The Impact of Public Basic Research on Industrial Innovation: Evidence from the Pharmaceutical Industry[1]

by

Andrew A. Toole
Rutgers University
Dept. of Agricultural, Food and Resource Economics
55 Dudley Road
New Brunswick, NJ 08901
Phone: + 1 (732) 932-9155 ext. 215
Fax: +1 (732) 932 8887
E:Mail: toole@aesop.rutgers.edu

Draft: May 9, 2008

[1] I would like to thank John Goddeeris, Jeff Wooldridge, Wally Mullin, F.M. Scherer, Henry Grabowski, Paul Romer, and Tim Bresnahan for their valuable comments on an earlier draft of this paper. Of course, all remaining errors are my own.

ABSTRACT

While most economists are convinced that public basic research is a fundamental driver of industry innovation and economic growth, our understanding of this relationship is surprisingly limited. Using data on basic research investments by the U.S. National Institutes of Health (NIH), this paper investigates the timing and significance of the contribution of public basic research to pharmaceutical innovation. The analysis finds that public basic research makes its primary contribution in the earliest part of the drug discovery stage of industry research, a period called the "drug concept" phase. NIH investments in basic research influence pharmaceutical innovation fourteen to twenty-one years prior to new drug application to the FDA. In terms of magnitude, the analysis finds that a 10% increase in public investment in basic research ultimately leads to a 6.4% increase in the number of new drugs on the market.

1 Introduction

This paper examines the relationship between publicly funded basic research and product innovation in the pharmaceutical industry. In 2004, the United States invested $26.1 billion in basic research (NSF 2007). Historically, federal sources have financed over half of all U.S. national expenditure on basic research and more than double the amount spent by for-profit industry. In the life sciences, the U.S. National Institutes of Health (NIH) is the world's largest public enterprise supporting basic research. In 2004, NIH obligations for basic research were $14.8 billion, which represents about 83% of total federal obligations for basic research in the life sciences. Surprisingly, especially in light of the magnitude of public support for basic research, there are very few empirical studies evaluating the contribution of this research to industrial innovation.

Part of the explanation for this paucity of economic research is related to the fact that academic institutions such as universities, colleges, and other not-for-profit institutions perform the bulk of federally funded basic research, not industry.[2] This fact is recognized in a new generation of models introduced by Jaffe (1989) in which industrial invention or productivity is related to academic research. Working within the production function framework, Jaffe made a fundamental change in the traditional specification by grouping industries and academic departments into technological areas. This is critically important because research externalities are poorly captured using firm or industry boundaries as opposed to technological or scientific areas. Jaffe's analysis found that academic research makes a significant contribution to corporate

[2] According to the NSF (2003, Table 27), industrial firms and industry administered federally funded research and development centers (FFRDCs) have performed about 11% of federally sponsored extramural basic research between 1970 and 2000. Universities, colleges, not-for-profit institutions, and the FFRDCs these institutions manage have performed about 88% of this research.

patenting. Adams (1990) and Acs et al. (1991) subsequently contributed to this line of research.[3] Adams explored the relationship between fundamental knowledge stocks and manufacturing productivity growth. Acs et al. used Jaffe's model to explore how academic research affects an indicator of small business innovation.

This paper complements and extends existing research in two distinct ways. First, the analysis focuses on the pharmaceutical industry. "Drugs and Medical Technology" is one of the technological areas explored in Jaffe's (1989) research. This focus allowed the model to more closely match the innovative process as well as control for specific characteristics that are not shared by other industries. The pharmaceutical industry is a natural choice for analysis because its innovative process is relatively structured and science oriented. The regulatory requirements imposed by the U.S. Food and Drug Administration (FDA) formalize the pharmaceutical innovative process and provide both a common structure across firms and a good measure of product innovation.

Second, the analysis uses new micro-level data on NIH basic research investments that allow pharmaceutical innovation as well as private and public R&D investments to be matched by medical areas. This approach groups drugs and research according to the biological system in question. For example, an antihypertensive drug would be grouped in the cardiovascular area along with private and public research related to the cardiovascular system. Similar to Jaffe (1989), these data make it possible to analyze the relationship between innovation and R&D investments within technology areas over time.

Using a panel database of new drugs observed by medical areas over eighteen years, the results indicate that NIH funded basic research makes a positive and significant contribution to

[3] See McMillan and Hamilton (2003) and Cockburn and Henderson (2001) for a broader literature review. Also, see Branstetter and Ogura (2005) for a recent analysis using patent citation data.

pharmaceutical product innovation. The elasticity estimate in the preferred model implies that a permanent 10% increase in NIH public basic research investment ultimately leads to a 6.4% increase in the number new molecular entities (NMEs), an important category of new drug therapies defined by the FDA. The results also support the hypothesis that publicly funded basic research has its impact in the "drug concept" stage of industry research – the earliest stage of drug discovery preceding compound synthesis. In other words, NIH investments in basic research influence pharmaceutical innovation fourteen to twenty-one years prior to new drug application to the FDA. Further, as argued by Romer (1990) in the endogenous growth literature, there are aggregate long-run increasing returns to scale in pharmaceutical innovation. When both private and public investments are increased proportionately by 1%, there is a 1.44% increase in the number of NMEs after the appropriate time lags.

The rest of the paper adopts the following organization. Section 2 outlines the pharmaceutical innovative process and reviews the research relationships within this process. Section 3 presents the empirical model used in the analysis. This is followed by a description of the data sources and measures in section 4. Section 5 presents the empirical results and discussion. Concluding remarks are found in section 6.

2 Pharmaceutical Product Innovation and Public Basic Research

Innovation in the pharmaceutical industry takes place when private firms introduce new drug therapies into the marketplace. Before any private firm is allowed to sell a new drug therapy in the United States, it must receive approval from the U.S. Food and Drug Administration (FDA). As such, FDA policies and requirements have a fundamental impact on the nature and structure of the pharmaceutical innovative process. The FDA also classifies new pharmaceutical products. One important category is labeled new molecular entities (NMEs).

These new products have never been previously approved for the treatment of disease and are expected to have the largest therapeutic and economic potential.[4] The pharmaceutical innovations analyzed in this paper come from the group of NMEs.[5] This is important to keep in mind. One should not confuse NMEs with other pharmaceutical products that are discovered through long-term experience or post-market clinical observation. While these other products, such as Upjohn's Rogaine cream for hair growth, may be therapeutically and economically important, new indications or uses of approved drugs do not qualify as new additions to the group of new molecular entities.

To understand the role and contribution of public basic research to new drug innovation it is necessary to recognize that the stage of industry research determines the nature of the research problems faced by industry scientists and the nature of the external knowledge relevant to these problems. The structure of the pharmaceutical innovative process is typically described as beginning with drug discovery, progressing to pre-clinical studies, then to human clinical development, and eventually to application to the FDA for approval. This process is relatively uniform and "linear" compared to most other industries due in part to the regulatory requirements imposed by the FDA.

Figure 1 shows a simplified "chain-linked" depiction of the pharmaceutical innovative process (Kline and Rosenberg 1986). Unlike standard depictions, this figure explicitly recognizes the "drug concept" phase of industry research as the very first point in the innovative

[4] Even within this group of potentially important innovations, there are significant differences in actual or realized therapeutic and economic impact. See Scherer and Harhoff (2000) and Grabowski and Vernon (1996) for an analysis of the distribution of sales revenue from NMEs. Cockburn (2005) provides a good overview of the issues.
[5] See Appendix A for a precise description of the NMEs used in the analysis. Because new dosage formulations, diagnostics and so forth are excluded from this analysis but are included in the FDA's category of NMEs, some researchers refer to the innovations analyzed in this paper as "new chemical entities."

process.[6] Evidence from case studies and discussions with industry scientists suggest new drug concepts are a combination of publicly available biomedical knowledge and the firm's own basic research. Public basic research provides a foundation of knowledge which creates both new opportunities for addressing therapeutic outcomes and new information for chemical screening. The new opportunities stem mainly from advances in our understanding of metabolic processes in normal and disease states while, in the chemical screening step, more clearly defined therapeutic targets are combined with structural design methods that utilize computers and automated screening technologies. By monitoring the advances in public basic research, the pharmaceutical industry absorbs and extends public knowledge with an eye toward the ultimate commercial products that may be produced (see Cockburn and Henderson 2001, Gambardella 1995 fro more background).

