Academia versus Industry as a Wellspring of New Ideas in Drug Discovery: the Case of Oncology

by

Jorge Cesar Conde

B.A. Biology, The Johns Hopkins University, 1998
M.B.A., Harvard Business School, 2005

SUBMITTED TO THE HARVARD – MIT DIVISION OF HEALTH SCIENCES AND TECHNOLOGY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN HEALTH SCIENCES AND TECHNOLOGY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2006

© 2006 Jorge Cesar Conde. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part.
Academia versus Industry as a Wellspring of New Ideas in Drug Discovery: the Case of Oncology

by

Jorge Cesar Conde

Submitted to the Department of Health Sciences and Technology in partial fulfillment of the requirements for the degree of Master of Science in Health Sciences and Technology

ABSTRACT

The United States population is aging, and the need for novel approaches to treat and manage disease continues to grow. Among the diseases that will impact this population, cancer remains a therapeutic area with significant unmet need. Pharmaceutical and biotechnology industries must continue to meet revenue and income growth expectations and will become increasingly dependent on novel drugs in their pipelines. In order for the pharmaceutical and biotechnology industries to meet the demands of both patients and shareholders, productivity in the research and development process will need to improve significantly.

In order to understand how best to improve the drug discovery and development process, it is important to identify potential sources of innovation throughout the process. Among these, an important consideration is to understand the paths that molecules take through the discovery and development process. This thesis used the marketplace of oncology drugs in development to test the hypothesis that novel molecules largely originate in academia, are developed by biotechnology companies and ultimately are licensed to pharmaceutical companies for commercialization. This thesis analyzed a database of 364 unique oncology small molecules and biologics entering Phase I clinical development between 1991 and 2002.

Thesis Supervisor: Fiona Murray
Title: Associate Professor, MIT Sloan School of Management
Dedication

To the Drs. Conde:
Thank you for your unending inspiration and support

To Cesar and Enrique:
I must be lucky: my best friends,
my most trusted advisors
and my biggest critics
happen to be my two brothers
Acknowledgements

The author wishes to thank the following people for their research assistance, thoughtful discussion, and aid in completing this thesis: Thomas Roberts, Bernardo Goulart, Jeffrey Clark and Stan Finkelstein for generously providing access to the databases on which this thesis is based, as well as POPI and the Sloan Foundation for funding their original work, Lourdes Sosa for contributing her thoughts and work to this thesis, Anthony Sinskey for his advice, and Marsha Warren for her continued guidance.

Professor Fiona Murray (MIT Sloan) has been an exceptional thesis advisor, providing guidance, encouragement, and invaluable insights. Professor Murray has generously provided research materials and many hours of her personal time. Without her help, this thesis would not have been possible.
# TABLE OF CONTENTS

1. Productivity in drug development
   1.1 The need for new drug approvals.......................... 7
   1.2 The biotechnology and pharmaceutical industries........ 10
   1.3 Overview of universities and government funding....... 12
   1.4 The drug development and approval process............... 14
   1.5 Rates of attrition........................................... 17
   1.6 Why do products fail?........................................ 18
   1.7 The war on cancer............................................ 20
   1.8 The role of innovation........................................ 22

2. Hypotheses – the market for ideas in oncology
   2.1 An effective division of labor?.............................. 25
   2.2 From academia to biotech to pharma...................... 27
   2.3 Biotech as innovator......................................... 29
   2.4 In-house discovery and development...................... 29
   2.5 A traditional path to drug development?.................. 30

   3.1 Discussion of source data.................................... 31
   3.2 Database development and rationale...................... 33
   3.3 Measure of validation....................................... 37
   3.4 Molecule Type.................................................. 39
   3.5 Source, sponsor and licensor classification.............. 40

   4.1 Sources and sponsors of innovation...................... 41
   4.2 Validation....................................................... 43
   4.3 Biologics and small molecules............................. 45
   4.4 A snapshot of the oncology drug development landscape.... 48
   4.5 Division of labor – a map of the oncology drug development landscape... 51
   4.6 Potential paths to development.............................. 55

5. Discussion and Conclusions
   5.1 Summary of findings.......................................... 57
   5.2 An ideal path?................................................ 60
   5.3 Limitations....................................................... 61
   5.4 Areas for future research................................... 61

