The Relationship Between Pharmaceutical R&D Spending and NME Development

Taylor Wang

University of California, Berkeley | Department of Economics

Spring 2020

Supervisor: Benjamin Handel

Abstract

In recent years, pharmaceutical companies have used the justification of high R&D expenses to defend the rising prices of prescription drugs. However, as overall R&D expenditure in the industry grows at a substantial pace, whether or not the quantity of NMEs (new molecular entities) has grown meaningfully is questionable. This paper aims to identify a potential relationship between R&D expenditure and NME production utilizing panel data from 1999 to 2016. The results lacked statistical significance indicating no appreciable relationship between R&D expenditure and NME production. Results are robust to different model specifications.

1 Introduction

Pharmaceutical companies spend billions of dollars on research and development to develop the drugs that millions of Americans use every day. While just the cost of successfully developing a pharmaceutical product has a range of estimates from $50 Million to $2 Billion (Adams & Brantner, 2006, p. 420), pharmaceutical firms must also shoulder the cost of products that fail within the pipeline, which numbers tens and even hundreds to every successful drug that makes it to market. It is this tremendous cost of R&D that firms often cite as the primary reason for their startlingly high prescription drug prices (Gronde et al., 2007). Though already incredibly elevated, R&D expenditure in recent years continues to grow at a tremendous rate.

Figure 1: Trends in pharmaceutical R&D expenses over time

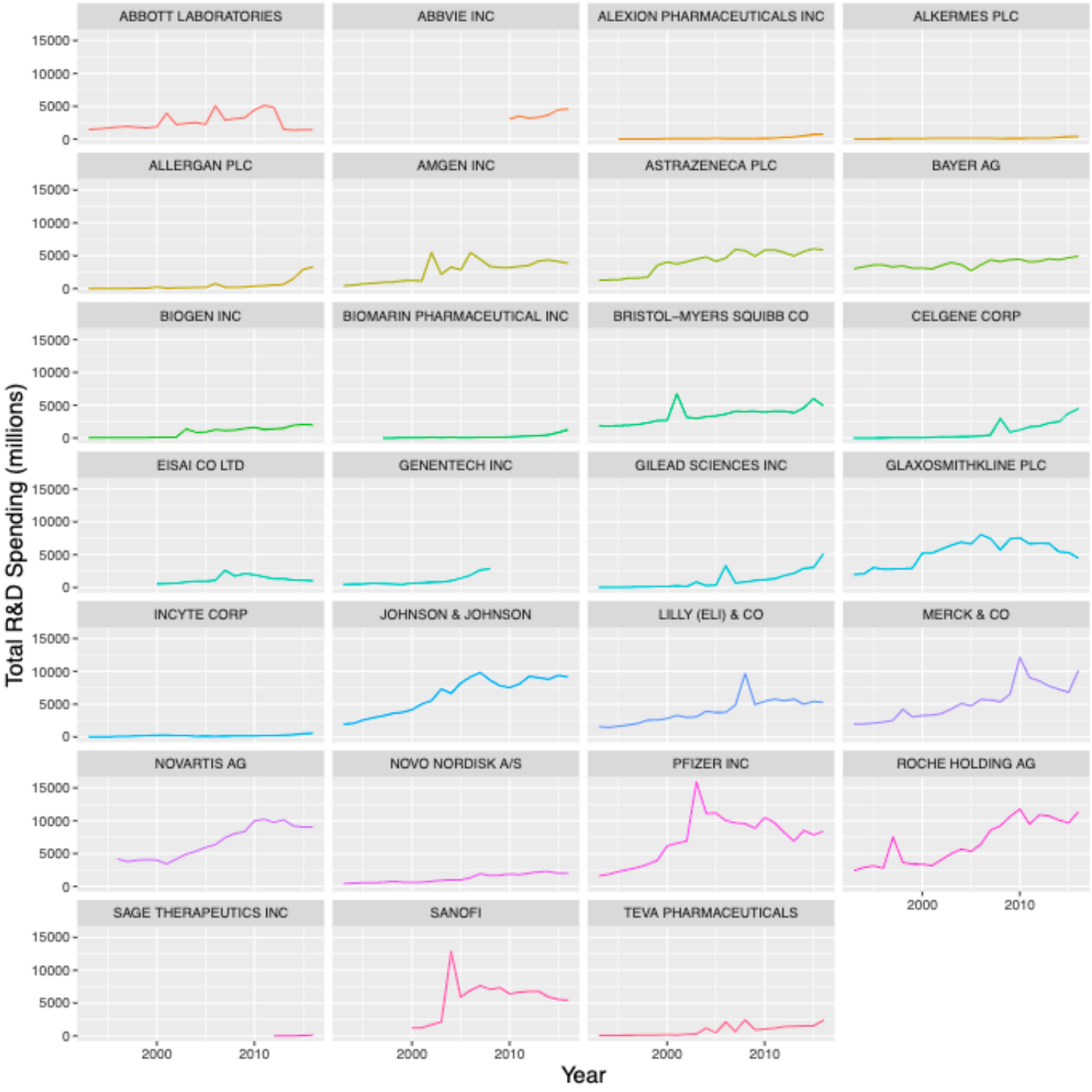

Figure 2: Trends in pharmaceutical R&D expenses over time by Company

This then leads to the question of whether the amount of spending put into research and development is associated with an increase in development of pharmaceutical products. When looking at preliminary numbers for the quantities of NMEs[1], the number of drugs developed has not changed dramatically over time while the level of R&D expenses has. Therefore, I hypothesized that there is no relationship between the two. In the figure below, each point represents the sum of total R&D expenses (for twenty-seven firms) in that particular year while the size of the point represents the number of NMEs developed that year. We would expect the size of the points to grow as total R&D expenses grow as well, but based on the figure, there does not seem to be a clear relationship.

