

Discussion Paper No. 11-063

**The Impact of Public Basic Research
on Industrial Innovation:
Evidence from the
Pharmaceutical Industry**

Andrew A. Toole

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 11-063

**The Impact of Public Basic Research
on Industrial Innovation:
Evidence from the
Pharmaceutical Industry**

Andrew A. Toole

Download this ZEW Discussion Paper from our ftp server:

<http://ftp.zew.de/pub/zew-docs/dp/dp11063.pdf>

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von
neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung
der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

Discussion Papers are intended to make results of ZEW research promptly available to other
economists in order to encourage discussion and suggestions for revisions. The authors are solely
responsible for the contents which do not necessarily represent the opinion of the ZEW.

Non-Technical Summary

While most economists believe that public scientific research fuels industry innovation and economic growth, systematic evidence supporting this relationship is surprisingly limited. In a recent study, Acemoglu and Linn (2004) identified market size as a significant driver of drug innovation in the pharmaceutical industry, but they did not find any evidence supporting science-driven innovation from publicly funded research. Their results are troubling for at least two reasons. First, a significant body of qualitative evidence suggests publicly funded biomedical research contributes to new drug innovation. Second, it calls into question the rationale supporting the enormous public investments into biomedical research.

This paper generalizes existing case study research by examining the evidence for a systematic relationship between the U.S. National Institutes of Health (NIH) investments into biomedical research performed in academic laboratories and pharmaceutical industry innovation. Novel data on NIH biomedical research awards from 1955 through 1996 are combined with the pharmaceutical industry's own R&D investment and a market size proxy to estimate a panel data model of pharmaceutical innovation by therapeutic market over time.

The statistical analysis shows that NIH funded basic research, potential market size, and industry R&D all have economically and statistically significant effects on the entry of new drugs. The elasticity estimate in the preferred model implies that a 1% increase in the stock of public basic research ultimately leads to a 1.8% increase in the number new molecular entities (NMEs), an important category of new drug therapies defined by the U.S. Food and Drug Administration (FDA). For an average NME, the results also indicate the lag between public investment and NME applications to the FDA is seventeen to twenty-four years. The primary contribution of public basic research to new technological opportunities seems to occur in the years preceding private drug discovery.

The analysis also finds a positive return to public investment in basic biomedical research. Using market sales data for an average NME, the direct return for the six therapeutic markets analyzed is about forty-three percent. One must interpret the magnitude of this estimate cautiously. The estimate does not reflect the plurality or totality of channels through which basic biomedical research is likely to impact social outcomes. It is limited to the contribution of basic research to NME innovation and represents only a fraction of the social return to public basic research.

Das Wichtigste in Kürze

Trotz der allgemeinen Auffassung von Ökonomen, dass öffentlich geförderte Forschungsprojekte zur Verbesserung von Innovationen und wirtschaftlichem Wachstum führen, gibt es hierfür kaum systematische Evidenz. In einer kürzlich durchgeführten Studie identifizieren Acemoglu und Linn (2004) Marktgröße als einen wesentlichen Faktor für die Entwicklung neuer Medikamente in der Pharmaindustrie, während sie keinen signifikanten Einfluss wissenschaftsbasierter Innovationen, die durch öffentliche Fördergelder unterstützt wurden, finden. Diese Ergebnisse geben aus zwei Gründen Anlass zur Sorge. Erstens weist ein wesentlicher Teil der qualitativen Evidenz darauf hin, dass öffentlich geförderte biomedizinische Forschung zur Herstellung neuer Medikamente beiträgt. Zweitens stellt es die hohen öffentlichen Investitionen in biomedizinische Forschung in Frage.

Dieses Papier verallgemeinert existierende Forschung zu Fallstudien, indem ein systematischer Zusammenhang zwischen Investitionen des U.S. National Institutes of Health (NIH) in biomedizinische Forschung, die in wissenschaftlichen Laboren durchgeführt wurde, und Innovationen in der Pharmaindustrie untersucht wird. Neue Daten über die von NIH geförderte biomedizinische Forschung von 1955 bis 1996 werden mit Daten zu den FuE Investitionen der Industrie sowie einem Proxy für Marktgröße verknüpft. Dies ermöglicht die Schätzung eines Paneldatenmodells für pharmazeutische Innovationen in spezifischen therapeutischen Märkten über die Zeit.

Die Auswertung zeigt, dass die durch das NIH geförderte Grundlagenforschung, die ungefähre Marktgröße, sowie die Forschungs- und Entwicklungsausgaben der Industrie einen ökonomischen und statistisch signifikanten Einfluss auf den Markteintritt neuer Medikamente haben. Die geschätzte Elastizität im bevorzugten Model impliziert, dass eine Erhöhung der öffentlichen Grundlagenförderung um 1% zu einem Anstieg von neuen Wirkstoffen (new molecular entities NMEs) um 1,8% führt. NMEs sind laut der U.S. Food and Drug Administration (FDA) eine wichtige Kategorie neuer Medikamententherapien. Die Ergebnisse zeigen, dass zwischen öffentlicher Investition und der Beantragung eines NME bei der FDA für einen durchschnittlichen NME 17 bis 24 Jahre liegen. Der wichtigste Beitrag der öffentlichen Grundlagenförderung zur Erkundung neuer Technologien scheint damit der Entdeckung neuer Medikamente durch die Industrie zeitlich voranzugehen.

Die Analyse zeigt auch einen positiven Ertrag öffentlicher Investitionen in die Grundlagenforschung im Bereich der Biomedizin. Auf Basis von Umsatzdaten für einen durchschnittlichen NME wird gezeigt, dass der direkte Ertrag für die sechs untersuchten therapeutischen Märkte ungefähr 43% beträgt. Die Höhe des geschätzten Ertrags muss jedoch mit Vorsicht interpretiert werden. Der geschätzte Betrag reflektiert weder die Vielfalt noch die Gesamtheit der Wege, durch welche Grundlagenforschung im Bereich der Biomedizin die Gesellschaft beeinflussen kann. Die Interpretation beschränkt sich auf den Beitrag der Grundlagenforschung für NME Innovationen und repräsentiert damit nur einen Bruchteil des gesellschaftlichen Ertrags öffentlicher Forschungsförderung.

The Impact of Public Basic Research on Industrial Innovation: Evidence from the Pharmaceutical Industry¹

A revised version is forthcoming in *Research Policy*.

Andrew A. Toole ^{a,b}

a) *US Dept. of Agriculture, Economic Research Service, Washington DC, United States*

b) *Centre for European Economics Research (ZEW), Mannheim, Germany*

This version: March 2011

Abstract

While most economists believe that public scientific research fuels industry innovation and economic growth, systematic evidence supporting this relationship is surprisingly limited. In a recent study, Acemoglu and Linn (2004) identified market size as a significant driver of drug innovation in the pharmaceutical industry, but they did not find any evidence supporting science-driven innovation from publicly funded research. This paper uses new data on biomedical research investments by the U.S. National Institutes of Health (NIH) to examine the contribution of public research to pharmaceutical innovation. The empirical analysis finds that both market size and NIH funded basic research have economically and statistically significant effects on the entry of new drugs with the contribution of public basic research coming in the earliest stage of pharmaceutical drug discovery. The analysis also finds a positive return to public investment in basic biomedical research.

Keywords: R&D, NIH, social return, biomedical, research lags, public science, new molecular entities

JEL Classification: O31; O32; L65; H51

Address: Andrew A Toole
U.S. Department of Agriculture
Economic Research Service
355 E Street SW
Washington, DC 20024
USA
Phone: +1 (202) 694-5208
E-Mail: atoole@ers.usda.gov

¹The views expressed in this article are the author's and do not necessarily represent the views of the U.S. Department of Agriculture or the Economic Research Service.

...successful management of industrial research is dependent on rapid access to the latest discoveries in academic laboratories...

Edward M. Scolnick, M.D.
Former President of Merck Research Labs
1990 Industrial Research Institute Medalist Address

1 Introduction

In his 1990 Medalist Address to the Industrial Research Institute, Dr. Edward M. Scolnick, former President of Merck Research Laboratories, emphasized the importance of academic research as a source of new ideas fueling innovation in the pharmaceutical industry. In particular, he highlighted the contribution of university research funded by the U.S. National Institutes of Health (NIH) to the discovery of new drugs using a variety of specific examples such as the discovery of Captopril and Proscar.