6. Sources and Bibliography........................................ 63
1. Productivity in drug development

1.1 The need for new drug approvals

The pharmaceutical and biotechnology industries, as well as the US government and academic research centers, have and continue to play a central role in the discovery, development and commercialization of drugs to treat and manage disease and improve patient quality of life. Life expectancy in the United States has continued an almost uninterrupted rise since the beginning of the 20th century, increasing from 45 years in 1900 to 77.6 years in 2005 (NCHS, 2005). According to the National Center for Health Statistics, “drugs, both prescribed and over-the-counter, are an increasingly important component of health care. New drugs and new uses for older drugs are improving health outcomes and quality of life, curing some conditions, preventing or delaying disease, and hastening recovery” (NCHS, 2004). Although it would be difficult to quantify the precise impact drugs have had on increased life expectancy, a Columbia University researcher estimates that new medicines account for up to 40% of life expectancy increases (Lichtenberg, 2003). According to the pharmaceutical industry trade association, Pharmaceutical Research and Manufacturers of America (PhRMA), the introduction of new medicines have helped drive down AIDS deaths, reduce major cardiovascular events and increase cancer survival rates (PhRMA, 2006). Despite these advances, however, entire areas of medicine, particularly oncology and neurology, continue to be severely underserved by therapeutic-based treatments. In addition, an aging population, increasing obesity rates and the rise of drug-resistant infectious agents will continue to drive the need for novel drugs.

Shifting demographics are not the only driver for product innovation – new medicines will be needed to maintain industry growth rates for shareholders, while reducing overall healthcare costs to payors and consumers. Despite the advent of high throughput screening technologies,
failure rates in drug development remain high. The cost and risk associated with developing novel drugs often results in successful drug candidates that are priced at a premium to compensate for drug development failures, which is often the justification for the high cost of treatment with patent-protected drugs. Although increased R&D productivity would help address these issues, there is evidence that the FDA approval rates have actually decreased. In 2002, only 17 new chemical entities (NCEs) were approved by the FDA, the lowest rate in a decade. Including the approval of biological license applications (BLAs), in 2002 the total number of approvals was at its lowest point since 1994 (Kola et al, 2004). Prous Science, a drug database and information service, reported that between 1990 and 2000, the year with the lowest number of NCEs approved with a novel mechanism of action was 2000 (Prous, 2002). These trends may be at least partially explained by the fact that industry is targeting increasingly complex diseases or that the regulatory bar has been heightened as new drugs are compared to improved standards of care (Kola et al, 2004). Regardless, companies will increasingly need to find ways to improve the R&D process, either through the discovery of higher quality drug candidates or improvements in the drug screening process to reduce the risk of attrition.

<table>
<thead>
<tr>
<th>2002 sales</th>
<th>Anticipated sales from current products in 2012</th>
<th>Annual real growth target</th>
<th>Sales gap for new products to fill in 2012</th>
<th>Estimated number of NCEs required to fill gap (over ten years)</th>
<th>Year 2012 required NCE output</th>
</tr>
</thead>
<tbody>
<tr>
<td>$45 billion</td>
<td>$30 billion</td>
<td>5%</td>
<td>$43.5 billion</td>
<td>75-90</td>
<td>9.5-11</td>
</tr>
<tr>
<td>$30 billion</td>
<td>$20 billion</td>
<td>5%</td>
<td>$29 billion</td>
<td>50-60</td>
<td>6.5-7.5</td>
</tr>
<tr>
<td>$20 billion</td>
<td>$13.3 billion</td>
<td>5%</td>
<td>$19.3 billion</td>
<td>33-40</td>
<td>4.3-5</td>
</tr>
<tr>
<td>$15 billion</td>
<td>$10 billion</td>
<td>8%</td>
<td>$22 billion</td>
<td>40-45</td>
<td>5.5-6.0</td>
</tr>
<tr>
<td>$15 billion</td>
<td>$10 billion</td>
<td>6%</td>
<td>$17 billion</td>
<td>30-35</td>
<td>4.0-4.5</td>
</tr>
<tr>
<td>$15 billion</td>
<td>$10 billion</td>
<td>5%</td>
<td>$14.5 billion</td>
<td>25-30</td>
<td>3.25-3.75</td>
</tr>
<tr>
<td>$15 billion</td>
<td>$10 billion</td>
<td>4%</td>
<td>$12 billion</td>
<td>20-25</td>
<td>2.5-3.0</td>
</tr>
<tr>
<td>$10 billion</td>
<td>$6.67 billion</td>
<td>5%</td>
<td>$9.67 billion</td>
<td>16.5-20</td>
<td>2.15-2.25</td>
</tr>
</tbody>
</table>

*Adapted from REE 7. All figures in US $. NCE, New Chemical Entity.