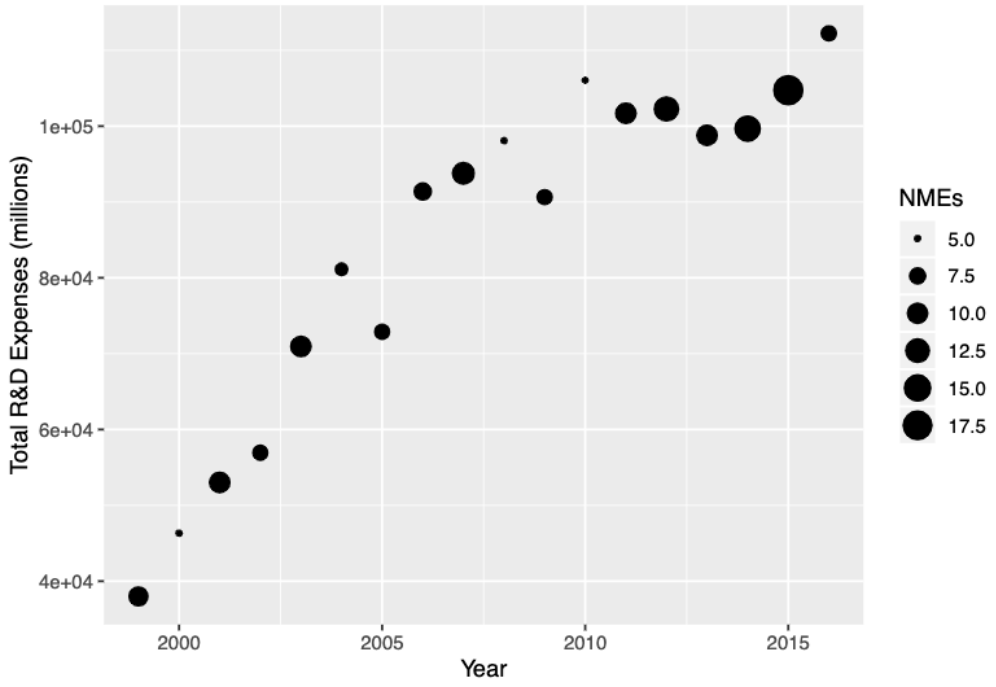

Figure 3: Total R&D expenditure from 1999-2016 by NME development

[1] New Molecular Entities

While the specific number changes from year to year, has there been a substantial increase in NME development that mirrors the rise of R&D expenditure? Many studies have assessed this particular relationship in a variety of different contexts (Munos, 2009; Graves & Langowitz, 1993; Scherer, 2001; Hashimoto et al., 2008; Langowitz & Graves, 1992) but so far there is no recent literature analyzing the effect of R&D expenditure on NME development for large firms.

This paper attempts to address this question by analyzing a subset of twenty-seven large pharmaceutical companies and their respective R&D expenditures and product development. Working with a logistic regression model, my dependent variable was whether or not an NME was developed, and my main independent variable was R&D expenditure. Other internal and external factors were controlled for along with entity and time fixed effects. Additionally, various sensitivity analyses were considered as further robustness checks.

The remainder of this paper is structured as follows: Section 2 highlights other literature within the field and where a space to contribute lies; Section 3 provides an overview of the data and their sources; Section 4 describes the methodology used to estimate the relationship; Section 5 presents the results of the paper; Section 6 covers robustness checks based on different model specification; and Section 7 summarizes and discusses the results.

2 Literature Review

The issue of international innovation within pharmaceutical productivity has been thoroughly analyzed within the health economic sphere. Keyhani et al. (2010) takes a cross-section of the issue and finds that relative to comparable countries, “the US contribution to global discovery of NMEs was roughly proportional to its contribution to… prescription drug spending” (p. 1077) indicating that the United States was neither outperforming nor

underperforming globally. On the other hand, Munos (2009) takes a longitudinal approach and finds a flatlining of productivity over time. Assessing over 1200 NMEs over the period of 60 years, Munos found that the number of NMEs being produced was relatively stagnant (2009). With no significant increase in the number of NMEs developed over time, it seems that the impact of R&D expenditure is negligible. However, there are many confounding factors affecting the rate of NME development such as "revolutionary scientific discoveries in the 1970s" and the overall "rise of biotechnology" (Cockburn, 2004). As the industry has changed fundamentally over time, the direct impact of R&D expenditure is surely different as well. While the historic, industry level trend may be that the number of NMEs are not increasing, without understanding the impact of R&D and the channel between R&D and development, it is hard to discern the true cause of that stagnancy. And while both Keyhani et al. and Munos are informative, neither uses econometric models leading to a lack of practical insight as to how companies should operate within the industry or how effective government policies can stimulate productivity. My focus now shifts to whether there exists true innovative productivity, as driven by R&D, within the industry. Such findings, either positive or negative, can provide guidance to both the industry and the government in fostering greater productivity.