The present paper generalizes existing case study research by examining the evidence for a systematic relationship between NIH investments into biomedical research performed in academic laboratories and pharmaceutical industry innovation. The belief that academic research creates new knowledge fueling technological opportunities has a long history in economics (Griliches 1979, 1992; Klevorick et al. 1995). Growth theorists use the nonrivalrous nature of new knowledge to explain growth in income per capita and to introduce the possibility of increasing returns to scale (Aghion and Howitt 2005; Jones 2005). In the empirical literature, Jaffe (1989) analyzed the production of corporate patents by region over time and found that academic research made a significant contribution. Jaffe's findings were reinforced when applied to a single year of data on innovations by Acs et al. (1991). Adams (1990) found that academic knowledge made a significant contribution to manufacturing productivity growth with a lag of up to thirty years on spillovers.²

Focusing on the pharmaceutical industry, Acemoglu and Linn (2004) built a theoretical model to explain the entry of new drugs into medical therapeutic markets. The

² There are a number of other contributions to this literature including Mansfield (1991, 1998), Narin et al. (1997), Beise and Stahl (1999), and Arundel and Geuna (2004). Salter and Martin (2001) provide a survey.

model highlighted the influence of market size on innovation, but it also included the possibility that changes in technological opportunities from the supply-side could augment innovation. Their empirical tests found strong evidence that potential market size stimulates new drug entry, but found no evidence that NIH investments into biomedical research stimulate innovation. Moreover, Acemoglu and Linn did not control for the pharmaceutical industry's own investments in research and development (R&D).

The finding that NIH investments have no systematic relationship with pharmaceutical innovation is troubling for at least two reasons. First, it is inconsistent with existing qualitative and quantitative evidence. Among the sectors analyzed by Jaffe (1989), drugs and medical technology showed the strongest influence of academic research on corporate patenting. In two different surveys, Mansfield (1991, 1998) found the pharmaceutical industry had the highest percentage of new products based on recent academic research. Cohen et al. (2002) reported that public research influenced new project ideas in the pharmaceutical industry more than in any other manufacturing industry. Looking at science papers cited in U.S. drug and medical patents, Narin et al. (1997) found that 79% originated from public science institutions. Cockburn and Henderson (1998), using co-authorship data, showed that firm-level “connectedness” to public research was positively related to performance in drug discovery. Second, it calls into question the contribution of public investments into biomedical research.³ The NIH is the largest public enterprise supporting biomedical research. In 2010, the NIH invested over \$20 billion in biomedical research performed at universities and other not-for-profit research institutions. New drug innovation should be one of the important channels for reaping the benefits of these enormous public investments in biomedical research.

³ The contribution of public research to drug innovation is an important part of the debate on pharmaceutical profits and drug pricing. Based mostly on case study evidence in reports such as NIH (2000), advocates on both sides of the debate acknowledge a positive NIH contribution, but they interpret the evidence differently. For further background, refer to Sampat and Lichtenberg (2011) and Reichert and Milne (2002).

Using novel data on NIH biomedical research awards from 1955 through 1996, this analysis examines the possibility that changes in technological opportunities from public investments in basic biomedical research contribute to pharmaceutical industry innovation. The NIH data are combined with the pharmaceutical industry's own R&D investment and a market size proxy to estimate a panel data model of pharmaceutical innovation by therapeutic market over time. The statistical analysis shows that NIH funded basic research, potential market size, and industry R&D all have economically and statistically significant effects on the entry of new drugs. The elasticity estimate in the preferred model implies that a 1% increase in the stock of public basic research ultimately leads to a 1.8% increase in the number new molecular entities (NMEs), an important category of new drug therapies defined by the U.S. Food and Drug Administration (FDA). For an average NME, the results also indicate the lag between public investment and NME applications to the FDA is seventeen to twenty-four years. The primary contribution of public basic research to new technological opportunities seems to occur in the years preceding private drug discovery.

The analysis also finds a positive return to public investment in basic biomedical research. Using market sales data for an average NME, the direct return for the six therapeutic markets analyzed is about forty-three percent. One must interpret the magnitude of this estimate cautiously. The estimate does not reflect the plurality or totality of channels through which basic biomedical research is likely to impact social outcomes. It is limited to the contribution of basic research to NME innovation.⁴ Even for NME innovation, the rate of return calculation is based on estimates of sales revenue that do not capture consumer surplus, intergenerational improvements in health, or indirect returns acting through industry R&D. So, while the return is positive, the calculations represent only a fraction of the social return to public basic research investment.

⁴ Salter and Martin (2001) survey and classify the variety of ways that public investment in basic research can have economic benefits. Also refer to McMillan and Hamilton (2003), Cockburn and Henderson (2001), and Malo (2009).

The rest of the paper is organized into the following sections. 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 a new drug therapy can be marketed in the United States, it must receive approval from the U.S. Food and Drug Administration (FDA). As such, FDA policies and requirements fundamentally impact the nature and structure of the pharmaceutical innovative process. The FDA also classifies new pharmaceutical products. The pharmaceutical innovations analyzed in this paper come from their group of new molecular entities (NMEs), which is the category of new products with the greatest therapeutic and economic potential.⁵ 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 molecular entities. For this reason, the analysis focuses on NIH investments into basic research, although public clinical research investments were included in some of the robustness checks.

The nature and structure of the pharmaceutical innovative process determines how and when public basic research influences new drug innovation. This process is typically

⁵ 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 (1994, 1996) for an analysis of the distribution of sales revenue for NMEs. Cockburn (2006) provides a good overview of the issues.

described as beginning with drug discovery, moving to pre-clinical studies, human clinical development, and eventually to application for approval from the FDA. It is relatively structured and sequential compared to most other industries due to the regulatory requirements imposed by the FDA. Case study evidence suggests that public basic research has its primary influence on industry drug discovery. 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 for more background).

While there are a number of fascinating examples from the case study literature, the story of Captopril is a classic example of 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 that 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 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 Cockburn and Henderson 1998, 2001; Dustan et al. 1996; Gambardella 1995; Henderson 1994; Maxwell and Eckhardt 1990; NIH 2000; OTA 1993; Scolnick 1990; Silverstein et al. 1995; Zycher et al. 2008).

Captopril was one of twenty-one case histories of important drugs examined by Cockburn and Henderson (1998). Based on their qualitative research, they noted that drug discovery is characterized by a high degree of public and private interaction in research. The pharmaceutical innovative process 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 discovery to market and back to drug discovery (Gelijns and Rosenberg 1994). However, describing and documenting the number, frequency, and modes of interaction is quite difficult. Aside from individual case histories, firm-level survey data provide the best insights into how pharmaceutical scientists access and use public research. From a survey of U.S. R&D managers, Cohen et al. (2002) reported 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. Arundel and Geuna (2004) found the same ranking using survey data on large R&D performing firms located in Europe.⁶ Aside from publications and reports, these other mechanisms of accessing public research are interactive and almost certainly involve bi-directional flows of knowledge between public and private research scientists as described by Cockburn and Henderson (1998).

Another important aspect of the pharmaceutical innovative process is the length of time from drug discovery to FDA application. Since public basic research is expected to

⁶ They used 1993 data from the policies, appropriation and competitiveness in Europe (PACE) survey, which excludes France.

make its contribution in the drug discovery 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 detailed data held by the Tufts Center for the Study of Drug Development. 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 FDA application. Based on these lags, investments into public basic research relevant to pharmaceutical innovation are expected to occur thirteen or more years before the date of FDA application. The empirical analysis will examine the lag between public basic research investment and NME applications.

3 Empirical Specification and Estimation

The empirical analysis models the entry of new molecular entities into therapeutic markets over time. Acemoglu and Linn (2004) used this approach to test the importance of market size in innovation.⁷ In addition to potential market size, the specification used here introduces new measures of NIH funded basic research and also controls for the pharmaceutical industry’s own research and development investment. The main estimating equation for new drug entry has the following form:

$$(1) E[NME_{ct} | \zeta_c, \mathbf{X}_{ct}] = \exp(\beta_0 + \beta_1 \ln(PubBasStock_{ct-j}) + \beta_2 \ln(IndRDStock_{ct-1}) + \beta_3 \ln(M_{ct}) + \zeta_c + \mu_t)$$

where E is the expectations operator, NME_{ct} is a count of pharmaceutical product innovations in medical therapeutic market c and FDA application year t (all applications were subsequently approved). The ζ_c are time-constant therapeutic market intercepts capturing

⁷ Earlier studies examining pharmaceutical innovation used similar approaches such as Baily (1972), Wiggins (1981), Henderson and Cockburn (1996).

heterogeneity and \mathbf{X}_{ct} represents all explanatory variables dated at time t and earlier. $PubBasStock_{ct-j}$ is the stock of NIH investment into public basic research for therapeutic market c , lagged j years prior to NME application. The empirical analysis will examine alternative lag lengths j . $IndRDS_{ct-1}$ is the stock of R&D investment by private pharmaceutical firms into therapeutic market c dated one year before FDA application. M_{ct} is a measure of the potential market size in market c at time t based on demographic information. The μ_t are annual time dummy variables capturing shocks common to all markets. For those explanatory variables appearing in log form, the parameters estimates have a constant elasticity interpretation.