Figure 1: Pipeline gap (Kola et al. 2004 p. 711)

The business case for the need for new drug approvals was analyzed by Ismail Kola and John Landis in a 2004 review article in Nature Reviews Drug Discovery (Kola et al, 2004). A survey by Accenture Consulting (Figure 1, above) analyzed the number of new NCEs required
to maintain target revenue growth rates for large pharmaceutical companies (Accenture, 2001). On the basis of Accenture’s calculations, a company with $20 billion in annual sales in 2002 would need four approved NCEs per year to maintain a 5% annual growth target through 2012 (assuming that 2002 products contribute $13.3 billion to 2012 revenue). Considering that, for the entire industry, only 17 NCEs were approved in 2002, a significant increase in productivity would need to be achieved by any one company to consistently meet a goal of four NCEs per year. Historically, large pharmaceutical companies have dealt with the challenge of meeting growth targets through mergers and acquisitions. Consolidation, however, is a short-term solution as the combined company must maintain a similar growth rate off of a larger revenue base. The need for new drug approvals is further increased as blockbuster drugs continue to come off patent.

In order to increase the availability of new drugs, the R&D process will clearly need to undergo significant improvements. Among these, will certainly be the need to enhance the quality of drug candidates entering clinical trials in order to reduce the current rates of attrition in drug development. To put it simply, new drug approvals will ultimately need to be driven by new ideas in the form of novel discoveries and innovative approaches to the treatment of disease. Many have argued that the source of new ideas has traditionally been the industrial R&D laboratories within pharmaceutical companies and, more recently, biotechnology companies. Others cite the productivity of academia in providing new ideas and discoveries to industry that have eventually led to the development and commercialization of new drugs. So, where do the new ideas come from and when and how do they make the transition between academic, biotechnology and pharmaceutical companies? In other words, how is the market for new ideas in pharmaceutical drug discovery and development structured?
1.2 The biotechnology and pharmaceutical industries

The biotechnology and pharmaceutical industries consist of a broad and diverse array of companies developing therapies and diagnostics to improve human health. The pharmaceutical trade association (PhRMA) has nearly 50 member companies and Ernst & Young estimates that there are over 1,400 public and private biotechnology companies in the United States (E&Y, 2006). Taken together, these industries account for the largest aggregate source of funding for biomedical research in the United States. PhRMA estimates that pharmaceutical member companies spent nearly $40 billion in R&D in 2005, with biotechnology and other companies contributing an additional $51 billion. R&D investment is considered an important driver for future growth and spending has increased dramatically over the last 20 years. Today, pharmaceutical companies and profitable biotechnology companies are estimated to spend between 15 and 20% of total sales on R&D every year (PhRMA, 2006).
Figure 2: Investment in R&D (PhRMA, 2006)

The biotech and pharma industries have focused enormous resources on the discovery and development of novel cancer therapies and diagnostics. PhRMA estimates that there are approximately 683 drugs currently in clinical development for cancer indications, which is 150 more than the second therapeutic area (neurologic disorders) and over twice the number of drugs in development for cardiovascular disorders (PhRMA, 2006). It is difficult to estimate precisely the amount of money that is allocated to cancer R&D across the biotech and pharma industries. However, if the proportion of cancer drugs in development as compared to all other disease areas is any indication, cancer could represent up to $25 billion of annual R&D expenditures, or 28% of total R&D spending.