Innovative Productivity, defined as the "returns to scale with respect to the size of the R&D effort" (Graves & Langowitz, 1993, p. 595) can demonstrate how productively pharmaceutical firms are operating and whether their R&D spending is translating directly to output. In the assessment of that productivity, Brown and Svenson (1998) define inputs as standard raw materials that go into the process and outputs as "patents, new products, new processes…" (p. 31) while productivity represents the relationship between those inputs and outputs. In the existing literature, the established findings on the relationship between R&D

expenditure and number of new molecules produced are a bit at odds. Some found a logically unsurprising relationship between R&D expenditure and production (Jensen, 1987; Langowitz & Graves, 1992; Scherer 2001; Mansfield 1981). Others found the relationship to be more inconclusive (Bottazzi et al., 2001; Coad & Rao, 2008; Griliches, 1979). The only common consensus being that it is incredibly difficult to isolate the effects of R&D expenditures on "productivity", defined in any number of ways. Thus, the link between R&D spending and the number of NMEs produced within a given year is certainly up for debate, and their relationship in our current time period is vastly different from assessments in the past due to the aforementioned changes throughout the industry itself. While discussion on industry level trends in production and the connection to expenditure are rather prolific, a gap remains when estimating the current, direct relationship between R&D expenditure and the number of pharmaceutical products that a firm is able to produce within the largest global competitors.

Hashimoto et al. takes an initial approach at analyzing more recent connections between pharmaceutical expenditure and productivity, looking at just the Japanese pharmaceutical industry. They concluded that "R&D efficiency of the Japanese industry… has surely gotten worse almost monotonically" (2005). Using Data Envelopment Analysis, Hashimoto et al. concluded that innovative productivity within the entire industry was declining, using R&D expenditure as the relevant input and patents, pharmaceutical sales, and operating profit as the relevant outputs. Looking at four different sets of panel data, they used different lags for R&D to account for the effect of expenditure to be fully realized. Overall, the methodology within the model was valid but was limited in regard to the fact that it only assessed the Japanese pharmaceutical industry. While informative, the assessment of the Japanese pharmaceutical market cannot be extrapolated to the global market as Japan operates in different ways as

compared to other firms located internationally. But the conclusion from Hashimoto et al. was consistent with Munos' overview of historic trends in pharmaceutical innovation.

Looking at modern global firms, Langowitz and Graves conducted an analysis of innovative productivity in which they assessed 31 pharmaceutical companies over a 19-year period to assess the relationship between R&D expenditures and the number of NCEs (new chemical entities) produced. Finding a clear positive relationship between the two, they controlled for average expenditures 6 and 3 years prior, FDA approval time and firm size (1992). However, the use of NCEs over NMEs is limiting as it only allows for products that contain no active moiety that has ever been approved. NMEs account for a greater amount of new production. Additionally, Acemoglu and Linn assessed the impact of "market size on the entry of nongeneric drugs and new molecular entities" (p. 1049) and found a statistically significant, positive relationship indicating that market size serves as a valuable control in pharmaceutical production (2004). Ultimately, both Acemoglu & Linn and Langowitz & Graves assess useful covariates that can be included in further analyses of productivity.

For this paper, I seek to analyze the power players of the pharmaceutical industry and their relative innovative productivity using similar inputs and methodology to Langowitz and Graves (1992) though in a more modern context. Though insightful, their study was limited to the years prior to the exponential increase in R&D expenditure and assessed the impact on NCEs rather than NMEs. Therefore, I will use their suggested controls of FDA approval time and firm size as well as a control for potential market size from Acemoglu and Linn (2004) as additional, relevant inputs. These controls, in addition to the inputs and outputs from Hashimoto et al. allow for a renewed look at the current relationship between R&D expenditures and productivity and provides a space for contributing further to the literature.

3 Data

3.1 Data Sources

My empirical analysis is based off of NME data collected from the FDA's archives as well as company specific financial and internal data from the COMPUSTAT database. The sample consists of several large pharmaceutical companies that are listed as members of PhRMA[2] or have subsidiaries listed within the organization. Assessing the largest pharmaceutical companies within the PhRMA organization as the units of observation provides a relatively homogeneous sample of companies who are all committed to investing a significant proportion of their total revenue to research and development.

3.2 Company Selection

The companies within this sample include Abbott Labs, AbbVie, Alexion, Alkermes PLC, Allergan, Amgen, AstraZeneca, Bayer, Biogen, BioMarin, Bristol-Myers Squibb, Celgene, Eisai Inc, Eli Lilly, Genentech, Gilead, GlaxoSmithKline, Incyte Corporation, Johnson & Johnson, Merck & Co., Novartis, Novo Nordisk, Pfizer, Roche Holding AG, Sage Therapeutics, Sanofi, and Teva Pharmaceuticals.

These companies were selected for two reasons - annual revenue and membership within the PhRMA organization as of December 2019. In selecting the top pharmaceuticals based on firm total revenue, I sought to find the companies that not only serve as the principal players in the international pharmaceutical market, but also those that all have the capital to invest within the development of new pharmaceutical products. The use of PhRMA members was considered

[2] The Pharmaceutical Research and Manufacturers of America

due to the fact that PhRMA has strict requirements regarding the percent proportion of R&D expenditure out of total sales. All members are required to meet a minimum requirement of a "three-year average global R&D to global sales ratio of 10 percent or greater [and a] three-year average global R&D spending of at least $200 million per year" (PhRMA). Therefore, all parties are committed to innovation and development and would be highly unlikely to make most of their profit by simply producing generic drugs. Not all PhRMA members nor all firms within the Top 25 largest pharmaceutical firms by revenue were used due to either only meeting one of the two selection criteria or from a lack of data. Some private and international companies were not available within the COMPUSTAT database, and others were not members of the PhRMA organization.

