MARKET SIZE IN INNOVATION: THEORY AND EVIDENCE FROM THE PHARMACEUTICAL INDUSTRY*

DARON ACEMOGLU AND JOSHUA LINN

This paper investigates the effect of (potential) market size on entry of new drugs and pharmaceutical innovation. Focusing on exogenous changes driven by U. S. demographic trends, we find a large effect of potential market size on the entry of nongeneric drugs and new molecular entities. These effects are generally robust to controlling for a variety of supply-side factors and changes in the technology of pharmaceutical research.

I. INTRODUCTION

This paper constructs a simple model linking innovation rates to current and future market size, and documents the empirical relationship between market size and innovation in the pharmaceutical industry. Our empirical work, which exploits changes in the market size for different drug categories driven by U. S. demographic trends, finds economically significant and relatively robust effects of market size on innovation.

Although many historical accounts of important innovations focus on the autonomous progress of science and on major breakthroughs that take place as scientists build on each other's work, economists typically emphasize profit incentives and the size of the target market. For example, in his seminal study, *Invention and Economic Growth,* Schmookler argued that: ". . . invention is largely an economic activity which, like other economic activities, is pursued for gain" [1966, p. 206]. To emphasize the role of market size, Schmookler titled two of his chapters "The amount of invention is governed by the extent of the market."

The role of profit incentives and market size in innovation is also important both for the recent endogenous technological change models, which make profit incentives the central driving

* We thank Joshua Angrist, David Autor, Ernst Berndt, Olivier Blanchard, Alfonso Gambardella, Amy Finkelstein, Stanley Finkelstein, Jerry Hausman, Paul Joskow, Frank Lichtenberg, Lawrence Katz, Michael Kremer, David Popp, Jeffrey Wooldridge, two anonymous referees, and participants at the University of Chicago Graduate School of Business, Massachusetts Institute of Technology, the NBER Summer Institute, Princeton University, and the Wharton School of the University of Pennsylvania for useful comments. We are especially grateful to Frank Lichtenberg for generously sharing the Food and Drug Administration drug approval and the Computer Retrieval of Information on Scientific Projects data with us. This research is partially funded by the National Science Foundation and the Russell Sage Foundation.

The Quarterly Journal of Economics, August 2004

force of the pace of aggregate technological progress (e.g., Romer [1990], Grossman and Helpman [1991], and Aghion and Howitt [1992]), and for the induced innovation and directed technical change literatures, which investigate the influence of profit incentives on the types and biases of new technologies (see, e.g., Kennedy [1964], Drandakis and Phelps [1965], Samuelson [1965], Hayami and Ruttan [1970], and Acemoglu [1998, 2002, 2003]). A recent series of papers by Kremer, for example [2002], also builds on the notion that pharmaceutical research is driven by market size and argues that there is generally insufficient research to develop cures for third-world diseases such as malaria, because those who suffer from these diseases have a limited ability to pay.

In this paper we investigate the effect of market size on drug entry and pharmaceutical innovation. A major difficulty in any investigation of the impact of market size on innovation is the endogeneity of market size—better products will have larger markets. Our strategy to overcome this problem is to exploit variations in market size driven by U. S. demographic changes, which should be exogenous to other, for example scientific, determinants of innovation and entry of new drugs.[1] To estimate potential market size, we construct age profiles of users for each drug category, and then compute the implied market size from aggregate demographic and income changes given these (time-invariant) age profiles.[2] We measure entry and innovation using the Food and Drug Administration's (FDA) approval of new drugs.[3]

Our results show that there is an economically and statistically significant response of the entry of new drugs to market size. As the baby boom generation aged over the past 30 years, the markets for drugs mostly consumed by the young have declined, and those for drugs consumed by the middle-aged have increased. The data show a corresponding decrease in the rate of entry of

1. For many drugs non-U. S. markets may also be relevant. Nevertheless, the U. S. market is disproportionately important, constituting about 40 percent of the world market [IMS 2000]. Below, we report results using changes in OECD market size as well as U. S. market size.

2. Loosely speaking, "market size" corresponds to the number of users times their marginal willingness to pay. Therefore, market size can increase both because the number of users increases and because their marginal willingness to pay changes. We focus on changes driven by demographics to isolate exogenous changes in market size.

3. These data were previously used by Lichtenberg and Virahbak [2002], who obtained them under the Freedom of Information Act. We thank Frank Lichtenberg for sharing these data with us.

new drugs in categories mostly demanded by the young and an increase for drugs mostly consumed by the middle-aged. Our estimates suggest that a 1 percent increase in the size of the potential market for a drug category leads to a 6 percent increase in the total number of new drugs entering the U. S. market. Much of this response comes from the entry of generics, which are drugs that are identical or bioequivalent to an existing drug no longer under patent protection.

More important, there is a statistically significant response of the entry of nongeneric drugs, which more closely correspond to new products and "innovation": a 1 percent increase in potential market size leads to approximately a 4 percent increase in the entry of new nongeneric drugs. We also look at the relationship between market size and entry of new molecular entities. These drugs, which contain active ingredients that have not been previously marketed in the United States, provide a measure of more radical innovations (there are 442 new molecular entities compared with 2203 new nongenerics during our sample period). We find that a 1 percent increase in potential market size is associated with a 4–6 percent increase in the entry of new molecular entities. These results together show an important effect of potential market size on pharmaceutical innovation.

The effect of market size on the entry of new drugs is generally robust. We obtain similar results when we use different measures of market size, when we exploit changes in OECD market size, and when we control for a variety of supply-side factors including advances in biotechnology.

We also investigate whether it is current market size or past or future market sizes that have the largest effect on entry of new drugs. On the one hand, because changes in demographics can be anticipated in advance, drug entry may respond to future market size. On the other hand, because there is typically a 10–15 year gap between research and FDA approval (e.g., DiMasi, Hansen, and Lasagna [1991]), entry may respond to past market size. We find that all nongenerics respond to current market size, while current and five-year leads of market size have the strongest effects on new molecular entities and generics. These results suggest that pharmaceutical research responds to anticipated changes in market size with a lead of 10–20 years.

The magnitude of the effect of potential market size on drug entry is quite large. This may be partly because our key variable measures potential market size rather than actual market size

(i.e., what the market size would be if the number and incomes of individuals in a particular age group change without a change in the age profile of use and expenditure). Results using another data source suggest that a 1 percent increase in potential market size is associated with approximately a 4 percent increase in actual market size, so the estimates for nongenerics and new molecular entities are consistent with a proportional effect of actual market size on innovation as predicted by our theoretical model.[4]

There are a number of other studies related to our work. First, Schmookler [1966] documents a correlation between sales and innovation, and argues that the causality ran largely from the former to the latter. The classic study by Griliches [1957] on the spread of hybrid seed corn in U. S. agriculture also provides evidence consistent with the view that technological change and technology adoption are closely linked to profitability and market size. Pakes and Schankerman [1984] investigate this issue using a more structural approach, linking R&D intensity at the industry level to factor demands and to growth of output. In more recent research, Scott Morton [1999] and Reiffen and Ward [2004] study the decision of firms to introduce a generic drug and find a positive relationship between entry and expected revenues in the target market. None of these studies exploit a potentially exogenous source of variation in market size, however.

Second, some recent research has investigated the response of innovation to changes in energy prices. Most notably, Newell, Jaffe, and Stavins [1999] show that between 1960 and 1980, the typical air conditioner sold at Sears became significantly cheaper, but not much more energy-efficient. On the other hand, between 1980 and 1990, there was little change in costs, but air conditioners became much more energy-efficient, which, they argue, was a response to higher energy prices. In a related study, Popp [2002] finds a strong positive relationship between patents for energy-saving technologies and energy prices.

Third, there is substantial research focusing on innovation in the pharmaceutical industry. Henderson and Cockburn [1996], Cockburn and Henderson [2001], and Danzon, Nichelson, and Sousa Pereira [2003] study the determinants of success in clinical

4. It is also possible that the marginal innovations induced by an increase in market size are less productive, so a 4–6 percent increase in the number of new drugs may correspond to a smaller increase when weighted by effectiveness or other measures of productivity.

trials, focusing mainly on firm and project size. Galambos and Sturchio [1998], Cockburn, Henderson, and Stern [1999], Gambardella [2000], Malerba and Orsenigo [2002], and Ling, Berndt, and Frank [2003] discuss various aspects of the recent technological developments in the pharmaceutical industry.

Most closely related to this study are Lichtenberg and Waldfogel [2003], Finkelstein [2004], Cerda [2003], and DellaVigna and Pollet [2004]. Lichtenberg and Waldfogel [2003] document a relative decline in mortality among individuals with rare diseases following the Orphan Drug Act, and argue that this is related to the incentives created by the Act to develop drugs for these conditions. Finkelstein exploits three different policy changes affecting the profitability of developing new vaccines against six infectious diseases: the 1991 Center for Disease Control recommendation that all infants be vaccinated against hepatitis B, the 1993 decision of Medicare to cover the costs of influenza vaccinations, and the 1986 introduction of funds to insure manufacturers against product liability lawsuits for certain kinds of vaccines. She finds that increases in vaccine profitability resulting from these policy changes are associated with a significant increase in the number of clinical trials to develop new vaccines against the relevant diseases.[5] Cerda's Ph.D. dissertation from the University of Chicago is an independent study of the effect of demographics on innovation in the pharmaceutical sector. Although Cerda uses a somewhat different empirical methodology, he reaches similar conclusions to our study. Finally, DellaVigna and Pollet investigate whether the stock market responds to demographics-driven changes in the size of the market for a number of products.

The rest of the paper is organized as follows. We outline a simple model linking innovation to market size in the next section. Section III briefly explains our empirical strategy, and Section IV describes the basic data sources and the construction of the key variables. Section V presents the empirical results and a variety of robustness checks. Section VI contains some concluding remarks, and Appendix 1 gives further data details.

5. Lichtenberg [2003] also presents evidence suggesting that the types of new drugs changed toward drugs more useful for the elderly after Medicare was established.

II. Theory

This section outlines a simple model that illustrates the impact of market size on innovation. We consider a small open economy consisting of a set I of infinitely lived individuals. Time is continuous $t \in [0,\infty)$. There are two types of goods in this economy. First, there is a basic good y, which can be consumed, used for the production of other goods, or for research expenditure. Individual i has an exogenously given endowment $y_i(t)$ at time t. Second, there are J drugs, $x_1, \ldots, x_J$, each with a potentially time-varying "quality," $q_1(t), \ldots, q_J(t)$. Each individual demands only one type of drug. Hence, we partition the set I of individuals into J disjoint groups, $G_1, \ldots, G_J$ with $G_1 \cup G_2 \cup \ldots \cup G_J = I$, such that if $i \in G_j$, then individual i demands drug j. More specifically, if $i \in G_j$, then his preferences are given by

$$\int_0^\infty \exp(-rt)[c_i(t)^{1-\gamma}(q_j(t)x_{ji}(t))^\gamma]\,dt, \tag{1}$$

where r is the discount rate of the consumers and the interest rate faced by the economy, $\gamma \in (0,1)$, $c_i(t)$ is the consumption of individual i of the basic good at time t, and $x_{ji}(t)$ is the consumption of drug j. This Cobb-Douglas functional form and the assumption that each individual only consumes one type of drug are for simplicity and do not affect the main results.[6]

Normalizing the price of the basic good to 1 in all periods, and denoting the price of drug j at time t by $p_j(t)$, the demand of individual $i \in I$ for drug j is $x_{ij}(t) = \gamma y_i(t)/p_j(t)$ if $i \in G_j$, and $x_{ij}(t) = 0$ if $i \notin G_j$. Summing across individuals, total demand for drug j is

$$X_j(t) = \frac{\gamma Y_j(t)}{p_j(t)}, \tag{2}$$

where $Y_j(t) \equiv \Sigma_{i \in G_j} y_i(t)$ is the total income of the group of individuals consuming drug j.