Equation (1) is estimated using the Poisson quasi-maximum likelihood estimator (QMLE). As a member of the linear exponential family of distributions, Poisson QMLE produces consistent estimates of the population parameters as long as the conditional mean is correctly specified (Gourieroux, Monfort, Trognon 1984; Wooldridge 1997). Two alternative tests of the conditional mean specification in equation (1) were performed, White's information matrix test and Ramsey's RESET test, and neither test found any evidence of misspecification.⁸ For hypothesis testing, the analysis does not assume dynamic completeness. The cross-sectional dimension of the panel dataset used to estimate equation (1) is small relative to the time series dimension. While not affecting consistency, the asymptotic variance estimator should be adjusted for potential serial correlation in the score (see Wooldridge 1997 for details). The Newey-West heteroscedasticity/serial-correlation [H/SC] robust standard errors are reported allowing for up to third-order serial correlation (Wooldridge 1991, 1994).

As discussed in Section 2, interaction and feedback are intrinsic parts of the

⁸ For the preferred model (Table 3, model 5), the robust Chi-squared statistic for White's information matrix test under the generalized linear model assumption is: $\text{Chi}(24)=24.37$ with a $p\text{-value}<.4407$. The robust Chi-squared statistic for Ramsey's RESET test is: $\text{Chi}(2)=0.547$ with a $p\text{-value}<.7607$. For details on the testing procedure see Wooldridge (1991, 1997).

pharmaceutical innovative process. Feedback in equation (1) takes place when shocks to NME applications at time t influence industry and public basic research investment at future dates ($t+1$ and beyond). For the typical lag structure characterizing pharmaceutical innovation, 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).⁹ An appropriate estimator will allow for feedback as well as address the time-constant unobserved heterogeneity across therapeutic markets represented by the ζ_c in equation (1).

The approach taken here allows for feedback by using the pooled Poisson QMLE and accounts for correlated heterogeneity using pre-sample information as suggested by Blundell, Griffith and Van Reenen (1995).¹⁰ This is called the “entry stock” Poisson estimator. 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.¹¹ The sample available for this analysis is relatively small (108 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 therapeutic market were calculated using data from 1964 to 1979. The central nervous system market had the highest pre-sample mean of 4.94 drugs per year

⁹ Formal statistical tests show that industry R&D investment is not strictly exogeneous. As reported in model (3) in Table 4, using the fixed effects Poisson QML estimator, the lead of industry R&D investment is positive and significant with a t-statistic of 2.35 and a p-value < .019.

¹⁰ Pooled Poisson QMLE does not require strict exogeneity for consistency, but allows for arbitrary feedback over time.

¹¹ 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 “entry stock” estimator to account for correlated fixed effects.

while the respiratory market had the lowest with 0.625 drugs per year.

Another source of heterogeneity across therapeutic markets is the burden of disease. The empirical specification uses pre-sample mortality and hospital discharge rates across therapeutic markets observed in 1979 to account for heterogeneity in the burden of disease. Cardiovascular had the highest mortality rate of 4.24 deaths per thousand people and respiratory had lowest rate of 0.312 deaths per thousand people. Central nervous system diseases had the highest hospital discharge rate of 23.72 per thousand people and infectious disease had lowest rate of 7.72 per thousand people.

The specification of equation (1) incorporates the characteristic lags in the pharmaceutical innovative process. Real industry R&D investment flows over a twelve year period, which is the mean period from compound synthesis to FDA application, were aggregated into an industry stock dated one year prior to application (t-1). Using a stock formulation preserves degrees of freedom, avoids collinearity problems, and models the cumulative nature of research investment. It was constructed using the perpetual inventory method described by Hall et al. (1988) assuming a twenty percent depreciation rate ($\delta=0.20$).¹² The R&D industry stock for NMEs applications in therapeutic market c at time t is given by:

$$IndRDStock_{ct-1} = IndRDflow_{ct-1} + \sum_{j=1}^{11} (1-\delta)^j IndRDflow_{ct-j-1}$$

The annual time dummy variables capture shocks common to all the therapeutic markets in the pharmaceutical innovative process. For instance, over the sample period from 1980 to 1997, legislative changes such as the 1984 Patent Term Restoration and Competition Act and the Prescription Drug User Fee Act of 1992 may have influenced when new drug

¹² 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. As in prior research, other depreciation assumptions lead to very similar coefficient estimates (see Hall et al. 1988). Using an unrestricted twelve-year distributed lag gives a statistically significant long-run R&D elasticity that is about 8% smaller than reported using the stock formulation. All results for public basic research are robust to different formulations of the industry lag.

applications were submitted to the FDA. The annual dummy variables account for any common influence of these events across therapeutic markets on NME applications.

4 Data Sources and Measurement

The empirical analysis uses a panel database on new drug entry by therapeutic market observed from 1980 through 1997. This database combines six major sources of information: industry data from the Pharmaceutical Research and Manufacturers Association (PhRMA), new drug entry data from the Food and Drug Administration (FDA), public biomedical science investment data from the National Institutes of Health (NIH), mortality and hospital discharge data from the National Center for Health Statistics, population data from the Census Bureau, and income data from the Current Population Survey. The database is organized into six broad therapeutic markets: endocrine/neoplasm (cancer), central nervous system, cardiovascular, gastro-intestinal/genito-urinary, anti-infectives, and respiratory.

The first step in the data collection process was to obtain the most disaggregated and comprehensive publicly available information on pharmaceutical industry R&D investment. These data were collected from PhRMA covering the period 1965-1999. PhRMA collected their data through an annual survey of their member firms, but the public release data are aggregated across drug discovery/clinical development phases and into therapeutic markets. The industry data determined the level of aggregation used in this analysis and all other sources of information discussed below were organized to match the six broad therapeutic markets available from the industry data series. The annual R&D data include monies spent on successes, failures, new molecular entities (NMEs), line extensions, new dosage forms, and so forth. (roughly 80% of these funds went toward NMEs).¹³ The flows were adjusted for inflation using the Producer Price Index for pharmaceutical preparations with 2000 as the base year. For the regression model, the real flows were cumulated into a twelve year stock

¹³ More information on the PhRMA data can be obtained through their Website at <http://www.phrma.org>.

as described in Section 3.

Through a Freedom of Information Act request, the FDA supplied data on new molecular entities (NMEs) approved for marketing between 1964 and 2001. Two steps were performed to prepare the raw NME data for analysis. First, the NMEs had to be allocated into the six therapeutic markets. This required eliminating some of the listed NMEs that did not match the markets analyzed or would double count innovations such as new dosage formulations, certain biologics, diagnostics, and materials. It also required the use of medical references on pharmacology and treatment indications to allocate each NME to a therapeutic market. This process is described in greater detail in Appendix A. Second, for the regression analysis, each approved NME was dated by year of application to the FDA. This was done to minimize any FDA regulatory or budgetary influences on the timing of pharmaceutical innovation.¹⁴

For public investments into biomedical research, the NIH supplied a novel database containing all grant and contracts awards by the U.S. Department of Health and Human Services, the umbrella organization of the NIH, between 1955 and 1994. These data were further updated using the NIH CRISP database (Computer Retrieval of Information on Scientific Projects) covering the years 1972-1996. A multistage procedure was used to separate these data by character of research (basic, clinical, other) and further allocate grants and contracts to therapeutic markets. (Refer to Appendix A for a detailed description of the procedure.) The annual investment flows were deflated using the NIH Biomedical Research and Development Price Index (BRDPI) maintained by the Bureau of Labor Statistics (base year is 2000). The real annual flows were cumulated into a stock assuming no obsolescence of knowledge capital. As with industry R&D, the stock formulation preserves degrees of freedom, avoids collinearity problems, and models the cumulative nature of research. In any

¹⁴ The NMEs analyzed are (or were) protected by one or more patents. In this regard, they are an indicator of important patents since they represent a subset of all patented pharmaceutical compounds that have met the FDA standards for safety and efficacy.

single year, only about 25-30% of the NIH extramural research budget is available for new competing awards with the other 70-75% already committed to ongoing projects.¹⁵

The proxy variables for potential market size by therapeutic market and year were constructed based on the methodology used by Acemoglu and Linn (2004). Their method is attractive because variation in market size comes from potentially exogenous demographic changes holding drug expenditure by therapeutic market and age group constant. For this analysis, the proxy variables are also based on reasonably exogenous demographic changes in population and income. Because drug expenditure shares as used by Acemoglu and Linn (2004) were not available, potential need is measured using pre-sample shares of “at risk” population in 1979.

