bhattacharya and Mikko Packalen
NBER Working Paper No. 13862
March 2008
JEL No. I1,L31,O33

ABSTRACT

This paper examines whether the composition of medical research responds to changes in disease incidence
and research opportunities. The paper also provides new evidence on induced pharmaceutical innovation.
In both cases we use the change in the demographic structure of the market (measured by age structure
and obesity prevalence) to test the induced innovation hypothesis. Technological opportunity is calculated
from estimates of structural productivity parameters. The extent of inventive activity is measured from
the MEDLINE database on 16 million biomedical publications. We match these data with data on
disease incidence. We show that medical research responds to changes in disease incidence and research
opportunities. We also find that pharmaceutical innovation responds to aging- and obesity-induced
changes in potential market size.

Jay Bhattacharya
117 Encina Commons
Center for Primary Care
and Outcomes Research
Stanford University
Stanford, CA 94305-6019
and NBER
jay@stanford.edu

Mikko Packalen
University of Waterloo
Department of Economics
200 University Avenue West
Waterloo, ON N2L 3G1
Canada
packalen@uwaterloo.ca

1 Introduction

One key virtue of for-profit allocation is that decisions made by for-profit firms must necessarily respond to changes in the market. That is, for-profit producers must shape their products to fit the demands of their consumers or clients, or else risk failure. There is abundant evidence that for-profit producers (for instance in the pharmaceutical industry) innovate according to market demand (see Acemoglu and Linn, 2004).

Non-profit allocation, on the other hand, imposes looser budget constraints on firms (Lakdawalla and Philipson, 2006). In principle at least, loose constraints could insulate non-profit producers from the raw demand of the market, and hence divorce production decisions from demand. This sort of reasoning has often been applied to academic publishing, and is a good working definition of an academic ivory tower. While there may be good reasons to insulate some markets with non-profit production from the vagaries of the market, academic medicine is not such a market. Despite the evident importance that the non-profit producers of academic medicine respond to the market (that is, the epidemiology of patient health), there is little extant evidence that they do. In fact, to our knowledge there is little extant evidence on whether non-profit allocation induces a link between market demand and innovation in any market.

We show that the direction of medical research does respond to changes in the expected number of people who benefit from the research and changes in technological opportunity. The identification of the technological opportunity effect relies on our analysis of a formal model of optimal allocation of research effort. The analysis enables us to estimate the structural productivity parameters that govern technological opportunity. The disease incidence effect is identified using aging- and obesity-driven exogenous variation in disease incidence. To our knowledge our study is the first study on non-profit innovation that identifies the technological opportunity effect and the first study on non-profit innovation that uses exogenous variation to identify the induced innovation effect. We also present evidence on aging- and

obesity-induced pharmaceutical innovation. While the finding of aging-induced pharmaceutical innovation replicates a finding in the existing literature the finding of obesity-induced pharmaceutical innovation is new.

The empirical analysis is facilitated by the construction of a match between a medical vocabulary and data on disease incidence. This match enables us to use the massive indexed MEDLINE database on 16 million biomedical publications to measure innovation in medicine. To our knowledge our study is the first to take advantage of the panel nature of this database which is rich in its information content and thereby has great potential for future innovation research.

2 Background

The literature on the determinants of non-profit innovation is near non-existent. Lichtenberg (1999) and Lichtenberg (2006) are exceptions. Lichtenberg (1999) finds a positive correlation between public biomedical funding and both the disease prevalence and the disease severity. Lichtenberg (2006) finds a positive correlation between cancer incidence and the number of biomedical publications. Unlike these studies, we use panel data and exogenous variation in the disease incidence to identify the induced innovation effect, and we also determine the technological opportunity effect.

Our contribution builds on two connected strands of literature on for-profit innovation. 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), which we discuss below, Finkelstein (2004) and Lichtenberg and Waldfogel (2003). Technological opportunity was examined first by Scherer (1965) and Schmookler (1966), and more recently by Popp (2002) which we discuss below.

The studies most closely related to ours are Popp (2002) and Acemoglu and Linn (2004). Popp (2002) uses data on energy prices and patenting activity across energy technologies

over time and finds a positive relationship between innovation and both energy prices and technological opportunity.[1] We too examine both induced innovation and the technological opportunity effect. But in contrast with Popp (2002) our main focus is on non-profit innovation and we use data on biomedical publications to measure inventive activity.

Acemoglu and Linn (2004) use the change in age demographics to identify the effect of market size on pharmaceutical innovation and find a positive empirical relationship.[2] In comparison, we use changes in both age demographics and obesity prevalence to identify the induced innovation effect. Also, whereas Acemoglu and Linn (2004) measure the extent of pharmaceutical innovation from new drug introductions, which reflects for-profit production, we measure it from biomedical publications, which reflects non-profit production. An additional difference between our work and Acemoglu and Linn (2004) is that our main focus is not on induced pharmaceutical innovation but on the determinants of medical research.

The methodology we use to estimate technological opportunity builds on the methodology developed in the studies on patenting by Caballero and Jaffe (1993) and Jaffe and Trajtenberg (1996). In comparison with these studies we derive the estimating equation from a model of optimal allocation of research effort. This enables us to estimate the structural productivity parameters. Moreover, the probability that a given knowledge cohort is used in research depends also on the quality of other existing knowledge cohorts.

Our finding of obesity-induced innovation is also related to the empirical studies on preference externalities by Waldfogel (2003) and George and Waldfogel (2003). In these studies the preference externality arises from racial characteristics whereas in our case the preference externality is determined by the consumers' decisions about body weight (to the extent it is a decision). In a companion paper (Bhattacharya and Packalen, 2008) we calculate the welfare effect of the induced innovation externality of obesity.

---

[1] Newell, Jaffee and Stavins (1999) exploit the changes in energy prices and changes in the cost and energy efficiency of air conditioners to examine the effect of energy prices on the direction of technological change.

[2] DellaVigna and Pollet (2007) exploit the changes in the age demographics to study stock market returns.

3   Theory

We first present a model of medical research. In the second subsection we discuss why technological opportunity in medical research is also a measure of pharmaceutical innovation.

3.1   A Model of Medical Research

We distinguish between two types of medical research: drug-related medical research and other medical research. In both cases, following the induced innovation hypothesis, the optimal research effort is increasing in disease incidence. In the case of drug-related medical research the research effort is also in part determined by technological opportunity as new research opportunities are presented by the discovery of new active ingredients. The pharmaceutical research that leads to the discovery of new active ingredients precedes the type of medical research examined here. The model of allocation of other medical research is a special case of the model of allocation of drug-related medical research. Therefore, we only consider the allocation of drug-related medical research in the formal analysis.

We divide drug-related medical research into mutually exclusive categories by the disease examined in the research and by the active ingredient used in the research. We also lump together the ingredients by their year of discovery (cohort), denoted by f. We first consider the optimal allocation of drug-related research effort in the disease i in year t across cohorts f. We assume that the benefit of research on disease i that uses ingredients from a cohort f depends on a measure of the baseline productivity of the cohort of ingredients f in research on the disease i, denoted by a_{if}, on the elapsed time $t - f$ since the initial discovery of the ingredients, on the extent of the research effort in year t that uses an ingredient from the cohort f in research on the in disease i, denoted N_{itf}, on the expected number of people with the disease i in year t, denoted by M_{it}, and on other factors, denoted by ε_{itf}.

Specifically, we assume that the benefit from research on the disease i in year t that uses

an ingredient from the cohort of ingredients f is

$$M_{it} \times \left\{ \alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)} \right] + \varepsilon_{itf} \right\} \times \ln\left(N_{itf}\right). \tag{1}$$

In this expression the factor $\left[1 - e^{-\beta_2(t-f)} \right]$ represents the lag after the discovery of the ingredients in the cohort f before the full potential of this cohort of ingredients in medical research on the disease i is revealed. The factor $e^{-\beta_1(t-f)}$ represents the eventual decay in the usefulness of the cohort of ingredients f in medical research as the properties of this cohort of ingredients become established knowledge.[3] We assume that the variable ε_{itf} is observable to the medical researchers and satisfies $E\left[\varepsilon_{itf}\right] = 0$.

The range of cohorts f is $\{f_0, f_0 + 1, ..., t\}$ in year t. The total benefit from research on the disease i in year t is the sum of the benefit (1) over all cohorts:

$$M_{it} \sum_{f=f_0}^{t} \left\{ \alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)} \right] + \varepsilon_{itf} \right\} \times \ln\left(N_{itf}\right). \tag{2}$$

We assume that the allocation of research across cohorts of ingredients within a disease is optimal. Denoting $p_{itf}^* \equiv \dfrac{N_{itf}^*}{\sum_{f=f_0}^{f} N_{itk}^*}$ the first-order conditions for the optimum imply that[4]

$$p_{itf}^* = \frac{\alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)} \right] + \varepsilon_{itf}}{\sum_{f=f_0}^{t} \left\{ \alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)} \right] + \varepsilon_{itf} \right\}}. \tag{3}$$

We now consider the optimal allocation of research effort across diseases. We assume

---

[3] We do not model explicitly the effect that the amount of research in the preceeding 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.

[4] Denoting the optimal level of effort by N_{itf}^*, the first-order condition for the optimum is $M_{it} \times \alpha_{if} \times \dfrac{e^{-\beta_1(t-f)} \times \left[1-e^{-\beta_2(t-f)}\right] + \varepsilon_{itf}}{N_{itf}^*} = M_{it} \times \alpha_{if'} \times \dfrac{e^{-\beta_1(t-f')} \times \left[1-e^{-\beta_2(t-f')}\right] + \varepsilon_{itf'}}{N_{itf'}^*}$ for all (i, t, f, f'). Denoting $c_{itf} \equiv \alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)}\right] + \varepsilon_{itf}$ this condition can be rewritten as $c_{itf} \times N_{itf'}^* = N_{itf}^* \times c_{itf'}$ for all (i, t, f, f'). Taking the sum of both sides of the equation $c_{itf} \times N_{itf'}^* = N_{itf}^* \times c_{itf'}$ over all $f' \in \{f_0, ..., t\}$ 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 (3) in the text.

that the overall benefit from research in year t is the sum of the benefit from research in each disease (see the expression (2)) over all diseases. The overall benefit is therefore

$$\sum_i M_{it} \sum_{f=f_0}^{t} \left\{ \alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)}\right] + \varepsilon_{itf} \right\} \times \ln\left(N_{itf}\right). \tag{4}$$

If the allocation of resources across cohorts of ingredients within a disease is optimal, using the definition of p^*_{itf} and the definition $N_{it} \equiv \sum_{f=f_0}^{t} N_{itf}$ the expression (4) for the overall benefit from research can be rewritten as

$$\sum_i M_{it} \sum_{f=f_0}^{t} \left\{ \alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)}\right] + \varepsilon_{itf} \right\} \times \ln\left(N_{it} \times p^*_{itf}\right). \tag{5}$$

If also the allocation of resources across diseases is optimal we have

$$N_{it} = \sum_i N_{it} \times \frac{M_{it} \sum_{f=f_0}^{t} \alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)}\right] + \varepsilon_{itf}}{\sum_i M_{it} \sum_{f=f_0}^{t} \left\{ \alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)}\right] + \varepsilon_{itf} \right\}} \tag{6}$$

for all (t, i) by the first-order conditions for the optimum.

The factor $\sum_{f=f_0}^{t} \left\{ \alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)}\right] + \varepsilon_{itf} \right\}$ in the equation (6) is a measure of technological opportunity in research on the disease i in year t. Denoting this measure of technological opportunity by K_{it} we can rewrite the equation (6) 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

$$\ln N_{it} = \ln K_{it} + \ln M_{it} + \alpha_t, \tag{7}$$

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

The empirical predictions of this model are two-fold. First, by the equation (7) the model predicts a proportional relationship between research effort in a disease and technological opportunity for research on the disease and a proportional relationship between the research effort in a disease and the disease incidence. 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. Second, the model predicts the relationship (3) between the parameters α_{if}, β_1 and β_2 that govern technological opportunity K_{it} and the probability that a cohort of ingredients is used in research. We use this predicted relationship to estimate the parameters α_{if} and the parameters β_1 and β_2 in order to construct an estimate of technological opportunity K_{it}.