At any point in time, there is one firm with the best-practice

6. The Cobb-Douglas assumption implies that the share of income spent on drugs is constant. This assumption can easily be relaxed by considering a utility function with an elasticity of substitution different from one, as in the factor market models with directed technical change (see, e.g., Acemoglu [1998, 2002]).

technology for producing each type of drug, and it can produce one unit of this drug with quality $q_j(t)$ using one unit of the basic good. If there is an innovation for drug line j currently with quality $q_j(t)$, this leads to the discovery of a new drug of quality $\lambda q_j(t)$, where $\lambda > 1$. For the purposes of the model, we think that any innovation is approved (for example, by the FDA) and can be sold to consumers immediately (and is under patent protection indefinitely).

There is free entry into R&D, and each firm has access to an R&D technology that generates a flow rate δ_j of innovation for every dollar spent for research on drug j. So if R&D expenditure at time t is $z_j(t)$, the flow rate of innovation (and of entry of new drugs) for drug j is

$$n_j(t) = \delta_j z_j(t). \tag{3}$$

Differences in δ_j's introduce the possibility that technological progress is scientifically more difficult in some lines than others.

A key feature of this R&D technology for our focus is that research is *directed* in the sense that firms can devote their R&D to developing particular types of drugs. The pharmaceutical industry, especially in the recent past, is a prime example of an industry where companies with fairly sophisticated R&D divisions or specialized R&D firms can undertake research for specific drug lines (e.g., Gambardella [2000] and Malerba and Orsenigo [2002]).[7]

The demand curves in (2) have an elasticity equal to one, so an unconstrained monopolist would charge an arbitrarily high price. However, the firm with the best drug in line j is competing with the next best drug in that line. An arbitrarily high price would allow the next best firm to capture the entire market. Therefore, the firm with the best drug sets a *limit* price to exclude the next best firm—i.e., to ensure that consumers prefer its product rather than the next best drug supplied at the lowest possible price (i.e., equal to marginal cost, which is 1). If a consumer buys from the best-practice firm with quality $q_j(t)$ and price $p_j(t)$ and

7. Naturally, there exist examples of research directed at a specific drug type leading to the discovery of a different product, such as the well-known example of Viagra, which resulted from research on hypertension and angina, and was partly accidentally discovered from the detection of side effects in a clinical study (see, e.g., Kling [1998]). The working paper version [Acemoglu and Linn 2003] shows that the results here generalize even when there is a large component of random R&D, whereby research directed at drug j can result in the discovery of other drugs.

chooses her optimal consumption as given by (2), her instantaneous utility at time t is $(q_j(t))^{\gamma}(1-\gamma)^{1-\gamma}\gamma^{\gamma}(p_j(t))^{-\gamma}y_i(t)$; and if she purchases from the next best firm, at quality $q_j(t)/\lambda$ and price equal to marginal cost, 1, she obtains utility $\lambda^{-\gamma}(q_j(t))^{\gamma}(1-\gamma)^{1-\gamma}\gamma^{\gamma}y_i(t)$. The limit price, which equalizes these two expressions, is

$$p_j(t) = \lambda \text{ for all } j \text{ and } t. \tag{4}$$

The profits of the firm with the best product of quality $q_j(t)$ in line j at time t are

$$\pi_j(q_j(t)) = (\lambda - 1)\gamma Y_j(t). \tag{5}$$

Here $\lambda\gamma Y_j(t)$ corresponds to the market size (total sales) for drug j. Notice that profits of drug companies are independent from quality, $q_j(t)$, which is a feature of the Cobb-Douglas utility.

Firms are forward-looking, and discount future profits at the rate r. The discounted value of profits for firms can be expressed by a standard dynamic programming recursion. $V_j(t|q_j)$, the value of a firm that owns the most advanced drug of quality q_j in line j at time t, is[8]

$$rV_j(t|q_j) - \dot{V}_j(t|q_j) = \pi_j(q_j(t)) - \delta_j z_j(t)V_j(t|q_j), \tag{6}$$

where $\pi_j(q_j(t))$ is the flow profits given by (5), and $z_j(t)$ is R&D effort at time t on this line by other firms.[9] Intuitively, the value of owning the best technology in line j, $rV_j(t|q_j)$, is equal to the flow profits, $\pi_j(q_j(t))$, plus the potential appreciation of the value, $\dot{V}_j(t|q_j)$, and takes into account that at the flow rate $n_j(t) = \delta_j z_j(t)$ there will be an innovation, causing the current firm to lose its leading position and to make zero profits thereafter.

Free entry into R&D to develop better quality drugs implies zero profits; i.e.,

$$\text{if } z_j(t) > 0, \text{ then } \delta_j V_j(t|q_j) = 1 \text{ for all } j \text{ and } t \tag{7}$$

(and if $z_j(t) = 0$, $\delta_j V_j(t|q_j) \leq 1$ and there will be no equilibrium R&D for this drug).

An equilibrium in this economy is given by sequences of prices $p_j(t)|_{j=1,\ldots,J}$ that satisfy (4), consumer demands for

8. Throughout, we assume that the relevant transversality conditions hold and discounted values are finite.

9. Because of the standard replacement effect first emphasized by Arrow [1962], the firm with the best technology does not undertake any R&D itself (see, for example, Aghion and Howitt [1992]).

drugs $x_i(t)|_{i \in I}$ that satisfy (2) and R&D levels $z_j(t)|_{j=1, \ldots, J}$ that satisfy (7) with $V_j(\cdot)$ given by (6).

An equilibrium is straightforward to characterize. Differentiating equation (7) with respect to time implies that $\dot{V}_j(t|q_j) = 0$ for all j and t, as long as $z_j(t) > 0$. Substituting this equation and (7) into (6) yields the levels of R&D effort in the unique equilibrium as

$$z_j(t) = \max \left\{ \frac{\delta_j(\lambda - 1)\gamma Y_j(t) - r}{\delta_j}; 0 \right\} \text{ for all } j \text{ and } t. \tag{8}$$

Equation (8) highlights the market size effect in innovation: the greater is $Y_j(t)$, i.e., the greater is the market size for a particular drug, the more profitable it is to be the supplier of that drug, and consequently, there will be greater research effort to acquire this position. In addition, a higher productivity of R&D as captured by δ_j also increases R&D, and a higher interest rate reduces R&D since current R&D expenditures are rewarded by future revenues.

Another important implication of this equation is that there are no transitional dynamics. At any point in time, R&D for a particular drug line is determined by the current market size—past and anticipated future market sizes do not affect current research effort. This is an implication of the linear R&D technology, which ensures that whenever there are profit opportunities, there will immediately be sufficient R&D to arbitrage them, ensuring that $\dot{V}_j(t|q_j) = 0$. The intuition for the lack of response to anticipated changes in future market size highlights an important effect in quality ladder models of technological progress: firms would like to own the best-practice product at the time the market size actually becomes larger. Investing in R&D far in advance of the increase in market size is not profitable, since another firm would improve over this innovation by the time the larger market size materializes. In fact, with the linear model here, z_j can change discontinuously, so investing even a little in advance of the actual increase in the size of the market is not profitable.

Combining equations (3) and (8) gives entry of new drugs as

$$n_j(t) = \max \{\delta_j(\lambda - 1)\gamma Y_j(t) - r; 0\}. \tag{9}$$

This equation relates innovation or entry of new products to market size (total expenditure of consumers in this line of drug).

It also encompasses the alternative view of the determinants of innovation discussed in the Introduction, that the cross-drug distribution of R&D is determined by technological research opportunities or perhaps by other motives unrelated to profits. If there are large and potentially time-varying differences in δ_j's, then these may be the primary factor determining variation in R&D across drug lines, and market size may have only a small effect. Whether or not this is so is an empirical question.

The working paper version of our paper [Acemoglu and Linn 2003] presented a number of generalizations of this framework. First and most importantly, we modified the R&D technology captured in equation (3) to allow for within-period decreasing returns, so that

$$n_j(t) = \delta_j z_j(t)\phi(z_j(t)),$$

where $\phi'(z) \leq 0$ (the model studied above is the special case with $\phi'(z) \equiv 0$). Most of the results here generalize, but the model also implies a potential response to anticipated changes in market size. In particular, let us assume that $Y_j(t) = Y_j$ for all t. Then it is straightforward to show that steady-state R&D will be given by

$$z_j^S = \max\left\{\frac{(\delta_j\phi(z_j^S)(\lambda - 1)\gamma Y_j - r)}{\delta_j\phi(z_j^S)}; 0\right\},$$

which is similar to (8). If there is an unanticipated change in Y_j, there continues to be no transitional dynamics (i.e., z_j immediately jumps to its new steady-state value). But it can be shown that if there is an *anticipated* increase in market size, there will be entry of new drugs in advance of the actual increase. Nevertheless, the same forces emphasized here imply that investing in R&D too far in advance would not be profitable because another firm is likely to innovate further before the actual increase in market size materializes. In terms of our empirical work, even if demographic changes are anticipated 20 or 30 years in advance, we may expect significant entry and innovation responses much later, perhaps 5 or 10 years in advance.

Second, we extended this model to incorporate entry of both generic and nongeneric drugs and showed that market size has a positive effect on entry of both types of drugs, and that, under plausible circumstances, the effect of market size on generic entry should be larger than on nongenerics.

III. Empirical Strategy

III.A. Empirical Specification and Estimation Issues

As $r \to 0$, equation (9) implies that $n_j(t)$ is proportional to $\delta_j m_j(t)$, where $m_j(t) \equiv \lambda\gamma Y_j(t)$ is the market size for drug line j at time t. We measure entry of new drugs (or innovation), $n_j(t)$, as new drug approvals by the FDA in broad drug categories as described below. This measure, denoted by N_{ct} for drug category c at time t, includes entry of generic drugs. Although generic drugs do not correspond to "innovation," their entry is driven by the same profit incentives as innovation. After presenting results using all drug approvals, our analysis focuses on the relationship between market size and entry of nongenerics and new molecular entities. Nongenerics include all drugs that are not identical or bioequivalent to an existing drug, while new molecular entities are drugs classified by the FDA as containing an active ingredient previously not marketed in the United States. Throughout, instead of actual market size, we use potential market size driven by demographic changes, which we denote by M_{ct}. The construction of this variable is discussed below.

Adding other potential determinants, time effects, and rearranging, equation (9) yields a Poisson model for the conditional mean of new drugs (see Wooldridge [2002]):

$$E[N_{ct}|\zeta_c,\bar{X}_c] = \exp(\alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \zeta_c + \mu_t), \tag{10}$$

where E is the expectations operator, N_{ct} is the number of new drugs in category c in time period t, M_{ct} is potential market size, X'_{ct} is a vector of controls, including a constant, ζ_c's are a full set of category fixed effects that correspond to the δ_j terms above, μ_t's are a full set of time effects capturing any common time component, and finally, $\bar{X}_c$ is the vector $((M_{c1}, \ldots, M_{cT}) \vdots X'_{c1} \vdots \ldots \vdots X'_{cT} \vdots (\mu_1, \ldots, \mu_T))$, with T denoting the number of time periods in our sample. This specification ensures that time effects have proportional impacts on entry of new drugs. Note also that this equation allows the coefficient of $\log M_{ct}$ to differ from 1, which could be the case if actual market size differs systematically from the potential market size, $\log M_{ct}$, or if preferences are not Cobb-Douglas (see Acemoglu [1998, 2002]).