While there are a number of fascinating examples, the story of captopril illustrates nicely how public basic research can contribute to pharmaceutical innovation. Captopril prevents high blood pressure by inhibiting the conversion of angiotensin I to angiotensin II and it was the first compound in a new class of drugs called angiotensin-converting-enzyme inhibitors or "ACE inhibitors." Its discovery built on two lines of publicly funded research performed in academic settings. The first line involved the identification and description of the renin-angiotensin system. While this public research dates back to at least 1934, the key scientific papers that identified angiotensin I and angiotensin II were published in the mid 1950s. The second line of public research originated in Brazil. Here, research into the cause of death from snake venom identified a natural substance which acts on its victim by fatally lowering blood pressure. In 1965, it was shown that this natural substance blocks the conversion of angiotensin I to

[6] This phase is usually implicit in the drug discovery phase or is sometimes recognized as the "basic research" phase of drug discovery.

angiotensin II. Armed with this public knowledge, the scientists at Squibb were able to synthesize the first ACE inhibitor in the early 1970s. Captopril was subsequently approved for marketing by the FDA in 1981 (for additional cases and discussion see Maxwell and Eckhardt 1990; Scolnick 1990; OTA 1993; Henderson 1994; Cockburn and Henderson 1998, 2001; Gambardella 1995; Silverstein et al. 1995; Dustan et al. 1996).

Captopril was one of twenty-one case histories of important drugs examined by Cockburn and Henderson (1998). Based on their qualitative research, they point out that drug discovery is characterized by a high degree of public and private interaction in research. Figure 1 shows the pharmaceutical innovative process as is a learning process with multiple points of public-private interaction as well as information feedback taking place along the whole innovative chain from drug concept to market and back to drug concept. However, describing and documenting the number, frequency, and modes of interaction is quite difficult. Aside from individual case histories, the Carnegie Mellon Survey on Industrial R&D provides the best insight into how pharmaceutical scientists access and use public research. From their survey of pharmaceutical R&D managers, Cohen et al. (2002) report that public research is quite important as a source for new R&D projects and project completion. In terms of how public knowledge is accessed, the top four mechanisms were publications and reports, meetings and conferences, informal interactions, and consulting. In light of the dynamic interaction and feedback within the pharmaceutical innovative process, the basic research underlying new drug concepts almost certainly involves bi-directional flows of knowledge between public and private research scientists. Since estimating the magnitude of the contribution of public basic research to pharmaceutical innovation is one of the objectives of this research, it is appropriate to interpret

the empirical estimates as capturing a "net flow" of knowledge from public sector research to industry innovation.

Another important aspect of the pharmaceutical innovative process is the length of time from drug concept to FDA application. Since public basic research is expected to make its contribution in the drug concept stage of industry research, the empirical analysis will need to incorporate time lags. DiMasi (2001) provides estimates of the average phase lengths between major milestones based on the detailed data held by the Tufts Center for the Study of Drug Development. These average time lags are shown at the bottom of Figure 1. The average "pre-human testing phase" is three to four years and covers the period from drug synthesis to first testing in humans. The average clinical development period is eight to nine years and covers the period from first testing in humans to the application date to the FDA. Based on these lags, public basic research should influence pharmaceutical drug concept development thirteen years or more before the date of application to the FDA. The empirical analysis will examine the relevant lag between public basic research and NME application. Based on case histories and the structure of the pharmaceutical innovative process, the main hypothesis examined in this paper is:

Hypothesis: Public basic research makes a positive and significant contribution to NME innovation with its primary contribution coming in the drug concept stage of industry research.

3 An Empirical Model of Pharmaceutical Innovation

The empirical objective is to estimate the timing, statistical significance, and magnitude of the contribution of public basic research to pharmaceutical innovation. The modeling approach taken here uses a modified version of the knowledge production function first

implemented by Pakes and Griliches (1984) and used to study knowledge spillovers from academic research to corporate patenting by Jaffe (1989) and to manufacturing innovation by Acs et al. (1991).[7] Following Jaffe (1989), the pharmaceutical production function is specified by technology area instead of by firm. This is an attractive specification because inter-sectoral R&D relationships occur predominantly, albeit not exclusively, between researchers working in the same technology area. To a first approximation, one would expect industry scientists working on cardiovascular drug concepts to interact with academic cardiologists and draw on public biomedical knowledge in cardiology as opposed to other technology areas such as neurology or gastroenterology. This is essentially a technology mapping that, as Griliches (1992) points out, avoids having to make additional assumptions about the "technology distance" between alternative research performers (i.e. industrial and academic scientists).

The technology-level pharmaceutical production relation has the following form:

$$(1)\quad NME_{it} = \exp\left(\alpha_0 + \gamma_{t-k} \sum_{k=1}^{12} \ln(IndRD_{it-k}) + \sum_{j=9}^{25} \delta_{t-j} \ln(PubBas_{it-j}) + \beta_1 \ln(M_{it}) + \varsigma_i + \mu_t \right) + \varepsilon_{it}$$

where i represents one of seven broad medical areas and t represents time over an eighteen year period from 1980 through 1997.[8] NME_{it} is a count of pharmaceutical product innovations in medical area i and FDA application year t. $IndRD_{it-k}$ are aggregate R&D investment flows by private pharmaceutical firms into medical area i beginning in the year before FDA application and extending twelve years back in time. $PubBas_{it-j}$ are public basic research investment flows

[7] The production function framework has been applied to the pharmaceutical industry by Baily (1972), Wiggins (1981), Henderson and Cockburn (1996), and several others.

[8] Equation (1) is a reduced-form representation. This analysis does not attempt to posit a structural model incorporating the particular modes and frequency of interaction between public basic research performers and industry scientists.

into medical area i starting nine years prior to FDA application (the mean date of first testing in humans) and extending back twenty-five years prior to application date. M_{it} is a measure of the potential market size in medical area i at time t based on demographic information; the ς_i 's are time-constant medical area intercepts capturing heterogeneity; μ_t 's are annual time dummy variables capturing shocks common to all areas; and ε_{it} is a random disturbance term. For those explanatory variables appearing in log form, the parameters estimates have a constant elasticity interpretation.

The innovation counts of NMEs by application year and medial area are assumed to follow a Poisson process. However, the Poisson property requiring the conditional mean to equal the conditional variance is overly restrictive. Instead, the analysis draws from the generalized linear models (GLM) literature and assumes a proportional relationship between the conditional mean and conditional variance.[9] This formulation is implemented using the Poisson quasi-maximum likelihood estimator (QMLE). As Wooldridge (2002) describes, the Poisson QMLE has attractive robustness properties. For equation (1), the Poisson QMLE is consistent and efficient as long as the conditional mean is correctly specified, the GLM variance assumption holds, and the model is dynamically complete.[10]

As discussed in section II, interaction and feedback are intrinsic parts of the pharmaceutical innovative process. Feedback in equation (1) takes place when shocks to NME applications at time t (changes in ε_{it}) influence industry and public basic research investment at

---

[9] This assumption implies that the ratio $\frac{Var[NME_{it} \mid \mathbf{X}_{it}]}{E[NME_{it} \mid \mathbf{X}_{it}]} = \sigma^2$ is constant, but allowed to differ by a positive "dispersion" parameter, σ^2 (see Wooldridge 2002 for further details).

[10] Throughout the analysis the time series processes are assumed to be essentially stationary and weakly dependent as defined in Wooldridge (1994). Specification tests show that all three of these assumptions hold for equation (1). The test results are available from the author.

future dates (t+1 and beyond). For the lag structure given in equation (1), shocks to NME applications in 1980 could influence industry and public research investment in 1981, and affect subsequent innovation in 1992 and beyond. When feedback is present, it invalidates the strict exogeneity assumption required for consistency of standard panel data estimators. This means the usual fixed and random effects count data estimators are not appropriately applied to equation (1). An appropriate estimator will allow for feedback as well as address the time-constant unobserved heterogeneity across medical areas represented by the ς_i 's in equation (1).

The approach taken here is to allow for feedback by using the pooled Poisson QMLE and to account for correlated heterogeneity using pre-sample information as suggested by Blundell, Griffith and Van Reenen (1995).[11] Blundell, Griffith and Windmeijer (2002) present theoretical and Monte Carlo results showing this approach works well in count data models when the sample size is small and the explanatory variables are highly persistent.[12] The sample available for this analysis is quite small (126 observations) and both the industry and public basic research investment flows are very persistent. This approach exploits the fact that there is a long pre-sample history of pharmaceutical NME innovation. In the empirical analysis, the pre-sample averages of NMEs for each medical area are calculated using data from 1964 to 1979. The central nervous system medical area has the highest pre-sample mean of 4.94 drugs per year while the dermatologic area has the lowest with 0.44 drugs per year.

