Spending on New Drug Development

Christopher P. Adams[♦]

Van V. Brantner[♣]

February 5, 2007

Abstract

This paper replicates DiMasi et al (2003, 2004) estimates of expenditure on new drug development using publicly available data. The paper estimates that average expenditure on drugs in human clinical trials is around $25m per year, with $18m per year on drugs in Phase 1, $33m on drugs in Phase 2 and $19m per year on drugs in Phase 3 of the human clinical trials. The paper's estimated expenditure on new drug development is somewhat greater than suggested by the survey results presented in DiMasi et al (2003, 2004). The paper combines a twelve year panel of research and development expenditure for 183 publicly traded firms in the pharmaceutical industry with panel of drugs in human clinical trials for each firm over the same period. The paper estimates drug expenditure by estimating the relationship between research and development expenditure and the number of drugs in development for 1,682 company/years (183 firms multiplied by the number of years for which we have financial and drug development information). The paper also estimates expenditure on drugs in various therapeutic categories. The paper confirms earlier work by the authors that $802m may under estimate the average amount needed to bring a new molecular entity to market. It also confirms that there is significant variation in spending by therapeutic category.

1. Introduction

DiMasi et al (2003, 2004) estimate the cost of new drug development for all drugs and for drugs in certain therapeutic categories, respectively. The authors estimate the average cost of new drug development to be $802m per new drug. This number has become a central part of the policy debates on numerous policy issues regarding the pharmaceutical industry including the Medicare Prescription Drug Act, drug importation, generic entry and vaccine development. Drug companies argue the high cost of drug development justifies the high prices paid by governments, insurers and customers. Given the importance of the $802m number to the debate it is important to know whether it is correct and what it means.

DiMasi et al (2003) calculate the cost of new drug development with data from two sources. The authors survey ten large pharmaceutical firms and ask those firms to report the expenditure in human clinical trials for 68 drugs chosen at random from the Tuft's drug development database called the CSDD. The authors then use information on average success rates and successful durations from the CSDD data to calculate the cost of bringing a new drug to market. Recently, Light and Warburton (2005) point out numerous problems with DiMasi et al (2003). In particular, because "cost data used was proprietary and confidential, readers cannot know how each company collected its data, or what was counted as research costs, and no independent verification of the accuracy of the information is possible" (p. 1031). This paper provides an independent verification of the survey cost data by using an alternative publicly available data source on research and development expenditure. In a companion paper, Adams and Brantner (2006) verify the second part of DiMasi et al (2003) paper by using publicly available data to estimate success rates and average successful durations.

By comparing aggregate annual expenditure on research and development across firms and over time to the number of drugs in human clinical trials for each firm and each year, we can determine the "marginal expenditure" on an additional drug in development. If Drug Firm A spends an additional $50m in 1992 relative to 1991 but in 1992 Drug Firm A has two additional drugs in development we argue this provides an estimate of average annual expenditure by Drug Firm A, i.e. $25m per drug per year. Similarly, if Drug Firm B spends $100m more than Drug Firm A in 1992 but Drug Firm B has an additional four drugs in development in 1992, then we estimate drug expenditure to be $25m per drug per

year. Note that this is an estimate of the correlation between expenditure and the number of drugs in development. We are not attempting to estimate the impact of an additional dollar of expenditure on the number of drugs in development or the impact of additional drug on the amount of expenditure.

There are a number of advantages to this approach. First, we are using publicly available data so our results can be verified by other researchers. Second, we are using data from 183 publicly traded firms rather than ten firms selected by the study's authors. Our selection criteria is that the firms have research and development expenditure information in the CompuStat data base, be in the pharmaceutical industry (see Danzon et al (2004)) and have drugs in the Pharmaprojects data set (see Adams and Brantner (2006)). These firms range in size from 100 employees to almost 180,000 employees with sales ranging from $2m annually to almost $45b annually. Third, we are using contemporaneous reports of research and development expenditure where the reports are scrutinized by both the market and the SEC. In their comment on DiMasi et al (2003), Light and Warburton (2005) argue that

> considering the clear interest of pharmaceutical companies in higher (rather than lower) estimates of drug development costs, and sampled firms' likely awareness of the intended use of the survey data, it is not unlikely that companies would deliberately and systematically overstate costs in their survey responses. (p. 1031)

We argue that such biases are less likely here given the large number of firms and the checks on the reports including audits.