The estimation of (10) would lead to biased estimates, however, since the nonlinearity in (10) makes it impossible to estimate the fixed effects, the ζ_c's, consistently. To deal with this

problem, we follow Hausman, Hall, and Griliches [1984], and transform (10) to obtain a multinomial distribution for N_{ct} of the form,

$$(11)\quad E[N_{ct}|\zeta_c,\bar{X}_c,\bar{N}_c] = \frac{\exp(\alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \mu_t)}{\Sigma_{\tau=1}^{T} \exp(\alpha \cdot \log M_{c\tau} + X'_{c\tau} \cdot \beta + \mu_\tau)} \bar{N}_c,$$

where $\bar{N}_c = \Sigma_{\tau=1}^{T} N_{c\tau}$ is the number of drugs approved in category c over the entire sample. This transformation removes the drug category dummies, and the coefficient of interest, α, can be estimated consistently. We estimate this equation using quasi-maximum likelihood (QML). Wooldridge [1999] shows that QML has good consistency properties, even when the true model is not Poisson, for example, when there is a different distribution of the error term.[10]

Below, we also estimate a linear model of the form,

$$(12)\quad \log \tilde{N}_{ct} = \alpha \cdot \log M_{ct} + d_{ct} + X'_{ct} \cdot \beta + \zeta_c + \mu_t + \varepsilon_{ct},$$

where the left-hand-side variable is defined as $\tilde{N}_{ct} = N_{ct}$ if $N_{ct} \geq 1$ and $\tilde{N}_{ct} = 1$ if $N_{ct} = 0$, and d_{ct} is a dummy that equals 1 when $N_{ct} = 0$. This procedure, first used by Pakes and Griliches [1980], is simple and flexible, but the estimates are biased, since d_{ct} is endogenous.

In addition, we estimate equations with lags and leads of log M_{ct} to determine whether there are significant delays and anticipation effects. Delayed effects are possible, since, as reported by DiMasi, Hansen, and Lasagna [1991], drug approval may be as much as fifteen years after the time of initial research. Anticipation effects are possible, since changes in demographics can be anticipated a long time in advance (see the discussion in Section II).[11]

10. Define the vector $\eta \equiv (\alpha \vdots \beta \vdots \mu)$. Then the QML estimate $\hat{\eta}$ maximizes the log likelihood function $\Sigma_{c=1}^{C} L_c(\hat{\eta})$, where C is the number of categories, $L_c(\hat{\eta}) \equiv \Sigma_{t=1}^{T} N_{ct} \log p_t(\hat{\eta})$, and $p_t(\hat{\eta}) \equiv \exp(\alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \mu_t)/\{\Sigma_{\tau=1}^{T} \exp(\alpha \cdot \log M_{c\tau} + X'_{c\tau} \cdot \beta + \mu_\tau)\}$.

The (Huber-White) robust asymptotic variance-covariance matrix is calculated as $\hat{A}^{-1}\hat{B}\hat{A}^{-1}/C$, where $\hat{A} \equiv C^{-1} \Sigma_{c=1}^{C} \bar{N}_c \nabla_\eta p'_t(\hat{\eta}) D_c(\hat{\eta}) \nabla_\eta p_t(\hat{\eta})$, $\hat{B} \equiv C^{-1} \Sigma_{c=1}^{C} \nabla_\eta p'_t(\hat{\eta}) D_c(\hat{\eta}) \hat{u}_c \hat{u}'_c D_c(\hat{\eta}) \nabla_\eta p_t(\hat{\eta})$, $p(\hat{\eta}) \equiv [p_1(\hat{\eta}), \ldots, p_T(\hat{\eta})]'$, and $D_c(\hat{\eta}) \equiv [\text{diag}\{p_1(\hat{\eta}), \ldots, p_T(\hat{\eta})\}]^{-1}$. Here ∇_η denotes the gradient with respect to η and $\hat{u}_c$ is the vector of residuals calculated as $\hat{u}_c = N_c - p(\hat{\eta})\bar{N}_c$, with $N_c \equiv [N_{c1}, \ldots, N_{cT}]$. See Wooldridge [1999] for more details.

11. An additional issue is that the FDA approval process may be faster for more profitable drugs, and thus potentially for drugs with greater market size (see Dranove and Meltzer [1994]). Our data do not enable us to investigate this issue.

III.B. Potential Market Size and Identification

Throughout, we exploit the potentially exogenous component of market size driven by demographic trends, combined with differences in the age profiles of expenditure and use for different types of drugs. We obtain the age profiles from micro drug consumption data, and the changes in U. S. demographics from the Current Population Survey (CPS) data. Our (income-based) measure of potential market size is

$$M_{ct} = \sum_{a} u_{ca} \cdot i_{at}, \tag{13}$$

where i_{at} is the income of individuals in age group a at time t in the United States, and u_{ca} gives the age profile for drug category c. We compute u_{ca} as the average expenditure share of drugs in category c in the total income of those in age group a. This income-based measure corresponds closely to the market size in the theoretical model, which is a combination of the number of consumers and their incomes. We also check the robustness of our results with an alternative population-based measure, calculated using the U. S. population for age group a at time t for i_{at}, and the average number of drugs in category c used per person in age group a for u_{ca}. It is important that the over-time source of variation in both measures is not from changes in individual use, but purely from demographic changes captured by i_{at}; i.e., u_{ca}'s are not time-varying.[12] Consequently, changes in prices and drug quality, which may result from innovations and affect consumption patterns, will not cause over-time variation in M_{ct}. Our baseline measure uses five-year age groups and time periods corresponding to five-year intervals. We also check the robustness of our results using single year age groups and ten-year intervals.

The major threat to the validity of our empirical strategy is from potentially time-varying omitted variables (the drug category fixed effects take out any variable that is not time-varying). Omitted variables related to market size or profit opportunities

12. Because of data availability, we cannot use estimates of u_{ca} that predate our sample period. Therefore, our estimates of age profiles may have been affected by the availability of new drugs during the sample. This should not create a spurious relationship between potential market size, M_{ct}, and entry of new drugs, since all variation in M_{ct} is driven by aggregate demographic changes, and all of our regressions control for drug category fixed effects. In any case, the numbers in Table I suggest that age profiles do not change much over time (in fact, if preferences are Cobb-Douglas as in (1) and stable, the expenditure measure of u_{ca} should be constant).

may induce a bias in the implied magnitudes, but will not lead to spurious positive estimates of the effect of market size (in other words, the presence of such variables is essentially equivalent to mismeasurement of the appropriate market size). More threatening to our identification strategy would be omitted supply-side variables that are potentially correlated with our market size measure. To show that this is not the source of our findings, we check for residual serial correlation and control for potential supply-side determinants of innovation and entry.[13]

IV. Data and Descriptive Statistics

The demographic data come from the March CPS, 1965–2000. We compute i_{at} in equation (13) for five-year age groups, ranging from 0–4 to 90+. Individual income is constructed by dividing family income equally among the members of the family. For the purposes of the diagrammatic presentation, we aggregate the age groups into three broad categories, 0–30, 30–60, and 60+, corresponding to young, middle-aged, and elderly users. Income and population movements of the five-year age groups within each of these broad groups are relatively similar.

Figure I shows population shares, and Figure II shows the corresponding income shares (i.e., income of the corresponding age group divided by total income in that period) for the three broad age groups. To facilitate comparison with Figure III, Figure II starts in 1970. Both figures show a large amount of variation across age groups over time. In particular, it is possible to trace the baby boomers, as the fraction of individuals in the age bracket 0–30 in the 1970s, and those in the age bracket 30–60 in the 1980s and the 1990s.

The FDA classifies all prescription drugs into 20 major drug categories, which are further subdivided into 159 categories. These categories are based on a combination of therapeutic intent and chemical structure. We drop four of the twenty major categories from this classification: Anesthetics, Antidotes, Radiophar-

13. Another source of endogeneity may be that innovations in certain drug categories extend the lives of the elderly, thus increasing their M_{ct}. Lichtenberg [2002, 2003] provides evidence that new drugs extend lives. This source of endogeneity is not likely to be quantitatively important, however, since the variation resulting from extended lives in response to new drugs is a small fraction of the total variation in M_{ct}. Nevertheless, we also report estimates that instrument M_{ct} with past demographics, purging it from changes in longevity.

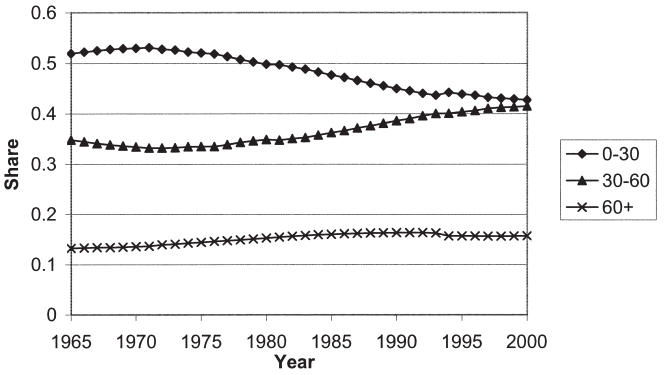

FIGURE I
Share of Population by Age Group from CPS, 1965–2000
Share of population is the population of the corresponding age group divided by total population, computed from the March CPS.

maceuticals, and Miscellaneous.[14] We then subdivide some of these categories according to the conditions and diseases that the drugs are used to treat.[15] For example, within the Hematologics major category, we separate Anemia drugs from Anticoagulants because they treat different diseases. We also subdivide broader groups when the age distribution of expenditure is sufficiently heterogeneous. For example, the indications of drugs in Estrogens/Progestins and Contraceptives overlap somewhat, but the age structure of users is quite different: 20–30 year-olds use Contraceptives most, while 50–60 year-olds use Estrogens/Progestins most. In one case, we combine categories from different major classes, Antifungals and Dermatologics, because the drugs have similar indications and age distributions. The result is a

14. We drop the Anesthetics, Radiopharmaceuticals, and Miscellaneous categories because most of the items in these categories were not developed for a distinct market. Radiopharmaceuticals are used for diagnostic purposes, and the Miscellaneous category mainly contains surgical and dental tools. The Antidote category is dropped because there are few drugs approved and there is little use of these drugs in the surveys. See Appendix 1 for further details on the construction of our categories.

15. Other authors, for example Lichtenberg [2003], have used a more detailed classification system based on diseases. We were unable to construct a comprehensive mapping of the prescription drugs listed in the micro data surveys to the detailed disease classes. Our classification system relies on the FDA categories, but then subdivides those according to disease and age distribution.

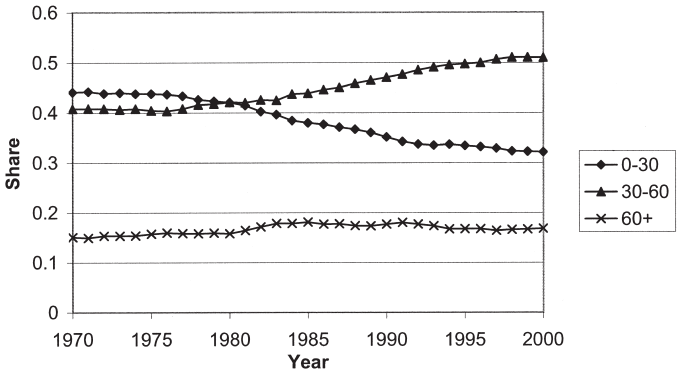

FIGURE II
Share of Income by Age Group from CPS, 1970–2000
Share of income is income of the corresponding age group divided by total income, computed from the March CPS. Individual income is obtained by dividing total family income equally among family members.

classification system with 33 categories, which are listed in Appendix 2.[16]

Our main data source for drug use is the Medical Expenditure Panel Survey (MEPS), which is a sample of U. S. households over the years 1996–1998. The survey has age and income data for each household member, and covers about 28,000 individuals each year. There is also a list of prescription drugs used by each person (if any), and the amount spent on drugs, which includes copayments and payments by insurance companies and government programs (e.g., Medicaid and worker's compensation).[17] In all, there are about 500,000 medications prescribed. We compute drug expenditure and use by five-year age groups, then divide these by the corresponding income and population numbers to construct the income-based and the population-based measures of

16. The working paper version [Acemoglu and Linn 2003] used a system with 34 categories constructed purely based on differences in the age profiles of expenditure within the major FDA categories. Results using this alternative classification are reported in Table III. Further details on the construction of the 33 categories used here and on our alternative classification system are available upon request.