3.3 New Molecular Entities

A new molecular entity "contain(s) active moieties that have not been approved by (the CDER[3]) previously" (CDER) and is used as a measurement of the number of new, non-generic products produced by pharmaceutical companies. To identify the number of new molecular entities produced by a company in a given year, I pulled lists of approved NMEs each year from the FDA archives and then used the FDA Orange Book to assign each to their applicant and developer company. If multiple companies co-developed a product, all were credited with 1 NME each for that entity. As most companies produce either 0 or 1 NMEs in any given year (see Figure 4 below), RNME was then defined as an indicator variable of whether or not a company developed any NMEs in a particular year, with 1 indicating 1 or more NMEs developed and 0 indicating none. NMEs were used as the outcome of choice rather than NCEs (new chemical

[3] Center for Drug Evaluation and Research

entities) due to the fact that the requirements for NCEs are too restrictive. NCEs are drugs that contain no active moiety that has ever been approved before by the FDA, thereby significantly restricting what constitutes productivity.

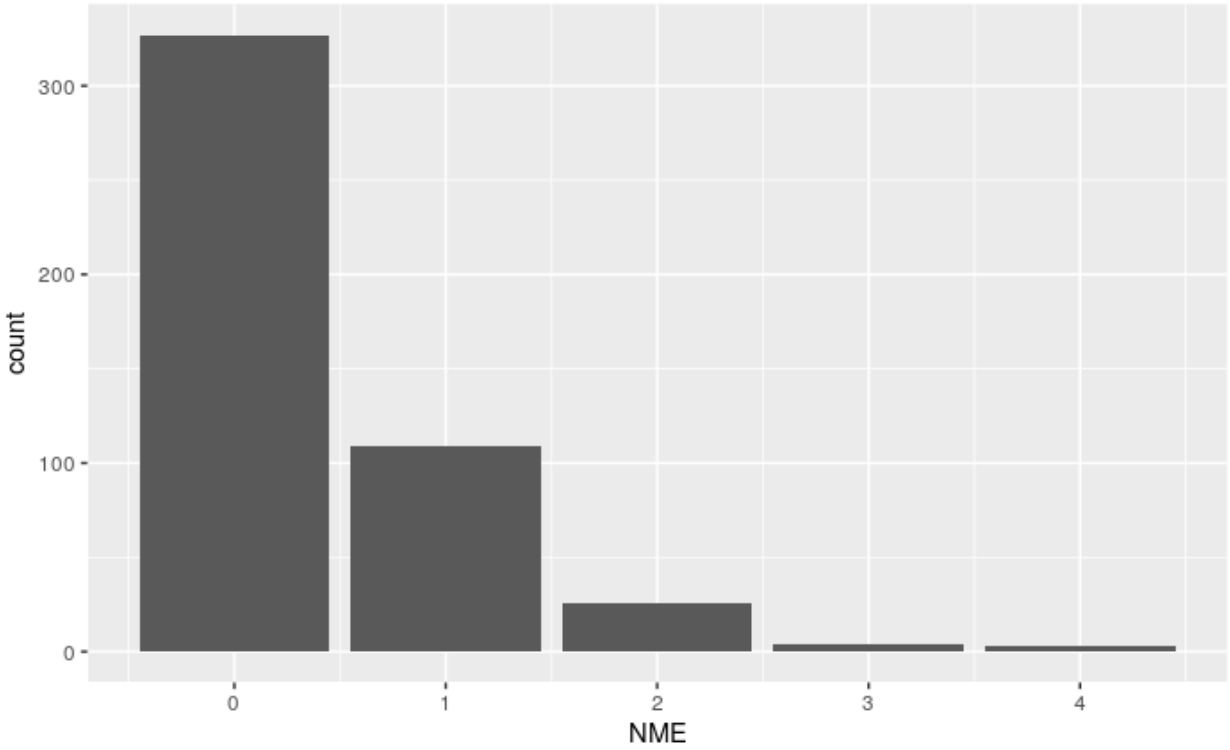

Figure 4: Distribution of number of NMEs developed per year within sample period

3.4 FDA Median Approval Time

FDA median approval time data was also collected from the FDA archives. The standard approval time for every year was logged from the database to the extent of the availability of the data. Every year, the FDA varies slightly in the median approval time for a standard new drug application (NDA) from the time that the application is submitted. The median approval time was used for standard applications because it provides a more uniform metric for regulatory stringency than priority applications. Additionally, a more significant number of approved NDAs are standard rather than priority. While an imperfect proxy for regulatory stringency, FDA

median approval times provide the closest, standardized proxy for all NDAs that is inherently company invariant.

3.5 Descriptive Statistics

From the 27 units of observation, I compiled a panel data set from 1999 to 2016 that includes the number of NMEs approved (the main outcome variable) as well as R&D expenses (the main independent variable) along with controls for median FDA approval time and median FDA review time as proxies for regulatory stringency and employees as a proxy for firm size. However, as not all the companies have the full time period extent of data available, there is an unbalanced panel of 441 total observations. Table 1 below provides general descriptive statistics on the key variables involved. All values are adjusted for inflation.