Of course there are also serious concerns about the approach we use here. First, the data is aggregate research and development expenditure. It not only includes expenditure on drugs in human clinical trials it also includes development expenditure on drugs yet to reach trials. To identify the amount spent in human clinical trials we must infer the information from cross sectional and time-series variation in expenditure that is associated with variation in the number of drugs in development. Such variation may lead to spurious estimates. For example, if one firm specializes in cancer drugs and we compare the specialty firm's expenditure on cancer drugs to that of a firm that has just one or two cancer drugs, we may estimate that expenditure on the extra drug is negative - a theoretical impossibility. We attempt to control for some cross sectional variation between firms by conditioning on net sales. Second, we are estimating changes for the "marginal drug" which may be more

3

expensive than the average drug.[1] The relationship between expenditure on the marginal drug and expenditure on the average drug depends on what assumption the reader is willing to make regarding how expenditure per drug changes with the number of drugs. If expenditure per drug is constant then the marginal and the average are the same. On the other hand if expenditure per drug is increasing with the number of drugs in development then marginal expenditure will be higher than average expenditure. That said, marginal expenditure may be a more useful measure for determining the incentive effects of policy changes.

DiMasi et al (2003) uses a similar approach to verify their own estimates. The authors use firm level R&D expenditure reported by PhRMA and estimate lagged expenditure on firm level counts of approved drugs. The authors estimate average expenditure per approved drug to be between $354m and $558m. These numbers are similar to their estimate of $403m using the survey data. Other researchers have simply divided aggregate R&D expenditure by the total number of approvals per year. The concern with these approaches is that less than one in four drugs in human clinical trials actually make it to the market and the process can take between six and twelve years with substantial variation across drugs (Adams and Brantner (2003)).

The rest of the paper proceeds as follows. Section 2 discusses the data used in this study and provides some background information on new drug development. Section 3 presents the results. Section 4 concludes.

2. Data and Background

This paper combines data from two data sources. Information on each firm's research and development expenditure comes from the Standard Poor's CompuStat Industrial file and Global Vantage Industrial Commercial file used by Danzon et al (2004).[2] Information on drugs in development comes from a Pharmaprojects data set used by Adams and Brantner (2006) and Abrantes-Metz et al (2005). The two data sets overlap for the years

---

[1] Thanks to Eric Durbin for pointing this out.
[2] All monetary values are in 1999 dollars using the domestic manufacturing Producer Price Index.

1989 to 2001. The data sets are matched using the name of the pharmaceutical firm.[3] Pharmaprojects updates its information on the firms developing each drug after a merger, so we used text searches of the database and searches of a related data set called the Manufacturing Index to determine the ownership of drugs over time.[4]

According to Danzon et al (2004) there are 383 firms in their original data. Once we match these firms to firms in the Pharmaprojects data we are left with 183 firms. It is not clear exactly why there are firms that do not match. The two data sets do not exactly overlap in time and that may explain some of it. Another explanation is that the Pharmaprojects does not capture name changes or mergers among smaller firms (see Footnote 4). Table 1 presents some basic summary statistics for this sample of firm/year combinations. Table 1 shows there are an average of four drugs in development for each firm for each year 1989 to 2001. Note this measure is not a very good measure of the stock of drugs in development because we only observe drugs entering one of the stages of human clinical trials after 1989. In the average firm/year $264m is spent on research and development, $2,355m is made in sales and there are 11,000 employees. Note that medians are substantially lower than the means suggesting that the distributions are all skewed towards zero.

Table 1: Firm/Year Summary Statistics

Variable	Obs	Mean	Median	Std. Dev.	Max
Number of Drugs	2,245	4	2	6	45
R&D Expenditure ($m)	1,682	264	37	551	4,678
Net Sales ($m)	1,701	2,355	110	5,438	44,611
Employees ('000)	1,537	11	1	25	179

Charts A1 – A3 present the distribution of the number of drugs in human clinical trials per firm year, the amount of R&D expenditure per firm year, and a scatter plot of the two, respectively. The first two charts show that the distributions of drugs and expenditures are heavily skewed to zero. The third chart seems to show a positive correlation between the amount of R&D expenditure per firm per year and the number of drugs in development per firm per year.