Holding industry R&D investment constant, the pre-sample mean of NMEs captures the industry's unobserved heterogeneity across medical areas affecting future NME innovation.

[11] Pooled Poisson QMLE does not require strict exogeneity for consistency, but allows for arbitrary feedback over time.

[12] A quasi-differenced GMM approach is another alternative. However, their Monte Carlo evidence shows this approach performs poorly in small samples with highly persistent regressors. That result was confirmed with the data used in this paper. See Blundell, Griffith, Van Reenen (1999) for another application of the "pre-sample mean estimator" to account for correlated fixed effects.

Successful past innovation is likely to fuel the development of "capabilities" or "competencies" which influence future innovation. Typically, these competencies are seen as productive resources that stimulate successful innovation in the future. However, in a series of papers, Rebecca Henderson and Iain Cockburn have argued that the pharmaceutical industry underwent a fundamental shift in the late 1970s from a "random" screening mode to a "rational" drug design mode of innovation (Henderson 1994, Henderson and Cockburn 1994, 1996). To the extent that rational drug design methods required the industry to develop new competencies to facilitate innovation, past competencies may be negatively related to future innovation. Given that pre-sample NMEs capture unobserved competencies through 1979 in equation (1), the sign of the coefficient estimate will reflect the net influence of these separate possibilities.

The burden of disease, captured in this analysis using pre-sample mortality rates across medical areas, plays an important and somewhat complicated role in the pharmaceutical innovative process.[13] First, it is a source of heterogeneity because each of the medical areas in equation (1) is an aggregate over a different number of unobserved research programs addressing specific disease conditions.[14] Since both public and private R&D investments are allocated across disease conditions based (in part) on mortality rates, those medical areas with greater mortality rates will also be the medical areas with a greater number of active research programs. Consequently, heterogeneity in mortality rates across medical areas will reflect differences in the "complexity" of research in these areas since aggregating over a larger number of interconnected research programs creates a more diverse knowledge base that is harder to navigate.

Second, heterogeneity in mortality rates is correlated with the key explanatory variables

[13] The empirical specification includes pre-sample mortality rates observed in 1979 to account for heterogeneity in the burden of disease. In 1979, the cardiovascular medical area had the highest mortality rate of 4.24 deaths per thousand people and dermatologic had lowest rate of 0.010 deaths per thousand people.
[14] See Henderson and Cockburn (1996) for an analysis of industry patenting and R&D at the research program level.

in equation (1). Greater mortality from disease in a medical area signals greater social need and (presumably) a greater willingness to pay for new drug therapies. All else constant, this should stimulate more private and public R&D investment. Prior research shows that both industry R&D investment and public biomedical research investment are positively associated with mortality rates (Toole 2007, Gross et al. 1999).

The relationship between mortality and potential market size is more complicated. Lichtenberg (2003) finds that pharmaceutical innovation is associated with reduced mortality. If new drug therapies lower mortality, successful innovation will reduce future market size for NMEs as long as greater life expectancy does not lead to greater need for NMEs drug therapies going forward. Here the distinction between NMEs and other drugs products is important to keep in mind because longer life expectancy could increase the demand for non-NME drug therapies while simultaneously reducing the demand for NMEs. On the other hand, if extending life creates future need for NMEs, lowering mortality could increase the future market size and re-enforce any beneficial demographic changes. In any case, the complex interaction between mortality rates and the potential market for pharmaceutical innovation highlights the importance of controlling pre-sample heterogeneity in mortality rates. Holding mortality rates constant will also help identify the influence of demographic changes in market size as highlighted by Acemoglu and Linn (2004).

The specification of equation (1) incorporates the characteristic lags in the pharmaceutical innovative process. Industry research investment flows extend over twelve years, which is the mean period from compound synthesis to FDA application. For the regression results presented in section V, these flows are aggregated into an industry stock dated one year prior to application (t-1). The stock is constructed using the perpetual inventory method

described by Hall et al. (1988) assuming a twenty percent depreciation rate (δ=0.20).[15] The R&D industry stock for NMEs applications at time t is given by:

$$IndStk_{t-1} = IndRD_{t-1} + \sum_{j=1}^{11} (1-\delta)^j \, IndRD_{t-j-1}$$

The time period over which public basic research is allowed to influence NME innovation is closely tied to the lag structure of the pharmaceutical innovative process. The mean lag from the first testing of a compound in humans to the FDA application date is eight to nine years. Since the drug candidate is already "in hand" at this point and the conduct of clinical trials is already determined in clinical trial protocols, the influence of public basic research on NME innovation must come before this point in time. Equation (1) allows public basic research to influence innovation from nine years prior to application to twenty-five years prior to application.[16] Since no prior information on the form of the lag between public basic research and NME innovation is available, the empirical specification leaves the public investment flows completely unrestricted. While this approach avoids mis-specifying the structure of the lag distribution, it does suffer from the common problems associated with multicollinearity in knowledge production function style models, namely unappealing sign changes on the coefficients across years and low precision due to inflated standard errors.[17] However, as Wooldridge points out, even when individual lag coefficients are not precisely estimated, the estimate of the long-run impact is often quite good (Wooldridge 2009, p 345). To test the hypothesis about the contribution of public basic research, the analysis will rely mainly on the

[15] This is a geometric form of deprecation for industry investment that assumes it loses its "productive capacity" at a constant rate of 20% per year. The results are robust to using the annual flows of industry R&D which do not impose any structure on the lag distribution.
[16] An earlier version of this research specified public basic research as a stock using the perpetual inventory method (Toole 2000). That alternative works fine, however, identifying the lag between public basic research and pharmaceutical innovation required a different approach. The earlier version also imposed the strict exogeneity assumption by using the fixed effects Poisson estimator.
[17] See Cincera (1997), Hall, Griliches and Hausman (1986), Pakes and Griliches (1984) for further background.

statistical significance of the long-run impact as measured by the sum of the annual elasticity estimates over the lag distribution.

The annual time dummy variables capture shocks common to all the medical areas in the pharmaceutical innovative process. For instance, over the sample period from 1980 to 1997, there were several legislative changes that may have affected pharmaceutical innovation such as the Orphan Drug Act of 1983, the 1984 Patent Term Restoration and Competition Act, the Prescription Drug User Fee Act of 1992, and the 1997 FDA Modernization Act. These legislative changes may have influenced when firms submitted their new drug applications to the FDA. The annual dummy variables will account for the common influence of these events across medical areas on NME innovation.

4 Data Sources and Measurement

In order to estimate the timing and significance of the contribution of public basic research to pharmaceutical innovation, I gathered data from six sources: the Food and Drug Administration (FDA), the Pharmaceutical Research and Manufacturers Association, the National Institutes of Health, the Centers for Disease Control (CDC), the U.S. Census Bureau, and the Current Population Survey. These data are organized into a panel dataset comprised of seven medical areas observed over an eighteen year period. A medical area is a technological area that groups compounds and research related to similar biological systems. The pharmaceutical industry has reported its R&D expenditure by medical areas since the early 1960s. The seven medical areas considered in this study are: endocrine/neoplasm (cancer), central nervous system, cardiovascular, gastro-intestinal/genito-urinary, anti-infective, dermatologic, and respiratory. The remainder of this section provides an overview of the data used in the analysis while the details about the database construction are provided in Appendix

A. Table 1 presents descriptive statistics of the variables used in the regression analysis.

As described earlier, pharmaceutical innovation is measured as a count of approved new molecular entities (NMEs) based on the data obtained from the FDA. The NMEs analyzed in this paper are a subgroup of all approved NMEs and are sometimes referred to as "new chemical entities." This is a category of unique therapeutic products introduced by the industry each year and includes the captopril compound discussed in section 2. Approved NMEs are also a measure of important patents since they are the subset of all patented pharmaceutical compounds that have met the FDA standards for safety and efficacy. Using a measure of important patents addresses one of the criticisms of using patent counts as an indicator of innovative output - the wide variation in the economic value of individual patents.[18]

Public data on pharmaceutical industry R&D investment were collected from the Pharmaceutical Research and Manufacturers Association's (PhRMA) Annual Survey Reports (various years). These public data are only available by medical area after aggregation across individual PhRMA member firms. The investment figures are also aggregated across both the drug discovery/pre-clinical and the clinical development phases. These figures include expenditure of research failures as well as successes and cover all pharmaceutical investment whether directed toward an NME or toward a non-NME drug product (roughly 80% of these funds go toward NMEs). These figures are a composite measure of industry R&D investment including physical capital, labor costs, manufacturing costs, and a fair share of overhead. The industry investment flows were adjusted for inflation using the Biomedical Research and Development Price Index with 2000 as the base year. For the regression model, the flows are cumulated into a twelve year stock as described in section 3.