Table 1: Key Descriptive Statistics

	Mean	*Std. Dev.*	*Median*	*Min*	*Max*
Year	2007.585	5.166118	2008	1999	2016
Employees (thousands)	43.95398	41.89611	31.2	0.09	135.696
Revenue (millions)	22250.18	20611.85	17997.9	1.1442	75358.35
R&D Expenditure (millions)	3423.429	3109.03	2906.319	28.3025	15894.85
R&D Intensity	0.346220	0.256362	0.274127	0.065750	2.049634
NMEs	0.367347	0.664848	0	0	4
FDA Median Approval Time	12.36168	1.624222	12	10	15.4
RNME	0.287982	0.453337	0	0	1
Observations	*441*				

4 Empirical Strategy

4.1 Model

Given a binary dependent variable and the focus of impact of R&D expenditure on NME development, a linear probability model (LPM) was the primary consideration. With an LPM, the interpretation of the association between R&D expenditure and NME development is simple with a one unit increase in R&D being associated with a β_1% increase in the probability that RNME = 1. However, with the possibility that the values of RNME be unreasonable as in less than 0 or greater than 1, I found that utilizing a logit model would be better suited for our binary dependent variable. The benchmark equation is:

$$RNME_{it} = \beta_0 + \beta_1 RD_{it} + \beta_2 AT_t + \beta_3 RDI_{it} + \beta_4 Emp_{it} + \beta_5 Rev_{it} + \alpha_i + \gamma_t + u_{it}$$

The outcome variable $RNME_{it}$ serves as an indicator variable for whether or not an NME is developed, RD_{it} represents research & development expenditure (in the millions), AT_t represents the median FDA approval time for new drugs (in months), RDI_{it} represents R&D expenditure as a proportion of total expenditure (in the millions), Emp_{it} represents the number of employees in the firm (in the thousands), and Rev_{it} represents total revenue (in the millions). Errors are robust. Entity fixed effects (α_i) and time fixed effects (γ_t) serve as a check to control for variation across companies and variation in NMEs over time respectively.

4.2 Further Robustness Checks

A potential robustness check includes comparison between the LPM, Probit, and Logit models to provide a check of the direction and relative magnitude of the effect of R&D expenditure. Additionally, a check using 3-year to 6-year lag accounts for the delay in the impact of R&D expenditure on future development. Three- to six-year lags were chosen to capture the

bulk effect of R&D as was suggested by Langowitz and Graves (1992). An additional robustness check considers taking the natural log of different independent variables to standardize the values of R&D expenditure and R&D intensity. As many companies vary in size and revenue, taking the natural log normalizes the data and has a more reasonable interpretation. Finally, a simple subgroup analysis on the companies that primarily specialize in prescription drugs, rather than consumer products or delivery systems, were analyzed to determine whether the production of other products could impact the direct R&D and RNME association.

5 Results

5.1 Main Results

The coefficients of the key regressions are displayed below in Table 2.

Table 2: Key Regression Results

Variables	(1) RNME	(2) RNME	(3) RNME	(4) RNME	(5) RNME
R&D Expenditure	-8.29e-05	-8.29e-05	-1.64e-05	5.99e-05	3.66e-05
	(9.66e-05)	(9.66e-05)	(0.000111)	(0.000129)	(0.000146)
Constant	-1.940***	-1.765	-2.117	-1.415	-1.461
	(0.517)	(2.675)	(2.659)	(2.682)	(2.625)
Median Approval Time		-0.0127	0.0251	0.0352	0.0387
		(0.215)	(0.213)	(0.213)	(0.209)
R&D Intensity			-1.704*	-2.018*	-1.868
			(1.001)	(1.189)	(1.285)
Employees				-0.0135	-0.0166
				(0.00952)	(0.0125)
Revenue					1.07e-05
					(2.46e-05)
R&D Exp. Odds Ratio	.9999171	.9999171	.9999836	1.00006	1.000037
	(9.66e-05)	(9.66e-05)	(0.000111)	(0.000129)	(0.000146)

ctrl Year FE	yes	yes	yes	yes	yes
ctrl Company FE	yes	yes	yes	yes	yes
Pseudo R-Squared	0.1976	0.1976	0.2026	0.2045	0.2049
Observations	441	441	441	441	441

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

All regressions were run with the clustered Huber Sandwich Estimator as well as year and company fixed effects. Columns 2 through 5 sequentially estimate different model specifications using controls for median FDA approval time, R&D intensity, Employees and Revenue as seen on the left-hand side. The coefficient for the main variable of interest was not statistically significant in any of the specifications that were run. The results are consistent with the hypothesis that R&D expenditure has no significant effect on the number of new molecular entities produced by a given company in a given year. Notably, when running the model without entity and time fixed effects as seen in Table 8 in the Appendix, there is a statistically significant relationship between R&D expenditure and NME production for several of the regressions. This difference demonstrates the importance of fixed effects and the impact that it has on standard error. After taking fixed effects, there is not much variation in the regression therefore leading to no significant relationship between the two main variables.

Table 2 above also indicates the same regressions but with an estimation of the odds ratio rather than the coefficients. Aside from the statistical significance, the magnitude of our odds ratio is also incredibly small indicating that the impact is rather negligible. The direct interpretation of my findings is that a one unit increase in the value of R&D expenditure (an increase by $1 million) indicates that the odds of developing an NME increases by 1.000037.

This value is hardly different from 1 and is not statistically significant. The other estimates have similar odds ratios, none of which indicate a particularly large impact of R&D expenditure on NME production.