As we can only construct a measure of technological opportunity for drug-related medical research, the relevant empirical prediction of the corresponding analysis for the case of other medical research is simply a proportional relationship between the disease incidence and the research effort in the disease. As in the case of drug-related research, different assumptions about the preferences would imply that the relationship is positive but non-proportional. We allow for this possibility in the empirical analysis.

3.2   Pharmaceutical Innovation

Acemoglu and Linn (2004) present a formal analysis of induced innovation that predicts a positive relationship between the extent of pharmaceutical innovation and the potential market size for new drugs. While their focus is on the effect of market size on the flow of pharmaceutical innovation, Acemoglu and Linn (2004) also note a positive empirical relationship between the stock of drugs and potential market size.

Our measure of research opportunity in drug-related medical research is determined by the extent of past pharmaceutical innovation. We can therefore examine the relationship between the extent of pharmaceutical innovation and the potential market size indirectly by comparing the changes in the relative potential market sizes across diseases and the changes in technological opportunity across diseases. The induced innovation hypothesis predicts a positive relationship between these variables.[5]

---

[5]An advantage of our indirect approach over the direct approach of Acemoglu and Linn (2004) is that variations in the measure of technological opportunity capture differences in the relative importance of

4   Empirical Strategy

We first discuss in turn the estimation of technological opportunity, the identification of the technological opportunity effect, and the identification of the market size effects. The regression models that we use to estimate the effects are then presented in the last subsection.

4.1   Estimation of Technological Opportunity

In the model of medical research a measure of technological opportunity in drug-related medical research on the disease i in year t is given by the expression

$$K_{it} \equiv \sum_{f=f_0}^{t} \left\{ \alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)} \right] + \varepsilon_{itf} \right\}, \tag{8}$$

where the parameters α_{if} specify the baseline productivity of each cohort, the parameter β_1 governs the eventual decay in the research potential of a cohort of ingredients, the parameter β_2 governs the rate at which the full potential of a cohort of ingredients for research is revealed to researchers, and ε_{itf} denotes other factors that influence the productivity of research that uses ingredients from the cohort f. We assume that $E[\varepsilon_{itf}] = 0$ and that ε_{itf} is independent and identically distributed.

The model also predicts the relationship (3) between the probability p^*_{itf} that an ingredient from the cohort f is used in research on the disease i in year t and the parameters α_{if}, β_1 and β_2. Denoting $\alpha_{it} \equiv 1/ \left[\sum_{f=f_0}^{t} \left\{ \alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)} \right] + \varepsilon_{itf} \right\} \right]$ the relationship (3) may be rewritten as

$$p^*_{itf}/\alpha_{it} = \alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)} \right] + \varepsilon_{itf}. \tag{9}$$

---

different drugs. A possible disadvantage of our approach is that the timing of the effect that the discovery of a new ingredient has on the measure of the technological opportunity does not necessarily coincide with the timing of the stream of profits from the discovery of the new ingredient.

When $t - f_0$ is large, we have that $\sum_{f=f_0}^{t} \varepsilon_{itf} \approx 0$. Applying this simplification modifies the definition of α_{it} as follows:

$$\alpha_{it} \equiv 1/\left[\sum_{f=f_0}^{t} \alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)} \right] \right] \tag{10}$$

and also modifies the relationship (3) as follows:

$$p_{itf}^* = \frac{\alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)} \right] + \varepsilon_{itf}}{\sum_{f=f_0}^{t} \alpha_{if} \times e^{-\beta_1(t-f)} \times \left[1 - e^{-\beta_2(t-f)} \right]} \tag{11}$$

The econometric challenge is to estimate the parameters α_{if}, β_1 and β_2. We first estimate the parameters β_1 and β_2 using non-linear least squares applied to the equation (9) while assuming arbitrary fixed values for the parameters α_{if} and α_{it}.[6] To estimate the parameters α_{if} we then use the following iterative procedure:

- We start by calculating initial estimates of α_{it} by plugging in the estimates of the parameters β_1 and β_2 as well as arbitrary (starting) values of α_{if} into the expression (10).[7]

- Using the estimates of the parameters α_{it} and the estimates of the parameters β_1 and β_2, we estimate the parameters α_{if} by least squares applied to the equation (9) and holding α_{it}, β_1 and β_2 fixed.

- We recompute the estimates of α_{it} by plugging in the estimates of α_{if} and the estimates of β_1 and β_2 into the expression (10). If the new value of the estimate of α_{it} is sufficiently close to the old value, we declare convergence. If not, we iterate the previous step until convergence.

---

[6] We assume that $\alpha_{if} = 1$ and $\alpha_{it} = 1$ 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.

[7] We assume that $\alpha_{if} = 1$ for all i, f.

This iterative procedure yields estimates of the parameters α_{if}. We then generate our estimate of technological opportunity using the estimates $\hat{\alpha}_{if}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ and the formula

$$\hat{K}_{it} \equiv E[K_{it}|\hat{\alpha}_{if}, \hat{\beta}_1, \hat{\beta}_2] \approx \sum_{f=f_0}^{t} \hat{\alpha}_{if} \times e^{-\hat{\beta}_1(t-f)} \times \left[1 - e^{-\hat{\beta}_2(t-f)} \right]. \tag{12}$$

4.2 Identification of the Technological Opportunity Effect

As the variation in technological opportunity over time is likely to be correlated with the variation in the unobserved determinants of medical research over time, we employ fixed effects approaches in which the technological opportunity effect is identified by comparing the changes in technological opportunity with the changes in the research effort across diseases over time.

The model of medical research predicts that within a disease the distribution of research across cohorts of ingredients is independent of the total amount of research on the disease. With a different functional form for the overall benefit from medical research this would not hold. Consequently, changes in the level of drug-related research on a disease could be correlated with changes in the estimated technological opportunity in research on the disease even if there was no causal relationship from technological opportunity to the level of the drug-related research effort.

However, the unobserved effects that influence the level of drug-related research effort and the unobserved effects that influence the level of other medical research effort are likely to be correlated. Therefore, if there is indeed reverse causality from the level of research effort to the measure of technological opportunity, there will be a positive relationship also between the estimated technological opportunity and the level of other medical research. In contrast, if there is no reverse causality from the level of research effort to the measure of technological opportunity, there will not be a positive relationship between the estimated technological opportunity and the level of other medical research. We can therefore test for

10

the presence of the reverse causality by including the estimate of technological opportunity also into the analyses of the determinants of other medical research.

4.3 Identification of Market Size Effects

As is well recognized in the literature on induced innovation, the causal effect of the potential market size on the extent of innovation cannot be inferred from the relationship between observed innovation and the observed market size due to the endogeneity of the observed market size. Acemoglu and Linn (2004) circumvent this problem by examining the relationship between the changes in pharmaceutical innovation and the changes in the potential market size that are caused by the exogenous changes in the age demographics of the population. The key conditions to the success of this identification strategy are that the effect of aging on disease incidence varies across diseases, that the age demographics of the population have changed over time, and that the changes in the age demographics are mostly caused by changes in fertility and are therefore exogenous to the rate of pharmaceutical innovation.

We follow this general identification strategy in our analyses of induced pharmaceutical innovation but in constructing the potential market size we take into account both the effect that the change in the age demographics has had on the disease incidence over time and the effect that the obesity epidemic has had on the disease incidence over time. As the potential market size is measured by the disease incidence, we naturally use the same methodology in the analyses of induced innovation in medical research.

The effect that an obesity-induced change in the potential market size has had on the extent of innovation may be different than the effect that an aging-induced change in the potential market size has had on the extent of innovation if one type of effect has been better understood than the other or if the change in the age demographics was more expected than the obesity epidemic. We allow for this possibility by decomposing the changes in the potential market size into aging-induced changes and obesity-induced changes.

11

4.4   The Empirical Models

We use two types of fixed effects approaches. In one case we include disease fixed effects and year fixed effects. Using this strategy the identifying variation comes from the variation in the regressor (either technological opportunity or potential market size) relative to all other diseases. In the other case we include disease fixed effects and the interactions of disease class fixed effects and year fixed effects. Using this strategy the identifying variation comes from the variation in the regressor (either technological opportunity or potential market size) relative to all other diseases in the same disease class. The parameters of interest will be different in these two specifications if the elasticity of substitution of research effort is different between diseases within each disease class than it is between diseases in different disease classes.

The first empirical specification that we use to examine induced pharmaceutical innovation is

$$\ln \hat{K}_{it} = \beta_M \ln M_{it}^{TOTAL} + \text{fixed effects} + \varepsilon_{it}. \tag{13}$$

Here K_{it} is the estimated measure of technological opportunity for drug-related research on the disease i in year t.[8] The variable M_{it}^{TOTAL} is the potential market size for the disease i in year t. We measure the potential market size by the incidence of the disease i in year t:

$$M_{it}^{TOTAL} \equiv \sum_{j=1}^{5} \sum_{k=1}^{3} \mu_{i,j,k} \times s_{j,t}^{AGE} \times s_{j,k,t}^{BMI}. \tag{14}$$

In the expression (14) the parameter $\mu_{i,j,k}$ is the incidence of the disease i among people in the age group j who are in the Body-Mass-Index (BMI) group i, the parameter $s_{j,t}^{AGE}$ is the share of people in the age group j in year t, and the parameter $s_{j,k,t}^{BMI}$ is the share people in

---

[8]See the subsection 3.2 for why the measure of technological opportunity in drug-related medical research is also a measure of pharmaceutical innovation.

the age group j who are in the BMI group k in year t.[9,10] 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.[11]

In the second empirical specification for induced pharmaceutical innovation we decompose the changes in the potential market size into aging-induced changes and obesity-induced changes. This second empirical specification is

$$\ln \hat{K}_{it} = \beta_A \ln M_{it}^{AGING} + \beta_O \ln M_{it}^{OBESITY} + \text{fixed effects} + \varepsilon_{it}, \tag{15}$$

where

$$M_{it}^{AGING} \equiv \sum_{j=1}^{5} \sum_{k=1}^{3} \mu_{i,j,k} \times s_{j,t}^{AGE} \times s_{j,k,t_0}^{BMI} \tag{16}$$

is the potential market size in year t for the disease i when the body weight distribution in year t is set to be the same as the body weight distribution is in the initial year t_0 in the sample and only the age distribution varies across time, and where

$$M_{it}^{OBESITY} \equiv \sum_{j=1}^{5} \sum_{k=1}^{3} \mu_{i,j,k} \times s_{j,t_0}^{AGE} \times s_{j,k,t}^{BMI}. \tag{17}$$

is the potential market size in year t for the disease i when the age distribution in year t is set to be the same as the age distribution is in the initial year t_0 in the sample and only the body weight distribution varies across time.[12] A positive estimate of the parameter β_A is

---

[9]The parameters $\mu_{i,j,k}$, $s_{j,t}^{AGE}$ and $s_{j,k,t}^{BMI}$ are estimated from data on the disease incidence and from data on demographics prior to the estimation of the induced innovation model. The data sources are discussed in the next section. As we use disease and year fixed effects we can ignore population and population growth in estimating the potential market size.

[10]In estimating the disease incidence we allow for the parameter 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 text. As we don't measure changes in the insurance rates across time we do not examine the effect that changes in the insurance rates across time may have on potential market size and innovation.

[11]For the age group 0-18 we do not distinguish the disease incidence by body weight and therefore assume that $s_{1,1,t}^{BMI} = 1$, $s_{1,2,t}^{BMI} = 0$ and $s_{1,3,t}^{BMI} = 0$ for all t.

[12]The decomposition arises as follows. Let M_{it_0} denote the incidence of the disease i in the initial year t_0. Let R_{it}^{AGING} denote the effect of aging alone on the incidence of the disease i so that if only aging affected the incidence of the disease i the disease incidence 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 the disease i so that if only aging and obesity affected the incidence of the disease i the disease incidence would be $M_{it} = M_{it_0} R_{it}^{AGING} \tilde{R}_{it}^{OBESITY}$ in year t. Let

therefore evidence of aging-induced innovation and a positive estimate of the parameter β_O is evidence of obesity-induced innovation.