[18] There is still significant variation in the therapeutic and economic impact of these NMEs. Refer to the references in footnote #1 for more background.

Proxy variables for public basic research knowledge were constructed using data on extramural funding by the U.S. National Institutes of Health (NIH). Comprehensive award level data allowed the construction of basic research flows by medical area from 1955 through 1996. The NIH funds a large and diverse body of construction, education, training, research, and development activities. Detailed activity and program codes, scientific review group codes, and funding agency identifiers are used to isolate the relevant basic research funding. To facilitate this process, I asked industry scientists to describe the nature of the basic research relevant to the pharmaceutical innovative process. They identified research directed at the discovery and characterizations of physiologically active substances and the definition of metabolic pathways related to normal and disease function. This definition covers only discovery/pre-clinical research. While some important new drugs are discovered through casual observation or directed clinical research, these are not the drug products considered in this analysis since new uses of post-approval drugs do not quality as new molecular entities. The NIH investment flows were adjusted for inflation using the Biomedical Research and Development Price Index with 2000 as the base year.

The variable measuring the potential market size for NME innovation was constructed based on the methodology used in Acemoglu and Linn (2005). Data on U.S. mortality and hospital admissions were obtained from the U.S. Centers for Disease Control for the pre-sample year 1979 and grouped into medical areas based on the International Classification of Diseases, 9th Clinical Modification. The 1979 data identified an "at risk" group in the U.S. population by age and gender for each of these medical areas. The medical area shares of the "at risk" group were fixed at the 1979 levels. Using data from the U.S. Census and the Current Population Survey, population and income data were collected by age and gender for each year in the

sample (1980-1997) and apportioned across medical areas using the 1979 shares. The main advantage of this method is that the market size variable is arguably strictly exogenous since NME innovations cannot feedback to change the size of the at risk group by affecting the burden of disease, at least as it is measured by mortality and hospital admissions. One disadvantage of this method is that there is very little cross-sectional variation and identifying the influence of demographic changes requires the pre-sample heterogeneity in the "at risk" shares to be held constant.

5 Estimation Results

This section presents the Poisson QMLE estimation results for equation (1). The discussion is organized into four parts. The first part examines the time lag between public basic research investment and pharmaceutical innovation. As discussed in Section 2, public basic research is expected to influence the drug concept stage of the pharmaceutical innovative process, which takes place in the earliest part of drug discovery. As such, public basic research should impact pharmaceutical innovation sometime earlier than twelve years prior to NME application – before compound synthesis takes place. The second part examines the statistical results from the regression models in greater detail. The third part examines the economic importance of public basic research based on the magnitude of its long-run contribution as given by the sum of the annual elasticity estimates over the full length of the lag distribution. The final part presents some alternative specifications as robustness checks.

Table 2 presents the main findings for the analysis. The table shows three versions of equation (1) which only differ in the number of public basic research lags included in the specifications. Since these annual flows are highly collinear and enter the specification in logs, the individual coefficient estimates generally show sign changes and low individual statistical

significance. Nevertheless, in all three specifications public basic research investment nineteen years prior to NME application is statistically significant. The preferred model, which tightens the lag specification using robust Wald tests, appears in model (3) in column 4.[19] This specification identifies a range of eight years over which public basic influences pharmaceutical innovation. This group of years is highly significant with a robust Wald statistic of $X^2(6)=468.53$ and a p-value<0.0000. The estimates suggest that NIH public basic research investment fourteen to twenty-one years prior to NME application contribute to pharmaceutical innovation. This period is consistent with the hypothesis that public basic research contributes to the drug concept stage of pharmaceutical innovation as described in section 2. It is also consistent with available evidence from individual drug case histories. The average lag between the initial public sector "enabling discovery" and FDA application was 20.5 years for the fourteen drugs studied by Cockburn and Henderson (1998).[20]

Notwithstanding the well known difficulties of estimating precise lag relationships, it is still informative to probe the lag structure a little deeper in order to gain suggestive insights into the interaction between public basic research and private pharmaceutical innovation. Figure 2 plots the lag coefficient estimates for public basic research found in model (3).[21] The lag distribution shows that the number of NME applications in year t depends on two "peaks" of public basic research influence occurring at fifteen and nineteen years before NME application. This pattern is consistent with two related facts. First, the period of time separating the peaks at

[19] Moving from model (1) to model (2) imposes exclusion restrictions on public basic research lags PubBas(t-9) through PubBas(t-13). The robust Wald statistic for joint significance is $X^2(5)=0.39$ with a p-value<0.9955. Moving from model (2) to model (3) imposes exclusion restrictions on public basic research lag PubBas(t-22) through PubBas(t-25). The robust Wald statistic for joint significance is $X^2(4)=4.94$ with a p-value<0.2037.
[20] Cockburn and Henderson (1998) report the average lag from enabling discovery to FDA approval to be 23.3 years. The 20.5 year period ends at FDA application, not approval. This estimate was calculated by subtracting the average FDA review period for drugs approved in the 1980s, which was 2.8 years as reported by DiMasi (2001).
[21] This is the mean lag distribution from pooling all medical areas. While it is likely that public basic research in each medical area has a somewhat different lag relationship with pharmaceutical innovation, the small sample size necessitates restricting the lag coefficients across medical areas in the empirical analysis.

fifteen and nineteen years is consistent with the average duration of NIH competing research project grants, which is four years. Second, prior research analyzing how industry R&D investment responds to NIH basic research funding also found a U-shaped time profile (Toole 2007). Because equation (1) characterizes a dynamic relationship, this lag profile gets shifted forward for NME applications in year t+1 and so forth. Consequently, a dynamic view across a number of NME application years suggests overlapping "waves" of influence in which private industry scientists are monitoring, responding, and interacting with academic basic researchers and/or their research findings on a continuous basis.

Although the empirical results presented here cannot pin down the shape of the lag distribution precisely, the peaks in the lag distribution may represent points in time when public basic research contributes significant information to industry innovation, especially since the underlying flows are quite smooth. If these are real effects, one would expect to see a consistent relationship over time between the public basic research investment and the NME applications. Figures 3 and 4 illustrate this relationship over the sample period with the number of NME applications on the right-hand axis and lagged real NIH investment into basic research on the left-hand axis. Figure 3 shows a striking similarity in the co-movements of the trends in NIH funding lagged fifteen years prior to NME application and the observed number of NME applications. For Figure 4, which uses NIH funding lagged nineteen years, the co-movements in these trends are not as tight, but are still suggestive. So, while the nature of the interaction between public and private researchers remains complex and largely unknown, these exploratory results provide some suggestive insights into the time profile of the interaction between public basic research and private pharmaceutical innovation.

For the statistical results, all three regression models in Table 2 show similar findings.

The long-run contribution of public basic research, which is given by the sum of the annual elasticity estimates over the full length of the lag distribution, is positive and highly significant in all three regression models (p-values<0.000 as reported at the bottom of Table 2). The public basic research flows are also jointly significant in model (3) once the specification of the lag distribution is tightened. Among the other covariates, the elasticity of NMEs with respect to the stock of industry R&D is highly significant with a p-value<0.000 in all models. The potential market size variable, which accounts for income and population changes within medical areas, is positive and significant in all the regression models in Table 1. Both pre-sample heterogeneity in mortality rates and NMEs are negative and significant.[22] As described in section 2, heterogeneity in mortality rates is likely capturing differences in "complexity" across medical areas stemming from aggregation over different numbers of interconnected research programs. The negative effect of past NMEs supports the view elaborated by Henderson (1994) and Henderson and Cockburn (1994) that the shift to a rational drug design mode of drug discovery in the late 1970s required the development of new "integrative competences" for successful pharmaceutical innovation.

Regarding the estimated magnitude of private R&D investment, model (3) shows the elasticity of NME applications with respect to the stock of industry R&D is about 0.81. Holding all else constant and using the sample averages, this estimate implies a marginal physical product of 0.0027 NMEs per $1 million and a marginal cost per NME of $370 million in real 2000 dollars.[23] In the literature, DiMasi et al. (1991, 2003) found the average out-of-pocket cost per

[22] The control for pre-sample heterogeneity in hospital admission rates across medical areas was always insignificant.

[23] The marginal product calculated as: mp = elasticity x (avg. NME / avg. industry flows). The industry R&D flows assume a twelve year investment period and are adjusted by a proportionality factor of 0.80 to account for industry investment directed at NMEs as opposed to product extensions. The proportionality factor is taken from the PhRMA annual survey reports.