17. Respondents list the pharmacy or medical provider where they obtained the prescription drug, which are then contacted to validate this information and to gather additional information on prescription drug payments.

u_{ca}.[18] Appendix 2 reports these numbers aggregated to the three broad age groups used in the figures. It shows a large amount of variation in the age profiles of expenditure across the 33 drug categories. The elderly spend more on many categories than do younger individuals, but there are numerous exceptions. For example, Antibiotics are used most by individuals in the youngest group, while Contraceptives and Antivirals are used most by 30–60 year-olds.

To investigate the stability of the age profile of users, we supplement the MEPS data with the National Ambulatory Medical Care Survey (NAMCS), which is an annual survey of doctors working in private practice and includes drug use data for the years 1980, 1981, 1985, and 1989–2000. Observations are at the doctor-patient-visit level; there are about 40,000 visits per year. Doctors are selected randomly, surveyed for a week, and patient-visits are then selected randomly from all the visits that week (further details on this survey are given in Acemoglu and Linn [2003]). We use the same classification system with the NAMCS as with the MEPS. Because the NAMCS does not contain expenditure information and its sampling scheme makes it less representative and less reliable than the MEPS, we focus on the MEPS for our main analysis and use the NAMCS mainly to check the stability of the age profiles of users.

Table I gives correlations between various measures of drug use. The first two rows of Panel A show a high degree of correlation between age profiles of use from the NAMCS surveys at various dates, both unweighted or weighted by total use of each category in the survey. These results indicate that the age profiles are similar between the 1980s and the 1990s.[19] The third and fourth rows report mean correlations by drug. These are constructed by computing the within category correlation between the measures and then averaging it across all categories. These correlations also show a substantial degree of persistence over time, especially when we look at the weighted correlation in row 4. The difference between the weighted and the unweighted cor-

18. Because income data from the CPS are more reliable, we use income estimates from the CPS to construct expenditure shares. Using the income estimates from the MEPS leads to very similar results (see the Appendix).

19. Nevertheless, as we will see below, there is evidence of an increase in use per person in categories that have also experienced an increase in market size due to demographic changes.

TABLE I
CORRELATIONS BETWEEN DIFFERENT DRUG USE MEASURES

	Panel A: NAMCS use per person		
	1980/1990	1990/2000	1980/2000
Correlation	0.897	0.861	0.861
Weighted correlation	0.906	0.843	0.856
Mean correlation by drug	0.709	0.651	0.626
Weighted mean correlation by drug	0.820	0.825	0.790
	Panel B: MEPS expenditure share		
	1996/1997	1997/1998	1996/1998
Correlation	0.961	0.965	0.929
Weighted correlation	0.962	0.973	0.937
Mean correlation by drug	0.698	0.686	0.575
Weighted mean correlation by drug	0.865	0.881	0.796

	Panel C: NAMCS/MEPS use and MEPS use/expenditure	
	NAMCS/MEPS use	MEPS use/expenditure
Correlation	0.869	0.954
Weighted correlation	0.891	0.956
Mean correlation by drug	0.804	0.902
Weighted mean correlation by drug	0.935	0.940

The numbers refer to the correlation of use per person or average expenditure share between the indicated dates and data sets. Observations are for five-year age groups by drug category (there are $33 \times 19 = 627$ observations in each case). In weighted correlations, observations are weighted by total use or expenditure from the MEPS or NAMCS. Mean correlation by drug computes correlations separately by drug category, and then calculates the average.

relations reflects the relatively imprecise estimates of use per person for the smaller categories.

Panel B performs the same calculation for expenditure shares from the three waves of the MEPS (weighted correlations now use total expenditure in each category as weights), and similarly shows substantial persistence in the age profiles of expenditure. Notably, there is now an even larger difference between weighted and unweighted mean correlations by drug, presumably because the MEPS, which is a more representative sample of the U. S. population than the NAMCS, has only a few observations in some of the smaller drug categories. This motivates our focus below on weighted regressions. Finally, Panel C shows high correlations both between

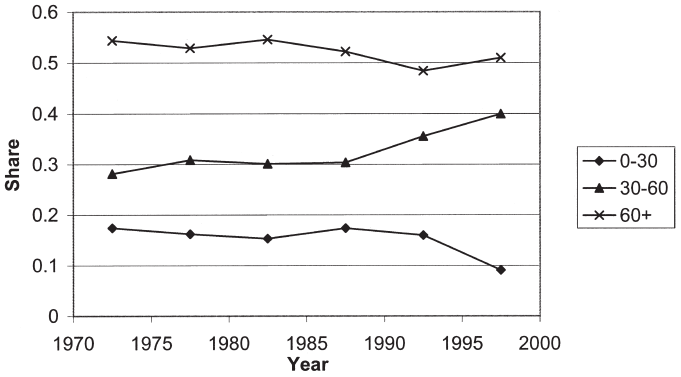

FIGURE III
Share of FDA Approvals by Age Group, 1970–2000

Share of FDA approvals is computed as approvals of drugs in the corresponding broad age group divided by total approvals in that period, calculated from the FDA data set of New Drug Approvals. Each of the 33 drug categories is assigned to one of the three broad age groups according to which broad age group has the largest expenditure (see Appendix 2).

the NAMCS and the MEPS measures and between expenditure shares and use per person in the MEPS.

The last major data source is a list of FDA new drug approvals. We exclude over-the-counter drugs, the so-called orphan drugs,[20] and drugs that have the same identifying characteristics (i.e., same name, company, and category, or the same FDA approval number). We focus on the time period 1970–2000. Both the quality of the approvals data and the quality of our measures of potential market size deteriorate as we go back in time for a number of reasons. First, we can only match FDA categories for drugs that are still listed by the FDA; second, before 1970 we cannot separately identify generics and nongenerics; and finally, we are using age profiles from the 1990s. Our approvals data set for 1970–2000 comprises 5374 prescription drugs, including both generics and nongenerics (see the Appendix). Since 1970 there have been 2203 nongeneric approvals and 442 new molecular entities.

Figure III shows the share of drug approvals over time to compare with changes in income shares depicted in Figure II. To

20. These drugs treat rare conditions, affecting fewer than 200,000 people. An example is botox, first developed to treat adult dystonia, which causes involuntary muscle contractions. We drop these drugs because we have difficulty matching them consistently, and because they receive special inducements under the Orphan Drug Act.

construct Figure III, we allocate each of the 33 categories to the broad age group that has the largest expenditure in that category. The share of drug approvals is equal to the number of approvals in a given category in each five-year period divided by total approvals in that period.[21] Although this cut of the data uses only a small part of the information that the regression analysis below exploits, a positive association between changes in income shares and changes in drug approvals can be detected by comparing this figure with Figure II. For example, the income share of the 30–60 group increases over the sample, as does the entry of drugs most used by this group. The shares of income and entry of drugs for those 0–30, on the other hand, show a downward trend. Finally, both the shares of income and entry of drugs for the 60+ group are relatively constant over the sample period. We explore these patterns in greater detail in the regression analysis below.

V. Results

V.A. Basic Specifications

Table II provides the basic results from the estimation of equation (11) with quasi-maximum likelihood (QML). The top panel is for all approvals. Panels B and C look at nongenerics and new molecular entities (nongenerics containing new molecules), and Panel D reports results for generics. Throughout the paper the standard errors are corrected for heteroskedasticity using the Huber-White formula (see footnote 10). In this table we use the basic (income-based) measure of $\log M_{ct}$, constructed using expenditure data from the MEPS, and income from the CPS, the time periods correspond to five-year intervals, and observations are weighted by total expenditure in the corresponding drug category in the MEPS.

Column (1) of Panel A shows that the QML estimate of α for all new drugs is 6.15 with a standard error of 1.23, which is significant at the 1 percent level.

The remaining columns of Panel A investigate whether it is current market size or past or future market size that has the

21. There are large fluctuations in the total number of approvals, partly because of a number of institutional changes. For example, it was discovered in 1989 that some FDA officials were taking bribes to speed up the approval process for generic drugs. As a result, in the early 1990s the approval process for generics was greatly slowed. See, for example, *The Washington Post,* August 16, 1989. In fact, there is a large drop in generics approvals in the early 1990s, but only a small decline for nongenerics. We thank Ernst Berndt for suggestions on this issue.

TABLE II
EFFECT OF CHANGES IN MARKET SIZE ON NEW DRUG APPROVALS

	(1)	(2)	(3)	(4)
Panel A: QML for Poisson model, dep var is count of drug approvals				
Market size	6.15	6.84	−2.22	
	(1.23)	(4.87)	(4.12)	
Lag market size		−0.61		
		(3.85)		
Lead market size			10.16	7.57
			(4.28)	(1.99)
Panel B: QML for Poisson model, dep var is count of nongeneric drug approvals				
Market size	3.82	6.72	2.91	
	(1.15)	(7.63)	(5.31)	
Lag market size		−2.49		
		(5.97)		
Lead market size			−1.77	1.73
			(6.94)	(2.02)
Panel C: QML for Poisson model, dep var is count of new molecular entities				
Market size	3.54	5.79	−1.38	
	(1.19)	(6.66)	(5.16)	
Lag market size		−1.99		
		(5.28)		
Lead market size			7.35	5.75
			(5.11)	(2.37)
Panel D: QML for Poisson model, dep var is count of generic drug approvals				
Market size	11.81	8.55	1.28	
	(3.30)	(6.85)	(7.17)	
Lag market size		3.12		
		(5.94)		
Lead market size			13.24	14.65
			(8.66)	(3.71)
Number of observations	198	198	165	165

Huber-White robust standard errors are reported in parentheses. The dependent variable in Panel A is count of drug approvals, in Panel B the dependent variable is count of nongeneric drug approvals, in Panel C the dependent variable is new molecular entities, and in Panel D, it is generic drug approvals, all calculated from the FDA data set of New Drug Approvals (see Appendix 1). Market size is log potential market size calculated from the MEPS and the CPS, using five-year age groups (see text). Lag market size refers to one-period lag of market size, and Lead market size refers to one-period lead of market size. All regressions include drug and time dummies, and use the income-based measure of market size. Time intervals are five years. Estimates are weighted by total expenditure for the category in the MEPS. The Poisson model is estimated by quasi-maximum likelihood (QML), with the Hausman, Hall, and Griliches [1984] transformation. See equation (11) in the text.

strongest effect on entry of new drugs. Column (2) includes current and five-year lagged market size together;[22] column (3) includes current and five-year lead market size; and finally, column (4) looks at the relationship between lead market size and entry of new drugs.

The entry of all drugs appears to respond to current or five-year lead market size. When current and lagged market sizes are included together, the coefficient on current market size has a similar magnitude to column (1), while lagged market size is negative, and neither coefficient is significant, presumably because current and previous market sizes are highly correlated. When current and lead market sizes are included together, current market size is not significant, whereas lead market size is significant at 5 percent. Moreover, column (4) shows that lead market size has a somewhat larger effect than current market size (the estimate of α is now 7.57, with standard error 1.99).

The results in Panel A combine generics and nongenerics. Entry of generics and nongenerics may be driven by different processes. Moreover, generics, which are identical to existing drugs, do not correspond to "innovation." Panel B shows the relationship between potential market size and entry of new nongeneric drugs. The estimate of α is now 3.82, with standard error 1.15, which is also significant at 1 percent.

Perhaps more relevant for the relationship between market size and innovation is the response of new molecular entities. These are drugs classified by the FDA as containing new active ingredients, and thus correspond to more radical innovations (there are 442 new molecular entities and 2,203 new nongenerics in our data set). Panel C shows a significant relationship between market size and new molecular entities. The estimate of α is 3.54 (standard error = 1.19), and is again statistically significant at 1 percent. Interestingly, while all nongenerics are most responsive to current market size, the coefficient of lead market size is larger with new molecular entities, 5.75, though the standard error is also larger, 2.37. This evidence, though not conclusive, is consistent with a limited anticipation effect in the response of innovation to market size.[23]

22. We construct the lagged market size measures for 1960s using demographic information from the CPS, so the number of observations does not decline. If we only use the post-1970 data, the results are similar, though lagged entry is somewhat stronger.