The empirical model that we use to analyze of the effect of technological opportunity and the effect of the disease incidence on the level of medical research effort is[13]

$$\ln N_{it} = \beta_K \ln \hat{K}_{it} + \beta_A \ln M_{it}^{AGING} + \beta_O \ln M_{it}^{OBESITY} + \text{fixed effects} + \varepsilon_{it}. \qquad (18)$$

The variable N_{it} is a measure of medical research effort on the disease i in year t. As was discussed in the subsection 4.2, we include the technological opportunity variable also in the analyses of other medical research because if in the analyses of other medical research the estimate of the parameter β_K is close to zero it is an indication that a positive estimate of the coefficient β_K for drug-related research is not a result of reverse causality.

5   Data

We postpone the discussion of the descriptive statistics until the beginning of the next section. To estimate the disease incidence for each age and BMI group we use the Medical Expenditure Panel Survey (MEPS) data from years 1996-2005.[14] Each subject is followed in MEPS for two years. For each subject we aggregate the observations in each year into one

---

$R_{it}^{OBESITY}$ denote the effect of obesity alone on the incidence of disease i so that if only obesity affected the incidence of the disease i the disease incidence would be $M_{it_0} R_{it}^{OBESITY}$ 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 aging and obesity effects by using the variables $\ln \left(R_{it}^{AGING} \right)$ and $\ln \left(R_{it}^{OBESITY} \right)$ as regressors. Because the empirical specifications include disease and year fixed effects we can replace these variables with the variables $\ln \left(M_{it_0} R_{it}^{AGING} \right)$ and $\ln \left(M_{it_0} R_{it}^{OBESITY} \right)$ which in the text are denoted by $\ln \left(M_{it}^{AGING} \right)$ and $\ln \left(M_{it}^{OBESITY} \right)$, respectively.

[13]Because the empirical results show that for medical research the effect of aging-induced changes in the disease incidence and the effect of obesity-induced changes in the disease incidence are so different we do not show the results for the specification in which the two effects are restricted to be the same (the specification in which the disease incidence is measured by M_{it}^{TOTAL}).

[14]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.

observation. 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. The resulting MEPS data includes 262,958 observations on 149,737 subjects.[15]

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.[16] 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.[17]

To measure medical research effort as well as technological opportunity in drug-related medical research we use the MEDLINE database on approximately 16 million biomedical publications, generally from 1950 to the present. Publications in the database are indexed by the 2007 version of the Medical Subject Headings (MESH) vocabulary. MESH is a hierarchical medical vocabulary of over 20000 different terms.

Because the MEPS data on the disease incidence is indexed by the ICD-9 classification system and the publications are indexed by the MESH vocabulary we construct a match between the ICD-9 codes and the MESH vocabulary. We limit the match effort to diseases for which the MEPS data includes at least 100 observations.[18] 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

---

[15]Except for subjects in the age group 0-18 we exclude subjects without either age or BMI information.

[16]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 2005.

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

[18]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.

(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 matched diseases and their matched MESH entry/entries as well as the unmatched diseases are listed in the Appendix 1. The match yields 127 separate matches between a disease or a group of diseases and a MESH entry/entries.[19] The 127 diseases belong to 12 disease classes. 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.[20]

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 estimate of the research effort. 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 the Appendix 1 the match for such diseases is marked with an asterisk and the year prior to which any observations are excluded.

To measure the extent of the research effort related to a disease we count the number of publications that are matched to the disease. A publication may be indexed to multiple diseases among the 127 matched diseases. We allow for this possibility by counting publications that are matched to more than one disease the same we would count the matches if

---

[19]The matched diseases account for 377,482 of the 745,355 disease mentions in the MEPS data.

[20]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.

each match was from a separate publication.

We identify active ingredients from the Federal Drug Administration (FDA) data on drug approvals during 1939-2006. As we generally cannot distinguish between active ingredients and their derivatives in the biomedical publications data, we consider the first word in the list of approved active ingredients to be the ingredient name that we use in our study. This yields a list of 1448 ingredients.

We search both the title and the abstract of every publication in the MEDLINE database for each ingredient in the list of ingredients. We set the cohort (year of discovery) of an ingredient to equal the year prior to the year in which the ingredient is first mentioned either in the abstract or in the title of a publication. In estimating technological opportunity, we measure the research effort in the disease i in year t that is related to the cohort f by the number of publications in year t which are matched to the disease i and mention an ingredient from the cohort f either in the title or in the abstract of the publication. A publication may be associated with multiple cohorts of ingredients. We allow for this possibility by counting publications which are matched to more than one cohort of ingredients the same way we would count the matches if each match was from a separate publication.

We use several strategies to identify and measure drug-related medical research. The first is to classify all publications that are matched to an 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. The second is to 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. The third is to classify all publications that have a MESH term indexed together with the "major topic" flag and the MESH qualifier term "drug therapy", "drug effects" or "pharmacology" as being drug-related research and count each such publication as one unit of research. We call these three measures as *DRUG 1*, *DRUG 2*, and *DRUG 3*, respectively.

We also use several strategies to identify and measure other medical research. The first is to classify all publications that are not matched to an ingredient as being other medical research and count each such publication as one unit of research. The second is to classify all publications that are not matched to an ingredient, that are not indexed are indexed with any of the MESH qualifier terms "drug therapy", "drug effects" or "pharmacology", and that are also not indexed with the MESH term "Chemicals and Drugs" as being other medical research and count each such publication as one unit of research. The third is to classify all research that is indexed with the MESH qualifier term "surgery" or "transplantation" as being other medical research and count each such publication as one unit of research. We call these three measures as *OTHER 1*, *OTHER 2*, and *OTHER 3*, respectively.

As the descriptive statistics discussed in the next section show there is a discontinuous jump in the share of publications with abstracts in the database from 1974 to 1975. Also, 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. When we determine the cohort of an ingredient we use the publications from years 1906-2005. In estimating the parameters that govern technological opportunity we limit the limit the range 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.[21] 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 technological opportunity we limit the range of the years t to 1970-2002.

---

[21]We multiply the initially estimated technological 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.

6 Results

We start with the descriptive statistics. All figures and tables are in the Appendix 2. Figure 1a shows the age and body weight distributions during the sample period. For both distributions the change has been gradual but the change in the body weight distribution began more recently than the change in the age distribution. Figure 1b shows the effect that the changes in the two distributions have had on the disease incidence for each disease from the beginning of the sample period (1975) to the end of the sample period (2005). For both variables there is considerable variation in the effect (from -10% to +20%). These identifying variations are also not too correlated for the effects to be separately identified in most cases.

Figure 2a depicts the count of all publications (*All Publications*) and the count of publications with an abstract (*Publications with an Abstract*) by the year of publication. The graph also shows the count of publications that are indexed to a disease (*Publications Indexed with a Disease*) and the count of publications that are indexed to a disease that is matched to an ICD-9 disease by our match (*Publications Matched*). A publication may be indexed to more than one disease and, consequently, our match may match a publication to more than one ICD-9 disease. Therefore, 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*).

Figure 2b depicts the count of matches to one of the 127 diseases for the three measures of drug-related medical research in each year and Figure 2c depicts the count of matches to one of the 127 diseases for the three measures of other medical research in each year. 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 related to a disease is expected to be inversely related to the count of publications that are related to the disease and the estimated effects are identified from the effects on the share of publications that are related to each disease.

We first estimate the parameters α_{if} and the parameters β_1 and β_2 that govern the measure of technological opportunity using the iterative procedure described in the subsection 4.1. The estimates of the parameters β_1 and β_2 are $\hat{\beta}_1 = 0.0628$ (s.e. 0.0045) and $\hat{\beta}_2 = 0.003$ (s.e. 0.0004). Figure 2d shows that the predicted probability that is calculated based on the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ as a function of the ingredient age $t - f$ tracks the mean of the observed probability closely except for when the ingredient age is 35 and over. The share of publications that use ingredients aged 35 and over is artificially inflated by the fact that the MEDLINE database consists mostly of publications published after 1950 and therefore our methodology of assigning the year of discovery of each ingredient assigns the year of discovery between 1950 and 1965 for a disproportionate number of ingredients as can be seen from Figure 2e.[22]

We first use the constructed measure of technological opportunity in drug-related medical research as a proxy for pharmaceutical innovation to examine the induced innovation hypothesis for pharmaceutical innovation. Because we expect the variance of the dependent variable to be inversely related to the count of publications on the disease the observations are weighted by the total count of matches to an ingredient cohort (measure *DRUG 1*) for the disease during the sample period.[23] The results are shown in Table 1.

Columns 1 and 2 show that when the potential market size effect is identified from changes relative to diseases within each disease class there is strong evidence for the induced innovation hypothesis both when the effects of aging and obesity are identified jointly and when the effects are identified separately.[24] Columns 3 and 4 show that when the potential market size effect is identified from changes relative to all other diseases the induced inno-

---

[22]Moreover, the precision of the estimates for ingredient ages 35 and over is influenced by the small number of observations in the data on ingredients aged 35 and over.

[23]That is, each observation is weighted by $\sum_{t=1975}^{2005} N_{it}^{DRUG\ 1}$.

[24]The interpretation of the point estimates is straightforward. For example, the point estimate in the column 1 implies that a one percent increase in potential market size increases pharmaceutical innovation by three percent. The point estimates are consistent with the findings in Acemoglu and Linn (2004) who focus on aging-induced pharmaceutical innovation.

vation hypothesis is supported when the effects are identified jointly but the coefficients are no longer statistically significant when the effects are identified separately. Figures 3a and 3b depict the empirical relationship for the specification analyzed in column 1. The positive relationship is clearly not a result of outliers. The relationship is even more robust when we only consider the 50% of the diseases with the most publications. This supports the use of weighted regressions in the analysis.

Before we examine drug-related medical research and other medical research separately, we first examine the determinants of all medical research. The count of all publications, denoted by N_{it}^{ALL}, that is used in this analysis corresponds to the measure *Publication-Disease Matches* in Figure 2a. The results are shown in Table 2.[25]

Columns 1 and 2 show that aging-induced increases in the disease incidence have increased the medical research effort in the disease. In contrast, there is no evidence of a corresponding effect for obesity-induced changes in the disease incidence. Columns 3 and 4 show at best a weak relationship between technological opportunity for drug-related research on a disease and the amount of total research on the disease. Columns 3 and 4 also show that the inclusion of technological opportunity variable renders the effect of aging-induced changes in the disease incidence statistically insignificant. However, as can be seen from Figure 3c, which depicts the fixed effects specification analyzed in column 3, with the exception of the outlier disease 299 there is a robust positive relationship between aging-induced changes in the disease incidence 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 the disease incidence and the overall research effort in the disease is again statistically significant.

Because the change in the age distribution has had such an unusual effect on the predicted

---

[25]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 $\tilde{K}_{it} = 0$ or $N_{it}^{ALL} = 0$.

disease incidence for the disease 299 (see Figure 1b) and because the dramatic increases in the number of diagnoses and research interest in the children's mental health diseases have been well recognized but without agreement over the causes of this, in the subsequent analyses we exclude the children's mental health diseases.[26]

The results for drug-related research are shown in Table 3.[27] There is robust evidence across the three measures of drug-related research and the two fixed effects specifications both for the hypothesis that technological opportunity is a determinant of the allocation of drug-related research effort across diseases and for aging-induced changes in the composition of research. The robustness of the technological opportunity estimate presented in the column 1 is illustrated in Figure 3d. In contrast, there is no evidence in any of the specifications for a positive relationship between obesity-induced changes in the disease incidence and the amount of drug-related research on the disease. If anything, the results suggest that there may be a negative relationship between obesity-induced changes in the disease incidence and the extent of drug-related research on a disease. We return to this issue after discussing the corresponding results for other research.