---

[3] This matching was done by hand in order for it to be as accurate as possible.
[4] This was done for all mergers involving firms in the Forbes' top twenty of pharmaceutical industry over the period as well as any other major mergers in the pharmaceutical industry.

Chart 1 presents a summary of the research and development process for new drugs. The first stage of drug discovery is commonly called "preclinical development." In this stage pharmaceutical firms analyze thousands of drugs to determine whether one may have an affect on a disease or condition. As candidates are discovered these drugs are tested on animals to determine whether the drug may be safe and effective in human beings. It is estimated that drugs spend over four years in preclinical testing. DiMasi et al (2003) do not have direct survey information on preclinical expenditure because pharmaceutical firms do not track preclinical expenditure by particular drug candidates. Given this and given that the Pharmaprojects data is based on public information and is not very reliable regarding drugs in preclinical development; we do not estimate expenditure on preclinical development.
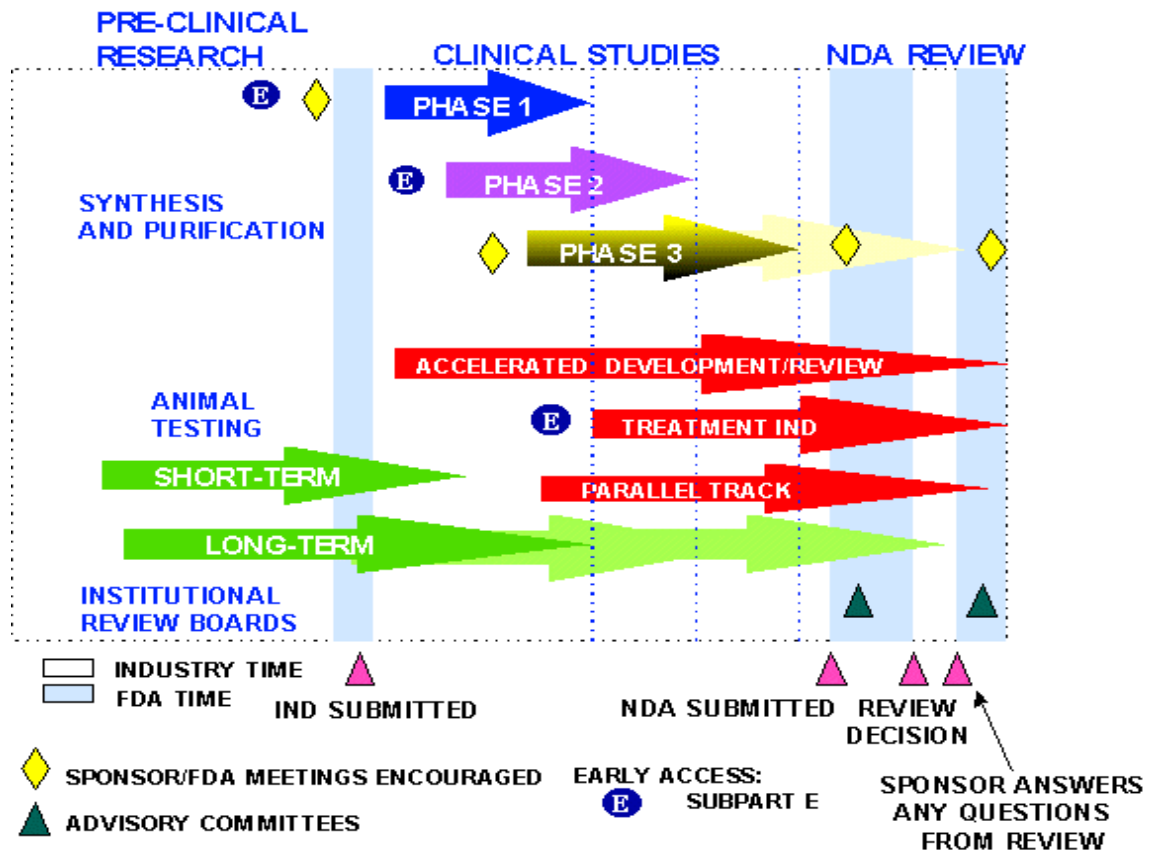
Chart 1: CDER Chart of Drug Review Procedure