23. Here "limited" does not refer to the strength of the effect, but to the fact that the response to market size is five years before the change in market size, not

Finally, for completeness, Panel D shows the effect of potential market size on the entry of generics. The estimate of α in the baseline specification of column (1) is 11.81 (standard error = 3.30). This effect is considerably larger than those that are obtained in the cross-sectional studies, such as Scott Morton [1999] and Reiffen and Ward [2004]. This might partly reflect the complex entry dynamics of generic drugs, though we did not find evidence of such dynamics in our investigations. Since generics are not our main focus, we do not pursue potential explanations for this large effect further.

The magnitude of the effects of market size on entry of nongenerics and new molecular entities in Table II is also large, in particular, larger than the proportional effects predicted by our model.[24] Note, however, that our key right-hand-side variable measures potential market size rather than actual market size, and these two measures might differ because of changes in expenditure shares. To investigate whether the difference between actual and potential market size might affect the magnitude of our estimates, we use the NAMCS, where we can measure actual market size in terms of total use in each drug category between 1980 and 2000 (recall that the NAMCS data do not contain expenditure information). A simple regression of actual market size in each drug category on our measure of population-based potential market size and category and period dummies yields a coefficient of 4.06 (standard error = 1.60), which suggests that

---

further in advance. Two-period (ten-year) lead of market size has a large, but imprecise effect on new molecular entities (significant at 5 percent), while further leads are insignificant.

24. Our estimates refer to the effect of market size on the flow of new drugs, which may differ from the effect on the stock of drugs. To check for this possibility, we estimate our basic models using $\Sigma_{\tau=1}^{t} N_{c\tau}$, i.e., the stock of drugs at time t, as the dependent variable. Consistent with our finding of limited residual serial correlation below, this procedure leads to slightly smaller estimates, 1.84 (standard error = 0.98) for nongenerics, and 3.10 (standard error = 1.06) for new molecular entities. The evidence is therefore consistent with a broadly similar and somewhat smaller response of the stock of drugs to market size than the response of the flow of new drugs.

It also has to be borne in mind that these estimates are informative about the effect of market size on the *composition* of research, and the relationship between total pharmaceutical market size and aggregate research could be quite different. If we estimate (11) for nongenerics without time effects, we obtain a coefficient of 0.22, with a standard error of 0.09 (for new molecular entities, the estimate is 0.41, with a standard error of 0.06). This is consistent with the view that the response of the composition of R&D to market size is quite different from the response of total R&D. Nevertheless, the difference between the results with and without time effects is at least partly due to the presence of other time-varying factors affecting entry of new drugs.

between 1980 and 2000 actual market size went up by 4 percent for every 1 percent increase in potential market size.[25] Assuming that this relationship also applies to the entire sample and to the income-based measure of market size, our estimates of 4–6 percent response to potential market size are consistent with a proportional relationship between entry of new drugs and actual market size.[26]

V.B. Robustness

Table III investigates the robustness of the effect of potential market size on the entry of nongeneric drugs (Panel A) and new molecular entities (Panel B). Although the entry of new molecular entities may be responding somewhat more to the five-year lead of market size than to current market size, we report results for current market size in Panel B for compatibility with Panel A. Results for the effect of five-year leads of market size on new molecular entities are similar, and generally somewhat stronger.

Column (1) replicates the baseline results from Table II for comparison. In column (2) time periods are ten years instead of the five-year intervals. The estimate of α for nongenerics is somewhat larger, 4.81 (standard error = 1.31), while the estimate for new molecular entities is similar to the baseline, 3.91 (standard error = 1.29). Both estimates are significant at 1 percent.

Column (3) looks at the effect of changes in market size driven purely by population changes (in this case, regression weights are total use in the corresponding category in the MEPS). The estimates in both panels are larger than the baseline, and continue to be significant at 1 percent. Since the income-based

25. This result implies that use per person went up in categories experiencing an increase in market size due to demographic changes, which may itself be partly due to increases in the number and quality of drugs in these categories.

26. In any case, our estimate of a 4 percent increase in the rate of entry of new nongenerics in response to a 1 percent increase in market size is not implausible. There are a total of 2203 nongeneric approvals between 1970 and 2000, thus, on average, 10 approvals in every five-year interval in each of our 33 categories. Therefore, our estimate implies that a 2.5 percent increase in market size should lead to the entry of about one new drug. Total pharmaceutical sales were approximately $130 billion in 1999 [IMS 2000], which implies an average annual expenditure of $3.9 billion per category. A 2.5 percent increase therefore corresponds to $97.5 million, or about $1.5 billion over fifteen years, which is the life of a typical nongeneric drug. Since entry costs for nongenerics are around $800 million (in 2000 dollars, DiMasi, Hansen, and Grabowski [2003]), entry of one new drug in response to an increase of approximately $1.5 billion in revenue is within the range of plausible responses. Naturally, this calculation is very rough and only suggestive, since it ignores the difference between average demand and the demand that a marginal entrant will capture.

TABLE III
ROBUSTNESS CHECKS

	Baseline QML	10-year intervals	Population-based market size	NAMCS market size	OECD market size	Unweighted regressions	Market size uses single-age groups	Market size uses previous classification	Linear model	Drop cardiac
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: dependent variable is count of nongeneric drug approvals										
Market size	3.82	4.81	5.35	4.53	3.27	1.81	3.67	3.68	3.37	3.90
	(1.15)	(1.31)	(1.63)	(1.12)	(0.86)	(1.61)	(1.18)	(1.07)	(1.75)	(1.38)
Approvals	2203	2203	2203	2203	2203	2203	2203	2309	2203	2078
Panel B: dependent variable is count of new molecular entities										
Market size	3.54	3.91	5.13	4.16	3.28	4.62	3.35	2.73	3.54	3.17
	(1.19)	(1.29)	(1.22)	(1.01)	(0.84)	(1.98)	(1.23)	(1.74)	(1.40)	(1.46)
Approvals	442	442	442	442	442	442	442	492	442	397
Number of observations	198	99	198	198	198	198	198	204	198	192

Huber-White robust standard errors are in parentheses. The dependent variable is count of nongeneric drug approvals in Panel A and count of new molecular entities in Panel B, computed from the FDA data set of New Drug Approvals. All columns except (3)–(5) use the income-based measure of market size; columns (3)–(5) use the population-based measure. All regressions include drug and time dummies, and, except for column (6), are weighted. Market size in column (4) is computed using the NAMCS data for drug use, and in column (5) market size is computed using total OECD population, as explained in text. In column (7) market size is computed using single-age groups (see text). In column (8) market size is constructed as in Table II, except that the classification from Acemoglu and Linn [2003] is used (see text). The Poisson model is estimated using quasi maximum likelihood, with the Hausman, Hall, and Griliches [1984] transformation in columns (1)–(8) and (10). In column (9) the linear model in equation (12) is estimated. If a cell is empty, the log approvals variable is set equal to zero, and a dummy variable is added, equal to one when the cell is empty (see text).

measure is closer to the notion of market size suggested by theory, we continue to focus on this measure.[27] Column (4) shows similar results using the population-based measure of market size with our alternative data set, the NAMCS.

Column (5) uses an OECD market size measure combining West European and Japanese demographic information with the U. S. information.[28] Since we only have information on population for the other countries, we perform this exercise for the population-based measure of market size. The U. S. and OECD populations by age group have a high correlation, equal to 0.81. Using the OECD market size measure leads to similar, and somewhat surprisingly, more precise results. For all nongenerics, the estimate of α is 3.27 (standard error = 0.86), while for new molecular entities, the estimate is 3.28 (standard error = 0.84).

Column (6) investigates the effect of weighting on the estimates. The unweighted estimate of α for all nongenerics is smaller than the baseline, 1.81 (standard error = 1.61), and no longer statistically significant. For new molecular entities the estimate is larger and still significant: 4.62 (standard error = 1.98). In both cases, the standard errors are significantly larger, reflecting the fact that expenditure shares in smaller cells are less precisely estimated.

Column (7) uses an alternative measure of market size constructed with single-age groups for i_{at}'s and u_{ca}'s in equation (13). This procedure uses more information about the age profiles, but since there are fewer observations in some single-age groups, the estimates of u_{ca}'s are less precise. The estimates using this alternative measure are very similar to the baseline results; the estimate of α for nongenerics is 3.67 with a standard error of 1.18, and the estimate for new molecular entities is 3.35 (standard error = 1.23).

Column (8) uses the alternative classification system from Acemoglu and Linn [2003], which uses only differences in the age profiles of expenditure to subdivide the major FDA categories. This classification system contains 34 categories, and because it includes a number of small FDA detailed categories that are dropped from the current sys-

27. Instrumenting the income-based market size measure with the population-based market size measure leads to similar estimates to the baseline. The estimate for nongenerics is 4.07 (standard error = 1.37), and the estimate for new molecular entities is 3.84 (standard error = 1.41).

28. These data were obtained from the United Nations Web site, esa.un.org/unpp/.

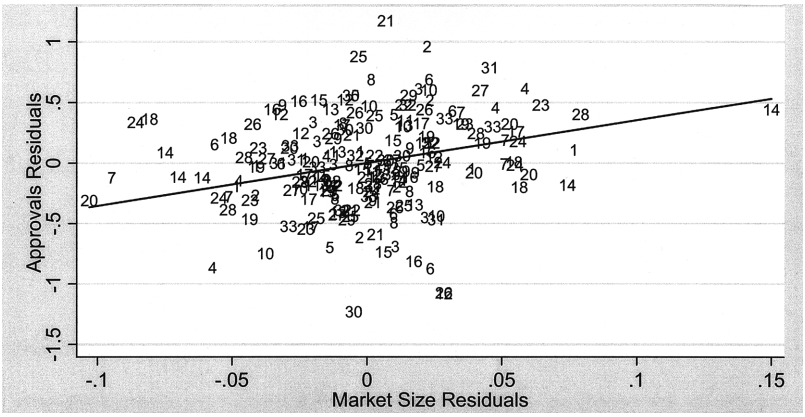

FIGURE IV

Approvals Residuals versus Market Size Residuals for New Molecular Entities

Approvals residuals and market size residuals are residuals from OLS regressions of log approvals and log income-based market size on category and time dummies, weighted by expenditure with five-year intervals. Fitted values are predicted approvals residuals obtained from the OLS regression in Table III, Panel B, column (9).

tem, there are now 106 more approvals (see Appendix 1). The estimate for nongenerics is similar to the baseline in column (1), 3.68 (standard error = 1.07), while the estimate for new molecular entities is smaller than the baseline and insignificant. However, with this classification system there is a stronger effect of lead market size on new molecular entities (and thus somewhat stronger evidence for anticipation effects). For example, the estimate of α with new molecular entities and lead market size is 6.81 (standard error = 1.29), which is significant at the 1 percent level (not reported).

Column (9) estimates the model in equation (12) with the Pakes-Griliches transformation using OLS. The results are similar to those in column (1). For nongenerics, the estimate of α is 3.37 with a standard error of 1.75, and for new molecular entities, it is 3.54 with a standard error of 1.40. We also use the linear model to document that the relationship we observe is not driven by outliers. Figure IV shows the relationship between the residuals of new molecular entities versus the residuals of market size, log M_{ct}, after drug category and time period dummies are removed. Observations are labeled by their drug category codes (see Appendix 2), and each code appears more than once, since there

are multiple periods. This figure shows that there are no major outliers (the figure for nongenerics is similar).[29]

Finally, column (10) checks the robustness of the results to dropping the Cardiac category, which includes the most diverse types of drugs. The exclusion of this category has little effect on the estimates.[30]

V.C. Potential Supply-Side Determinants of Innovation

The first part of Table IV investigates the robustness of the baseline results to controlling for potential supply-side determinants of innovation, such as changes in scientific incentives or opportunities captured by the δ_j's in the theoretical model. Panel A reports results for nongenerics, while Panel B looks at new molecular entities.