The results for other medical research are shown in Table 4.[28] For all specifications the estimate of the coefficient on the technological opportunity is much smaller than the corresponding estimate was for the three measures of drug-related research, and except for

---

[26]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 technological 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 the disease 299 and to a lesser extent also the two other children's mental health diseases (314 and 315) are outliers is that during the sample period there may have been a general disproportional increase in research to diseases that primarily affect the children. We plan to explore this possibility in future research.

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

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

first fixed effects specification for the most inclusive measure of other research *OTHER 1* (see column 1), which is the most likely of the three measures to include also some drug-related publications, the relationship is also statistically insignificant. As was discussed in the subsection 4.2, the finding of no relationship between the level of other research and the technological opportunity in drug-related research is evidence against the possibility that reverse causality is the reason for the observed positive relationship between the measure of technological opportunity and the extent of drug-related research.

The results for the measures *OTHER 1* and *OTHER 2* that are reported in columns 1 and 2 and in columns 3 and 4 also provide evidence of aging-induced changes in the composition of other medical research across diseases but show no evidence of obesity-induced changes in the composition of other medical research across diseases. The results for surgery-related research (the measure *OTHER 3*) in columns 5 and 6 show that the relationship between aging-induced changes in the incidence of a disease and the extent of surgery-related research on the disease is positive but not statistically significant. As was the case for drug-related-research, the results for surgery-related research suggest a possible negative relationship between obesity-induced changes in the incidence of a disease and the extent of surgery-related research on the disease. We now examine these relationships with additional analyses that are reported in Table 5.

In the analyses that are reported in columns 1 and 2 the logarithm of the ratio of the most restrictive measure of drug-related research and all research is set as the dependent variable.[29] As expected, the results indicate a positive relationship between technological opportunity in drug-related research on a disease and the share of research on the disease that is drug-related. Because aging-induced changes are expected to influence drug-related research and all research the same way, the finding that there is no statistically significant relationship between aging-induced changes in the disease incidence and changes in the share

---

[29]Each observation is weighted by $\sum_{t=1975}^{2005} N_{it}^{DRUG\ 3}$.

of research on a diseases that is drug-related research is as expected. However, the point estimate still leaves open the possibility that drug-related research reacts to changes in the disease incidence more strongly than all medical research. As in the earlier analyses of drug-related research the obesity-induced changes in the disease incidence have a negative but statistically insignificant relationship with the changes in the dependent variable. This suggests that an obesity-induced increase in the incidence of a disease may decrease the share of research on the disease that is drug-related. A likely explanation is that for the diseases for which the disease incidence is higher for the obese than it is for the normal weight research effort is substituted from general research on the disease to obesity-specific research on the disease as the obesity rate increases.

In the analyses reported in columns 3 and 4 the logarithm of the ratio of surgery-related research and all research is set as the dependent variable.[30] The results show a negative and statistically significant relationship between the share of surgery-related research on a disease and the measure of technological opportunity in drug-related research on the disease. This is both evidence against the aforementioned reverse causality explanation for the positive relationship between drug-related research and the measure of technological opportunity and evidence that an increase in technological opportunity in drug-related research shifts research effort away from other types research to drug-related research. The finding of no relationship between aging-induced changes in the disease incidence and the ratio of research that is surgery-related is as expected. The negative relationship between obesity-induced changes in the disease incidence and the share of research on the disease that is surgery-related again suggests the possibility that an obesity-induced increase in the incidence of a disease shifts resources away from general research on the disease to obesity-specific research on the disease.

---

[30]Each observation is weighted by $\sum_{t=1975}^{2005} N_{it}^{OTHER\ 3}$.

7 Conclusion

We present evidence on how non-profit innovation and for-profit innovation respond to changes in the observable determinants of the optimal allocation of inventive activities. Our results show that the composition of medical research across diseases responds to changes in technological opportunity and aging-induced changes in the disease incidence. While we find that obesity-induced changes in the disease incidence have not changed the composition of medical research across diseases, the results suggest that an obesity-induced increase in the incidence of a disease may have shifted research away from drug-related research and surgery-related research on the disease and likely toward obesity-specific research on the disease. Our results also show that the composition of pharmaceutical innovation across diseases responds to both aging- and obesity-induced changes in the relative disease incidence.

The empirical analysis was in part facilitated by our analysis of a formal model of optimal allocation of medical research effort. The analysis enabled us to identify the structural productivity parameters that govern technological opportunity. Our research has also demonstrated the research potential of the massive and information rich MEDLINE biomedical publications database for future research on the economics of innovation.

Our results on medicine and pharmaceuticals and the existing research on induced innovation and technological opportunity in for-profit innovation suggest that there does not exist a fundamental difference between for-profit and non-profit allocation mechanisms in terms of how the allocation of inventive activity responds to changes in the characteristics that determine the optimal allocation: academic medicine is not an ivory tower. Outcomes under decentralized non-profit allocation mechanisms are therefore not necessarily inferior to the outcomes under for-profit allocation mechanisms. We believe that this finding is important for its implications to economic policy and is certainly worthy of closer attention in future research.

References

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

Bhattacharya, J. and M. Packalen, 2008, "The Other Ex-Ante Moral Hazard in Health," Working Paper, Stanford University and University of Waterloo.

Caballero, R. J. and A. B. Jaffe, 1993, "How High are The Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth," in Blanchard O. J. and S. Fisher, eds., *NBER Macroeconomics Annual 1993*. Cambridge: MIT Press, pp. 15-74.

Cameron, A. C., Gelbach, J. B. and D. L. Miller, 2007, "Bootstrap-Based Improvements for Inference with Clustered Errors," NBER Technical Working Paper No. 344.

DellaVigna, S., and J. M. Pollet, 2007, "Demographics and Industry Returns," *American Economic Review*, 97, pp. 1667-702.

Finkelstein, A., 2004, "Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry," *Quarterly Journal of Economics*, 119, pp. 527-64.

George, L. and J. Waldfogel, 2003, "Who Affects Whom in Daily Newspaper Markets?" *Journal of Political Economy*, 111, pp. 765-84.

Jaffe A. B. and M. Trajtenberg, 1996, "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*, 93, pp. 12671-7.

Hicks, J. R., 1932, *Theory of Wages*. London: Macmillan.

Lakdawalla, D. and T. Philipson, 2006, "The Nonprofit Sector and Industry Performance," *Journal of Public Economics*, 90, pp. 1681-98.

Lichtenberg, F. R., 1999, "The Allocation of Publicly Funded Biomedical Research," in *Medical Care Output and Productivity: Studies in Income and Wealth*, LXIII, Berndt, E. and D. Cutler, eds., Chicago: University of Chicago Press.

Lichtenberg, F. R., 2006, "Importation and Innovation," NBER working paper No. 12539.

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

Newell, R. A., Jaffee, A. and R. Stavins, 1999, "The Induced Innovation Hypothesis and Energy-Saving Technological Change," *Quarterly Journal of Economics*, 114, pp. 907-40.

Popp, D., 2002, "Induced Innovation and Energy Prices," *American Economics Review*, 92, pp. 160-80.

Scherer, F. M., 1965, "Firm Size, Market Structure, Opportunity, and the Output of Patented Inventions," *American Economic Review*, 55, pp. 1097-125.

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

Waldfogel, J. 2003, "Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated Product Markets," *Rand Journal of Economics*, 34, pp. 557-68.

Appendix 1: ICD-9/MESH Match

1.   **INFECTIOUS AND PARASITIC DISEASES**

GROUP	ICD-9 entry/entries		MESH entry/entries
011	011	Pulmonary tuberculosis	Tuberculosis, Pulmonary [C01.252.410.040.552.846.899]
034	034	Streptococcal sore throat and scarlet fever	Scarlet Fever [C01.252.410.890.823]
052	052	Chickenpox	Chickenpox [C02.256.466.175]
053	053	Herpes zoster	Herpes Zoster [C02.256.466.423]
054	054	Herpes simplex	Herpes Simplex [C02.256.466.382] *1982-
070	070	Viral hepatitis	Hepatitis [C06.552.380]
075	075	Infectious mononucleosis	Infectious Mononucleosis [C15.604.515.516]
110	110	Dermatophytosis	Tinea [C17.800.838.208.883]
	111	Dermatomycosis, other and unspecified	Tinea Versicolor [C01.703.295.936]
112	112	Candidiasis	Candidiasis [C01.703.160]
132	132	Pediculosis and phthirus infestation	Lice Infestations [C03.858.211.300]
133	133	Acariasis	Mite Infestations [C03.858.211.394]
NO MATCH	038	Septicemia	
	074	Specific diseases due to Coxsackie virus	

2.   NEOPLASMS

GROUP	ICD-9 entry/entries	MESH entry/entries
150	150-159 Malignant neoplasm of digestive organs and peritoneum 211 Benign neoplasm of other parts of digestive system	Digestive System Neoplasms [C04.588.274] Abdominal Neoplasms [C04.588.033] Anal Gland Neoplasms [C04.588.083]
162	162 Malignant neoplasm of trachea, bronchus, and lung 163 Malignant neoplasm of pleura	Respiratory Tract Neoplasms [C04.588.894.797]
171	171 Malignant neoplasm of connective and other soft tissue 214 Lipoma 215 Other benign neoplasm of connective and other soft tissue	Soft Tissue Neoplasms [C04.588.839] *1976-
172	172 Malignant melanoma of skin 173 Other malignant neoplasm of skin 216 Benign neoplasm of skin	Skin Neoplasms [C04.588.805]
174	174 Malignant neoplasm of female breast 175 Malignant neoplasm of male breast 217 Benign neoplasm of breast	Breast Neoplasms [C04.588.180]
179	179 Malignant neoplasm of uterus, part unspecified 180 Malignant neoplasm of cervix uteri 181 Malignant neoplasm of placenta 182 Malignant neoplasm of body of uterus 183 Malignant neoplasm of ovary and other uterine adnexa 184 Malignant neoplasm of other and unspecified female genital organs 218 Uterine leiomyoma 219 Other benign neoplasm of uterus 220 Benign neoplasm of ovary 221 Benign neoplasm of other female genital organs 185 Malignant neoplasm of prostate 186 Malignant neoplasm of testis 187 Malignant neoplasm of penis and other male genital organs 222 Benign neoplasm of male genital organs 188 Malignant neoplasm of bladder 189 Malignant neoplasm of kidney and other and unspecified urinary organs 223 Benign neoplasm of kidney and other urinary organs	Genital Neoplasms, Female [C13.351.937.418] Genital Neoplasms, Male [C04.588.945.440] Urologic Neoplasms [C12.758.820]
200	200-208 Malignant neoplasm of lymphatic and hematopoietic tissue	Leukemia [C04.557.337] Lymphoma [C04.557.386]
230	230-234 Carcinoma in situ	Carcinoma in Situ [C04.557.470.200.240]
NO MATCH	(none)	

3. ENDOCRINE, NUTRITIONAL AND METABOLIC DISEASES, AND IMMUNITY DISORDERS

GROUP	ICD-9 entry/entries		MESH entry/entries
240	240	Simple and unspecified goiter	Goiter [C19.874.283]
	241	Nontoxic nodular goiter	
242	242	Thyrotoxicosis with or without goiter	Hyperthyroidism [C19.874.397]
243	243	Congenital hypothyroidism	Hypothyroidism [C19.874.482]
	244	Acquired hypothyroidism	
250	250	Diabetes mellitus	Diabetes Mellitus [C18.452.394.750]
265	265	Thiamine and niacin deficiency states	Vitamin B Deficiency [C18.654.521.500.133.699]
	266	Deficiency of B-complex components	
272	272	Disorders of lipoid metabolism	Lipid Metabolism Disorders [C18.452.584]
274	274	Gout	Gout [C05.550.114.423]
275	275	Disorders of mineral metabolism	Hemochromatosis [C18.452.565.500.480]
			Hepatolenticular Degeneration [C18.452.648.618.403]
			Hypophosphatemia, Familial [C18.452.750.400.500]
			Hypercalcemia [C18.452.174.451]
			Hypocalcemia [C18.452.174.509]
276	276	Disorders of fluid, electrolyte, and acid-base balance	Hypokalemia [C18.452.950.565]
			Hypernatremia [C18.452.950.452]
			Acidosis [C18.452.076.176]
			Alkalosis [C18.452.076.354]
279	279	Disorders involving the immune mechanism	Agammaglobulinemia [C15.378.147.142]
			DiGeorge Syndrome [C16.131.300] *1977-
			Dysgammaglobulinemia [C15.378.147.333]
			Wiskott-Aldrich Syndrome [C15.378.100.100.970]
NO MATCH	256	Ovarian dysfunction	