For the purposes of this thesis, pharmaceutical companies are defined as companies that are fully integrated, profitable and global in scope. These companies are members of PhRMA.
(with certain exceptions) and are generally considered to be part of the pharmaceutical sector by industry analysts. Biotechnology companies are considered to include the group of companies that are primarily involved in R&D and have not yet achieved profitability or commercial scale. Although these are rough guidelines for determining how a company is classified, it is important to note that larger established biotechnology companies, like Genentech, Millennium Pharmaceuticals and others are considered biotechnology companies in this analysis, a convention that is generally followed by industry and Wall Street research analysts. The rationale for not using a strict classification system is to reflect and challenge the heritage and reputation of biotechnology companies as innovative companies focused on biologics, while pharma is considered to be less innovative and focused primarily on small molecule development. In this case, biologics are defined as natural, modified or engineered protein-based products, including monoclonal and radio-labeled antibodies, peptides and therapeutic proteins. Small molecules are defined as active chemical compounds that can be purified from natural sources or are chemically synthesized.

1.3 Overview of universities and government funding

In addition to the efforts by the biotech and pharma industries, the United States government is an important factor driving innovation in research and development through the use of funding and grants. The National Institutes of Health (NIH) is the primary source of funding for biomedical research and development in the United States. With an annual budget of just under $30 billion, the NIH is the largest single provider of R&D funding and is second only to aggregate R&D spending by the entire pharmaceutical industry (NIH, 2005). Although NIH funding has been growing at a compounded annual growth rate of nearly 15%, growth is expected to slow in the coming years due to competing pressures and growing federal budget deficits (UBS, 2003). The reversal of this trend was most recently seen in the proposed 2006
budget of $28.8 billion, which showed less than a 1% increase over the previous year’s budget. The majority of the NIH budget is distributed to external parties of which academic institutions and research centers the primary beneficiaries. Of the $28.8 billion proposed budget, approximately 53% is to be awarded directly as grants, while an additional 10% allocated each to research centers and contract R&D (NIH, 2005).

As described above, cancer is an important focus of the NIH where approximately 17% of the budget will be spent on cancer research and programs. These funds are administered primarily through the National Cancer Institute (NCI), which was established in 1937. The NCI currently has an annual budget of nearly $5 billion and, like the NIH, awards research grants, R&D contracts and small business awards (SBIR) that support feasibility studies (Phase I) and development (Phase II). Among its areas of focus, the NCI is actively promoting research in early detection, cancer biology and molecular epidemiology (NCI, 2005). The role of NCI funding, and of life sciences funding in general, in stimulating research that is ultimately commercialized has been studied extensively and has been found to play an important role in driving innovation and economic growth (Romer, 1985; Lucas, 1993; and Audretsch et al, 2006).

Universities and research centers are the primary beneficiaries of NIH and NCI grants. The relationship between the federal government and universities has changed over the last two decades, which in turn has significantly impacted the way universities interact with industry. In the early 1980s, changes in federal policy enabled small business and nonprofit organizations (including universities) to patent discoveries made with government funded grants. Over the last 20 years, research universities have continued to develop increasingly close ties to the pharmaceutical and biotechnology industries through licensing, strategic alliances and other forms of technology transfer (Owen-Smith et al, 2001). Like government sponsored R&D funding, this increase in activity between universities and industry has become a driving force in
the development of high technology industries (Saxenian, 1994; Powell, 1998) and economic development (Feldman & Florida 1994).

### 1.4 The drug development and approval process

Regulation of pharmaceuticals has a long history in the United States, dating back to 1820 when a group of physicians in Washington DC established the U.S. Pharmacopeia, the first compendium of standard drugs for the US. Despite these early efforts to monitor the use and supply of pharmaceutical products, an additional 85 years elapsed before specific legislation was passed to create a government agency to oversee pharmaceutical regulation. The earliest incarnation of this agency was established by the Pure Food and Drugs Act, passed by Congress in 1906. The precursor to the modern drug development and approval process, however, originated with the passage of the Food, Drug, and Cosmetics Act of 1938, which mandated that every new prescription drug marketed in the United States must first be proven safe and approved by the newly-empowered Food and Drug Administration (FDA, 2004).