First, recall that the major threat to our identification strategy is changes in the δ_j's (since permanent differences in δ_j's are already taken out by the drug category fixed effects). If the δ_j's vary over time, they are also likely to be serially correlated. Adding lags of log N_{ct} to our basic specifications is therefore a simple way to check the importance of these concerns.

Columns (1) and (2) of Table IV report the results of QML estimation of a lagged dependent variable specification of the form,

$$E[N_{ct}|\zeta_c,\bar{X}_c,\bar{N}_c] = \frac{\exp(\alpha \cdot \log M_{ct} + \psi \cdot \log \tilde{N}_{ct-1} + \nu \cdot d_{ct-1} + \mu_t)}{\Sigma_{\tau=1}^{T} \exp(\alpha \cdot \log M_{c\tau} + \psi \cdot \log \tilde{N}_{c\tau-1} + \nu \cdot d_{c\tau-1} + \mu_\tau)} \bar{N}_c, \quad (14)$$

where we use the notation in (11), and in particular, $\bar{N}_c = \Sigma_{\tau=1}^{T} N_{c\tau}$ is the number of drugs approved in category c over the entire sample. Since N_{ct-1} can be equal to 0, we follow a procedure similar to that of Pakes and Griliches [1980], used in column (9) of Table III above, and define $\tilde{N}_{ct-1} = N_{ct-1}$ if $N_{ct-1} \geq 1$ and $\tilde{N}_{ct-1} = 1$ if $N_{ct-1} = 0$, and add the dummy d_{ct-1} that equals 1 when $N_{ct-1} = 0$. The estimate of α for nongenerics in

29. Dropping the one category that appears as a slight outlier, category 14 (Anorexiants/CNS Stimulants), has no effect on the estimate, which only changes to 3.66 (standard error = 1.46).

30. We have also experimented with dropping each of the other categories one at a time. The effect of market size on both nongenerics and new molecular entities remains significant at 5 percent in all cases, except when we drop Antibiotics. Without Antibiotics, the estimates are similar to the baseline results, but no longer significant. For nongenerics the estimate is 1.72 (standard error = 2.01), and for new molecular entities, 4.14 (standard error = 2.59).

TABLE IV

POTENTIAL SUPPLY-SIDE AND DEMAND-SIDE DETERMINANTS OF INNOVATION

	Lag dep var (1)	Lag dep var, IV (2)	Life-years lost (3)	Public funding (4)	Preexisting trends (5)	Major cat trends (6)	Major cat trends, linear model (7)	Drop Cancer, Vascular (8)	Drop Thyroid, Anemia (9)	Health insurance mkt size (10)	IV with previous mkt size (11)
Panel A: dependent variable is count of nongeneric drug approvals											
Market size	3.84	3.98	3.58	3.86	3.24	2.93	5.46	3.59	3.72	1.92	2.93
	(1.07)	(1.16)	(1.70)	(1.20)	(0.91)	(5.11)	(2.43)	(1.26)	(1.11)	(0.44)	(1.45)
Lagged dependent variable	0.12	−0.43									
	(0.10)	(0.33)									
Panel B: dependent variable is count of new molecular entities											
Market size	3.57	3.69	3.64	3.56	3.84	7.64	3.79	4.59	3.27	2.10	3.08
	(1.39)	(1.38)	(1.79)	(1.20)	(1.17)	(5.83)	(2.20)	(1.40)	(1.23)	(0.51)	(1.32)
Lagged dependent variable	−0.19	−1.34									
	(0.11)	(0.82)									
Number of observations	198	198	198	198	198	198	198	186	186	198	198

Huber-White robust standard errors are in parentheses. The dependent variable is count of nongeneric drug approvals in Panel A and count of new molecular entities in Panel B, computed from the FDA data set of New Drug Approvals. Market size is constructed as in Table II for columns (1)–(9) and (11). All regressions include drug and time dummies, use the income-based measure of market size, and are weighted by expenditure. In column (2) the lagged dependent variable is instrumented with the twice lagged first difference of the dependent variable (see text). Life-years lost is years prior to age 65 for each death in the United States, calculated from the Mortality Detail Files (see text). Column (3) includes a count of total life-years lost due to diseases in the corresponding category and time interval. Column (4) includes the amount of funding from NIH research grants in each category and time interval, calculated from the CRISP database. 1940/1970 trend is the log difference of drug approvals for category c between 1940 and 1970. In column (5) the 1940/1970 trend is interacted with period dummies (see text). Major drug category trends are linear time trends interacted with dummies for the sixteen major drug categories (see text). In column (10) market size includes information on health-care coverage (see text). In column (11) current market size is instrumented with the market size of the same cohort five years earlier.

this case is 3.84 (standard error = 1.07), and the coefficient on lagged entry is 0.12 (standard error = 0.10), thus small and insignificant.[31] The results for new molecular entities are similar.

The estimates in column (1) treat all right-hand-side variables as strictly exogenous. Because $\log \tilde{N}_{ct-1}$ is endogenous, these estimates are inconsistent. As long as there is no additional autocorrelation in the errors, instrumenting for $\log \tilde{N}_{ct-1}$ with $\Delta \log \tilde{N}_{ct-2}$ would lead to consistent estimates.[32] Since the estimating equation in (14) is nonlinear, we perform this instrumentation strategy by adding the residuals from the first-stage regression as an additional right-hand-side variable to the second stage (see Wooldridge [2002] Chapter 19.5). Column (2) reports the results of this exercise. Once again, the coefficients are very similar to the baseline estimates in all three panels, and show no evidence of significant effects of lagged entry. For example, for nongenerics, the estimate of α is 3.98 (standard error = 1.16), and for new molecular entities, the estimate is 3.69 (standard error = 1.38). Furthermore, as a direct check, we tested and found no evidence for serial correlation in the residuals from the estimation of equation (11) in Table II. Overall, the results show that the effect of potential market size on entry of nongenerics and new molecular entities is robust to controlling for lagged entry, and there is no evidence of residual serial correlation.

A plausible conjecture is that nonprofit incentives to develop drugs would be related to opportunities to save lives or cure major illnesses. Our second strategy controls for differences in the health benefits of new drugs across categories. New drugs in our data set include both drugs that are demanded by the consumers but do not "save lives," such as Prozac, Paxil, Vioxx, or Viagra, and those that actually save lives such as heart medicines or cancer treatments (see Lichtenberg [2002, 2003], on the effect of pharmaceutical innovations on declines in mortality). We esti-

31. In the estimates reported in columns (1) and (2) of Table IV, we use information on approvals before 1970 to construct lags, so the sample size remains the same as in the basic specifications. Because before 1970 we cannot distinguish generics and nongenerics (though we can identify new molecular entities), we use all approvals for the pre-1970 lags in Panel A.

Whether or not we use the pre-1970 data is not important for the results. Estimating the model in (14) without the pre-1970 data for nongenerics gives an estimate of α of 3.75 (standard error = 1.93), and the lagged dependent variable is again very small and insignificant. The same is true for new molecular entities.

32. See, for example, Arellano and Bover [1995] and Blundell and Bond [1998]. We cannot use other commonly used moment restrictions, since equation (14) cannot be first-differenced.

mate the number of life-years lost corresponding to each drug category using the Mortality Detail Files from the National Center for Health Statistics from 1970–1998. Following Lichtenberg [2002], for each death we subtract the person's age from 65, then calculate the total number of life-years lost for all the deaths resulting from diseases related to drugs in each category.[33] Column (3) adds life-years lost to the right-hand side of our baseline regression models as a proxy for this source of nonprofit incentive to undertake research. The estimate of the effect of market size on nongenerics is now 3.58 (standard error = 1.70), and the estimate for new molecular entities is 3.64 (standard error = 1.79). In all cases, the variable for life-years lost is not significant.

We also investigate the implications of differences in scientific funding for various drug categories. Using the Computer Retrieval of Information on Scientific Projects (CRISP) data set (details are contained in Lichtenberg [2001] and Acemoglu and Linn [2003]), we construct a variable measuring the total amount of federal funding for research in each drug category, and include this variable as a control. To the extent that government funding also responds to potential market size (for example, because drug companies have a greater tendency to apply for funding in areas where they plan to do research), this variable would be correlated with our market size measure. In practice, the correlation is low, and column (4) shows that the inclusion of this variable has little effect on our estimates. The estimate of the effect of market size is 3.86 (standard error = 1.20) for nongenerics and 3.56 (standard error = 1.20) for new molecular entities. The funding variable itself is positive, but small and insignificant (not reported in the table).

Next, to control for potential trends in scientific opportunities across drug categories, we add proxies for preexisting trends. We construct an estimate for preexisting trends as $\Delta_c = (\log N_{c,70} - \log N_{c,40})$, where $\log N_{c,70}$ is the log approvals for category c in 1970 and $\log N_{c,40}$ is the log approvals in 1940.[34] We then add a full set of interactions between Δ_c and the time dummies. This specification therefore allows drug categories that have grown

33. For example, if someone dies at age 32, this counts as 33 life-years lost; people dying older than 65 receive no weight in this calculation. Since many of our categories contain diseases or conditions that do not lead to death, we obtain several empty cells.

34. Because we cannot distinguish between nongenerics and generics before 1970, in Panel A, we use total approvals before 1970. Also, since there were no new molecular entities approved in the 1940 period, in Panel B we construct the preexisting trend using 1950 and 1970 approvals of new molecular entities.

differentially between 1940 and 1970 to also grow at different rates in the later periods. Column (5) reports the results of this exercise. The estimates are very similar to the baseline in all three panels. These results are perhaps not surprising, since pre-1970 approvals are considerably noisier and do not distinguish between generics and nongenerics, thus are only an imperfect control for preexisting trends.

An alternative, and substantially more demanding, strategy is to include in-sample linear time trends. To do so, we add linear time trends for each of the sixteen major FDA categories. We expect technological differences to be well approximated by the sixteen major drug categories, which are based on broad therapeutic intent. The estimates, reported in column (6), are quantitatively similar to the baseline, but no longer significant because the standard errors are larger, reflecting the fact that changes in market size due to demographic trends are relatively smooth, and thus highly colinear with the time trends (nevertheless, the estimate for new molecular entities is significant if we use lead market size). Since distinguishing between linear time trends and changes in market size may be more difficult with the nonlinear model in (11), we also estimate the model in (12) with linear time trends using OLS. The results, reported in column (7), are also similar to the baseline, and now the estimate for nongenerics is significant at 5 percent, and the estimate for new molecular entities is significant at 10 percent.

We also investigate the potential effects of recent advances in biotechnology, such as the use of recombinant DNA or other technological changes, which may correspond to changes in the δ_j's in terms of our model. In column (8) we drop the categories of Cancer and Vascular, which, according to the FDA approval list, have witnessed the entry of the greatest number of orphan drugs (presumably by biotechnology firms). In addition, there is anecdotal evidence that biotechnology firms were first active in producing insulin (the Thyroid and Glucose category) and in the Anemia category, so in column (9) we drop these two categories.[35] In both cases, the estimates are very similar to our baseline results.

Finally, to see whether the advent of biotechnology or other technological advances of the past two decades have changed the

35. Biotechnology firms have also been active in producing human growth factor, but since there are only a small number of individuals using these drugs in the MEPS, these drugs are not included in our approvals data set.

relationship between market size and entry of new drugs, we estimate our baseline models including an interaction between a post-1985 (or post-1990) dummy and market size. Our estimates show no evidence of significant interactions.[36]

The results in this subsection therefore document that a number of controls for supply-side factors have little effect on our main finding regarding the effect of market size on entry of nongeneric drugs and new molecular entities. Although these results are not conclusive on the effect of scientific or other supply-side factors in pharmaceutical research, they suggest that the effect of potential market size on entry and innovation is relatively robust.