4. DISEASES OF BLOOD AND BLOOD-FORMING ORGANS

GROUP	ICD-9 entry/entries	MESH entry/entries
280	280 Iron deficiency anemias 281 Other deficiency anemias 282 Hereditary hemolytic anemias 283 Acquired hemolytic anemias 284 Aplastic anemia 285 Other and unspecified anemias	Anemia [C15.378.071]
288	288 Diseases of white blood cells	Agranulocytosis [C15.378.553.546.184] Granulomatous Disease, Chronic [C15.378.553.774.535] Eosinophilia [C15.378.553.231] Leukocytosis [C15.378.553.475]
NO MATCH	(none)	

5.   MENTAL DISORDERS

GROUP	ICD-9 entry/entries		MESH entry/entries
295	295	Schizophrenic psychoses	Schizophrenia [F03.700.750]
296	296	Affective psychoses	Mood Disorders [F03.600]
	309	Adjustment reaction	Adjustment Disorders [F03.075]
299	299	Psychoses with origin specific to childhood	Child Development Disorders, Pervasive [F03.550.325] *1981-
300	300	Neurotic disorders	Anxiety Disorders [F03.080] Dissociative Disorders [F03.300] *1981- Factitious Disorders [F03.400] Somatoform Disorders [F03.875]
301	301	Personality disorders	Personality Disorders [F03.675]
302	302	Sexual deviations and disorders	Sexual and Gender Disorders [F03.800]
303	303	Alcohol dependence syndrome	Substance-Related Disorders [F03.900]
	304	Drug dependence	
	305	Nondependent abuse of drugs	
314	314	Hyperkinetic syndrome of childhood	Attention Deficit Disorder with Hyperactivity [F03.550.150.150]
315	315	Specific delays in development	Developmental Disabilities [F03.550.362] Communication Disorders [F03.550.350]
NO MATCH	308	Acute reaction to stress	
	306	Physiological malfunction arising from mental factors	

6.   DISEASES OF THE NERVOUS SYSTEM AND SENSE ORGANS

GROUP	ICD-9 entry/entries	MESH entry/entries
320	320 Bacterial meningitis 321 Meningitis due to other organisms 322 Meningitis of unspecified cause 323 Encephalitis, myelitis, and encephalomyelitis	Meningitis [C10.228.228.507] Encephalitis [C10.228.228.245] Myelitis [C10.228.228.618]
332	332 Parkinson's disease	Parkinsonian Disorders [C10.228.662.600]
340	340 Multiple sclerosis	Multiple Sclerosis [C10.114.375.500]
343	343 Infantile cerebral palsy	Cerebral Palsy [C10.228.140.140.254]
345	345 Epilepsy	Epilepsy [C10.228.140.490]
346	346 Migraine	Migraine Disorders [C10.228.140.546.399.750]
350	350-359 Disorders of the peripheral nervous system	Peripheral Nervous System Diseases [C10.668.829]
361	361 Retinal detachments and defects 362 Other retinal disorders	Retinal Diseases [C11.768]
363	360 Disorders of the globe 363 Chorioretinal inflammations and scars and other disorders of choroid 364 Disorders of iris and ciliary body	Uveal Diseases [C11.941]
365	365 Glaucoma	Glaucoma [C11.525.381]
366	366 Cataract	Cataract [C11.510.245]
367	367 Disorders of refraction and accommodation	Refractive Errors [C11.744]
368	368 Visual disturbances 369 Blindness and low vision	Vision Disorders [C23.888.592.763.941]
371	371 Corneal opacity and other disorders of cornea	Corneal Diseases [C11.204]
372	372 Disorders of conjunctiva	Conjunctival Diseases [C11.187] *1981-
373	373 Inflammation of eyelids 374 Other disorders of eyelids	Eyelid Diseases [C11.338]
375	375 Disorders of lacrimal system	Lacrimal Apparatus Diseases [C11.496]
380	380 Disorders of external ear	Otitis Externa [C09.218.705.496]
381	381 Nonsuppurative otitis media and Eustachian tube disorders 382 Suppurative and unspecified otitis media 383 Mastoiditis and related conditions	Otitis Media [C09.218.705.663]
386	386 Vertiginous syndromes and other disorders of vestibular system	Labyrinth Diseases [C09.218.568]
389	389 Hearing loss	Hearing Loss [C10.597.751.418.341]
NO MATCH	(none)	

7.   DISEASES OF THE CIRCULATORY SYSTEM

GROUP	ICD-9 entry/entries	MESH entry/entries
401	401-405 Hypertensive disease	Hypertension [C14.907.489]
410	410 Acute myocardial infarction 412 Old myocardial infarction	Myocardial Infarction [C14.280.647.500]
413	413 Angina pectoris	Angina Pectoris [C14.907.553.470.250.125]
414	414 Other forms of chronic ischemic heart disease 440 Atherosclerosis 441 Aortic aneurysm and dissection 442 Other aneurysm	Arteriosclerosis [C14.907.137.126] Aneurysm [C14.907.055]
426	426 Conduction disorders 427 Cardiac dysrhythmias	Arrhythmia [C14.280.067]
428	428 Heart failure	Heart Failure, Congestive [C14.280.434]
430	430-438 Cerebrovascular disease	Cerebrovascular Disorders [C14.907.253]
444	444 Arterial embolism and thrombosis 451 Phlebitis and thrombophlebitis 452 Portal vein thrombosis 453 Other venous embolism and thrombosis	Embolism and Thrombosis [C14.907.355] Phlebitis [C14.907.681]
454	454 Varicose veins of lower extremities 456 Varicose veins of other sites	Varicose Veins [C14.907.927]
455	455 Hemorrhoids	Hemorrhoids [C14.907.449]
458	458 Hypotension	Hypotension [C14.907.514]
NO MATCH	(none)	

8. DISEASES OF THE RESPIRATORY SYSTEM

GROUP	ICD-9 entry/entries		MESH entry/entries
460	460	Acute nasopharyngitis [common cold]	Nasopharyngitis [C07.550.350.700]
	462	Acute pharyngitis	Pharyngitis [C07.550.781]
	472	Chronic pharyngitis and nasopharyngitis	
461	461	Acute sinusitis	Sinusitis [C08.460.692.752]
	473	Chronic sinusitis	
463	463	Acute tonsillitis	Tonsillitis [C08.730.817]
	474	Chronic disease of tonsils and adenoids	
464	464	Acute laryngitis and tracheitis	Laryngitis [C08.360.535]
			Tracheitis [C08.907.763]
	476	Chronic laryngitis and laryngotracheitis	Epiglottitis [C08.730.368.351]
			Croup [C08.360.535.365]
466	466	Acute bronchitis and bronchiolitis	
	490	Bronchitis, not specified as acute or chronic	Bronchitis [C08.381.495.146]
	491	Chronic bronchitis	
477	477	Allergic rhinitis	Rhinitis [C08.460.799]
480	480	Viral pneumonia	
	481	Pneumococcal pneumonia [Streptococcus pneumoniae pneumonia]	
	482	Other bacterial pneumonia	
	483	Pneumonia due to other specified organism	
	484	Pneumonia in infectious diseases classified elsewhere	Pneumonia [C08.381.677]
	485	Bronchopneumonia, organism unspecified	
	486	Pneumonia, organism unspecified	
	514	Pulmonary congestion and hypostasis	
487	487	Influenza	Influenza, Human [C08.730.310]
492	492	Emphysema	Emphysema [C23.550.325]
493	493	Asthma	Asthma [C08.127.108]
511	511	Pleurisy	Pleurisy [C08.528.735]
NO MATCH	470	Deviated nasal septum	

9.   DISEASES OF THE DIGESTIVE SYSTEM

GROUP	ICD-9 entry/entries		MESH entry/entries
520	520	Disorders of tooth development and eruption	Tooth Abnormalities [C07.650.800]
	521	Diseases of hard tissues of teeth	Tooth Erosion [C07.793.763]
	524	Dentofacial anomalies, including malocclusion	Tooth Abrasion [C07.793.707]
			Malocclusion [C07.793.494]
522	522	Diseases of pulp and periapical tissues	Periapical Diseases [C07.320.830]
	523	Gingival and periodontal diseases	Dental Pulp Diseases [C07.793.237] *1980-
			Periodontitis [C07.465.714.533]
			Gingival Diseases [C07.465.714.258]
526	526	Diseases of the jaws	Jaw Cysts [C04.182.089.530] *1980-
			Granuloma, Giant Cell [C05.500.368]
527	527	Diseases of the salivary glands	Salivary Gland Diseases [C07.465.815]
528	528	Diseases of the oral soft tissues, excluding lesions specific for gingiva and tongue	Stomatitis [C07.465.864]
			Noma [C07.465.604]
530	530	Diseases of esophagus	Esophageal Diseases [C06.405.117]
531	531	Gastric ulcer	
	532	Duodenal ulcer	
	533	Peptic ulcer, site unspecified	Peptic Ulcer [C06.405.608]
	534	Gastrojejunal ulcer	Peptic Ulcer Hemorrhage [C06.405.227.700]
	578	Gastrointestinal hemorrhage	Gastrointestinal Hemorrhage [C23.550.414.788]
535	535	Gastritis and duodenitis	Gastritis [C06.405.205.697]
	555-558	Noninfective enteritis and colitis	Duodenitis [C06.405.469.275.600] *1980-
			Enteritis [C06.405.205.462]
			Colitis [C06.405.205.265]
536	536	Disorders of function of stomach	Achlorhydria [C06.405.748.045]
			Gastric Dilatation [C06.405.748.300]
			Dyspepsia [C23.888.821.236]
540	540	Acute appendicitis	
	541	Appendicitis, unqualified	Appendicitis [C06.405.205.099]
	542	Other appendicitis	
550	550-553	Hernia of abdominal cavity	Hernia [C23.300.707]
560	560	Intestinal obstruction without mention of hernia	Intestinal Obstruction [C06.405.469.531]
562	562	Diverticula of intestine	Diverticulum, Colon [C23.300.415.124]
			Diverticulum, Stomach [C23.300.415.500]
574	574	Cholelithiasis	Cholelithiasis [C06.130.409]
577	577	Diseases of pancreas	Pancreatitis [C06.689.750]
			Pancreatic Cyst [C06.689.500]
NO MATCH	571	Chronic liver disease and cirrhosis	

10.   DISEASES OF THE GENITOURINARY SYSTEM

GROUP	ICD-9 entry/entries		MESH entry/entries
590	590	Infections of kidney	Nephritis [C12.777.419.570]
592	592	Calculus of kidney and ureter	Nephrolithiasis [C12.777.419.600] Ureterolithiasis [C12.777.725.938]
595	595	Cystitis	Cystitis [C12.777.829.495]
600	600	Hyperplasia of prostate	Prostatic Diseases [C12.294.565]
	601	Inflammatory diseases of prostate	
	602	Other disorders of prostate	
607	607	Disorders of penis	Penile Diseases [C12.294.494]
610	610	Benign mammary dysplasias	Breast Diseases [C17.800.090]
	611	Other disorders of breast	
614	614	Inflammatory disease of ovary, fallopian tube, pelvic cellular tissue, and peritoneum	Adnexal Diseases [C13.351.500.056]
	620	Noninflammatory disorders of ovary, fallopian tube, and broad ligament	
615	615	Inflammatory diseases of uterus, except cervix	Uterine Diseases [C13.351.500.852] Vaginal Diseases [C13.351.500.894] Vulvar Diseases [C13.351.500.944]
	616	Inflammatory disease of cervix, vagina, and vulva	
	618	Genital prolapse	
	621	Disorders of uterus, not elsewhere classified	
	622	Noninflammatory disorders of cervix	
	623	Noninflammatory disorders of vagina	
	624	Noninflammatory disorders of vulva and perineum	
	625	Pain and other symptoms associated with female genital organs	
617	617	Endometriosis	Endometriosis [C13.351.500.163]
628	628	Infertility, female	Infertility, Female [C13.351.500.365.700]
NO MATCH	627	Menopausal and postmenopausal disorders	