Today, in order to receive FDA approval, a drug must first undergo extensive testing in animal models and human subjects in order to establish the drug’s safety and efficacy. The first step in the drug development process consists of discovery, where compounds are screened and modified to enhance favorable drug-like properties. Compounds that are determined to have the appropriate profile are moved into preclinical development, where they are tested in the laboratory and in animal models to test for potential adverse affects. Once preclinical testing is completed, a clinical candidate is selected and the entity developing the drug submits an Investigational New Drug (IND) application. The entity submitting the IND is referred to as the applicant or drug sponsor. A drug sponsor is defined as the person or entity that assumes
responsibility for the development and marketing of a new drug, including responsibility for compliance with all applicable FDA regulations (CDER, 1998). A sponsor is usually either an academic institution or a commercial entity, such as a pharmaceutical or biotechnology company. If the FDA does not have any objections to the IND application, the sponsor is allowed to advance the candidate into clinical trial testing in human subjects.

Generally speaking, the clinical trial process occurs in three distinct steps. Although trial size and design varies across indications, a Phase I trial generally consists of 20-100 healthy volunteers (cancer clinical trials are an important exception, because it is generally considered unethical to administer highly toxic drugs to healthy individuals). The goal of a Phase I trial is to determine the maximum tolerated dose and potential side effects of a drug candidate. A Phase II trial normally consists of 100-500 volunteers. These trials are known as dose-ranging studies, where the goal of the trial is to determine the most effective, safe dose. Subjects are also closely monitored for potential side effects. Finally, Phase III trials, known as pivotal trials, are conducted to determine efficacy and are usually the basis for an NDA filing. These trials usually consist of 1,000-5,000 volunteers. Once a drug is approved by the FDA, the company conducts post-marketing studies, known as Phase IV trials, to monitor for long-term side effects of the drug and enable the detection of rare events which may not have been observed in clinical trials (PhRMA, 2006).
Figure 2: The drug development process

Once the appropriate human clinical testing is complete, sponsors must submit a New Drug Application (NDA) to the FDA (prior to 2003, biologics were approved through a similar process known as a Biological License Application). An NDA consists of a description of the drug and its composition, an analysis of the clinical trial data, and a discussion of manufacturing and quality control protections. The FDA reviews the NDA and can either approve or reject the NDA, or can request additional information. An approved NDA allows the drug manufacturer to market and sell the drug in the U.S. market according to labeling approved by the FDA.

The path from drug discovery and development to regulatory approval and commercial launch is risky, time consuming and expensive. Although estimates vary, approval of a drug can take up to 10-15 years (DiMassi, 2001) at a cost of up to $800M a figure which includes the sunk costs of failed programs (DiMassi, 2003). It is estimated that for every 5,000-10,000 compounds that are screened, 250 will reach preclinical development, 5 will reach clinical development, and only 1 will eventually be approved (PhRMA, 2006).
1.5 Rates of attrition

In order to address the question as to why R&D productivity remains low across the pharmaceutical industry, it is important to consider the success rates across different indications. Figure 3 describes the success rates from Phase I to registration over a ten year period (1991-2000) for ten large pharmaceutical companies in the US and Europe. Across all indications, the average success rate is 11%, which is to say that, on average, only one in nine compounds entering clinical trials is approved by regulatory authorities in the US or Europe (Kola et al, 2004). It is important to note the wide range of success rates across different indications, from a high of approximately 20% for cardiovascular drug candidates, to a low around 5% for oncology and women’s health.

![Figure 3: Success rates from first-in-man to registration (Kola, et al. 2004. p. 711)](image)

The differences in attrition rates between therapeutic areas is due to inherent differences in how products in each area are validated, and can largely be explained by where in the clinical development process a drug candidate is most likely to fail. Overall, the failure rate of compounds in each stage of development is approximately 35% in Phase I, 55% in Phase II,
40% in Phase III and 23% at the registration stage (IMS, 2002). Oncology, however, has significantly higher failure rates in all stages, especially in the late stages of development – once substantial investments have already been made in a clinical candidate. The failure rates for oncology are: 40% (Phase I), 70% (Phase II), 59% (Phase III) and 30% (registration). Although the underlying causes of attrition will be discussed further below, oncology has traditionally been a particularly risky area for drug development as animal models for efficacy are not predictive of human response and the majority chemotherapeutics being tested are cytotoxins, and therefore likely to have result in systemic toxicities at therapeutic doses.