V.D. Changes in Health Insurance Coverage

Our market size measure only exploits changes in potential market size driven by demographic trends. Another source of variation in market size comes from changes in coverage of drug expenditure in private or public health insurance programs. During our sample period, there were significant changes in the coverage of drug expenditures in health insurance plans. For example, the percentage of 60+ year-olds with private insurance rose from 60 percent to 75 percent between 1974 and 1996 (authors' calculations). We now investigate the effect of including information on health insurance coverage in our market size measure.

We use the National Health Interview Survey (NHIS, 1974–1996) to construct a market size measure incorporating information on health insurance coverage as follows: $\tilde{M}_{ct} = \Sigma_a u_{ca} \cdot i_{at} \cdot f_{at}$, where f_{at} is the fraction of those of age a in period t with private health insurance, and u_{ca} and i_{at} are as defined above. Because there is no consistent information on prescription drug coverage, we assign prescription coverage to any individual with both doctor and surgical coverage. Prescription drug coverage is highly correlated with this measure in the years we can observe it. In column (10) we use $\log \tilde{M}_{ct}$ as our market size measure instead of $\log M_{ct}$. This leads to similar, and somewhat more precise, results. For nongenerics, the estimate of α is 1.92 with standard error 0.44, while for new molecular entities, it is 2.10 (standard error = 0.51). Despite the greater

36. For example, in a specification parallel to the model for nongenerics in column (1) of Table II, the estimate of α is 4.29 (s.e. = 1.66), and the interaction with the post-1985 dummy is −0.13 (s.e. = 0.18), thus small and insignificant. For new molecular entities, the estimate of α is 4.00 (s.e. = 1.46), and the post-1985 interaction is −0.10 (s.e. = 0.13).

precision of these results, we have more confidence in our baseline estimates, since the measure $\tilde{M}_{ct}$ effectively assigns 0 expenditure to those without insurance and relies on information on drug coverage imputed from doctor and surgical coverage.

V.E. Reverse Causality

Lichtenberg [2002, 2003] shows that new drugs have increased the average age at death. This introduces the potential for reverse causality whereby the market size for successful drugs may be endogenously larger, because their users live longer. This is unlikely to be a first-order concern, since drug-induced changes in population are small relative to the demographic changes that we are exploiting. Nevertheless, we address this issue by instrumenting for current population using the corresponding population from five years before. For example, we use the income of 50–54 year-olds in 1975 as an instrument for the income of 55–59 year-olds in 1980. The fraction of 50–54 year-olds is highly correlated with the fraction of 55–59 year-olds five years later, but is unaffected by new drugs that are developed in the intervening five years. As described above, the instrumental-variables procedure is performed by adding the residuals from the first-stage as an additional right-hand-side variable (see Wooldridge [2002]).

These instrumental-variables estimates, reported in column (11) of Table IV, are similar to the baseline results and show no evidence of reverse causality. For nongenerics, the estimate of α is 2.93 (standard error = 1.45), and for new molecular entities, the estimate is 3.08 (standard error = 1.32).

V.F. Patents

The results so far show a large and robust effect of potential market size on entry of nongeneric drugs and new molecular entities, and suggest a strong link between market size and innovation. In this subsection we briefly investigate the relationship between market size and another measure of pharmaceutical innovation, patents.[37]

We obtained data on pharmaceutical patents from Thomson

37. Firms typically apply for a patent prior to the clinical trial stage of drug development, or about 5–10 years before the drug is approved, and therefore lose a significant fraction of the life of the patent before it can begin marketing the drug. Part of the 1984 Hatch-Waxman Act allowed pharmaceutical companies to apply to the FDA for an extension of the life of their patents, if they could show that they lost marketing time while waiting for approval. The maximum extension is five years, and depends, among other things, on the length of the FDA approval

Derwent Inc., and with the help of a specialist at this company, we mapped these patents into our classification system.[38] Using these data, we found no significant relationship between market size and patents, which might be due to a variety of reasons.[39] First, this result may simply reflect the imperfect match between the patent data and the FDA categories, especially bearing in mind the potential use of certain chemical structures in multiple drug lines. Second, the significant costs and uncertainty involved in the development of new molecules and patentable products may be creating substantial attenuation (e.g., a drug intended for the 1990s may be patented in the 1980s or 1990s, depending on delays in the research process). Third, pharmaceutical companies may respond to profit incentives more during the later stages of the research process than during the earlier stages. Finally, because U. S. patents include those taken by foreign companies, they may be more responsive to OECD demand than to U. S. demand. To investigate the last possibility, we estimated the relationship between changes in OECD market size derived from European, Japanese, and U. S. demographic changes. In this case, we find a significant relationship between five- or ten-years leads of OECD market size and patents. With the five-year lead of market size, the estimate of α is 3.49 (standard error = 1.02) and with the ten-year lead, the estimate is 5.02 (standard error = 0.63).[40] Although this result suggests that OECD demand may be more important for patents, we are currently unable to make more progress in distinguishing between these various explanations, and the weaker results for patents remain a puzzle.

VI. Concluding Remarks

This paper investigates the response of entry of new drugs and pharmaceutical innovation to changes in potential market

process. Overall, companies have a maximum of fourteen years of patent protection after FDA approval.

We were unable to obtain data for a sufficient number of categories for another possible proxy for pharmaceutical innovation, clinical trials.

38. We could not use the data from the Hall-Jaffe-Trachtenberg patent data set (see Jaffe, Trachtenberg, and Henderson [2002]) because we were unable to map their classification based on chemical structure to our drug categories.

39. Finkelstein [2004] also finds weaker results for vaccine patents than for later stages of development.

40. Part of the difference between the U. S. and OECD results is driven by the fact that we are using income-based measures for the U. S. and the population-based measures for the OECD. Using the population-based measure for the U. S. lead market size with patents yields a coefficient of 2.34 (standard error = 4.63).

size driven by demographic changes. Our results indicate that a 1 percent increase in the potential market size for a drug category leads to approximately a 4 percent growth in the entry of new nongeneric drugs and new molecular entities.

The effect of market size on entry of new drugs and new molecular entities, if further proved to be robust, has important implications both for research on the pharmaceutical industry, and for the endogenous growth and directed technical change literatures. It provides evidence that, as posited by these models, R&D and technological change are directed toward more profitable areas. These findings also imply that there may be little pharmaceutical research toward drugs with small markets, especially toward those intended for groups with limited ability to pay, which is a key premise of recent work by Kremer [2002]. Based on this premise, Kremer suggests that there need to be selective government incentives for developing drugs against malaria and other third-world diseases.

We view this research as part of a broader investigation of the effects of profit incentives on innovation. Evidence from a single industry may be nonrepresentative, for example, because pharmaceuticals may be more research oriented than other industries. Future research investigating the response of innovation and entry of new products to market size both in specific industries and at the economywide level is necessary to substantiate the results presented here.

APPENDIX 1

A. *Medical Expenditure Panel Survey (MEPS) and Construction of the Drug Classification System*

The MEPS is an annual survey of randomly sampled households; we use the 1996, 1997, and 1998 surveys. We obtain each person's age, the name and national drug code of the prescription drug(s) used, and total expenditure (there are multiple records for people who use more than one prescription drug). Over the three years we have about 500,000 drugs used and about 85,000 individuals. Expenditure information includes out-of-pocket expenses, as well as amounts paid by insurance companies and government payments (e.g., Medicaid and worker's compensation). These data are collected from the pharmacies and medical providers listed by the respondents.

We begin with the 159 therapeutic categories, obtained from the FDA's National Drug Code (NDC) Directory. The names of these categories can be found in the second column of Appendix 2. The NDC Directory contains a file with the therapeutic category for most FDA-approved drugs currently on the market. We assign each drug in the MEPS to one of the 159 categories by matching it by national drug code with a drug in the NDC file. We cannot match about 10 percent of the drugs mentioned in the MEPS; these are usually not commonly used drugs, and make up less than 5 percent of the total drugs used.

Drug expenditure shares and use per person are calculated by computing drug expenditure and use by five-year age group, and then dividing by the income and population of the age group. We use the population numbers from the MEPS (so use per person is the weighted average of use per person of respondents in the MEPS), but income estimates from the CPS. We prefer the CPS income estimates because the MEPS income data are likely to contain greater measurement error; in the MEPS the sample is smaller, wage and salary incomes for almost half of the sample are imputed either based on broad income ranges or other information, nonwage incomes were generally imputed, and the imputation methods changed between the 1996–1997 and 1998 surveys. Nevertheless, the results are almost identical if we construct expenditure shares for individuals in the survey (i.e., without using CPS information in the same way as we do for use per person). For example, the estimate of the effect of market size on the entry of nongenerics is 4.08 (standard error = 1.31), and the estimate for new molecular entities is 3.59 (standard error = 1.35). We also checked the robustness of our results using an alternative market size measure constructed with single-age groups, and the results are reported in Table III. We prefer the measure using five-year age groups, since there are only a few observations in some single-age groups in the MEPS.

The FDA has assigned the 159 categories to one of twenty major therapeutic categories. As noted in the text, we drop four major categories: Anesthetics, Antidotes, Radiopharmaceuticals, and Miscellaneous.[41] Within each major category we first subdivide catego-

41. We also drop several minor categories when there are not sufficient observations to estimate a reliable age profile. We use about 1,000 observations as our cutoff rule. We obtain this number from observing that only categories with more than 1,000 observations have fairly smooth age profiles.

ries whose drugs have different indications (we determine drugs' indications by searching by name on the National Institute of Health Web site, www.nlm.nih.gov/medlineplus/druginformation.html). For categories that have not been subdivided based on indications, we then divide them if there is sufficient heterogeneity in the age profile of users for subcategories. Appendix 2 shows that we create subcategories when there is considerable age variation within broad categories. For example, within the Hormones major category, Estrogens/Progestins are used predominantly by 30–60 year-olds, while Contraceptives are used fairly evenly by 0–30 year-olds and 30–60 year-olds. This classification system differs somewhat from the working paper version, in which we divided major classes based entirely on age structure. The previous system includes several FDA categories that are dropped from the current one: the CNS Miscellaneous, Hyperlipidemia and Calcium Metabolism categories, which contain drugs used for heterogeneous conditions, and several categories that would have been subdivided based on drug indications, but had fewer than 1000 observations in the MEPS. The details of the previous classification system are in Acemoglu and Linn [2003], and Table III, column (8), reports results using this older classification.

B. Drug Approvals from the FDA

The list of FDA drug approvals were obtained by Lichtenberg and Virahbak [2002] under the Freedom of Information Act. We thank Frank Lichtenberg for generously sharing these data with us. Over-the-counter drugs and orphan drugs (of which only a few can be matched) are excluded. Biologics, which go through a separate approval process, are not in this data set.

We match drugs in the approval list to FDA categories by drug name and FDA approval number. Since 1970, 14,432 of 16,772 prescription drugs (86 percent) approved are matched, while before 1970, the match rate is about 51 percent. This motivates our focus on drug approvals between 1970 and 2000. Drugs that have the same approval number as a previously approved drug and drugs for which the corresponding FDA category is dropped because of insufficient observations in the MEPS are excluded. Finally, we drop drugs with the same name, MEPS category, and company as a previously approved drug.