The lack of a statistically significant relationship can be explained by the fact that often times the development of NMEs is contingent on factors beyond that of simply having the capital to invest, as well as the fact that our fixed effects capture a significant proportion of the variation in NME development. Regarding external factors, while having enough capital and funding is a foundational component of NME production, having other resources and people who contribute to the development process is what spurs innovation and new ideas. This mirrors a more recent trend within the pharmaceutical industry that pushes for a greater number of mergers and acquisitions of smaller, more innovation-driven biotechnology firms rather than dedicating internal talent and capita to innovating and creating new products (Shepherd 2017). Unfortunately, as of now I have neither the data nor the ability to test innovative talent within a pharmaceutical company. This mechanism could be tested in the future by quantifying external talent gained through acquisition of another firm or by quantifying the quality of innovation in firms that are being acquired.

5.2 Effect Size Confidence Intervals

As noted in the previous section, the odds ratio for each factor is presented, though they should all be interpreted with caution. Ultimately, as our sample size is relatively small, our results for the odds ratio for each factor is variable. None of the five estimates are statistically significant, indicating that the 95% confidence interval for each contains 1. As such, we can estimate with 95% confidence, that a 1 unit increase in R&D expenditure (in millions) will result in a true odds ratio between 0.9997512 and 1.000322. Employees and revenue have similar odds

ratio confidence intervals of [0.9596308, 1.008008] and [0.9999625, 1.000059], respectively. Median approval time and R&D intensity, however, have relatively wider confidence intervals. A one unit increase in median approval time (in months) will result in an odds ratio between 0.6902802 and 1.565256 with a half-width of 0.44. A one unit increase in R&D intensity will result in an odds ratio between 0.012444 and 1.915575 with an associated half-width of 0.95. The results are further displayed below.

Table 3: Estimated Odds Ratio Confidence Intervals

	OR	*Lower Bound*	*Upper Bound*
R&D Expenditure	1.000037	0.999751	1.000322
Median Approval Time	1.039454	0.690280	1.565256
R&D Intensity	0.154394	0.012444	1.915575
Employees	0.983522	0.959631	1.008008
Revenue	1.000011	0.999963	1.000059

6 Robustness Checks

6.1 Logit, Probit & Linear Probability Models

As a check in determining whether the form of my model directly resulted in an alternative outcome, I ran a probit model as well as a linear probability model to investigate whether either of the two models resulted in significantly different results from the main logit model. As shown below in Table 4, none of the coefficients were statistically significant and moreover, all values were incredibly small in magnitude. The linear probability model produced

a coefficient for R&D expenditure of 6.23e-07 which is not statistically significant. The probit and logit models had slightly larger coefficients but were still incredibly small and the odds ratio values were also not statistically significant.

Table 4: Logit, Probit & Linear Probability Model Regression Results

Variables	**LPM** RNME	**Logit** RNME	**Probit** RNME
R&D Expenditure	6.23e-07	3.66e-05	3.02e-05
	(2.44e-05)	(0.000146)	(8.20e-05)
Median Approval Time	0.00654	0.0387	0.0257
	(0.0348)	(0.209)	(0.122)
R&D Intensity	-0.160**	-1.868	-1.132*
	(0.0765)	(1.285)	(0.664)
Employees	-0.00335	-0.0166	-0.0105
	(0.00277)	(0.0125)	(0.00760)
Revenue	3.26e-06	1.07e-05	6.23e-06
	(4.76e-06)	(2.46e-05)	(1.41e-05)
Constant	0.217	-1.461	-0.827
	(0.460)	(2.625)	(1.575)
ctrl Year FE	yes	yes	yes
ctrl Company FE	yes	yes	yes
R-Squared	0.228	-	-
Pseudo R-Squared	-	0.205	0.205
Observations	441	441	441

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6.2 Log Regression

Due to the fact that many of the companies that I am analyzing vary in their total revenue and firm size, taking a log of several of the dependent variables provides better interpretability for the direct effect of R&D expenditure on NME development. Column 1 is our main logistic regression. Column 2 runs our logistic model while using the log of our main dependent variable, R&D expenditure. Column 3 adds in further logs of our firm descriptors and employee total

R&D revenue. While the magnitude of our results certainly changed from running the log of variables, the statistical significance of our results did not change. Our estimates for the impact of R&D expenditure on NME production are not statistically significant indicating that there isn't a significant relationship between expenditure and production.

Table 5: Log Regression Results

Variables	(1) RNME	(2) RNME	(3) RNME
R&D Expenditure	3.66e-05		
	(0.000146)		
Log R&D Expenditure		0.292	0.161
		(0.253)	(0.597)
Median Approval Time	0.0387	0.111	0.116
	(0.209)	(0.213)	(0.211)
R&D Intensity	-1.868	-1.932*	-1.990
	(1.285)	(1.036)	(1.377)
Employees	-0.0166	-0.0164	
	(0.0125)	(0.0125)	
Log Employees			0.345
			(0.788)
Revenue	1.07e-05	8.34e-06	
	(2.46e-05)	(2.15e-05)	
Log Revenue			-0.167
			(0.261)
Constant	-1.461	-4.417	-4.193
	(2.625)	(3.989)	(4.105)
ctrl Year FE	yes	yes	yes
ctrl Company FE	yes	yes	yes
Pseudo R-Squared	0.2049	0.2066	0.2043
Observations	441	441	441

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6.3 Lagged R&D Expenditure

Additionally, an important consideration is that the development of an NME in a given year is likely dependent on previous years' R&D Expenditure rather than just the expenditure in a given year. As a result, I use 3 to 6 year lags for R&D Expenditure to account for the possibility that prior years of R&D expenditure are likely to have a greater impact on the actual development of an NME that is in the pipeline rather than the R&D expenditure of that given year. Our results, again, are not statistically significant and do not vary tremendously in magnitude as compared to the baseline model without any lags in expenditure.