After preclinical development the sponsoring firm applies for an Investigation New Drug application (IND) with the FDA in order to test the candidate in humans.[5] There are three steps to human clinical trials. In Phase I, the drug is tested for safety on a small group (e.g. 20) of healthy volunteers. Phase II tests concentrate on safety but the test is on a larger group of patients with the condition (e.g. 200). Phase III are the large efficacy trials with upwards of 3,000 patients participating. Once the trials are completed the results of all three stages are presented to the FDA in the form of a New Drug Application (NDA).

Table 2 presents some basic summary statistics on the drugs owned by the firms in the sample. The table shows that these drugs seem to be fairly representative (see Adams and Brantner (2006) and DiMasi et al (2003)). Although Phase 1 drugs seems to have shorter durations and higher success rates than have been reported elsewhere.

Table 2: Summary Statistics for Drugs

		Obs	Mean
Duration			
	Phase 1	80	15.02
	Phase 2	74	31.20
	Phase 3	48	33.51
Success			
	Phase 1	49	0.85
	Phase 2	46	0.50
	Phase 3	34	0.74

3. Results

3.1 Mean Expenditure Estimates

Table 3 presents regression results for the amount of research and development expenditure on the number of drugs in human clinical trials. There are six regressions reported in the table. First are the basic regressions on the number of drugs in human clinical trials then on the number of drugs in each of the three phases of development. These regressions are then repeated adding measures of time and firm characteristics. All

---

[5] If the firm wants to eventually market the drug in the U.S. the firm must apply for an IND prior to undertaking human trials. That said, there are exceptions.

results report robust standard errors clustering on firm name. The number of drugs is the number of drugs in development for each firm/year combination. The variable time is simply the number of years from 1988. The variable "sales" is the amount of net sales for each firm/year.

Table 3: R&D Expenses OLS (robust standard errors)						
	1	2	3	4	5	6
All Phases	69.76		70.4		24.65	
	(6.06)		(6.13)		(3.24)	
Phase 1		83.26		80.71		17.53
		(17.44)		(17.49)		(9.15)
Phase 2		66.36		67.3		32.88
		(12.26)		(12.37)		(6.39)
Phase 3		61.15		64.43		19.44
		(22.05)		(13.35)		(10.13)
Time			-60.78	-58.33	-22.65	-23.11
			(11.13)	(13.35)	(5.80)	(6.57)
Time ^{2}			4.04	3.88	1.96	1.99
			(0.84)	(0.98)	(0.46)	(0.51)
Sales					0.08	0.08
					(0.01)	(0.01)
Constant	-32.87	-31.48	144.16	137.56	16.18	16.42
	(22.92)	(22.73)	(31.57)	(36.43)	(16.82)	(17.41)
Observations	1682	1682	1682	1682	1682	1682
R^2	0.57	0.57	0.58	0.58	0.89	0.89

Note: Standard errors are clustered on firm names.

Table 3 shows that average expenditure per drug in human clinical trials is between $70m and $25m per year. If sales are not accounted for then Phase 1 expenditure is estimated to be $83m per year, $66m for drugs in Phase 2 and $61m for drugs in Phase 3. These estimates seem very high and are certainly much higher than the amounts presented in DiMasi et al (2003). Once sales are accounted for, these estimates fall to $18m, $33m and $19m respectively.[6] The numbers presented in DiMasi et al (2003) are for the average drug over the length of the phase, while here we have estimated expenditure for one year. If we use the phase durations presented in Table 2 we can estimate expenditure for the whole phase. This procedure gives 1.25*18m = $22.5m for Phase 1, which is more than DiMasi et al (2003) estimate of $15.2m for Phase 1. For Phase 2 the same method produces an

---

[6] Note that these results are most properly thought of as correlations between the number of drugs and the amount of expenditure. There has been no effort made to account for endogeniety in the joint decisions to increase expenditure and take more drugs into clinical trials.

estimate of $86m which is much higher than the DiMasi et al (2003) estimate of $23.5m. Finally, for Phase 3 this method gives an estimate of $54m which is quite a lot less than the DiMasi et al (2003) estimate of $86.3m.