NME increased from $251 to $405 million (real 2000 dollars) between the two sample periods they analyzed.[24] Although their sample periods do not exactly match the sample period used here, the $370 million marginal cost figure falls within the expected range.

Notwithstanding the appreciable lag, public basic research makes an economically significant contribution to pharmaceutical product innovation. The long-run elasticity of NMEs with respect to public basic research ranges from 0.72 down to 0.64 as the lag distribution is tightened across the models in Table 2. Based on model (3), a permanent 10% increase in NIH public basic research funding will lead to a 6.4% increase in NME applications after twenty-one years. Holding all else constant and using the sample average of the NIH investment flows over the lag period, this estimate implies a marginal physical product of 0.008 NMEs per $1 million in real 2000 dollars.[25] Interestingly, this is the same marginal physical product (rounded to three decimals) found in earlier work which used a stock specification for public basic research and a nineteen year lag (Toole 2000).

Because the new ideas derived from public basic research are largely non-rivalrous and non-excludable, the magnitude of its contribution to pharmaceutical innovation has an important implication – there are increasing returns scale in NME innovation. Romer (1990) highlights how externalities in research and development can lead to increasing returns to scale in the context of endogenous growth models. Holding all else constant, the parameter estimates indicate that a proportional increase in both private R&D and public basic research leads to a more than proportionate increase in NME applications (that are ultimately approved). From

[24] In both studies, the authors define the population for their sample based on when the NMEs entered clinical trials. In their earlier study, which covers the period 1970 to 1982, the average out-of-pocket cost was given as $114 million in real 1987 dollars. The figure in the text was calculated by readjusting the base to the year 2000 using the Bureau of Labor Statistics producer price index (series ID WPU035). Their second study covered the period from 1983 to 2000.

[25] The marginal product calculated as: mp = elasticity x (avg. NME / avg. NIH flows). The NIH basic research flows correspond to the years in the lag distribution, fourteen to twenty-one years prior to NME application.

model (3) in Table 1, the sum of the coefficient estimates is 1.44 with a standard error of 0.168, with a p-value<0.000.

Table 3 reports four alternative specifications as robustness checks.[26] The specifications in columns (1) and (2) drop the medical areas with the smallest and largest NME applications, respectively. As seen in column (1), there are no notable changes due to dropping the smallest area (dermatologic NMEs). However, as seen in column (2), dropping the largest area (anti-infective NMEs) leads to three notable changes. First, the lag distribution for public basic research is different and begins later at seventeen years prior to NME application. The earlier years, lags 14-16, became jointly insignificant with a robust Wald statistic of $X^2(3)=1.01$ and a p-value<0.7999. The remaining lag coefficients are jointly significant with a robust Wald statistic of $X^2(5)=11.97$ and a p-value<0.0352. Drug innovations in the anti-infective medical area appear to have a shorter lag relationship with public basic research. This highlights the fact that the lag distribution found in Table 2 is not uniform across medical areas, but represents the mean lag across these areas. Second, the long-run elasticity of public basic research is no longer statistically significant. This is not especially surprising since dropping the anti-infective class amounts to dropping the medical area with the greatest variation in the dependent variable. The sample size for this analysis is too small to be robust to deleting one of the major medical areas. Third, heterogeneity as captured by the pre-sample mortality rates and average NMEs are no longer statistically significant.

The regression results in columns (3) and (4) of Table 2 show a test for endogeneity of industry R&D investment and report an OLS version of equation (1) based on the approach introduced by Pakes and Griliches (1980) for estimating a log-linear model when the dependent

[26] The specifications shown in Toole (2000) are additional robustness checks.

variable has zero values.[27] The number of hospital admissions and average length of hospital stay are used as instruments for industry R&D. As seen by the t-statistic on the first stage residuals in column (3), there is no evidence of endogeneity (see Wooldridge 2002 for details on the test).[28] For the OLS model in column (4), the public basic research coefficients are jointly significant with an F(6,6)=23.53 and a p-value<0.006. The long-run elasticity, however, is not significant in the OLS version.

6 Conclusion

This paper analyzed the relationship between publicly funded basic research and product innovation in the pharmaceutical industry. The empirical results are based on a panel data set of FDA approved new molecular entities grouped by medical area over the period 1980-1997. The analysis found that public basic research contributes to the drug concept stage of pharmaceutical drug discovery drug, which necessarily precedes compound synthesis. NIH investments in basic research influence pharmaceutical innovation fourteen to twenty-one years prior to new drug application to the FDA. The findings also suggest that a permanent 10% increase in public basic research investment leads to a 6.4% increase in the number of NMEs after the appropriate lags. This impact is economically significant and the combination of both public and private R&D investment leads to increasing returns to scale in the pharmaceutical innovative process.

Because of the unique structure of the pharmaceutical innovative process, it is difficult to generalize these findings to the other technological areas. However, future research should be

[27] This approach sets the log of NMEs equal to zero when there are no NME applications in a particular medical area and year and includes a dummy variable for these observations on the right-hand side of the equation. While this method is not efficient and is likely to be biased by the functional relationship between NMEs and the dummy variable, it provides a check on the significance of public basic research.

[28] Both of these instruments are insignificant when included in equation (1), but are individually and jointly significant in the first stage with an F(2,6)= 9.40 an a p-value<0.0142.

directed at improving our understanding of innovation and R&D investment in these other areas. Jaffe's (1989) research using multiple technology areas identifies some promising candidates.

Appendix A: Database Construction

Innovation Measure: Approved New Molecular Entities (NMEs)

Using the Freedom of Information Act, information on all approved new molecular entities (NMEs) was obtained from the U.S. Food and Drug Administration (FDA) for the period 1964 to 2000. Because the FDA defines this product category more broadly than the focus of this study, the definition of NMEs used here excludes diagnostic agents, certain biological agents, new dosage formulations, surgical and other materials such as contact lenses and devices. Some researchers, such as DiMasi et al. (1991, 2003), refer to this more narrowly defined group as new chemical entities. The remaining approved products were grouped by year of application and medical area. Although the year of application was supplied by the FDA, the compounds needed to be assigned to one of the seven medical therapeutic areas. This was accomplished by using the clinical pharmacology and treatment indication descriptions from the Physician's Desk Reference, Merck Index, and Martindale The Extra Pharmacopoeia. The final group of pharmaceutical innovations includes both self-originated and in-licensed compounds. In cases where empirical observation has revealed additional uses of a compound, such as Upjohn's Loniten (minoxidil) having the beneficial effect of hair growth, the compound was classified in its original therapeutic area (cardiovascular in the case of minoxidil).

Public Life Science Basic Research Investment

The National Science Foundation identifies the U.S. National Institutes of Health (NIH) as the primary funding agency for U.S. basic research investment into the life sciences (83% in 2004). Two data sources on NIH funded grants and contracts were obtained to estimate public basic research investment into each of the seven medical areas over time. First, an extract from the NIH IMPAC database provided data covering the years 1955 to 1985. Second, the NIH CRISP database provided data covering the years 1972 to 1996. For each grant and contract award, the data includes: the title, the identification number (activity code, institute code, and grant or contract number), the fiscal year of award, the award amount, and the scientific review group that recommended its approval. Beginning with the total set of awards for an individual year, the annual basic research flows were estimated using several data filters. The filters were designed to isolate basic research that is relevant to the pharmaceutical industry as defined in the text. Four main filter levels were defined and used. The first eliminated awards based on activity code. Generally speaking, this filter purged activities such as training, education, construction, demonstration, and institutional block grants from the data. The second filter eliminated awards based on institute code. All institutes and divisions were analyzed, however, there turned out to be several that contributed nothing or only a negligible amount to basic science relevant to the pharmaceutical industry. Examples of eliminated institutes include the National Institute on Dental Research, the National Institute on Environmental Health Sciences, the National Library of Medicine, and several others. The third filter eliminated scientific review groups that consider research that did not fit the definition of basic research. These groups were reviewed in every year and the filter is modified to be year specific. (This was necessary due to the splitting, adding, and discontinuance of scientific review groups over time.) The fourth filter level, which is really a group of filters, analyzed awards based on medical keywords - but not simply disease names - contained in the title of the grant or contract. The keyword filters were defined using the prefixes and suffixes commonly found in medical vocabulary. This was done to avoid the pitfalls encountered when one simply searches titles and abstracts that mention specific diseases. Overall, the four main filter levels separated the total universe of NIH grants and contracts into seven therapeutic areas of basic research, one group of unclassifiable basic research, and a remainder covering all other funded activities. The unclassifiable but relevant research was allocated using the shares of total counted research in each category. For a medical area total, the awards are summed across all grants. Annual totals were adjusted for price changes by using the Biomedical Research and Development Price Index maintained by the Bureau of Labor Statistics (base year 2000).