1.6 Why do products fail?

The primary goal of the clinical development process is to provide regulatory agencies with evidence that a given drug candidate is both safe and efficacious in a tested indication. Clinical trials are carefully designed and controlled to ensure that subjects are not exposed to unnecessary risks. In most clinical trials, a given drug candidate is being tested in a new indication, or using a different formulation or dosage. As such, it is difficult to predict precisely what effect a drug candidate will have on a subject or group of subjects. Kola et al examined the root causes of why compounds fail in clinical trials (Kola et al, 2004). Their results are detailed in Figure 3, below, which compares reasons for attrition in 1991 with causes for attrition in 2000. First, it is interesting to note that while in 1991 pharmacokinetics (PK) and bioavailability accounted for nearly 40% of all attrition, by 2000 the figure had dropped to less than 10%.
While it appears that there have been improvements for screening and testing a molecule for qualities that will affect PK or bioavailability before entering clinical trials, efficacy, clinical safety and toxicology continue to be the primary causes of attrition. In 2000, they accounted for approximately 60% of attrition, an increase from approximately 50% in 1991. This increase can partially be explained by the fact that, as PK and bioavailability profiles improved, drug candidates were able to progress further along clinical development, where safety and efficacy issues were increasingly likely to arise. The fact that the causes of attrition have moved downstream might result in drug candidates advancing to later stages of clinical development before failing, but only after having incurred significant development costs. Improving the ability to predict safety and efficacy is therefore likely to have a significant on the total cost and associated risk of developing novel drug candidates. Although the cause of attrition may vary significantly between therapeutic areas, efficacy is more likely to be a significant cause of attrition in areas in which animal models are not known to be especially predictive, such as oncology (Booth et al, 2003). For these reasons, oncology has been and continues to be an important focus area for both bench research and clinical development.
1.7 The war on cancer

In 1971, President Richard Nixon officially declared “war on cancer” with the enactment of the National Cancer Act. Earlier in the year, at his State of the Union Address, he requested an additional "$100 million to launch an intensive campaign to find a cure for cancer… [because] the time has come in America when the same kind of concentrated effort that split the atom and took man to the moon should be turned towards conquering this dread disease." The National Cancer Act gave the National Cancer Institute (NCI) unique autonomy at the NIH with special budgetary authority. In addition, Nixon’s administration converted the Army’s Fort Detrick biological warfare facility into the Frederick Cancer Research and Development Center. The original goal of the legislation and additional funding was to accelerate groundbreaking research to make the President’s promise – to find a cure for cancer by the end of the decade – a reality (Cancer Facts, 2002).

Although a cure for cancer was not discovered as the country entered the 1980s, the US government continued to support this war on cancer. Every President since Nixon has supported cancer research with increased funding. It is estimated that the NCI has funded over $60 billion in cancer research over the last 34 years (Cancer Facts, 2002). Over that period, the budget has grown 30-fold, from $150 million in 1971 to nearly $5 billion in 2005. According to the NCI, this investment has enabled advances in the prevention, treatment and maintenance of cancer. Overall, the mortality rate of cancer has continued to decline. In 2002, the mortality rate reached 193.6 per 100,000 in 2002, down from 213.5 in 1993. In addition, the rate of cancer incidence has been relatively stable since the mid 1990s. Prevention behaviors have shown some improvement, as adult and youth smoking is down. In addition, screening levels have remained high, especially in breast and cervical cancer screening (NCI, 2005). But,
despite these advances, cancer is still the second leading cause of death in the United States after heart disease. Certain factors, including increasing obesity and low socioeconomic status, have shown to be correlated with an increased risk of developing cancer. An estimated 1.4 million new cases of cancer were diagnosed and approximately 570,000 Americans died of cancer in 2005. Lung (163,510 deaths), colorectal (56,290 deaths), breast (40,870 deaths) and prostate (30,350 deaths) are the most prevalent cancers and, in some cases, their incidence has been rising. The NCI estimates that the overall cost for cancer is approximately $190 billion annually, of which $64 billion is for direct medical costs, $16 billion for morbidity costs and $109 billion for mortality costs. Treatment of breast, lung and prostate cancers accounts for over half of the direct medical cost (SG Cowen, 2005).