It is not clear what explains such a large discrepancy between our estimate of Phase 2 expenditure and DiMasi et al (2003) estimate. For Phase 1 and Phase 3, the DiMasi et al (2003) estimates lie within the 95% confidence interval around our estimates. But there is no overlap between the confidence interval around the DiMasi et al (2003) estimate of Phase 2 expenditure and the confidence interval around our estimate. One possibility and a more general concern is our method may be misallocating expenditure to drugs in different stages of development. This may occur for two reasons. A first issue is we assume that if a drug moves into a new phase in a particular year then the drug has been in that phase for the whole year. Still, given the expected difference in Phase 2 and Phase 3 expenditure this assumption is more likely to lead to an underestimate of Phase 3 expenditure than an overestimate of Phase 2 expenditure. Another issue is the relationship between financial years and the years assigned in the data. Again, this may reduce the accuracy of the estimates but it is unlikely to bias the estimates. Another possibility that may explain the differences is that there is under reporting of drugs in human clinical trials particularly in the earlier phases.[7]

The results suggest that annual expenditure on research and development is falling by about $22m per firm year, although the rate of decline has also fallen over time. It is not clear how to interpret this number as it is conditional on the number of drugs in development. The results also show that there is a strong relationship between sales and research and development expenditure with every $1 in sales associated with an extra $0.08 in R&D expenditure. Note also, adding sales to the regression substantially improves the model's ability to explain the data. The constant in this estimation cannot really be interpreted as we don't have a measure of the stock of drugs in development as of 1989. We are only able to observe new development starts in 1989 and later.

---

[7] It is not clear exactly how such under reporting would bias the results and in what direction.

Table 4: R&D Expenses Quantiles (standard errors)					
25%	*25%*	*50%*	*50%*	*75%*	*75%*
All Phases 7.65		12.87		16.14	
(0.10)		(0.16)		(0.17)	
Phase 1	7.51		14.13		14.88
	(0.43)		(0.47)		(0.52)
Phase 2	9.79		14.96		21.99
	(0.34)		(0.38)		(0.43)
Phase 3	4.80		8.74		10.59
	(0.42)		(0.48)		(0.59)
Time -2.36	-2.36	-3.85	-3.78	-4.41	-3.65
(0.64)	(0.80)	(0.91)	(0.87)	(0.97)	(1.02)
Time^2 0.17	0.17	0.30	0.29	0.38	0.33
(0.04)	(0.05)	(0.06)	(0.06)	(0.07)	(0.07)
Sales 0.07	0.07	0.08	0.08	0.10	0.10
(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant 1.81	1.97	6.86	6.68	13.82	11.81
(2.10)	(2.62)	(2.96)	(2.83)	(3.08)	(3.23)
Observations 1682	1682	1682	1682	1682	1682
Pseudo R^2 0.59	0.59	0.69	0.69	0.80	0.80

Table 4 presents results from the 25%, 50% and 75% quantile regressions. Comparing these results to the results from columns 5 and 6 in Table 3 we see that expenditure per drug per year is substantially less at the lower quantiles. At the bottom quartile expenditure per new drug in development is around $8m per year, with $13m at the median and $16m at the top quartile. These numbers compare to $25m per year for the mean. Similarly, estimated expenditures per phase of development are substantially lower at the 25%, 50% and 75% quantiles relative to the mean.

As before, we can transform our median estimates to compare to the median estimates presented in DiMasi et al (2003). This procedure gives 1.25*14.13m = $17.66m

for Phase 1, which is more than DiMasi et al (2003) estimate of $13.9m for Phase 1. For Phase 2 the same method produces an estimate of $39.9m which is much higher than the DiMasi et al (2003) estimate of $17m. Finally, for Phase 3 this method gives an estimate of $24.4m which is much less than the DiMasi et al (2003) estimate of $62m.