12. DISEASES OF THE SKIN AND SUBCUTANEOUS TISSUE

GROUP	ICD-9 entry/entries		MESH entry/entries
680	680	Carbuncle and furuncle	Furunculosis [C01.252.410.868.820.270]
681	681	Cellulitis and abscess of finger and toe	Cellulitis [C01.539.800.130]
	682	Other cellulitis and abscess	
684	684	Impetigo	Impetigo [C01.252.410.868.820.504]
690	690	Erythematosquamous dermatosis	Dermatitis, Seborrheic [C17.800.174.580]
	706	Diseases of sebaceous glands	Acne Vulgaris [C17.800.794.111]
691	691	Atopic dermatitis and related conditions	Dermatitis, Atopic [C17.800.174.193]
	692	Contact dermatitis and other eczema	Dermatitis, Contact [C17.800.174.255]
696	696	Psoriasis and similar disorders	Psoriasis [C17.800.859.675]
			Pityriasis [C17.800.859.600]
			Parapsoriasis [C17.800.859.575]
698	698	Pruritus and related conditions	Pruritus [C17.800.685]
			Prurigo [C17.800.674]
			Neurodermatitis [C17.800.174.660]
700	700	Corns and callosities	Callosities [C17.800.428.200]
703	703	Diseases of nail	Nail Diseases [C17.800.529] *1979-
704	704	Diseases of hair and hair follicles	Hair Diseases [C17.800.329] *1980-
705	705	Disorders of sweat glands	Sweat Gland Diseases [C17.800.946] *1976-
708	708	Urticaria	Urticaria [C17.800.862.945]
NO MATCH	707	Chronic ulcer of skin	
	695	Erythematous conditions	
	693	Dermatitis due to substances taken internally	

13. DISEASES OF THE MUSCULOSKELETAL SYSTEM AND CONNECTIVE TISSUE

GROUP	ICD-9 entry/entries		MESH entry/entries
710	710 728	Diffuse diseases of connective tissue Disorders of muscle, ligament, and fascia	Sjogren's Syndrome [C05.550.114.154.774] Scleroderma, Systemic [C17.300.799] Scleroderma, Localized [C17.300.787] Dermatomyositis [C05.651.594.297] Myositis [C05.651.594]
715	715 721	Osteoarthrosis and allied disorders Spondylosis and allied disorders	Osteoarthritis [C05.550.114.606]
722	722	Intervertebral disc disorders	Intervertebral Disk Displacement [C05.116.900.307]
726	726	Peripheral enthesopathies and allied syndromes	Bursitis [C05.550.251]
734	734	Flat foot	Flatfoot [C05.330.448]
735	735	Acquired deformities of toe	Hallux Valgus [C05.330.610] Hallux Varus [C05.330.612]
737	737	Curvature of spine	Spinal Curvatures [C05.116.900.800]
NO MATCH	717	Internal derangement of knee	

Appendix 2. Figures and Tables

Figure 1a.
Age and body weight distributions by year.
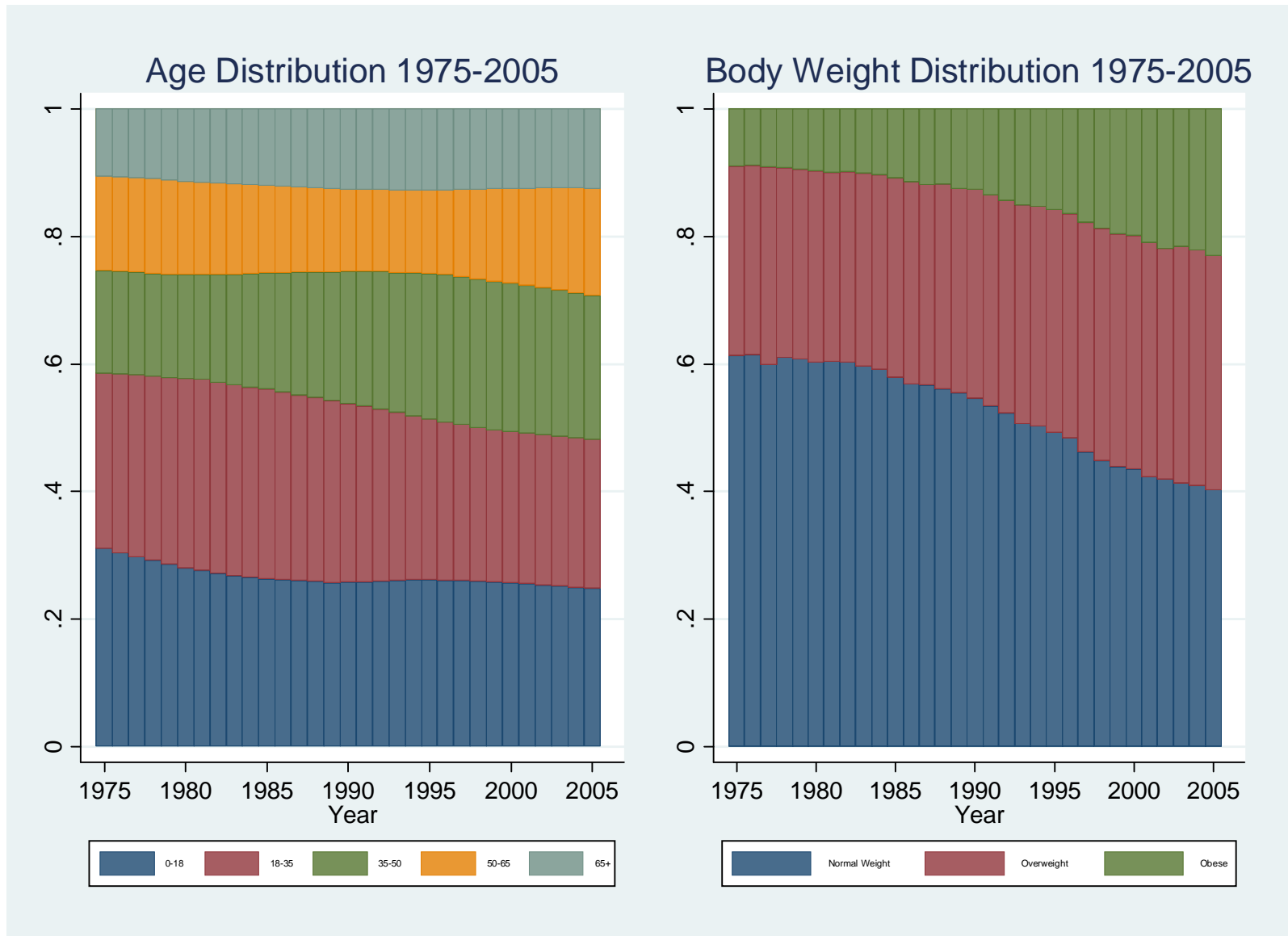
Figure 1b.
Effect of changes in age and body weight distributions on potential market size.
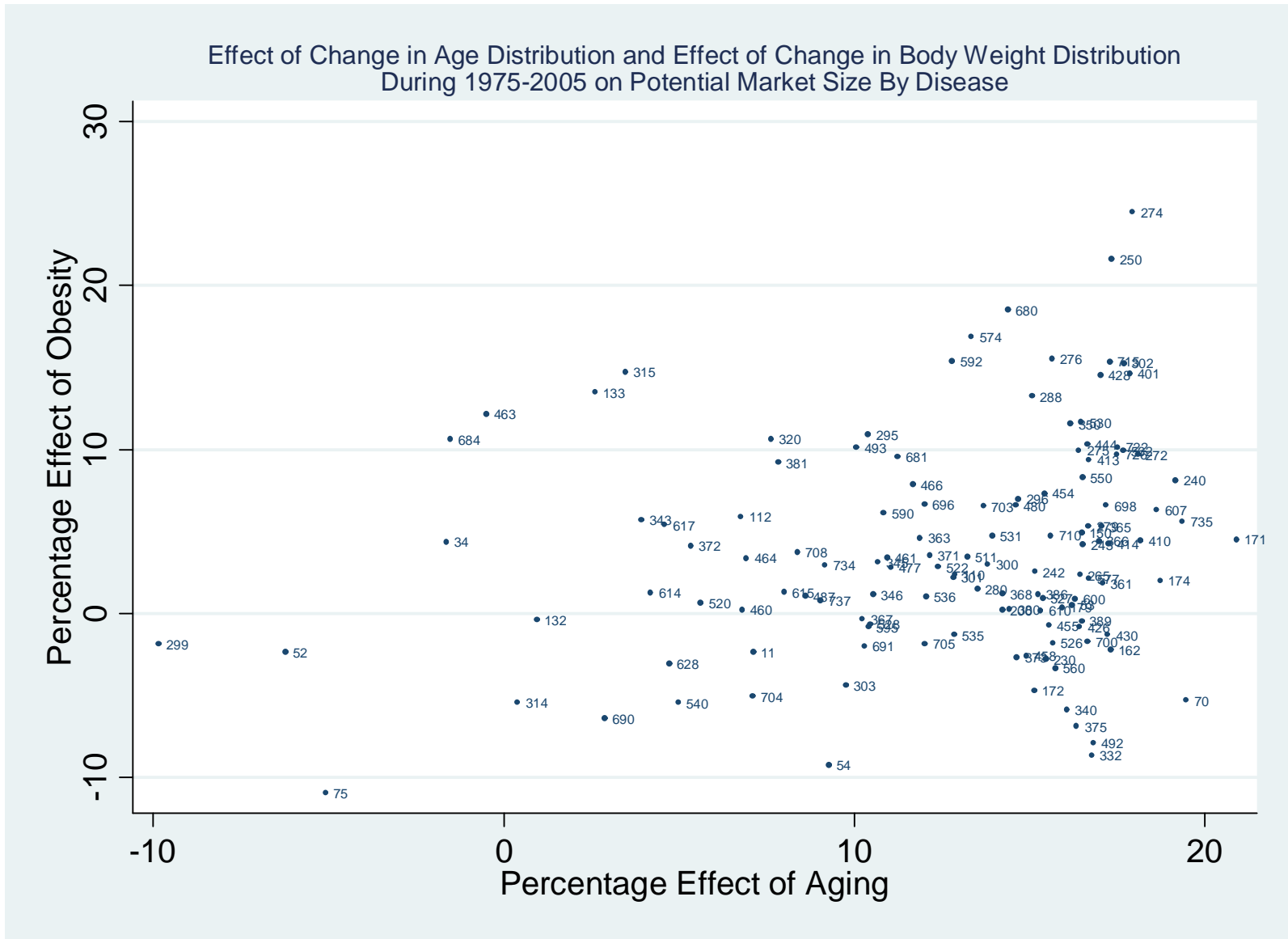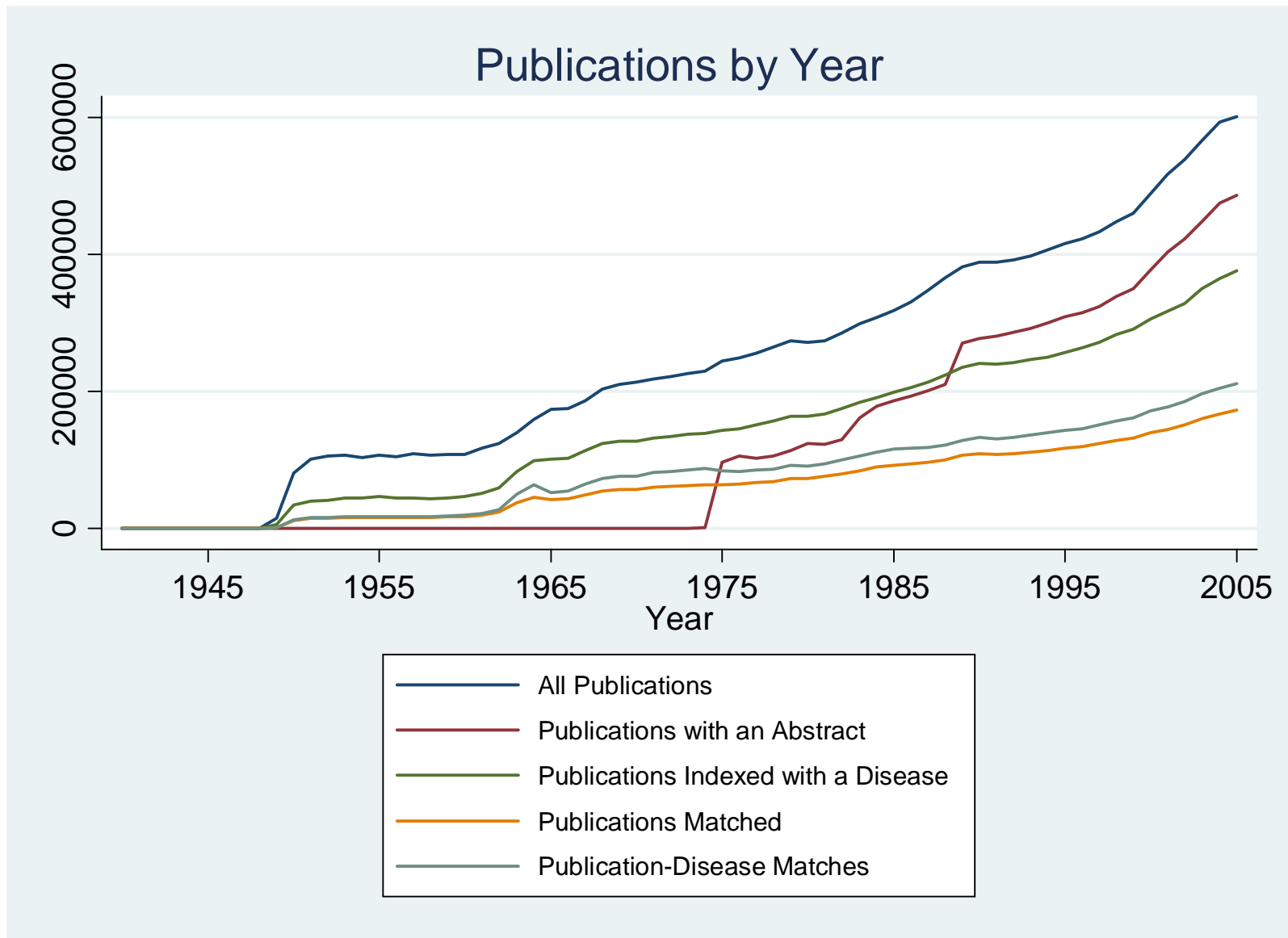
Effect of Change in Age Distribution and Effect of Change in Body Weight Distribution
During 1975-2005 on Potential Market Size By Disease