Construction of the potential market size proxy variables took place in three steps. First, the 1979 at risk population was determined using data supplied by the National Center for Health Statistics (NCHS) on mortality from the Multiple Cause of Death File and on hospital discharges from the National Hospital Discharge Survey. Each record was coded with indicators identifying the medical therapeutic market, age group, and gender.¹⁶ Second, total population and average real annual income by age group and gender were collected from the Census Bureau and the Current Population Survey for each year (1979-1997). The Census data along with the NCHS data were used to calculate the 1979 therapeutic market shares of at risk population by age group and gender.¹⁷ Further, total real income (purchasing power) by age group and gender was determined as the product of total population and average real income. In the third step, total real income by age group and gender for every

¹⁵ Data on NIH competing and non-competing research project grants can be obtained from the following Website: http://officeofbudget.od.nih.gov/spending_hist.html.

¹⁶ The medical therapeutic market indicator was determined from the International Classification of Diseases, 9th Clinical Modification. Age groups were defined to be consistent with age groupings used to report average income data in the Current Population Survey.

¹⁷ Annual deaths by therapeutic market were added back to the population figures in the sample period before apportioning using the at risk shares. This eliminates any feedback from new drugs to population size acting through mortality (Cerde 2007).

sample year was apportioned to the therapeutic markets using the 1979 shares of at risk population. The final potential market size proxies were calculated by summing real income across age groups and gender within a therapeutic market. The resulting potential market size variables are likely to be strictly exogenous since NME innovations cannot feedback to change the at risk population by affecting the burden of disease, at least as it is measured by mortality and hospital admissions. One limitation of these data is that chronic illnesses by therapeutic market are probably under-represented. Table 1 reports the descriptive statistics for NMEs, industry R&D, NIH investment, and potential market size by therapeutic market.

5 Estimation Results

Table 2 presents the initial Poisson results using a simple “build up” approach to estimate equation (1), but ignores the characteristic timing in the pharmaceutical innovative process. The Newey-West standard errors, which account for heteroskedasticity and up to third order serial correlation, are reported. The top row reports the coefficient estimates and standard errors for the stock of public basic research lagged one year before submission of NME applications. Looking across the table from model (1) to model (4), this stock remains statistically significant as control variables are added to the models. Model (2) adds the pre-sample controls for heterogeneity in order to use the Blundell et al. (1995) entry stock Poisson estimator. Controlling for heterogeneity between therapeutic markets increases the elasticity estimate for public basic research. The negative coefficient estimate on pre-sample NMEs indicates that greater past innovation reduces future pharmaceutical innovation. Holding any “recharge” from public basic research constant, it suggests that positive effects of capacity building from prior innovation are offset by other factors. This finding is consistent with prior research by Henderson and Cockburn (1996) who observed a pronounced downward trend in important patents after 1978. For the other heterogeneity controls, higher levels of past mortality reduce future NME innovation while greater levels of

hospital discharges are not significantly related to NME innovation in model (2). Hospital discharges are typically negative and significant in subsequent regression models.

Model (3) adds the control for potential market size. The market size control is positive and statistically significant, but larger in magnitude than previously found by Acemoglu and Linn (2004). This may reflect the use of more aggregated therapeutic markets that will capture comorbidity or co-occurrence of disease conditions within therapeutic markets. In model (4), the stock of pharmaceutical industry R&D is added to the specification. As expected, the coefficient estimates for public basic research and market size are reduced. Without industry R&D in the specification, effects of public basic research and potential market size are biased upward. As reported at the bottom of the table, the value of the log likelihood and the R^2 , which is calculated as the squared correlation between the actual and fitted values, indicate that model (4) is preferred. Overall, the results in Table 2 show the effect of public basic research on pharmaceutical innovation is positive and significant even without introducing the complex timing that is characteristic of pharmaceutical innovation.

The regression models reported in Table 3 explore the timing between investments into public basic research and pharmaceutical innovation. Because the results use the “partialling out” property of multiple regression analysis, it is important to remember that a basic research stock dated at time t includes all past annual investment flows up to and including time t . For instance, the public basic research stock dated at time $t-1$ includes eight additional years of NIH investment flows compared to the public basic research stock dated at $t-9$, and so forth.

Model (1) in Table 3 includes four alternative stocks of public basic research each

separated by eight years.¹⁸ Of these four, only the basic research stock dated at $t-17$ is statistically significant with the largest magnitude. Basic research investment flows captured by the other stocks do not significantly contribute to NME applications once the stock dated at $t-17$ is held constant. The results indicate the lag between public investment and NME application to the FDA is seventeen to twenty-four years. This lag period suggests that “enabling discoveries” in public basic research, a phrase coined by Cockburn and Henderson (1998), occur in the period leading into private industry drug discovery. While the data are not rich enough to determine the timing precisely, the results are consistent with available evidence from individual drug histories. Cockburn and Henderson (1998) found that the average lag between the initial public sector enabling discovery and FDA application was 20.5 years for the fourteen drugs they studied.¹⁹ Models (2) through (4) in Table 3 sequentially drop insignificant stocks to examine how the results change. Looking across the table, the coefficient estimate on the basic research stock fluctuates somewhat, but is always statistically significant. Model (5) is the preferred model.²⁰

Besides conforming to expectations based on qualitative evidence, model (5) is preferred on statistical criteria. Relative to model (4) in Table 2, which did not incorporate the characteristic timing in the pharmaceutical innovative process, the value of the log likelihood function and the R^2 favor model (5) in Table 3. It is important to emphasize that the results identify a time period of basic research investment and do not pinpoint investment in a particular year. The stock of public basic research dated at $t-17$ includes *all* past flows of

¹⁸ The eight year interval reflects the characteristic lag in the pharmaceutical innovative process as described by DiMasi (2001) and discussed in Section 2. The stock of public basic research dated at $t-9$ coincides with the beginning of clinical development and this determined the eight year interval size. Collinearity between the stocks limits the use of smaller time windows and prevents the analysis from pinning down the lag precisely. Given available data, the analysis can only identify an approximate time window.

¹⁹ 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).

²⁰ Appendix B shows corroborating results that use a “flow” approach to identify the lag period rather than the “stock” approach described above.

NIH investments and enabling discoveries necessarily build on a long history of investments into basic biomedical research, perhaps decades.

5.1 Economic Significance

The economic importance of public basic research in pharmaceutical product innovation is revealed by the magnitude of the coefficient estimate. Focusing on the preferred model (5) in Table 3, the elasticity estimate indicates that a 1% increase in the stock of public basic research dated at $t-17$ increases the expected number of NME applications (and ultimately marketed) by 1.8% in each therapeutic market. Holding all else constant and using the sample averages, this estimate implies a marginal physical product of about 0.0018 NMEs per \$1 million and a marginal cost per NME of about \$556 million (real 2000 dollars).

Concerning the estimated magnitude of private R&D investment, model (5) in Table 3 shows the elasticity of NME applications with respect to the stock of industry R&D is about 0.78. Holding all else constant and using the sample averages, this estimate implies a marginal physical product of about 0.0021 NMEs per \$1 million and a marginal cost per NME of about \$370 million in real 2000 dollars.²¹ In the literature, DiMasi et al. (1991, 2003) found the average out-of-pocket cost per NME increased from \$251 to \$405 million (real 2000 dollars) between the two sample periods they analyzed.²² Although their sample periods do not exactly match the sample period used here, the \$370 million marginal cost figure falls within the expected range.

²¹ The marginal product is calculated as: $mp = \text{elasticity} \times (\text{avg. NME} / \text{avg. industry flows})$. 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.

²² 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 drugs entering clinical trials from 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 drugs entering clinical trials from 1983 to 2000.