REFERENCES

Acemoglu, D., and J. Linn (2004). "Market size in innovation: Theory and evidence from the pharmaceutical industry," Quarterly Journal of Economic, 119(3), 1049-1090.

Acs, Z.J., D.B. Audretsch, and M.P. Feldman (1991). "Real effects of academic research: comment," American Economic Review, March 1991, 363-367.

Adams, J.D. (1990). "Fundamental stocks of knowledge and productivity growth," Journal of Political Economy, 98(4), 673-702.

Baily, M.N. (1972). "Research and development costs and returns: The U.S. pharmaceutical industry," Journal of Political Economy, 80(1), 70-85.

Blundell, R., R. Griffith, and J. Van Reenen (1995). "Dynamic count data models of technological innovation," The Economic Journal, 105(429), 333-344.

Blundell, R., R. Griffith, and J. Van Reenen (1999). "Market share, market value and innovation in a panel of British manufacturing firms," The Review of Economics and Studies, 66(3), 529-554.

Blundell, R., R. Griffith, and F. Windmeijer (2002). "Individual effects and dynamics in count data models," Journal of Econometrics, 108, 113-131.

Branstetter, L., Y. Ogura (2005). "Is academic science driving a surge in industrial innovation? Evidence from patent citations," NBER working paper no. 11561.

Cincera, M. (1997). "Patents, R&D, and technological spillovers at the firm level: Some evidence from econometric count models for panel data," Journal of Applied Econometrics, 12, 265-280.

Cockburn, I. (2006). "Is the pharmaceutical industry in a productivity crisis?" in Innovation Policy and the Economy, Volume 7, A. Jaffe, J. Lerner, and S. Stern (eds.), Cambridge, MA: MIT Press, 2006.

Cockburn, I., and R. Henderson (1998). "Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery, Journal of Industrial Economics, 156(2), 157-182.

Cockburn, I., and R. Henderson (2001). "Publicly funded science and the productivity of the pharmaceutical industry," in Innovation Policy and the Economy, Volume 1, A. Jaffe, J. Lerner, S. Stern (eds.), Cambridge, MA: MIT Press, 2001.

Cohen, W.M., R.R. Nelson, and J.P. Walsh (2002). "Links and impacts: The influence of public research on industrial R&D," Management Science, 48(1), 1-23.

David, P., D. Mowery, and E.W. Steinmueller (1992). "Analyzing the economic payoffs from basic research," Economics of Innovation and New Technology, 2, 73-90.

DiMasi, J.A. (2001). "New drug development in the United States from 1963 to 1999," Clinical Pharmacology & Therapeutics, 69(5), 286-296.

DiMasi, J.A., R.W. Hansen, and H.G. Grabowski (1991). "Cost of innovation in the pharmaceutical industry," Journal of Health Economics, 10, 107-142.

DiMasi, J.A., R.W. Hansen, and H.G. Grabowski (2003). "The price of innovation: new estimates of drug development costs," Journal of Health Economics, 22, 151-185.

Dustan, H.P., E.J. Roccella, and H.H. Garrison (1996). "Controlling hypertension," Archives of Internal Medicine, 156, 1926-1935.

Gambardella, A. (1995). Science and Innovation: The US pharmaceutical industry during the 1980s, Cambridge: Cambridge University Press, 1995.

Grabowski, H., and J.M. Vernon (1996). "Prospects for returns to pharmaceutical R&D under health care reform," in Competitive Strategies in the Pharmaceutical Industry, Robert B. Helms (editor), Washington, DC: The AEI Press, 1996.

Graves, S.B., and N.S. Langowitz (1993). "Innovative productivity and returns to scale in the pharmaceutical industry," Strategic Management Journal, 14, 593-605.

Griliches, Z. (1992). "The search for R&D spillovers," Scandinavian Journal of Economics, 94, Supplement, S29-S47.

Gross, C.P., G.F. Anderson, and N.R. Powe (1999). "The relation between funding by the National Institutes of Health and the burden of disease," The New England Journal of Medicine, 340(24), 1881-1887.

Hall, B.H., Z. Griliches, J.A. Hausman (1986). "Patents and R and D: Is there a lag?" International Economic Review, 27(2), 265-283.

Hall, B.H., C. Cummins, E.S. Laderman, and J. Mundy (1988). "The R&D master file documentation," NBER technical working paper no. 72, December 1988.

Henderson, R. (1994). "The evolution of integrative capability: Innovation in cardiovascular drug discovery," Industrial and Corporate Change, 3(3), 607-627.

Henderson, R., and I. Cockburn (1994). "Measuring competence? Exploring firm effects in pharmaceutical research," Strategic Management Journal, 15, 63-84.

Henderson, R., and I. Cockburn (1996). "Scale, scope and spillovers: The determinants of research productivity in the pharmaceutical industry," RAND Journal of Economics, 27(1), 32-59.

Jaffe, A.B. (1989). "Real effects of academic research," American Economic Review, 79(5), 957-970.

Kline, S.J., and N. Rosenberg (1986). "An overview of innovation," in The Positive Sum Strategy: Harnessing Technology for Economic Growth, Ralph Landau and Nathan Rosenberg (editors), National Academies Press, Washington DC.

Lichtenberg, F. (2003). "Pharmaceutical innovation, mortality reduction, and economic growth," in Measuring the Gains from Medical Research: An Economic Approach, K.M. Murphy and R.H. Topel (eds.), Chicago, IL: University of Chicago Press, 2003.

Mansfield, E. (1991). "Academic research and industrial innovation," Research Policy, 20, 1-12.

Mansfield, E. (1995). "Academic research underlying industrial innovations: Sources, characteristics, and financing," Review of Economics and Statistics, 55-65.

McMillan, G.S., and R.D. Hamilton III (2003). "The impact of publicly funded basic research: An integrative extension of Martin and Salter," IEEE Transactions on Engineering Management, 50(2), 184-191.

Maxwell, R.A., and S. Eckhardt (1990). Drug Discovery: A Case Book and Analysis, Clifton, NJ: Humana Press, 1990.

National Science Foundation, Division of Science Resources Studies. (2003) Federal Funds for Research and Development, Detailed Historical Tables: Fiscal Years 1951-2002. NSF 07-325. Arlington, VA.

National Science Foundation, Division of Science Resources Statistics. (2007) Federal Funds for Research and Development: Fiscal Years 2004-2006. NSF 07-323. John E. Jankowski, project officer. Arlington, VA.

Pakes, A., and Z. Griliches (1984). "Patents and R&D at the firm level: A first look," in R&D, Patents, and Productivity, Z. Griliches (ed.), Chicago: Chicago University Press, 1984.

Pharmaceutical Research and Manufacturers Association, "Annual Survey Report," PhRMA, Washington D.C., various years.

Romer, P.M. (1990) "Endogenous technological change," Journal of Political Economy, 98, S71-S102.

Scherer, F.M., D. Harhoff (2000). "Technology policy for a world of skew-distributed outcomes," Research Policy, 29, 559-566.

Silverstein, S.C., H.H. Garrison, and S.J. Heinig (1995). "A few basic economic facts about research in the medical and related life sciences," The FASEB Journal, 9, 833-840.

Toole, A.A. (2000). "The impact of public basic research on industrial innovation: Evidence from the pharmaceutical industry, SIEPR discussion paper no. 00-07.

U.S. Congress, Office of Technology Assessment, Pharmaceutical R&D: Costs, Risks, and Rewards, OTA-H-522 (Washington, DC: U.S. Government Printing Office, February,1993).

Wiggins, S.N. (1981). "Product quality regulation and drug introductions," Review of Economics and Statistics, 63, 615-619.

Wooldridge, J.M. (1994). "Estimation and inference for dependent processes," in the Handbook of Econometrics, Volume IV, R.F. Engle and D.L. McFadden (eds.), North-Holland, 1994.

Wooldridge, J.M. (2002). Econometric Analysis of Cross Section and Panel Data, Cambridge, MA: MIT Press, 2002.

Wooldridge, J.M. (2009). Introductory Econometrics: A Modern Approach, 4th edition, Mason, OH: South-Western Cengage Learning.