3.3 Implications for Cost Estimates

If we use the mean estimates for expenditure on drugs in development in place of the survey estimates used by DiMasi et al (2003) we can recalculate the over all "cost of drug development" or more accurately the net revenue needed to make investment in drug development profitable. Doing this calculation using the same durations and success rates as reported in Adams and Brantner (2006) we estimate new drug development cost to be $1,176m, which is much higher than the original estimate of $802m or even the Adams and Brantner (2006) estimate of $867m. If we use the durations and success rates reported in Table 2 then the estimate falls to $1,057m, but still higher than previous estimates. These high estimates may be due to measurement of expenditure on the marginal drug rather than the average drug. However, such an estimate may be more useful to policy makers as it is more likely to measure the impact of changes in policy on the development of new drugs. We could interpret these estimates as stating a firm would need expected net revenue of over $1 billion to develop one more drug for the market.[8]

3.4 Expenditure By Therapy

DiMasi et al (2004) presents estimates of drug development costs for a small number of major therapies. In attempt to replicate this work, Table 4 presents results similar to those presented in Table 3 but where the drug counts are by major therapeutic category. Table 4 presents the marginal cost of a drug by major therapy grouping. Note, that we would not expect these numbers to be negative. Therefore the observed negative coefficients suggest that the model is not estimated very well, although none of the negative

---

[8] Thanks to Mark Duggan for his thoughts about the differences between marginal and average drugs.

coefficients are statistically significantly different from zero. The table shows cardiovascular, dermatological, genitourinary, anticancer and neurological drugs are all have more expenditure per drug in human clinical trials than the average drug. New formulations of existing drugs have substantially smaller expenditure in human clinical trials than the average drug. The reader may be surprised that biotech drugs are estimated to have less than average expenditure. This may be due to imprecise measurement, or it may be due to the way these drugs are categorized in the data. That is, more important biological drugs may be categorized under their indication (anticancer or musculoskeletal) rather than as a product of the biotech industry.

Table 5: R&D Expenses OLS by Therapy for All Phases						
	1		2		3	
	Coefficient	Robust SE	Coefficient	Robust SE	Coefficient	Robust SE
Alimentary/Metabolic	-50.00	36.42	-48.75	35.41	-34.40	24.03
Blood and Clotting	-6.03	57.77	-1.49	54.72	47.73	41.61
Cardiovascular	116.64	30.82	119.89	30.57	35.23	19.41
Dermatological	228.18	109.23	236.75	107.11	92.23	46.84
Formulations	-10.66	11.18	-14.00	10.67	1.54	2.97
Genitourinary	23.06	56.01	15.44	54.40	101.20	42.14
Hormonal	138.87	59.48	138.81	58.29	-4.12	34.37
Immunological	47.12	57.15	44.22	56.62	-25.56	40.61
Anti-Infective	106.66	29.28	106.76	28.70	23.98	14.71
Anticancer	66.69	23.64	61.86	22.12	41.23	14.73
Musculoskeletal	9.62	36.41	11.19	36.01	-18.28	22.72
Neurological	121.01	29.42	123.53	29.38	41.69	17.98
Antiparasitic	281.86	156.80	279.89	148.92	179.07	103.01
Respiratory	151.46	77.65	152.59	76.74	-16.45	31.70
Sensory	-7.64	65.43	11.30	64.43	-3.10	58.53
Biotechnology	15.58	12.79	14.73	12.62	20.21	6.39
Miscellaneous	-8.15	14.28	-7.35	13.62	1.03	6.34
Time			-60.01	11.14	-21.81	5.92
Time ^{2}			4.49	.85	1.91	.46
Sales					0.07	0.01
Constant	8.68	23.21	151.05	32.14	31.27	17.43
Observations	1,682		1,682		1,682	
R^2	0.67		0.68		0.91	

Note: Standard errors are clustered on firm names.

One difference between the results reported here and the expenditure by therapeutic category reported in DiMasi et al (2004) is in relation to anti-infectives such as penicillin. According to DiMasi et al (2004) the expenditure in Phase 3 on these drugs is $137m while Phase 1 expenditure is $23m and Phase 2 expenditure is $20m. Here we find that average

expenditure on anti-infectives is less than for the average drug, and this also holds true for expenditure by phase (not reported). Note however, that our estimate for expenditure on this drug falls dramatically when we account for sales across firms, suggesting that larger firms are spending more per drug on developing anti-infectives.[9] In relation to cardiovascular drugs DiMasi et al (2004) suggests expenditure is slightly below average, our estimate puts expenditure on these drugs at slightly above average although the standard error includes the average.