Publications by Year

Legend:
- All Publications
- Publications with an Abstract
- Publications Indexed with a Disease
- Publications Matched
- Publication-Disease Matches

3 Measures of Drug-Related Research by Year

Legend:
- (DRUG 1) Matches to an Ingredient Cohort
- (DRUG 2) Matches to at Least One Ingredient
- (DRUG 3) Drugs Indexed as a Major Topic

3 Measures of Other Medical Research by Year

(OTHER 1) No Matches to Ingredients, Drugs Not Indexed as a Major Topic
(OTHER 2) No Matches to Ingredients, Drugs Not a Major Topic, No Indexed Drugs
(OTHER 3) Surgery
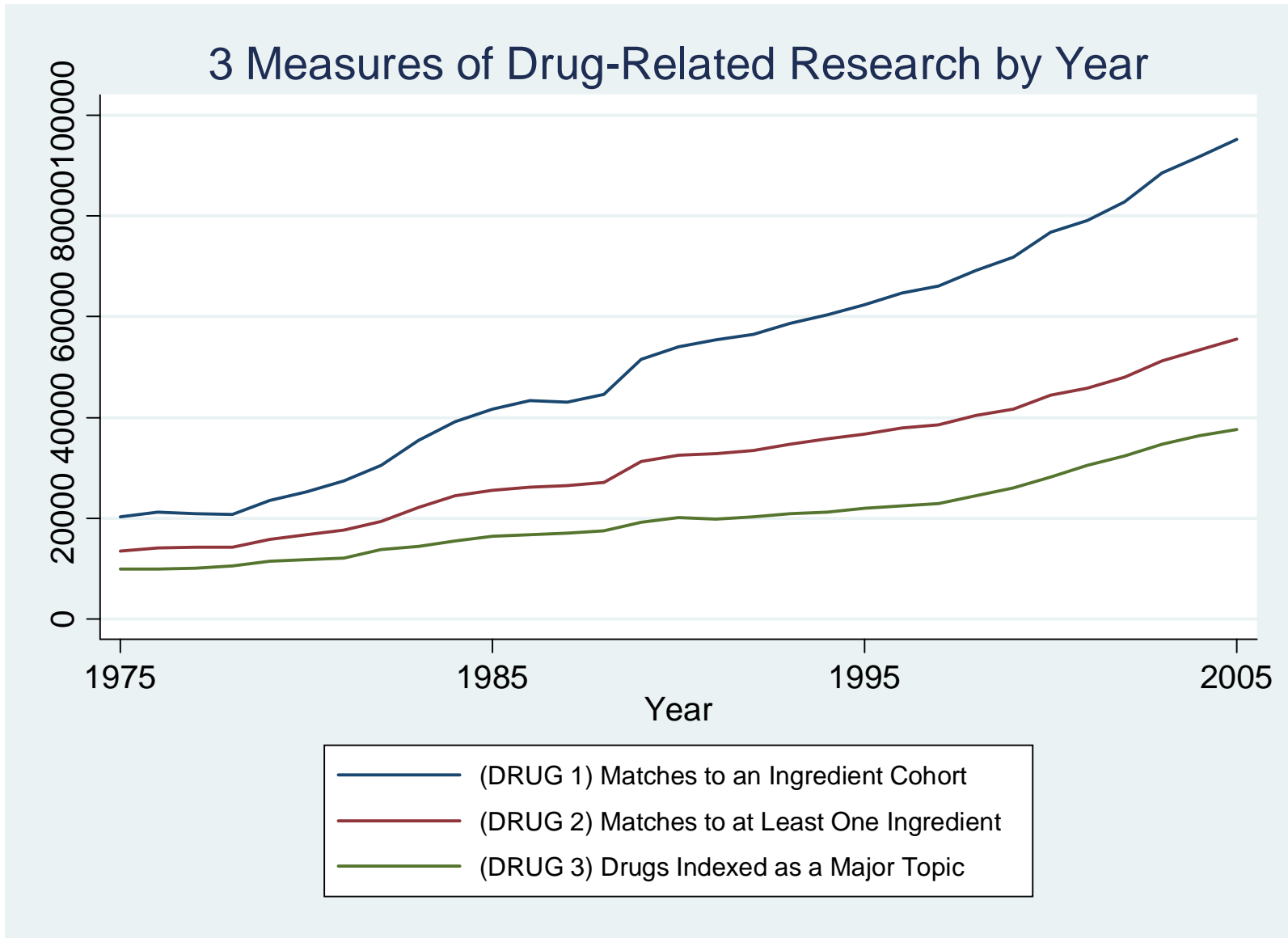
Figure 2d.
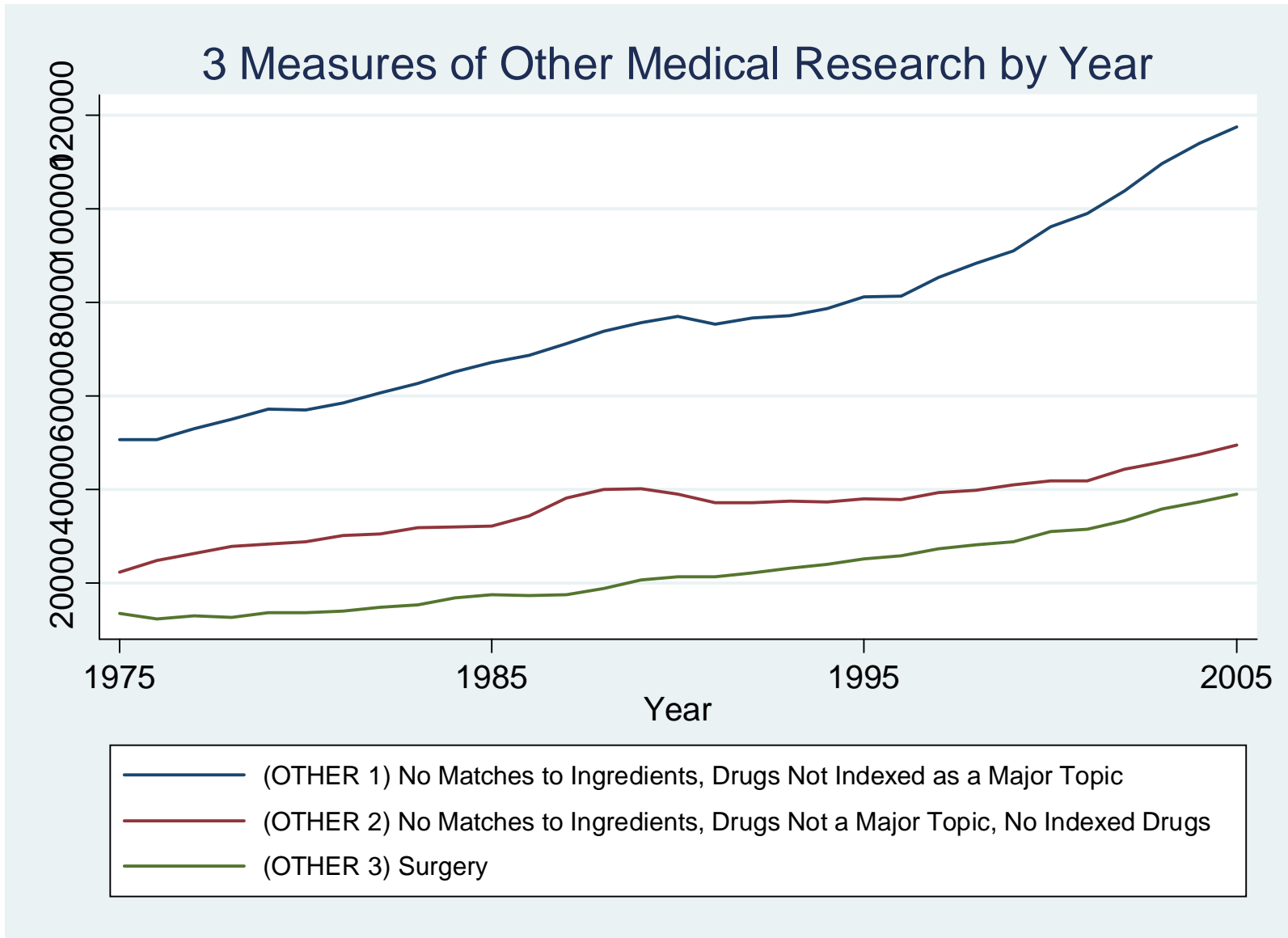**Predicted probabilities from stage 1 and the mean of observed probabilities.**
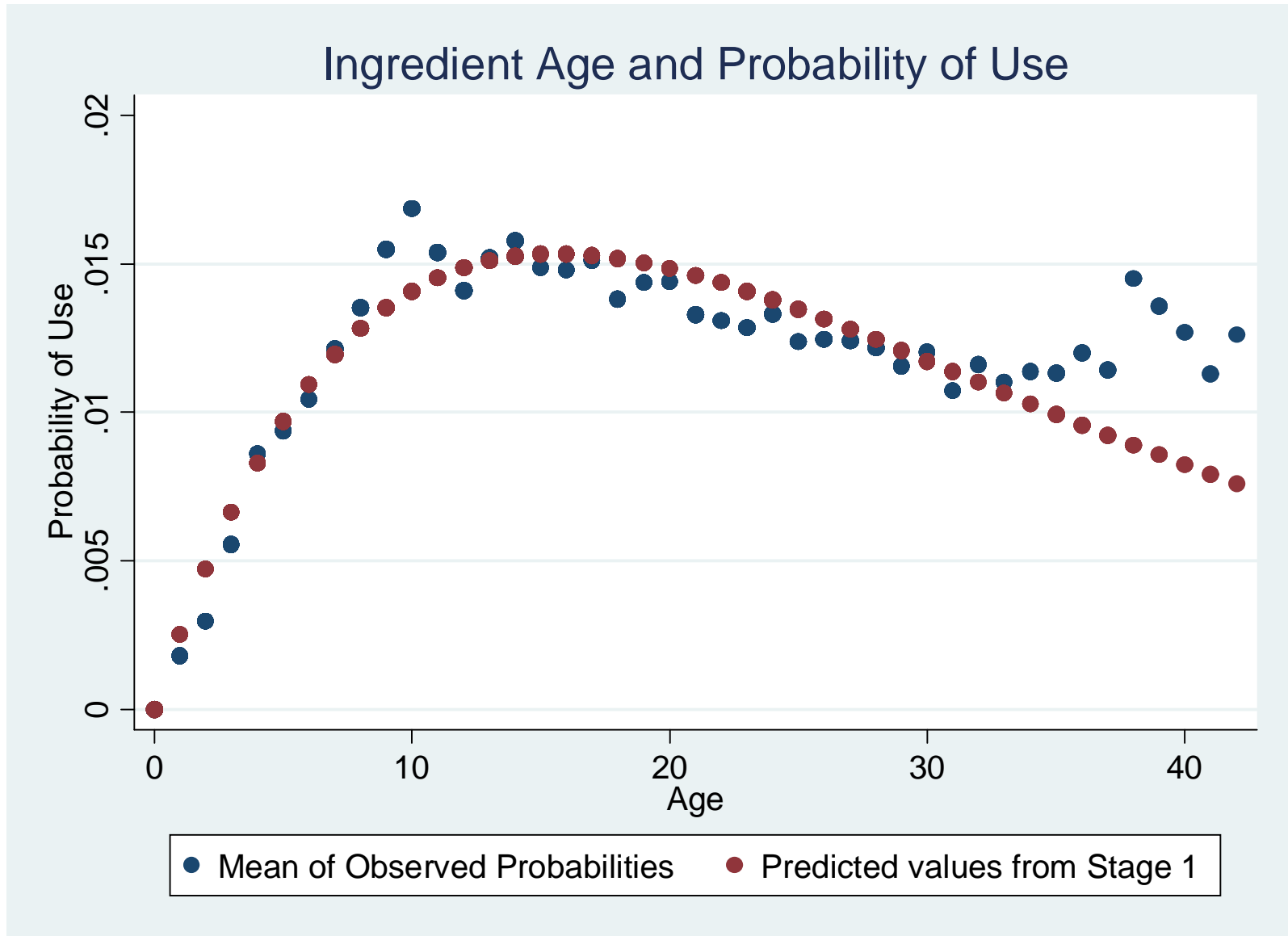
Figure 2e.
Number of ingredients by assigned year of discovery.
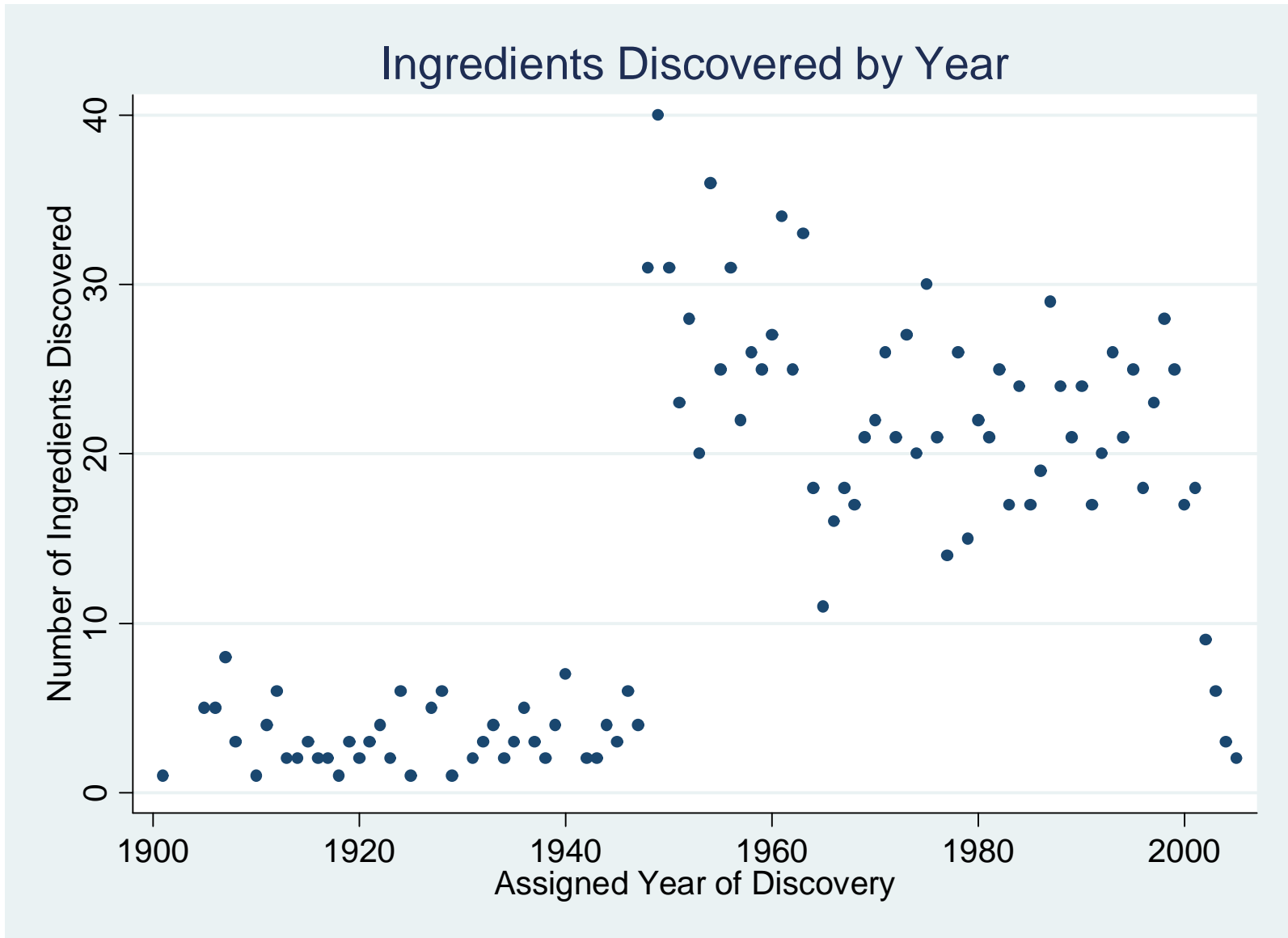
Figure 3a.
Residuals from regression of $\ln\left(\hat{K}_{it}\right)$ on Class×Year and Disease fixed effects vs.
residuals from regression of $\ln\left(M_{it}^{TOTAL}\right)$ on Class×Year and Disease fixed effects
for all **127** diseases.