Drawing on the research of Grabowski and Vernon (1994), it is possible to estimate a lower bound on the social return to public investment in basic biomedical research. Their study used a cohort sample of new molecular entities approved by the FDA between 1980 and 1984 and a twenty-year product life cycle model to estimate the time profile of sales revenue for an average NME.²³ Of course, sales revenue for an average NME is only a fraction of the expected benefit stream from public investments in basic biomedical research. The broader benefit stream would include components such as consumer surplus, cost savings from fewer hospital visits or shorter hospital stays, reductions in mortality, health benefits enjoyed by future generations, earnings from continued work force participation, and so forth. Moreover, NME innovation is only one channel for reaping the benefits of public investments in biomedical research and cannot capture the plurality or totality of channels through which basic research is likely to impact social outcomes. Consequently, one must interpret the magnitude of the return estimate given below with these caveats in mind.²⁴

Following Griliches (1958) and Trajtenberg (1990), the rate of return is estimated as a capitalized benefit-cost ratio at the time of NME introduction. The thought experiment asks: What is the social return if public basic research increased by \$1 million and the resulting marginal physical product of NMEs earned its share of the sales revenue for an average NME? The initial \$1 million dollar public investment is capitalized at real rate of 5% to the date of NME entry into the market, a period of nineteen years, and this is the denominator of the benefit-cost ratio. Similarly, the numerator of the benefit-cost ratio is the share of real sales revenues after discounting back to the date of NME entry. The rate of return for each therapeutic market, denoted by ρ_i , takes the following form:

$$\rho_i = r(B/C) = r \left[MP \sum_{\tau=1}^{20} \frac{sales_{\tau}}{(1+r)^{\tau}} / (1+r)^{19} \right]$$

²³ I would like to thank Henry Grabowski for providing the sales profile for an average NME.

²⁴ Research by Jones et al. (1998) and David et al. (1992) present a broader framework for understanding and analyzing the social returns to R&D and basic research investment. Also see the references in footnote #4.

where MP is the marginal physical product of public basic research and the other variables are self-explanatory.

For an individual therapeutic market, $\rho_i = .05(3.591/2.527) = 0.071$. Across all six markets considered in this analysis, the total direct return to public basic research is 0.426, or 43%. This says that a dollar of public basic research investment yielded about \$0.43 in annual benefits through NME innovation in perpetuity.^{25, 26} In addition to this direct return, there is an indirect return to public basic research acting through industry R&D.

Incorporating that component, however, would require a number of industry specific assumptions as discussed in Grabowski and Vernon (1994) such as tax treatment and contribution margins. It is important to remember that the long nineteen year lag between initial public investment and subsequent cash flows has a strong impact on the calculated return. Since payoffs from innovations made possible by investments in basic science are subject to long lags, the private incentives to make these investments will be very sensitive to changes in the real interest rate (Adams 1990). When combined with the limited appropriability of the benefits from basic research, the funding role of public sector appears justified based on this analysis.

5.2 Robustness

Table 4 reports the results using six alternative specifications and/or estimators as robustness checks. In previous research, Toole (2007) found that both public basic and clinical research stimulated additional pharmaceutical industry investment in R&D. If public clinical research has a direct effect on NME innovation, it could be an important omitted variable. Model (1) included the stock of public clinical research dated one year before NME

²⁵ The internal rate of return is an alternative to the capitalized benefit-cost ratio. The internal rate of return in each therapeutic market is about 6.3% and implies a total direct return of 37.8%.

²⁶ Under the assumption that NIH investment represents one-half of the total world stock of public basic research, the worldwide rate of return is essentially the same since Grabowski and Vernon (1994) estimate global NME sales as twice U.S. sales.

application as an additional covariate. This stock was not significant and the results remain essentially unchanged. Unreported results showed that changing the timing for the stock of public clinical research had no effect.

Model (2) tests the possibility that industry R&D investment is not sequentially exogenous as required for consistent estimators. A statistically significant coefficient on the residuals from a first-stage reduced form regression would indicate endogeneity. The test uses two instrumental variables. The first is the length of hospital stays per thousand people in the population by therapeutic market and year. The second is the proportion of deaths by race, white versus non-white, by therapeutic market and year. These instruments reflect developments in the patient population and are separate from decisions or outcomes within the pharmaceutical innovative process. Developments in the patient population are important inputs into the decision to invest in R&D. The instruments are strongly jointly significant in the first stage regression with an F-statistic of 62.87 and a p-value<.000 (Shea partial R² of excluded instruments is 0.7334). Having evaluated a drug candidate and committed to expensive Phase III clinical trials, unobserved shocks influencing the year of FDA application are likely to be related to firm-level and industry-level developments, not to contemporaneous changes in mortality or length of hospital stays in the patient population. This reasoning supports the exogeneity of the instruments.²⁷ As reported for model (2), the test finds no evidence that industry R&D violates the sequential exogeneity assumption.

Models (3) and (4) in Table 6 report the results using the fixed effects Poisson quasi-maximum likelihood estimator (QMLE). Model (3) tests for the possibility that industry R&D investment is not strictly exogenous as required for consistency of standard panel data estimators by including the lead of industry R&D as a regressor. The lead of industry R&D is positive and highly significant with a t-statistic of 2.35 and a p-value<.019. This is strong

²⁷ As shown in model (6), these instruments pass the Hansen's J overidentification test.

evidence that industry R&D is not strictly exogenous and reflects follow-on investments for post-approval clinical trials and product extensions. Even though the fixed effects estimator is inconsistent, model (4) reports the Poisson QMLE results. Compared to the preferred model, the estimated coefficient on industry R&D is smaller while the market size and public basic research stock have larger coefficient estimates.

Models (5) and (6) in Table 6 use Ordinary Least Squares and Two Stage Least Squares (2SLS) methods to estimate the linear analog of equation (1). The dependent variable is the log of NME applications and drops eight annual observations where NMEs are zero. The estimated elasticity on the stock of public basic research is about 1.7 and is statistically significant at the 1% level. The 2SLS estimator addresses potential endogeneity of industry R&D using the same instruments as discussed above for model (2). The results are very similar to the OLS regression. There is no statistically significant evidence that industry R&D is endogenous. Hansen's J test (the heteroskedasticity-robust version of the Sargan test) is $\chi^2(1) = 0.176$ (p-value = 0.6752) indicates that the IVs satisfied the exogeneity requirement.

6 Conclusion

Applying an econometric approach, this study found a systematic relationship between NIH investments into basic biomedical research performed in academic laboratories and pharmaceutical industry innovation. The preferred model implies a 1% increase in the stock of public basic research is associated with a 1.8% increase in the number of industry new molecular entity (NME) applications after a substantial lag. For an average NME, the lag between public investment and industry application is seventeen to twenty-four years. As interpreted through the lens of existing qualitative research, this time window identifies the occurrence of “enabling discoveries” in public basic research that build on and reflect a long history of prior research in both the public and private sectors. Combining the estimated

magnitude and lag of public basic research with sales data for an average NME, the total direct return to public basic research is estimated to be 43%. While this estimate is subject to a number of caveats (see Section 5), it does show that the returns to public investment in basic biomedical research are positive. The results also show that industry R&D investment and potential market size are economically and statistically significant determinants of NME innovation. Industry R&D investment, however, cannot be modeled as strictly exogenous since NME innovations stimulate follow-on R&D investment.

The econometric results are consistent with a large body of anecdotal and case study evidence that suggests a positive causal relationship between advances in public basic research and industry NME innovation. However, by itself, the econometric evidence presented here does not prove causality. Causation can be viewed as a cumulative phenomenon in which the combination of many studies using alternative approaches and evidence ultimately surpass some “belief threshold.” The results of this study add to the body of evidence.

It is important to keep in mind that the discovery stage of NME innovation, as described in the case study literature, is an interdependent, complementary, and often complex bi-directional process involving scientists in both academic and industrial laboratories. Because the information channels connecting public and private researchers are diverse, one cannot conclude that the elasticity estimate reflects only disembodied knowledge spillovers through “open science” mechanisms. This knowledge flow is probably occurring through a number of channels including meetings and conferences, informal contact, and consulting, and so forth (Cohen et al. 2002; Monjon and Waelbroeck 2003; Arundel and Geuna 2004).

The analysis did not find any direct relationship between public clinical research and NME innovation. However, it would be wrong to conclude that public clinical research plays

no role in pharmaceutical innovation. Public clinical research may be very important for non-NME innovations and off-label uses that may generate substantial health benefits (Cockburn 2006). Moreover, both public basic and clinical research have statistically and economically important indirect effects on pharmaceutical innovation by stimulating complementary follow-on private R&D investment (Toole 2007).

Notwithstanding the new findings from this research, a lot of work remains before we have a sufficient understanding of the complex relationship between public research and pharmaceutical innovation. One unresolved issue is the level of aggregation used to link product markets to academic research. As a first approximation, this analysis used a simple mapping that aligned product markets with biomedical research areas using broad therapeutic markets such as cardiovascular and so forth. However, it remains unclear how the contribution of public research varies when one considers disaggregated therapeutic markets. Progress on this issue will require the development of “technological distance” metrics that allow for more detailed mapping of biomedical research to disease markets (Griliches 1992; Jaffe 1989). Additionally, more research is needed to understand the mechanisms facilitating knowledge spillovers and other forms of public-private interaction stimulating innovation both within and between therapeutic markets.