Table 6: Lag Regression Results

Variables	(No Lag) RNME	3-Year Lag RNME	4-Year Lag RNME	5-Year Lag RNME	6-Year Lag RNME
R&D Expenditure	3.66e-05	2.05e-05	2.52e-05	-1.69e-05	0.000139*
	(0.000146)	(5.11e-05)	(7.92e-05)	(4.50e-05)	(7.48e-05)
Median Approval Time	0.0387	0.241	0.278	0.532*	0.200
	(0.209)	(0.266)	(0.285)	(0.316)	(0.296)
R&D Intensity	-1.868	-1.758*	-1.741	-1.303	-4.651**
	(1.285)	(1.019)	(1.286)	(0.886)	(2.332)
Employees	-0.0166	0.00703	0.00399	-0.0175	0.00244
	(0.0125)	(0.0212)	(0.0178)	(0.0167)	(0.0240)
Revenue	1.07e-05	8.40e-06	-7.71e-06	2.04e-05	2.82e-05
	(2.46e-05)	(2.89e-05)	(2.64e-05)	(3.11e-05)	(3.61e-05)
Constant	-1.461	-5.126	-4.983	-7.294*	-5.775
	(2.625)	(3.755)	(3.739)	(4.245)	(4.106)
ctrl Year FE	yes	yes	yes	yes	yes
ctrl Company FE	yes	yes	yes	yes	yes
Pseudo R-Squared	0.2049	0.2252	0.2131	0.2335	0.2379
Observations	441	348	317	310	280

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6.4 Subgroup Analysis

Furthermore, we can assess whether or not limiting the sample to companies that focus solely on producing pharmaceutical products may affect our results. The inclusion of companies such as J&J or Abbott who produce a multitude of consumer products may have skewed our results. When limiting the sample to companies whose focus is producing pharmaceutical products, we get the following results. Although the numerical values may differ compared to the original, our results are still not statistically significant with odds ratios at essentially 1. Ultimately, it is apparent that the exclusion of multifaceted companies who focus on a multitude of products besides drugs and generics has little impact within the context of our model.

Table 7: Subgroup Analysis Results

Variables	All Values RNME	Subgroup RNME
R&D Expenditure	3.66e-05	-3.11e-05
	(0.000146)	(0.000195)
Constant	-1.461	3.504
	(2.625)	(3.251)
Median Approval Time	0.0387	-0.326
	(0.209)	(0.250)
R&D Intensity	-1.868	-2.006
	(1.285)	(1.478)
Employees	-0.0166	-0.0206
	(0.0125)	(0.0211)
Revenue	1.07e-05	1.81e-05
	(2.46e-05)	(3.76e-05)
R&D Exp. Odds Ratio	1.000037	0.9999689
	(0.000146)	(.0001952)
ctrl Year FE	yes	yes
ctrl Company FE	yes	yes
Pseudo R-Squared	0.2049	0.2707
Observations	441	302

6.5 Limitations of the Model

Ultimately, despite the inclusion of internal firm controls, the model is likely to suffer from an endogeneity issue such as omitted variable bias from not controlling for market size as proposed by Acemoglu and Linn (2004). They found "a large effect of potential market size on the entry of nongeneric drugs and new molecular entities" (Acemoglu & Linn, 2004, p. 1049). Market size is likely to be correlated with our proxies for firm size as firms typically grow as the market for their products grow as well. Further research should take this endogeneity factor into account. Given more time, it would be appropriate to explore other possible omitted variables.

Additionally, it would be helpful to work with a larger dataset. Initially, I had planned to assess the impact over a period of 25 years (1993 to 2018) to fully capture the growing rise in R&D expenditure. Unfortunately, the proxy for regulatory stringency (FDA Approval Time) does not extend past 2016 and information on the specific breakdown of developers of NMEs is not available prior to 1999 on the FDA archives. While this data surely exists, given my resources it was unavailable. Also, not every NME in our 17-year period was accounted for as only 27 companies were analyzed. A larger overall data set would more accurately capture the effects before and after the rise of R&D expenditure.

Furthermore, the complex nature of R&D expenditure makes it inherently difficult to model. There are many factors besides funding that contribute to the success or failure of R&D. While financing is a major component, most pharmaceutical companies operate from a long-term perspective with a focus on blockbuster drugs that typically take several years to develop. My analysis doesn't take into account the relative impact of each NME developed, but rather the quantity which is a clear limitation. Further research in this subject could be better in addressing how large of an impact different long-term R&D models in the pharmaceutical industry have.

7 Conclusion

The focus of this paper was to identify the impact and productivity of research and development spending within the pharmaceutical industry. Therefore, the main question was whether or not there is a relationship between a company's R&D expenditure and that company's likelihood of developing an NME or new molecular entity. Over the past twenty years there has been a steady increase in the amount of R&D expenditure within the pharmaceutical industry but no such growth within the development of NMEs. I assessed this relationship by analyzing a group of twenty-seven high-revenue pharmaceutical firms over the period of 1999 to 2016 in a panel dataset. The findings were small in magnitude and not statistically significant indicating that there is no clear relationship between R&D expenditure and the likelihood of developing an NME. Thus, this paper adds to the earlier mentioned dichotomy of results regarding whether or not R&D expenditure truly impacts the development of NMEs.