Residuals from regression of $\ln\left(\hat{K}_{it}\right)$ **on Class×Year and Disease fixed effects vs.**
residuals from regression of $\ln\left(M_{it}^{TOTAL}\right)$ **on Class×Year and Disease fixed effects**
for the 63 diseases with the most observations.

Figure 3c.

Residuals from regression of $\ln\left(N_{it}^{ALL}\right)$ on $\ln\left(\hat{K}_{it}\right)$, $\ln\left(M_{it}^{OBESITY}\right)$, and Class×Year and Disease fixed effects vs. residuals from regression of $\ln\left(M_{it}^{AGING}\right)$ on $\ln\left(\hat{K}_{it}\right)$, $\ln\left(M_{it}^{OBESITY}\right)$, and Class×Year and Disease fixed effects for all 127 diseases.

Residuals for All Diseases

Figure 3d.
Residuals from regression of $\ln\left(N_{it}^{DRUG\ 1}\right)$ on $\ln\left(M_{it}^{AGING}\right)$, $\ln\left(M_{it}^{OBESITY}\right)$, and Class×Year and Disease fixed effects vs. residuals from regression of $\ln\left(\hat{K}_{it}\right)$ on $\ln\left(M_{it}^{AGING}\right)$, $\ln\left(M_{it}^{OBESITY}\right)$, and Class×Year and Disease fixed effects for the 63 Diseases with the most observations.

Table 1
Induced Pharmaceutical Innovation

	(1)	(2)	(3)	(4)
Dependent variable:	$\ln(\hat{K}_{it})$	$\ln(\hat{K}_{it})$	$\ln(\hat{K}_{it})$	$\ln(\hat{K}_{it})$
$\ln\left(M_{it}^{TOTAL}\right)$	3.03 [1.31] $p_{wild} = 0.002$		1.32 [0.70] $p_{wild} = 0.042$	
$\ln\left(M_{it}^{AGING}\right)$		3.25 [1.149] $p_{wild} = 0.008$		2.28 [1.38] $p_{wild} = 0.111$
$\ln\left(M_{it}^{OBESITY}\right)$		2.84 [1.43] $p_{wild} = 0.017$		0.92 [0.77] $p_{wild} = 0.188$
Fixed effects	Disease, Class×Year	Disease, Class×Year	Disease, Year	Disease, Year
Number of observations	3883	3883	3883	3883

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.

Table 2
Determinants of All Medical Research

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	$\ln(N_{it}^{ALL})$	$\ln(N_{it}^{ALL})$	$\ln(N_{it}^{ALL})$	$\ln(N_{it}^{ALL})$	$\ln(N_{it}^{ALL})$	$\ln(N_{it}^{ALL})$
$\ln\left(\hat{K}_{it}\right)$			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

In columns 5 and 6 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 Drug-Related Medical Research

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	$\ln(N_{it}^{DRUG\ 1})$	$\ln(N_{it}^{DRUG\ 1})$	$\ln(N_{it}^{DRUG\ 2})$	$\ln(N_{it}^{DRUG\ 2})$	$\ln(N_{it}^{DRUG\ 3})$	$\ln(N_{it}^{DRUG\ 3})$
$\ln\left(\hat{K}_{it}\right)$	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 4
Determinants of Other Medical Research

Dependent variable:	(1) $\ln(N_{it}^{OTHER\ 1})$	(2) $\ln(N_{it}^{OTHER\ 1})$	(3) $\ln(N_{it}^{OTHER\ 2})$	(4) $\ln(N_{it}^{OTHER\ 2})$	(5) $\ln(N_{it}^{OTHER\ 3})$	(6) $\ln(N_{it}^{OTHER\ 3})$
$\ln\left(\hat{K}_{it}\right)$	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 5
Determinants of Type of Medical of Research

Dependent variable:	(1) $\ln(\frac{N_{it}^{DRUG\ 3}}{N_{it}^{ALL}})$	(2) $\ln(\frac{N_{it}^{DRUG\ 3}}{N_{it}^{ALL}})$	(3) $\ln(\frac{N_{it}^{OTHER\ 3}}{N_{it}^{ALL}})$	(4) $\ln(\frac{N_{it}^{OTHER\ 3}}{N_{it}^{ALL}})$
$\ln\left(\hat{K}_{it}\right)$	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.