While it is reasonable to expect that other industries benefit from publicly supported university research, it is not appropriate to generalize the findings on the timing or economic significance of public basic research to other industries. The regulatory structure imposed by the FDA, while advantageous for the purposes of research, makes the characteristic lag structure in pharmaceutical industry unique. Furthermore, as one of the most science-intensive industries, the estimated magnitude may be large compared to other industries.

Disclaimer & Acknowledgements: The views expressed in this article are the author's and do not necessarily represent the views of the United States Department of Agriculture or the Economic Research Service. I would like to thank the anonymous referees for providing constructive comments. I am also grateful to Tim Bresnahan, John Goddeeris, Henry Grabowski, Wallace Mullin, Paul Romer, F.M. Scherer, and Jeffrey Wooldridge for their valuable comments on earlier drafts of this paper. Of course, all remaining errors are my own.

Appendix A: Data Sources and Measurement

Industry Research and Development (R&D)

Pharmaceutical industry R&D expenditure by therapeutic market and year were obtained from the U.S. Pharmaceutical Research and Manufacturers Association (PhRMA). PhRMA collected these data through their annual survey of members. The data were aggregated across firms as well as across the drug discovery and development phases before public release. The aggregate figures correspond to domestic U.S. company-financed R&D expenditures for human-use (dosage form) ethical drugs for the years 1965-1999. These figures include expenditure on physical capital, labor, manufacturing, as well as a share of overhead. Consistent data on investment shares into six therapeutic markets were available. These markets accounted for over 80% of the industry's investment. The therapeutic market shares were available for all years except three. For those years, the therapeutic market shares were imputed.

Pharmaceutical Innovation: Approved New Molecular Entities (NMEs)

In response to a Freedom of Information Act request, the U.S. Food and Drug Administration provided information on all approved new molecular entities (NMEs) for the period 1964 to 2001. The raw FDA data had to be adjusted to match the six therapeutic markets as defined by the industry data and duplicate entries in the FDA list had to be eliminated. Diagnostic agents, certain biological agents, surgical and other materials such as contact lenses and devices were eliminated as they did not match the therapeutic markets available for analysis. New dosage formulations were eliminated to avoid double counting. Some researchers, such as DiMasi et al. (1991, 2003), refer to this more narrowly defined group as new chemical entities. The approved NMEs were grouped by year of application and therapeutic market. Although the year of application was supplied by the FDA, the compounds needed to be assigned to one of the six medical therapeutic markets. 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 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 market (cardiovascular in the case of minoxidil).

Public Biomedical Research Investment

Two data sources were obtained on extramural biomedical research grant and contract awards by the NIH and other governmental agencies under the authority of the U.S. Department of Health and Human Services. (These other agencies include the FDA, the Center for Disease Control, the Agency for Health Care Policy and Research, and so forth.). The first data source was an extract from the NIH IMPAC database covering the years 1955 to 1994. 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. The second source was the NIH CRISP (Computer Retrieval of Information on Scientific Projects) database covering the years 1972 to 1996. Each grant and contract record in the CRISP database contains: the title, the identification number (activity code, institute code, and grant or contract number), investigator name, narrative description of project, organization receiving the award, address, administrative organization of the NIH or other agency, award amount, type of award, fiscal year of award, city, and state. Using the IMPAC database, the CRISP

records were supplemented to include the scientific review group that recommended approval. A scientific review group is a committee of peers within a scientific field that review grant applications and recommend applications for approval by the National Advisory Councils.

Identifying the relevant NIH research began with definitions of basic and clinical research. Public basic research was broadly defined as bench-level laboratory research directed at the discovery and characterization of physiologically active substances and the definition of metabolic pathways related to normal and disease function. Public clinical research was defined as patient-oriented research involving human subjects. It includes epidemiological research, but excludes social, behavioral, occupational, and health services research. This definition is more restrictive than the definition suggested by the NIH Director's Panel on Clinical Research (NIH 1997); however, the NIH definition of clinical research has been criticized for being too broad (Reicher and Milne 2002). The data process took place in two stages. This first stage separated all awards into three groups (mixed, clinical, and other) using the "type of award code" field. (These are codes like R01 for traditional research award or K08 for clinical investigator award.) A second step in this stage required taking the mixed group and separating out any remaining clinical and other awards from the basic research awards using keyword searches over the grant and contract titles. This finalized the breakout by basic, clinical, and other. The second stage took the basic and clinical groups and separated them into therapeutic markets and a general category. This was done in five steps. First, agencies that did not fund basic or clinical research relevant to the pharmaceutical industry were eliminated. For instance, agencies such as the Centers for Disease Control, the National Library of Medicine, the National Institute of Nursing Research, etc. were eliminated. Second, scientific review groups were matched to their respective therapeutic markets. For instance, the grants and contracts recommended by the neurology review group were placed into the central nervous system therapeutic market. Third, keyword filters were used to further sort those grants and contracts not matched by scientific review group. Fourth, the remaining uncategorized grants and contracts were allocated to therapeutic markets using the Institute codes. For instance, the remaining National Cancer Institute awards went to the endocrine/neoplasm market; the remaining National Eye Institute awards went to the central nervous system market. Fifth, for those Institutes which were too general to be classified (e.g. the National Institute of General Medicine), those awards were allocated in proportion to the grants and contracts successfully categorized. This process resulted in nominal investment flows by character and year for each of the therapeutic markets. Using the 1955-1985 sample growth rates by therapeutic market, the annual flow series were projected back five years to 1950. The 1950 initial stock of public basic research was estimated following Hall et al. (1988) by dividing the 1950 investment flow by the market specific growth rate. The nominal flows were adjusted for inflation using the Biomedical Research and Development Price Index and cumulated into a stock as described in the text.

Appendix B: Public Basic Research Lag Analysis Using Investment Flows

The “stock” method presented in the text found the lag between investment in public basic research and new molecular entity (NME) applications by industry firms to be between seventeen and twenty-four years. The stock argument used the partialling out property of multiple regression analysis to identify this lag. If the key enabling discoveries are taking place in the period identified using stocks, then investment flows should also identify the same lag period. Table B.1 repeats the analysis to determine the lag using investment flows summed across years. Consistent with the stock formulation, model (1) includes the sum of the investment flows for each of three periods: a clinical development period (t-1 to t-8), an extended pre-clinical discovery period (t-9 to t-16), and a pre-discovery period (t-17 to t-25). The results show the lag between investment in public basic research and NME applications is between seventeen and twenty-four years. For interpreting the magnitude of the coefficient, however, the stocks of basic research are appropriate because they account for the cumulative nature of research.

Table B.1: Pharmaceutical Innovation (1980-1997): "Flow" Method to Determine Public Basic Research Lag

Dependent Variable: QML Estimator:	(1) NME Applications Entry Stock Poisson	(2) NME Applications Entry Stock Poisson	(3) NME Applications Entry Stock Poisson	(4) NME Applications Entry Stock Poisson
PubBAS Flows(t-1 to t-8)	0.448 (0.573)			
PubBAS Flows(t-9 to t-16)	-0.148 (0.547)	-0.150 (0.539)		
PubBAS Flows(t-1 to t-16)			0.490 (0.826)	
PubBAS Flows(t-17 to t-24)	1.368 (0.417)***	1.331 (0.418)***	1.208 (0.364)***	1.267 (0.365)***
Ind R&D Stock(t-1)	0.767 (0.123)***	0.804 (0.111)***	0.775 (0.122)***	0.801 (0.112)***
Market Size(t)	5.434 (2.820)*	4.155 (2.148)*	5.423 (2.784)*	4.342 (2.042)**
Pre-sample NME (1964-79)	-1.594 (0.876)*	-1.102 (0.548)**	-1.618 (0.877)*	-1.187 (0.446)***

1979 Mortality (per 1000 pop.)	-1.252 (0.574)**	-0.967 (0.412)**	-1.262 (0.570)**	-1.014 (0.374)***
1979 Hosp. Discharges (per 1000 pop.)	-0.452 (0.236)*	-0.347 (0.182)*	-0.451 (0.233)*	-0.363 (0.174)**
Joint significance PubBAS(t-1 to t-8) and PubBAS(t-9 to t-16)	Chi(2)=1.26, p<0.5333			
Year Dummy Variables	Y	Y	Y	Y
Log Likelihood	-196.677	-196.87	-196.777	-196.891
R-squared	0.5762	0.570	0.5727	0.5696
Dispersion parameter	1.011	1.007	1.009	1.003
Observations	108	108	108	108

*** indicates significance at a 1% level (**, *) for 5% and 10% levels for two-sided tests. All explanatory variables are in natural logs except the time dummy variables. Newey-West H/SC robust standard errors (Bartlett weights, lag length=3)