Despite the investments in research, there is still a significant need for novel therapeutics for the treatment of cancer. Today, most cancers are treated with antineoplastic agents, which prevent the spread and proliferation of cancer cells, along with surgery and radiation. Although novel drugs with targeted mechanisms have begun to emerge over the last five years, older chemotherapeutics still dominate the cancer marketplace. Of the estimated $42 billion of oncology drugs sold in 2004, approximately 33% were chemotherapeutics and 51% were growth factors and supportive care therapies. The toxicity of chemotherapeutics is evidenced by the large and growing supportive care market. Supportive care therapies are treatments that are used in conjunction with or following chemotherapy or radiation therapy to protect against or compensate for treatment-related toxicities. Pegfilgrastim (Neulasta, developed by Amgen), for example, is a white blood cell growth factor that is prescribed to treat chemotherapy-associated myelosuppression (reduced white cell production). Another common supportive care treatment is to prescribe a chemoprotectant along with chemotherapy or radiation to minimize treatment-associated neurotoxicity (damage to nerve cells that can impact normal function). Although the need for supportive care therapies may be reduced as targeted therapies with improved safety
profiles are developed, targeted therapies currently account for a small percentage of total sales. Only 16% of oncology drug sales in 2004 were from monoclonal antibodies or other targeted therapeutics. And while targeted therapeutics are expected to be the future of drug treatment, the number of tumor types addressed by these novel drugs is still relatively small (SG Cowen, 2005).

1.8 The role of innovation

The goal of cancer drug therapy has always been to establish selective toxicity for cancer cells, targeting pathways or molecular targets that are critical for the survival and/or proliferation of the cancer. From a pharmacological standpoint, there are several ways that this level of selectivity can be accomplished: 1) attack targets that are unique to the cancer cell, 2) attack targets on the cancer cells that are similar but not identical to those on a host and 3) attack targets in the cancer cell that are shared by the host, but are not as critical to the host cells (Golan 2005). In order for the therapeutic to be effective, it must target the cancer cells while sparing (or being relatively less toxic to) host cells. Historically, cancer chemotherapies have employed the third strategy, by targeting rapidly dividing cells. While they have been shown to be effective in treating certain cancers, their long-term use is limited as they can also cause severe systemic toxicities in patients, as the therapy is toxic to host cells, particularly rapidly dividing cells in the digestive tract. In addition, cancers can quickly develop resistance to the drugs. As a result, there is a constant need to develop novel approaches to treatment. Today, there is much optimism surrounding the targeted therapeutics, which attack molecular targets that are over-expressed in certain cancers (Golan 2005). Clearly, early successes in the development of these targeted therapeutics have fueled hopes that, in the future, these therapies could play a central role in the treatment of cancer. Among these are Avastin.
(Genentech), an anti-VEGF monoclonal antibody approved for the treatment of colorectal cancer, and Gleevec (Novartis), a selective tyrosine kinase inhibitor developed for the treatment of chronic myeloid leukemia.

Although it is clear that innovation will be an important driver for the future of cancer therapies, there is continued debate surrounding how best to reduce attrition rates in cancer and improve development pipelines. An important component of this discussion is to understand the sources of innovation within the drug discovery and development process, which can be viewed as a value chain where different players contribute at different points along the process (see figure 5 below). The traditional view has been that role of academia was to create new knowledge, educate and publish advances in basic science. Discoveries with commercial potential were passed on to industry for additional research and preclinical development. Since the founding of the first biotechs in the late 1970s, these companies have been considered innovative and entrepreneurial in their approach to developing novel drugs and they were therefore focused primarily on the risky business of preclinical development and early clinical development. The large pharmaceutical companies, while capable of participating at any point along the value chain, have the infrastructure and expertise to undertake late stage clinical development, large scale manufacturing and commercialization.

Figure 5: Product discovery and development value chain
To date, many of the discussions concerning improvements in the development process have been at the organizational level and have focused on the decision-making process used to advance products through development. However, it is not clear today what the sources of innovation are likely to be the most effective and valuable in the future. In order to begin to address the issue of sources and drivers of innovation, a retrospective examination of products that have already moved through the clinic is required. A recent study analyzing a set of oncology molecules concluded that drug candidates sponsored by biotechnology companies have higher attrition rates in clinical trials when compared to pharma drugs (Scharfstein, 2004). Despite the higher failure rate, the biotech drugs were not found to be more novel and were not targeting riskier indications. This suggests that perhaps biotech companies are developing drugs that are somehow different than pharma in ways that are not easily measurable with simple metrics for novelty, or that biotech firms are less capable at managing clinical trials. An important question that remains to be understood in the context of the Scharfstein study is whether in fact, biotech firms are more likely to get ideas from academia that are less validated? Or perhaps that large pharma primarily develops ideas that come from internal research and that have better evidence supporting advancement to clinical development. In other words, does the market for oncology ideas lead to a “sorting” of different types of molecules with different levels of validation among different players?