These results support the argument in Adams and Brantner (2006) that there is substantial variation in drug development costs by therapeutic category. On average annual expenditure by drug is $25m. However, for genitourinary drugs like Viagra, the expenditure number is four times higher. It is twice as high for cancer drugs and neurological drugs. In contrast annual expenditure on new formulations of existing drugs is only a couple of million of dollars. Do therapeutic categories with high annual expenditures have high success rates and short durations to compensate? Not necessarily. Results presented in Adams and Brantner (2006) suggest genitourinary drugs do have very high success rates and relatively short durations. This means that their overall drug development costs may not be too different from "the average drug". However, cancer drugs and neurological drugs have relatively low success rates and long durations making them much more expensive than the average drug (Adams and Brantner (2006)).

4. Conclusion

Recent criticism of the study by DiMasi et al (2003) argues it is not possible to verify the results because the data is confidential. Further, Light and Warburton (2005) argue the sample of expenditure estimates may not be representative and may be biased upwards. This paper attempts to replicate DiMasi et al (2003) expenditure estimates using publicly available data. By matching information on drugs in development with research and development expenditure over the period 1989 to 2001 and across some 180 firms, we infer the additional annual expenditure on each new drug in development. Our results suggest expenditure on Phase 1 and Phase 2 is higher than suggested by DiMasi et al (2003) while expenditure on

---

[9] Note that the sample in DiMasi et al (2003, 2004) only includes large pharmaceutical firms.

Phase 3 is lower. If we combine our estimates in this paper with estimates on success rates and durations from Adams and Brantner (2006) we find that the "cost of drug development" (or the net revenue needed to make investment in new drugs profitable) is over $1 billion and higher than the DiMasi et al (2003) estimate of $802m. As we are estimating expenditure on the *additional* drug we may be estimating the revenue needed to invest in the marginal rather than the average drug. While this is likely to make our estimate higher, it may also make our estimate more useful when considering the consequences of policy changes such as price regulation. This paper also confirms results presented in Adams and Brantner (2006) that there is a substantial amount of variation in expenditure by therapeutic category.

References

Abrantes-Metz, R., Adams, C., and Metz, A. (2005). Pharmaceutical Development Phases: A duration analysis. *Journal of Pharmaceutical Finance, Economics and Policy*, 14(4): 19-42.

Adams, C. and Brantner, V. (2006). Estimating the Cost of New Drug Development: Is it really $802m? *Health Affairs*, March/April, 420-428.

Adams, C. and Brantner, V. (2003). New Drug Development: Estimating Entry from Human Clinical Trials. *FTC Working Paper 262*. July 7, 2003.

Danzon, P., Epstein, A. and Nicholson, S. (2004), Mergers and Acquisitions in the Pharmaceutical and Biotech Industries, November, The Wharton School, University of Pennsylvania.

DiMasi, J., Hansen, R., and Grabowski, H. (2003). The Price of Innovation: New Estimates of Drug Development Costs. *Journal of Health Economics,* 22: 151-185.

DiMasi, J., Hansen, R., and Grabowski, H. (2004). R&D Costs and Returns by Therapeutic Category. *Drug Information Journal*, 38: 211-223.

Light, D. and R. Warburton (2005). Extraordinary Claims Require Extraordinary Evidence. *Journal of Health Economics.* 24: 1030-1033.