REFERENCES

- Acemoglu, D., Linn J., 2004. Market size in innovation: Theory and evidence from the pharmaceutical industry. *Quarterly Journal of Economics* 119(3), 1049-1090.
- Acs, Z.J., Audretsch, D.B., Feldman, M.P., 1991. Real effects of academic research: Comment. *American Economic Review* 82(1), 363-367.
- Adams, J.D., 1990. Fundamental stocks of knowledge and productivity growth. *Journal of Political Economy* 98(4), 673-702.
- Aghion, P., Howitt, P., 2005. Growth with quality-improving innovations: An integrated framework, in: Aghion, P., Durlauf, S. (Eds), *Handbook of Economic Growth*. Elsevier: Amsterdam, volume 1A, 67-110.
- Arundel, A., Geuna, A., 2004. Proximity and the use of public science by innovative European firms. *Economics of Innovation and New Technology* 13(6), 559-580.
- Baily, M.N., 1972. Research and development costs and returns: The U.S. pharmaceutical industry," *Journal of Political Economy* 80(1), 70-85.
- Baum, C.F., Schaffer, M.E., Stillman, S., 2007. Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *The Stata Journal* 7(4), 465-506.
- Beise, M., Stahl, H., 1999. Public research and industrial innovations in Germany. *Research Policy* 28, 397-422.
- Blundell, R., Griffith, R., Van Reenen, J., 1995. Dynamic count data models of technological innovation. *The Economic Journal* 105(429), 333-344.
- Blundell, R., Griffith, R., Van Reenen, J., 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., Griffith, R., Windmeijer, F., 2002. Individual effects and dynamics in count data models. *Journal of Econometrics* 108, 113-131.
- Cerda, R.A., 2007. Endogenous innovations in the pharmaceutical industry. *Journal of Evolutionary Economics* 17, 473-515.
- Cockburn, I., 2006. Is the pharmaceutical industry in a productivity crisis? in: Jaffe, A., Lerner, J., Stern, S. (Eds), *Innovation Policy and the Economy*. MIT Press: Cambridge, MA, volume 7, 1-32.
- Cockburn, I., Henderson, R., 1998. Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. *Journal of Industrial Economics* 156(2), 157-182.

- Cockburn, I., Henderson, R., 2001. Publicly funded science and the productivity of the pharmaceutical industry, in: Jaffe, A., Lerner, J., Stern, S. (Eds), *Innovation Policy and the Economy*. MIT Press: Cambridge, MA, volume 1, 1-34.
- Cohen, W.M., Nelson, R.R., Walsh, J.P., 2002. Links and impacts: The influence of public research on industrial R&D. *Management Science* 48(1), 1-23.
- David, P., Mowery, D., Steinmueller, E.W., 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., Hansen, R.W., Grabowski, H.G., 1991. Cost of innovation in the pharmaceutical industry. *Journal of Health Economics* 10, 107-142.
- DiMasi, J.A., Hansen, R.W., Grabowski, H.G., 2003. The price of innovation: new estimates of drug development costs. *Journal of Health Economics* 22, 151-185.
- Dustan, H.P., Roccella, E.J., Garrison, H.H., 1996. Controlling hypertension. *Archives of Internal Medicine* 156, 1926-1935.
- Gambardella, A., 1995. *Science and Innovation: The US pharmaceutical industry during the 1980s*. Cambridge University Press: Cambridge, MA.
- Gelijns, A., Rosenberg, N., 1994. The dynamics of technological change in medicine. *Health Affairs* 13(3), 28-46.
- Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica* 52(3), 701-720.
- Grabowski, H., Vernon, J.M., 1994. Returns to R&D on new drug introductions in the 1980s. *Journal of Health Economics* 13, 383-406.
- Grabowski, H., Vernon, J.M., 1996. Prospects for returns to pharmaceutical R&D under health care reform, in: Helms, R.B. (Ed), *Competitive Strategies in the Pharmaceutical Industry*. The AEI Press: Washington, D.C., 194-207.
- Griliches, Z., 1958. Research costs and social returns: Hybrid corn and related innovations. *Journal of Political Economy* 66(5), 419-431.
- Griliches, Z., 1979. Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of Economics* 10(1), 92-116.
- Griliches, Z., 1992. The search for R&D spillovers. *Scandinavian Journal of Economics* 94, Supplement, S29-S47.
- Hall, B.H., Cummins, C., Laderman, E.S., Mundy, J., 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., Cockburn, I., 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.
- Jones, C, Williams, A., 1998. Measuring the social rate of return to R&D. *Quarterly Journal of Economics* 113(4), 119-135.
- Jones, C.I., 2005. Growth and Ideas, in: Aghion, P., Durlauf, S. (Eds), *Handbook of Economic Growth*. Elsevier: Amsterdam, volume 1B, 1063-1111.
- Klevorick, A.K., Levin, R.C., Nelson, R.R., Winter, S.G., 1995. On the sources and significance of interindustry differences in technological opportunities. *Research Policy* 24, 185-205.
- Malo, S., 2009. The contribution of (not so) public research to commercial innovations in the field of combinatorial chemistry. *Research Policy* 38, 957-970.
- Mansfield, E., 1991. Academic research and industrial innovation. *Research Policy* 20, 1-12.
- Mansfield, E., 1998. Academic research and industrial innovation: An update of the empirical findings. *Research Policy* 26, 773-776.
- Maxwell, R.A., Eckhardt, S., 1990. *Drug Discovery: A Case Book and Analysis*. Humana Press: Clifton, NJ.
- McMillan, G.S., Hamilton III, R.D., 2003. The impact of publicly funded basic research: An integrative extension of Martin and Salter. *IEEE Transactions on Engineering Management* 50(2), 184-191.
- Monjon, S., Waelbroeck, P., 2003. Assessing spillovers from universities to firms: Evidence from French firm-level data. *International Journal of Industrial Organization* 21, 1255-1270.
- Narin, F., Hamilton, K.S., Olivastro, D., 1997. The increasing linkage between U.S. technology and public science. *Research Policy* 26, 317-330.
- NIH Office of Director, 1997. Executive Summary: Director's Panel on Clinical Research Report, http://www.oenb.at/de/img/executive_summary--nih_directors_panel_on_clinical_research_report_12_97_tcm14-48582.pdf (last accessed February 24, 2011).
- NIH, 2000. NIH contributions to pharmaceutical development, in: *RX R&D myths: The case against the drug industry's R&D "scare card"*. (downloaded on 22 February 2011 from <http://www.citizen.org/documents/ACFDC.PDF>).

- Pharmaceutical Research and Manufacturers Association, Annual Survey Report. PhRMA: Washington D.C., various years.
- Reichert, J.M., Milne, C-P., 2002. Public and Private Sector Contributions to the Discovery and Development of 'Impact' Drugs. Tufts Center for the Study of Drug Development White Paper, May 2002.
- Salter, A.J., Martin, B.R., 2001. The economic benefits of publicly funded basic research: a critical review. *Research Policy* 30, 509-532.
- Sampat, B.N., Lichtenberg, F.R., 2011. What are the respective roles of the public and private sectors in pharmaceutical innovation? *Health Affairs* 30(2), 332-339.
- Scherer, F.M., Harhoff, D., 2000. Technology policy for a world of skew-distributed outcomes. *Research Policy* 29, 559-566.
- Scolnick, E.M., 1990. Basic research and its impact on industrial R&D. The IRI Medalist's Address, Industrial Research Institute: Arlington, VA, May 1990.
- Silverstein, S.C., Garrison, H.H., Heinig, S.J., 1995. A few basic economic facts about research in the medical and related life sciences. *The FASEB Journal* 9, 833-840.
- Toole, A.A., 2007. Does public scientific research complement private investment in research and development in the pharmaceutical industry? *Journal of Law & Economics* 50, 81-104.
- Trajtenberg, M., 1990. Economic analysis of product innovation: The case of CT scanners. Harvard University Press: Cambridge, MA.
- U.S. Congress, Office of Technology Assessment, Pharmaceutical R&D: Costs, Risks, and Rewards, OTA-H-522, Government Printing Office: Washington, D.C., 1993.
- Wiggins, S.N., 1981. Product quality regulation and drug introductions. *Review of Economics and Statistics* 63, 615-619.
- Wooldridge, J.M., 1991. On the application of robust, regression-based diagnostics to models of conditional means and conditional variances. *Journal of Econometrics* 47, 5-46.
- Wooldridge, J.M., 1994. Estimation and inference for dependent processes, in: Engle, R.F., McFadden, D.L. (Eds), *Handbook of Econometrics*, Elsevier: Amsterdam, volume IV.
- Wooldridge, J.M., 1997. Quasi-likelihood methods for count data, in: Pesaran, M.H., Schmidt, P. (Eds), *Handbook of Applied Econometrics*, Blackwell: Oxford, UK, volume 2, 352-406.
- Zycher, B., DiMasi, J.A., Milne, C-P., 2008. Private sector contributions to pharmaceutical science: Thirty-five summary case histories," Center for Medical Progress, Manhattan Institute for Policy Research.