Figure 1: Simplified "Chain-linked" Depiction of the Pharmaceutical Innovative Process

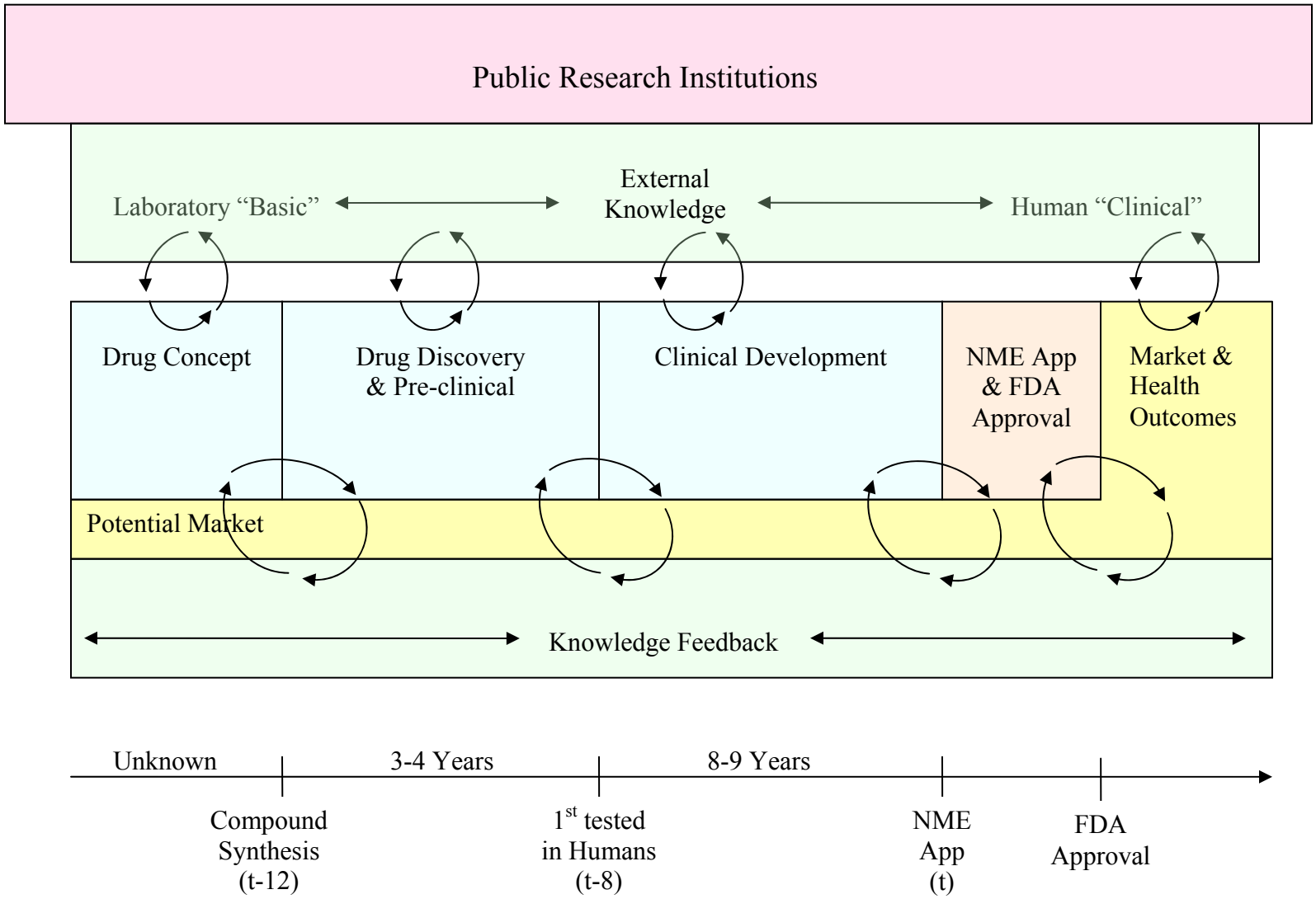

Figure 2: Estimated Lag Coefficients of NIH Funded Basic Research on Pharmaceutical NME Innovation

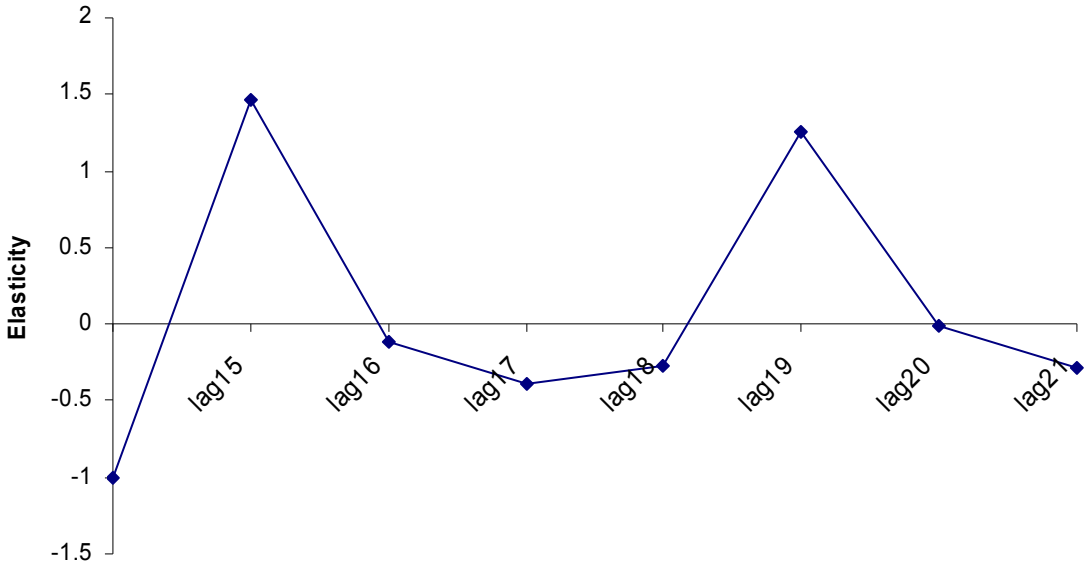

Figure 3: NME Applications and NIH Basic Research funding Lagged Fifteen Years Prior to Application

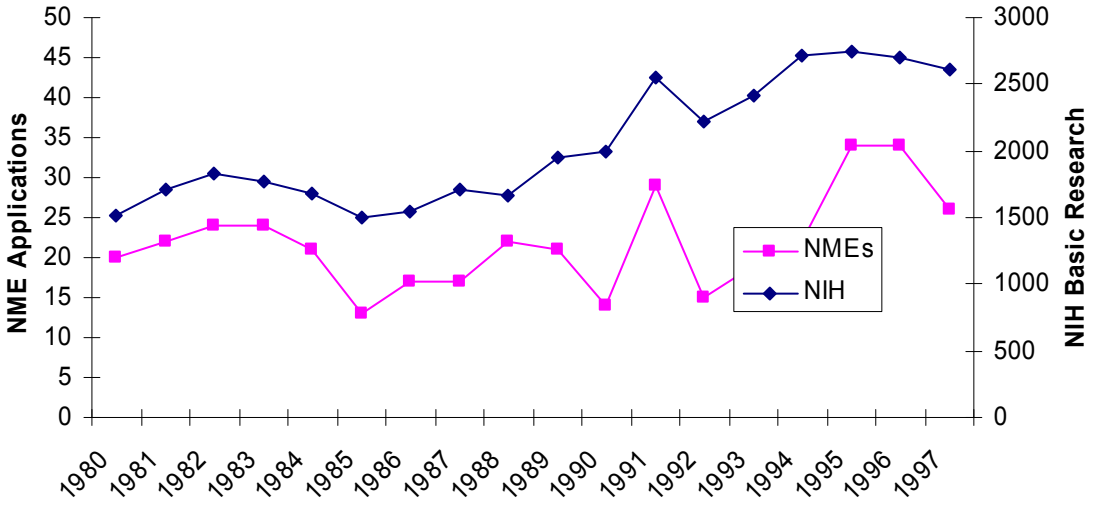

Figure 4: NME Applications and NIH Basic Research Funding Lagged Ninetten Years Prior to Application