APPENDIX 2: SUMMARY OF DISEASE CLASSIFICATION AND DRUG EXPENDITURE BY AGE GROUP

Class	Description	Expenditure share × 1000 {Share of expenditure by age group in total expenditure in brackets} 0–30	30–60	60+	Age group with largest expenditure
1	Antibiotics	0.95 {0.41}	0.62 {0.40}	0.90 {0.19}	0–30
2	Antivirals	0.03 {0.05}	0.36 {0.91}	0.05 {0.04}	30–60
3	Antiparasitics	0.01 {0.10}	0.05 {0.60}	0.08 {0.29}	30–60
4	Antifungals	0.26 {0.32}	0.23 {0.44}	0.38 {0.24}	30–60
5	Anemia	0.00 {0.07}	0.00 {0.47}	0.01 {0.47}	60+
6	Anticoagulants	0.00 {0.01}	0.06 {0.19}	0.77 {0.80}	60+
7	Glaucoma	0.00 {0.01}	0.03 {0.14}	0.58 {0.85}	60+
8	Acid/Peptic Disorders	0.17 {0.06}	0.89 {0.46}	2.87 {0.49}	60+
9	Antidiarrheals, Laxatives	0.00 {0.07}	0.01 {0.45}	0.03 {0.49}	60+
10	Cardiac	0.03 {0.01}	0.72 {0.32}	4.68 {0.67}	60+
11	Vascular	0.12 {0.02}	1.24 {0.34}	7.00 {0.64}	60+
12	Sedatives/Hypnotics, Antianxiety	0.05 {0.06}	0.25 {0.54}	0.58 {0.40}	30–60
13	Antipsychotics/Antimanics, Antidepressants	0.46 {0.13}	1.64 {0.70}	1.28 {0.18}	30–60
14	Anorexiants/CNS Stimulants	0.08 {0.52}	0.05 {0.45}	0.01 {0.03}	0–30
15	Vitamins/Minerals	0.00 {0.07}	0.01 {0.36}	0.05 {0.58}	60+
16	Electrolyte Replenishment/ Regulation, Water Balance	0.01 {0.03}	0.05 {0.26}	0.46 {0.71}	60+
17	Adrenal Corticosteroids	0.05 {0.26}	0.05 {0.40}	0.14 {0.34}	30–60
18	Androgens/Anabolic Steroids	0.00 {0.03}	0.01 {0.21}	0.07 {0.77}	60+
19	Estrogens/Progestins	0.31 {0.17}	0.71 {0.58}	0.97 {0.26}	30–60
20	Contraceptives	0.11 {0.47}	0.08 {0.52}	0.00 {0.01}	30–60
21	Blood Glucose Regulators, Thyroid/Antithyroid	0.08 {0.03}	0.75 {0.43}	2.90 {0.54}	60+
22	Topical Steroids	0.01 {0.21}	0.02 {0.36}	0.06 {0.43}	60+
23	Topical Anti-Infectives	0.01 {0.32}	0.01 {0.41}	0.02 {0.27}	30–60

APPENDIX 2: (*CONTINUED*)

Class	Description	Expenditure share × 1000 {Share of expenditure by age group in total expenditure in brackets} 0–30	30–60	60+	Age group with largest expenditure
24	Extrapyramidal Movement Disorders	0.00 {0.01}	0.03 {0.25}	0.29 {0.74}	60+
25	Skeletal Muscle Hyperactivity, Anticonvulsants	0.21 {0.19}	0.46 {0.64}	0.39 {0.18}	30–60
26	Oncolytics	0.20 {0.19}	0.29 {0.42}	0.84 {0.39}	30–60
27	Ocular Anti-Infective/Anti-Inflammatory	0.05 {0.29}	0.04 {0.32}	0.14 {0.39}	60+
28	Topical Otics	0.01 {0.41}	0.01 {0.30}	0.02 {0.28}	0–30
29	Vertigo/Motion Sickness	0.02 {0.19}	0.02 {0.38}	0.08 {0.42}	60+
30	Pain Relief	0.27 {0.09}	1.08 {0.55}	2.09 {0.35}	30–60
31	Antiasthmatics/Broncodilators	0.38 {0.23}	0.39 {0.37}	1.28 {0.40}	60+
32	Nasal Decongestants, Antitussives, Cold Remedies	0.05 {0.23}	0.07 {0.57}	0.08 {0.20}	30–60
33	Antihistamines, Inhalation/ Nasal	0.36 {0.25}	0.51 {0.53}	0.64 {0.22}	30–60

All data are from the MEPS, 1996–1998. Construction of the 33 categories is described in Appendix 1. Each category includes the indicated FDA subcategories. Expenditure share is the total expenditure on drugs in the category divided by the total income of users in that age group. Share of expenditure by age group is the fraction of total expenditure in the category accounted for by the age group. In this table shares of expenditure by category are calculated for 30-year broad age groups. Age group with largest expenditure is the broad age group with the greatest expenditure on the corresponding category.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

REFERENCES

Acemoglu, Daron, "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality," *Quarterly Journal of Economics,* CXIII (1998), 1055–1090.

——, "Directed Technical Change," *Review of Economic Studies,* LXIX (2002), 781–810.

——, "Labor- and Capital-Augmenting Technical Change," *Journal of European Economic Associations,* I (2003), 1–37.

Acemoglu, Daron, and Joshua Linn, "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," NBER Working Paper No. 10038, 2003.

Aghion, Philippe, and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica,* LX (1992), 323–351.

Arellano, Manuel, and Olympia Bover, "Another Look at the Instrumental-Variable Estimation of Error-Components Models," *Journal of Econometrics,* LXVIII (1995), 29–52.
Arrow, Kenneth J., "The Economic Implications of Learning-by-Doing," *Review of Economic Studies,* XXIX (1962), 155–173.
Blundell, Richard, and Stephen Bond, "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics,* LXXXVII (1998), 115–144.
Cerda, Rodrigo, "Drugs, Market Size and Population," University of Chicago, Ph.D. thesis, 2003.
Cockburn, Iain, and Rebecca Henderson, "Scale and Scope in Drug Development: Unpacking the Advantages of Size in Pharmaceutical Research," *Journal of Health Economics,* XX (2001), 1033–1057.
Cockburn, Iain, Rebecca Henderson, and Scott Stern, "The Diffusion of Science-Driven Drug Discovery: Organizational Change in Pharmaceutical Research," NBER Working Paper No. 7359, 1999.
Danzon, Patricia M., Sean Nicholson, and Nuno Sousa Pereira, "Productivity in Pharmaceutical-Biotechnology R&D: The Role of Experience in Alliances," NBER Working Paper No. 9615, 2003.
DellaVigna, Stefano, and Joshua Pollet, "Attention, Demographics and the Stock Market," University of California at Berkeley, mimeo, 2004.
DiMasi, Joseph A., Ronald W. Hansen, and Louis Lasagna, "Cost of Innovation in the Pharmaceutical Industry," *Journal of Health Economics,* X (1991), 107–142.
DiMasi, Joseph A., Ronald W. Hansen, and Henry G. Grabowski, "The Price of Innovation: New Estimates of Drug Development Costs," *Journal of Health Economics,* XXII (2003), 151–185.
Drandakis, E., and Edmund Phelps, "A Model of Induced Invention, Growth and Distribution," *Economic Journal,* LXXV (1965), 823–840.
Dranove, David, and David Meltzer, "Do Important Drugs Reach the Market Sooner?" *RAND Journal of Economics,* XXV (1994), 402–423.
Finkelstein, Amy, "Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry," *Quarterly Journal of Economics,* CXIX (2004), 527–564.
Galambos, Louis, and Jeffrey Sturchio, "Pharmaceutical Firms and the Transition to Biotechnology: A Study in Strategic Innovation," *Business History Review,* LXXII (1998), 250–278.
Gambardella, Alfonso, *Science and Innovation: The U. S. Pharmaceutical Industry During the 1980s* (Cambridge, UK: Cambridge University Press, 2000).
Griliches, Zvi, "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica,* XXV (1957), 501–522.
Grossman, Gene, and Elhanan Helpman, *Innovation and Growth in the Global Economy* (Cambridge, MA: MIT Press, 1991).
Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches, "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica,* LII (1984), 909–938.
Hayami, Yujiro, and Vernon Ruttan, "Factor Prices and Technical Change in Agricultural Development: The U. S. and Japan, 1880–1960," *Journal of Political Economy,* LXXVII (1970), 1115–1141.
Henderson, Rebecca, and Iain Cockburn, "Scale, Scope and Spillovers: The Determinants of Research Productivity in Drug Discovery," *Rand Journal of Economics,* XXVII (1996), 32–59.
IMS, "World Review" at www.ims-global.com/insight/report/market_growth/report0600.htm. 2000
Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson, *Patents, Citations, and Innovations* (Cambridge, MA: MIT Press, 2002).
Kennedy, Charles, "Induced Bias in Innovation and the Theory of Distribution," *Economic Journal,* LXXIV (1964), 541–547.
Kling, Jim, "From Hypertension to Angina to Viagra," *Modern Drug Discovery,* I (1998), 31–38.
Kremer, Michael, "Pharmaceuticals and the Developing World," *Journal of Economic Perspectives,* XVI (2002), 67–90.

Lichtenberg, Frank R., "The Allocation of Publicly Funded Biomedical Research," in *Medical Care Output and Productivity, Studies in Income and Wealth,* LXIII, Ernst Berndt and David Cutler, eds. (Chicago, IL: University of Chicago Press, 2001).
——, "Sources of U. S. Longevity Increases, 1960–1997," NBER Working Paper No. 8755, 2002.
——, "The Impact of New Drug Launches on Longevity: Evidence from Longitudinal, Disease-level Data from 52 Countries, 1982–2001," Columbia University, mimeo, 2003.
Lichtenberg, Frank R., and Suchin Virabhak, "Pharmaceutical Embodied Technical Progress, Longevity and Quality of Life: Drugs as Equipment for Your Health," NBER Working Paper No. 9351, 2002.
Lichtenberg, Frank R., and Joel Waldfogel, "Does Misery Love Company? Evidence from Pharmaceutical Markets Before and After the Orphan Drug Act," NBER Working Paper No. 9750, 2003.
Ling, Davina C., Ernst Berndt, and Richard G. Frank, "General Purpose Technologies, Capital-Skill Complementarity, and the Diffusion of New Psychotropic Medications among Medicaid Populations," Massachusetts Institute of Technology, mimeo, 2003.
Malerba, Franco, and Luigi Orsenigo, "Innovation and Market Structure in the Dynamics of the Pharmaceutical Industry: Towards a History Friendly Model," *Industrial and Corporate Change,* XI (2002), 667–703.
Newell, Richard, Adam Jaffe, and Robert Stavins, "The Induced Innovation Hypothesis and Energy-Saving Technological Change," *Quarterly Journal of Economics,* CXIV (1999), 907–940.
Pakes, Ariel, and Zvi Griliches, "Patents and R&D at the Firm Level: A First Look," *Economics Letters,* V (1980), 377–381.
Pakes, Ariel, and Mark Schankerman, "An Explanation into the Determinants of Research Intensity," in *R&D, Patents and Productivity,* Zvi Griliches, ed. (Chicago, IL: University of Chicago Press, 1984).
Popp, David, "Induced Innovation and Energy Prices," *American Economic Review,* XCII (2002), 160–180.
Reiffen, David, and Michael R. Ward, "Generic Drug Industry Dynamics," *Review of Economics and Statistics* (2004).
Romer, Paul M., "Endogenous Technological Change," *Journal of Political Economy,* XCVIII (1990), S71–S102.
Samuelson, Paul, "A Theory of Induced Innovation Along Kennedy-Weisacker Lines," *Review of Economics and Statistics,* XLVII (1965), 444–464.
Schmookler, Jacob, *Invention and Economic Growth* (Cambridge, MA: Harvard University Press, 1966).
Scott Morton, Fiona, "Entry Decisions in the Generic Drug Industry," *Rand Journal,* XXX (1999), 421–440.
Wooldridge, Jeffrey M., "Distribution-free Estimation of Some Nonlinear Panel Data Models," *Journal of Econometrics,* XC (1999), 77–97.
——, *Econometrical Analysis of Cross Section and Panel Data* (Cambridge, MA: MIT Press, 2002).