Table 1: Descriptive Statistics by Therapeutic Market (1980-1997)

Variable	All Markets	Endocrine/ Cancer	Central Nervous System	Cardio- vascular	Anti- infective	Gastro- Intestinal/ Genito- urinary	Respiratory
New Molecular Entities (counts)							
<i>Mean</i>	3.56	4.17	4.94	4.11	5.11	2.06	1.00
<i>Standard deviation</i>	2.59	2.92	2.44	2.08	2.54	1.21	0.97
<i>Minimum</i>	0	1	1	1	1	0	0
<i>Maximum</i>	10	10	10	9	9	4	3
Industry R&D Stock (12 year, real millions \$)							
<i>Mean</i>	7,054	8,824	8,445	10,858	8,454	2,898	2,843
<i>Standard deviation</i>	4,086	3,558	3,692	3,600	2,065	515	1,219
<i>Minimum</i>	1,217	4,285	4,784	4,761	5,738	1,876	1,217
<i>Maximum</i>	16,911	16,591	16,911	14,976	12,486	3,578	5,076
NIH Public Basic Stock (lagged 17 years, real millions \$)							
<i>Mean</i>	3,623	7,176	3,821	3,862	3,516	2,854	511
<i>Standard deviation</i>	2,702	3,452	1,690	1,668	1,554	1,232	316
<i>Minimum</i>	134	2,293	1,305	1,360	1,156	941	134
<i>Maximum</i>	13,484	13,484	6,880	6,822	6,158	5,036	1,157
Potential Market Size (real billions \$)							
<i>Mean</i>	105	98	147	160	35	139	49
<i>Standard deviation</i>	51	15	21	23	5	20	7
<i>Minimum</i>	29	75	116	124	29	108	40
<i>Maximum</i>	207	127	188	207	44	178	63

Table 2: Pharmaceutical Innovation (1980-1997): "Build up" Regression Models

Dependent Variable: QML Estimator:	(1) NME Applications Poisson	(2) NME Applications Entry Stock Poisson	(3) NME Applications Entry Stock Poisson	(4) NME Applications Entry Stock Poisson
PubBAS Stock(t-1)	0.586 (0.104)***	1.485 (0.339)***	4.299 (1.541)***	2.852 (1.212)**
Ind R&D Stock(t-1)				0.727 (0.130)***
Market Size(t)			12.790 (6.112)**	8.738 (4.905)*
Pre-sample NME (1964-79)		-0.715 (0.344)**	-4.007 (1.776)**	-2.814 (1.369)**
1979 Mortality (per 1000 pop.)		-0.308 (0.103)***	-2.794 (1.228)**	-1.990 (-0.968)**
1979 Hosp. Discharges (per 1000 pop.)		0.004 (0.012)	-1.075 (0.515)**	-0.727 (0.415)*
Year Dummy Variables	Y	Y	Y	Y
Log Likelihood	-217.751	-211.636	-208.029	-197.867
R-squared	0.347	0.4168	0.4744	0.5627
Dispersion parameter	1.184	1.140	1.128	1.010
Observations	108	108	108	108

*** indicates significance at a 1% level (**, *) for 5% and 10% levels for two-sided tests. All explanatory variables are in natural logs except the time dummy variables. Newey-West H/SC robust standard errors (Bartlett weights, lag length=3)

Table 3: Pharmaceutical Innovation (1980-1997): "Stock" Method to Determine Public Basic Research Lag

Dependent Variable: QML Estimator:	(1) NME Applications Entry Stock Poisson	(2) NME Applications Entry Stock Poisson	(3) NME Applications Entry Stock Poisson	(4) NME Applications Entry Stock Poisson	(5) NME Applications Entry Stock Poisson
PubBAS Stock(t-1)	1.433 (1.601)			0.959 (1.311)	
PubBAS Stock(t-9)	-1.291 (1.464)	-0.725 (1.331)			
PubBAS Stock(t-17)	2.367 (1.210)*	2.476 (1.183)**	1.874 (0.475)***	1.478 (0.644)**	1.844 (0.450)***
PubBAS Stock(t-25)	-0.285 (0.659)	-0.400 (0.677)	-0.229 (0.574)		
Ind R&D Stock(t-1)	0.737 (0.119)***	0.771 (0.109)***	0.774 (0.110)***	0.754 (0.118)***	0.780 (0.107)***
Market Size(t)	9.716 (6.977)	6.804 (5.991)	7.941 (5.467)	10.918 (3.793)***	9.532 (3.126)***
Pre-sample NME (1964-79)	-2.377 (1.727)	-1.442 (1.279)	-1.778 (1.070)*	-2.662 (1.075)**	-2.081 (0.640)***
1979 Mortality (per 1000 pop.)	-1.992 (1.337)	-1.378 (1.104)	-1.618 (0.980)*	-2.231 (0.750)***	-1.903 (0.565)***
1979 Hosp. Discharges (per 1000 pop.)	-0.823 (0.598)	-0.58 (0.517)	-0.676 (0.473)	-0.927 (0.320)***	-0.815 (0.268)***

Year Dummy Variables	Y	Y	Y	Y	Y
Log Likelihood	-196.272	-196.547	-196.634	-196.511	-196.685
R-squared	0.579	0.572	0.570	0.5734	0.570
Dispersion parameter	1.006	1.003	1.002	1.002	0.998
Observations	108	108	108	108	108

*** indicates significance at a 1% level (**, *) for 5% and 10% levels for two-sided tests. All explanatory variables are in natural logs except the time dummy variables. Newey-West H/SC robust standard errors (Bartlett weights, lag length=3)

Table 4: Pharmaceutical Innovation (1980-1997): Robustness Checks

Dependent Variable:	(1) NME Apps	(2) NME Apps	(3) NME Apps	(4) NME Apps	(5) ln(NME Apps)	(6) ln(NME Apps)
Estimator:	Entry Stock Poisson	Entry Stock Poisson	Poisson QMLE FE	Poisson QMLE FE	OLS Model	2SLS Model
Purpose:	Include Public Clinical	Test Endogeneity of Ind RD	Test Strict Exogeneity			Instrument for Ind RD
PubBAS Stock(t-17)	1.806 (0.473)***	1.851 (0.498)***	2.025 (0.607)***	1.897 (0.909)**	1.702 (0.533)***	1.719 (0.534)***
Public Clinical Stock(t-1)	0.630 (1.109)					
Ind R&D Stock(t-1)	0.769 (0.112)***	0.787 (0.134)***	0.326 (0.224)	0.688 (0.146)***	0.682 (0.130)***	0.647 (0.160)***
Market Size(t)	10.917 (4.171)***	9.592 (3.276)***	10.083 (3.779)***	9.873 (3.853)***	10.372 (4.074)***	10.414 (4.078)***
Pre-sample NME (1964-79)	-2.671 (1.284)**	-2.093 (0.674)***			-2.125 (0.755)***	-2.126 (0.761)***
1979 Mortality (per 1000 pop.)	-2.239 (0.858)***	-1.914 (0.594)***			-2.015 (0.709)***	-2.018 (0.713)*
1979 Hosp. Discharges (per 1000 pop.)	-0.927 (0.350)***	-0.820 (0.280)***			-0.888 (0.352)***	-0.893 (0.352)***
First Stage Residual		-0.022 (0.261)				
Lead of Ind R&D			0.452 (0.186)**			

Year Dummy Variables	Y	Y	Y	Y	Y	Y
Log Likelihood	-196.572	-196.683	-178.177	-179.073		
Dispersion parameter	1.001	1.003				
R-squared	0.5728	0.570			0.5283	0.5280
Hansen's J Overid Test						Chi(1)=0.176, P<0.6752
Observations	108	108	108	108	100	100

*** indicates significance at a 1% level (**, *) for 5% and 10% levels for two-sided tests. All explanatory variables are in natural logs except the time dummy variables. Newey-West H/SC robust standard errors (Bartlett weights, lag length=3) except for the Poisson QMLE fixed effects regressions (3) and (4). The two instruments for industry R&D are: (1) log of the proportion of deaths by race (white versus non-white) by therapeutic market and year; (2) log of total hospital days of care per 1000 population by therapeutic market and year. OLS and 2SLS regressions were performed using the “ivreg2” Stata command developed by Baum et al. (2007). Full output available upon request.