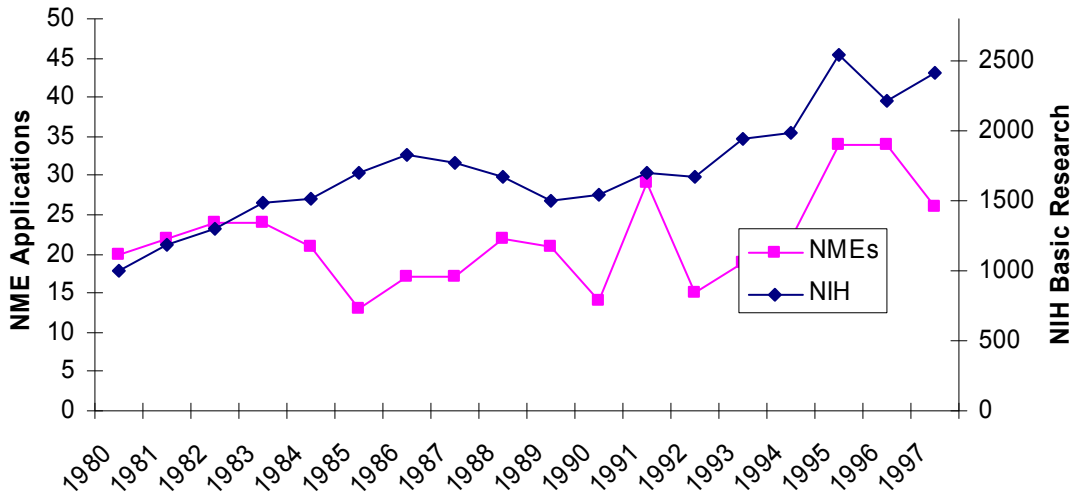

Table 1: Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
NME applications(t)	3.127	2.638	0.000	10.000
Ind R&D stock(t-1)	8.609	0.838	6.949	9.884
PubBas(t-14)	5.184	1.300	2.044	6.881
PubBas(t-15)	5.150	1.297	2.044	6.881
PubBas(t-16)	5.118	1.297	2.044	6.881
PubBas(t-17)	5.080	1.292	2.044	6.881
PubBas(t-18)	5.036	1.288	2.044	6.868
PubBas(t-19)	4.982	1.283	2.044	6.791
PubBas(t-20)	4.923	1.282	2.044	6.791
PubBas(t-21)	4.852	1.292	1.883	6.791
Market Size(t)	4.217	0.884	2.303	5.332
Pre-sample Mortality	-1.094	1.810	-4.631	1.447
Pre-sample Avg. NMEs	0.483	0.844	-0.827	1.597

Seven medical areas observed over eighteen years (obs = 126)

All variables are in logs, except the count of NME applications.

Table 2: Pooled Poisson QMLE Regression Models of Pharmaceutical Innovation

	(1) NME Applications(t)		(2) NME Applications(t)		(3) NME Applications(t)	
Ind R&D stock(t)	**0.786**	**(0.076)*** **	**0.799**	**(0.057)*** **	**0.805**	**(0.058)*** **
PubBas(t-9)	0.179	(0.538)				
PubBas(t-10)	-0.379	(0.792)				
PubBas(t-11)	0.545	(1.23)				
PubBas(t-12)	-0.104	(1.263)				
PubBas(t-13)	-0.010	(0.592)				
PubBas(t-14)	-1.141	(1.022)	-0.943	(0.726)	-0.999	(0.686)
PubBas(t-15)	1.392	(1.329)	1.348	(1.303)	1.471	(1.306)
PubBas(t-16)	-0.026	(1.621)	0.016	(1.61)	-0.114	(1.386)
PubBas(t-17)	-0.433	(0.569)	-0.487	(0.621)	-0.385	(0.365)
PubBas(t-18)	-0.299	(0.402)	-0.252	(0.486)	-0.277	(0.596)
PubBas(t-19)	**1.327**	**(0.809)***	**1.312**	**(0.727)***	**1.253**	**(0.522)****
PubBas(t-20)	-0.092	(1.034)	-0.104	(1.095)	-0.017	(1.148)
PubBas(t-21)	-1.211	(1.03)	-1.217	(1.001)	-0.288	(0.952)
PubBas(t-22)	0.907	(1.431)	0.935	(1.439)		
PubBas(t-23)	0.414	(1.222)	0.446	(1.178)		
PubBas(t-24)	-0.262	(0.975)	-0.285	(1.055)		
PubBas(t-25)	-0.092	(0.481)	-0.100	(0.468)		
Market Size(t)	**0.120**	**(0.04)*** **	**0.115**	**(0.048)*** **	**0.096**	**(0.034)*** **
Pre-sample Mortality	**-0.277**	**(0.045)*** **	**-0.264**	**(0.071)*** **	**-0.252**	**(0.058)*** **
Pre-sample Avg. NMEs	**-0.412**	**(0.106)*** **	**-0.367**	**(0.166)*** **	**-0.343**	**(0.144)*** **
Constant	**-9.392**	**(1.081)*** **	**-9.223**	**(1.497)*** **	**-9.124**	**(1.293)*** **
Yearly Dummy Variables	Significant		Significant		Significant	
Sum coeff. PubBAS	**0.715**	**(0.119)*** **	**0.667**	**(0.175)*** **	**0.644**	**(0.138)*** **
Joint Sig. PubBAS[(a)]	$X^2(6)=3.79$,	p <0.7049	$X^2(6)=4.86$,	p <0.5622	**$X^2(6)=468.5$,**	**p <0.0000**
Dispersion parameter	1.06		1.04		1.02	
Log Likelihood	-213.132		-213.357		-214.204	
Observations	126		126		126	

*** indicates significance at a 1% level (**, *) for 5% and 10% levels for two-sided tests.
Standard errors are clustered by medical area and robust to heteroscedasticity.
(a) some constraints were dropped due to collinearity.

Table 3: Pooled Poisson QMLE Regression Models of Pharmaceutical Innovation: Robustness Checks

	(1) NME Applications(t) Drop Smallest NME Area		(2) NME Applications(t) Drop Largest NME Area		(3) NME Applications(t) Instrument for Ind R&D		(4) NME Applications(t) Linear Model	
Ind R&D stock(t)	**0.815**	**(0.059)*** **	**0.656**	**(0.137)*** **	**0.743**	**(0.102)*** **	**0.698**	**(0.078)*** **
PubBas(t-14)	**-1.435**	**(0.633)** **			-1.006	(0.722)	-0.816	(0.765)
PubBas(t-15)	1.545	(1.442)			1.468	(1.325)	0.743	(1.129)
PubBas(t-16)	0.038	(1.618)			-0.085	(1.421)	0.315	(0.815)
PubBas(t-17)	-0.213	(0.443)	-0.122	(0.672)	-0.411	(0.377)	**-1.079**	**(0.542)* **
PubBas(t-18)	-0.271	(0.542)	-0.273	(0.816)	-0.287	(0.601)	0.745	(0.462)
PubBas(t-19)	**1.341**	**(0.569)** **	0.637	(0.459)	**1.243**	**(0.505)** **	0.332	(0.56)
PubBas(t-20)	-0.348	(1.244)	-0.104	(1.298)	0.006	(1.162)	-0.246	(1.094)
PubBas(t-21)	-0.003	(1.017)	0.211	(0.699)	-0.266	(0.921)	0.161	(0.855)
Market Size(t)	**0.103**	**(0.062)* **	**0.203**	**(0.122)* **	**0.084**	**(0.037)** **	0.027	(0.033)
Pre-sample Mortality	**-0.262**	**(0.079)*** **	-0.136	(0.18)	**-0.243**	**(0.05)*** **	**-0.158**	**(0.052)** **
Pre-sample Avg. NMEs	**-0.363**	**(0.21)* **	-0.063	(0.466)	**-0.332**	**(0.134)*** **	-0.055	(0.138)
Constant	**-9.092**	**(2.153)*** **	**-7.106**	**(2.873)** **	**-8.642**	**(0.956)*** **	**-5.433**	**(1.185)*** **
Yearly Dummy Variables	Significant		Significant		Not Significant		Not Significant	
First Stage Residual					0.125	(0.317)		
NME=0 Dummy							-0.300	(0.196)
Sum coeff. PubBAS	**0.656**	**(0.248)*** **	0.350	(0.507)	**0.662**	**(0.151)*** **	0.115	(0.105)
Joint Sig. PubBAS[(a)]	**$X^2(5)=20.82$, p <0.0009**		**$X^2(5)=11.97$, p <0.0352**		**$X^2(6)=644.2$, p<0.0000**		**F(6,6)=23.53, p <0.0006**	
R-Squared							0.6399	
Log Likelihood	-196.968		-172.19		-214.098			
Observations	108		108		126		126	

*** indicates significance at a 1% level (**, *) for 5% and 10% levels for two-sided tests.
Standard errors are clustered by medical area and robust to heteroscedasticity.
(a) some constraints were dropped due to collinearity.
Anti-infectives is the largest NME medical area, dermatologic is the smallest.
Instruments for the stock if industry R&D are the number of hospital admissions and average length of hospital stays.