One interesting aspect of my results was that the inclusion of R&D intensity as a control resulted in a change of sign for our key independent variable. While the regressions prior to including R&D intensity showed a nonsignificant positive coefficient for R&D expenditure, the regressions following the inclusion showed a nonsignificant negative coefficient. Further research could attempt to assess the relationship between R&D intensity and NME production and why the inclusion of such a variable would impact our results in such a way. Additionally, considering PhRMA has strict requirements for R&D intensity, it would be interesting to compare firms operating above and below that threshold to see if there is a significant difference in the outcomes.

8 References

Acemoglu, D., & Linn, J. (2004). Market size in innovation: theory and evidence from the pharmaceutical industry. *The Quarterly journal of economics*, *119*(3), 1049-1090.

Adams, C. P., & Brantner, V. V. (2006). Estimating the cost of new drug development: is it really $802 million?. Health affairs, 25(2), 420-428.

Bottazzi, G., Dosi, G., Lippi, M., Pammolli, F., & Riccaboni, M. (2001). Innovation and corporate growth in the evolution of the drug industry. *International journal of industrial organization*, 19(7), 1161-1187.

Brown, M. G., & Svenson, R. A. (1998). RTM Classic: Measuring R&D Productivity. *Research-Technology Management*, *41*(6), 30-35.

CDER. (2019, November 14). CDER's New Molecular Entities and New Therapeutic Biological Products. Retrieved November 20, 2019, from https://www.fda.gov/drugs/development-approval-process-drugs/new-drugs-fda-cders-new-molecular-entities-and-new-therapeutic-biological-products.

Coad, A., & Rao, R. (2008). Innovation and firm growth in high-tech sectors: A quantile regression approach. *Research policy*, 37(4), 633-648.

Cockburn, I. M. (2004). The changing structure of the pharmaceutical industry. *Health Affairs*, *23*(1), 10-22.

Graves, S. B., & Langowitz, N. S. (1993). Innovative productivity and returns to scale in the pharmaceutical industry. *Strategic Management Journal*, *14*(8), 593-605.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of economics*, 10(1), 92-116.

Gronde, T. V. D., Uyl-de Groot, C. A., & Pieters, T. (2017). Addressing the challenge of high-

priced prescription drugs in the era of precision medicine: A systematic review of drug life cycles, therapeutic drug markets and regulatory frameworks. PLoS One, 12(8), e0182613.

Hashimoto, A., & Haneda, S. (2008). Measuring the change in R&D efficiency of the Japanese pharmaceutical industry. *Research policy*, *37*(10), 1829-1836.

Jensen, E. J. (1987). Research expenditures and the discovery of new drugs. *The Journal of Industrial Economics*, 83-95.

Keyhani, S., Wang, S., Hebert, P., Carpenter, D., & Anderson, G. (2010). US pharmaceutical innovation in an international context. *American Journal of Public Health*, *100*(6), 1075-1080.

Langowitz, N. S., & Graves, S. B. (1992). Innovative productivity in pharmaceutical firms. *Research-Technology Management*, *35*(2), 39-41.

Mansfield, E. (1981). Composition of R and D expenditures: relationship to size of firm, concentration, and innovative output. *The Review of Economics and Statistics*, 610-615.

Munos, B. (2009). Lessons from 60 years of pharmaceutical innovation. *Nature reviews Drug discovery*, *8*(12), 959.

PhRMA Board of Directors Establishes New Membership Criteria. (2017, May 9). Retrieved December 2, 2019, from https://www.phrma.org/en/Press-Release/.

Scherer, F. M. (2001). The link between gross profitability and pharmaceutical R&D spending. *Health Affairs*, *20*(5), 216-220.

Shepherd, J. (2017). Consolidation and innovation in the pharmaceutical industry: the role of mergers and acquisitions in the current innovation ecosystem. *Journal of Health Care Law & Policy, Forthcoming.*

9 Appendix

9.1 Regression Results without Company & Year Fixed Effects

Table 8: Key Regression Results without Company & Year Fixed Effects

Variables	(1) RNME	(2) RNME	(3) RNME	(4) RNME	(5) RNME
R&D Expenditure	0.000186***	0.000188***	0.000165***	0.000173***	0.000141
	(3.75e-05)	(3.91e-05)	(3.97e-05)	(5.91e-05)	(8.99e-05)
Constant	-1.603***	-1.852**	-1.331	-1.324	-1.426
	(0.208)	(0.875)	(0.917)	(0.934)	(0.988)
Median Approval Time		0.0196	0.0236	0.0252	0.0295
		(0.0653)	(0.0641)	(0.0602)	(0.0619)
R&D Intensity			-1.557**	-1.617**	-1.499*
			(0.787)	(0.727)	(0.827)
Employees				-0.000734	-0.00444
				(0.00594)	(0.00850)
Revenue					1.26e-05
					(2.55e-05)
R&D Expenditure Odds Ratio	1.000186	1.000188	1.000165	1.000173	1.000141
	(3.82e-05)	(4.09e-05)	(4.26e-05)	(6.24e-05)	(0.000110)
ctrl Year FE	no	no	no	no	no
ctrl Company FE	no	no	no	no	no
Pseudo R-Squared	0.0581	0.0582	0.0727	0.0727	0.0736
Observations	441	441	441	441	441

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1