Appendix
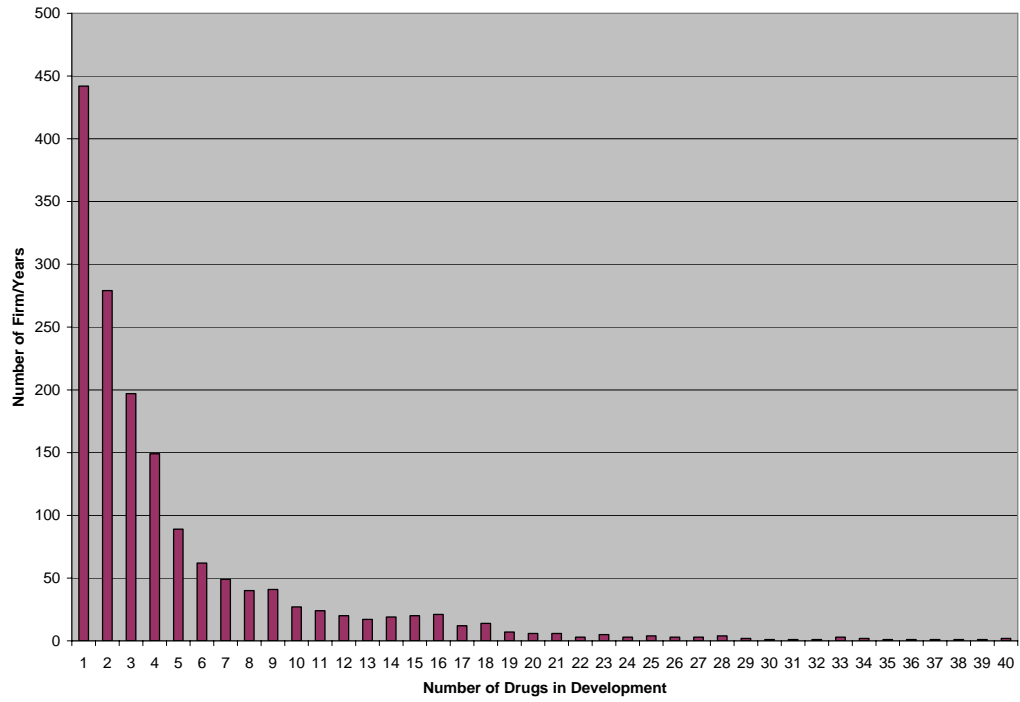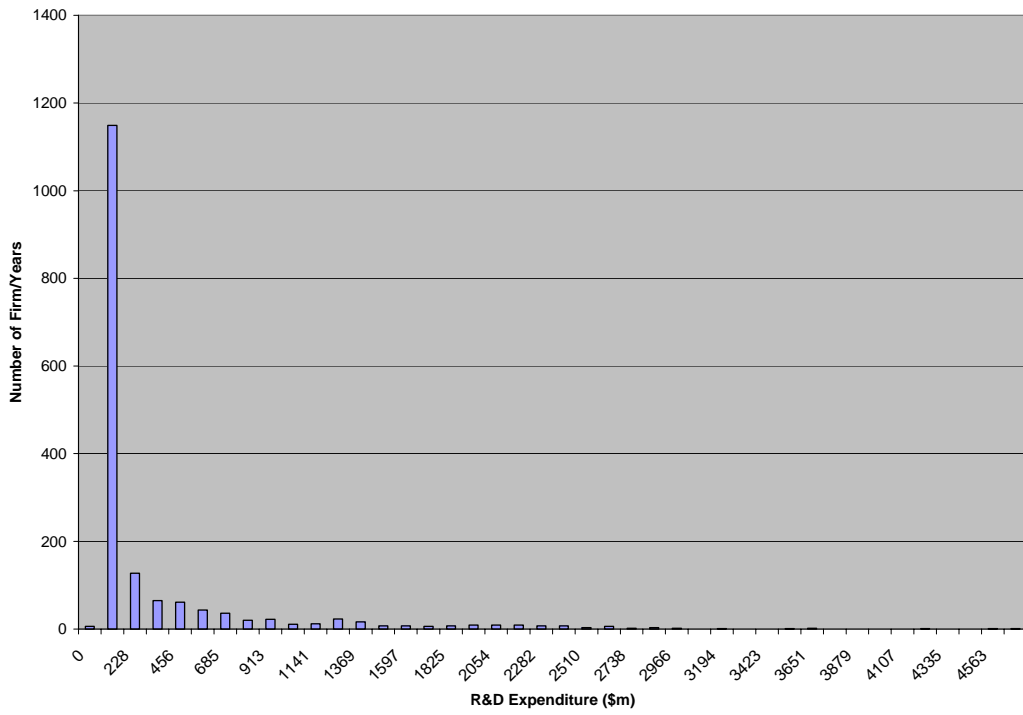
Chart A1: Drugs in Development

Chart A2: Annual R&D Expenditure

Chart A3: R&D Expenditure by Drugs in Development