This thesis will analyze whether the original source of drugs entering clinical development (academia, biotech, pharma) has any impact on attrition rates, and, if so, what the potential explanations for these differences may be. This thesis will attempt to elaborate how the marketplace for novel oncology ideas is structured, how these ideas, in the form of drug development candidates, are distributed across various sources and when and under what conditions they transition between different parties. This thesis will test the null hypothesis that
ideas are born in academia, licensed to biotech and developed by large pharma by examining a population of oncology drugs and determining the nature of these transitions.

2. Hypotheses – the market for ideas in oncology

2.1 An effective division of labor?

Academia and the biotech and pharma industries each play an active role in the drug discovery and development process. An important question, however, is whether each player along the value chain serves a predefined role which, when taken together, enhances the overall productivity of the process. In other words, is there an effective division of labor between these players? In order for that to be the case, one would need to assume that each of these players is ‘best’ in a specific role. Figure 6 below sketches out four potential paths a biologic or small molecule drug candidate could take throughout its development. While countless variations of these options are certainly possible, the four paths below will serve as a general framework for examining the contribution academia, biotech and pharma each make to the successful development of novel drugs.
This thesis will challenge the assumption that there is an effective division of labor between university, biotech and pharma players. The conventional view that most novel university discoveries are licensed to biotech companies, who in turn take on the risk of early development and sublicense to pharma to capture some downstream value (figure 6, path A) will be challenged using a database of oncology drug candidates. The marketplace for oncology ideas will be illustrated by the prevalence of each of the four paths in figure 6 using a database of biologic and small molecule drug candidates entering Phase I clinical development over the course of a decade (1991-2002). The data will be analyzed to determine whether the source of a drug candidate or the development path it takes has a meaningful impact on clinical trial success. This thesis will explore whether biotechnology companies necessarily get access to the 'best', most novel inventions from university and, further, if drug candidates sponsored by biotech companies enter Phase I clinical trials with more or less validated molecules than pharma. This analysis will also examine the assumption that biologic and/or targeted
therapeutics are developed primarily by biotech, while pharma primarily develops small molecule drugs, and in-licenses novel biologics from biotech.

2.2 From academia to biotech to pharma

Academic centers play an important role in the generation of original research, but with limited resources, academia must rely on industry for the eventual commercialization of their discoveries. The pharmaceutical and biotechnology industries had combined R&D expenditures of over $90 billion in 2006 alone. PhRMA estimates that the pharmaceutical industry spends approximately 25% of all R&D funding on preclinical testing, with the remaining 75% spent on clinical development, regulatory affairs and post-approval testing (PhRMA, 2006). The passage of the Bayh-Dole Act has enabled universities to participate in the development of novel molecules through agreements with biotechnology and pharmaceutical companies. As a result, over the last three decades, there has been a significant increase in the number of licensing agreements universities have established with companies (Edwards et al, 2003). Over the same period, biotechnology companies have matured in their own right from single technology platform plays to integrated research engines capable of contributing to the discovery and development of novel drugs. The dynamics of the rise of university participation and the maturation of the biotechnology sector have undoubtedly impacted the sharing of tasks that occurs between the players.

Given that background, a likely potential path for development would be for academia to act as the source of new ideas, for biotech to develop them and for pharmaceutical companies to test and commercialize them (figure 6, path A). Considering their traditional role as creators of new knowledge, it seems that academic institutions would be increasingly likely to act as the source of new ideas and therefore would be 'best suited' to participate on the discovery end of
the value chain. Since institutions do not have the capabilities to develop and commercialize drugs internally, they must seek to monetize their discoveries through licensing agreements. While the terms of a license to industry could vary significantly from case to case, academia has the incentives to maximize value by licensing its ideas out to the ‘best’ industry partner. Assuming that biotech companies are willing to focus and invest in innovative approaches, a biotech partner may be an ideal choice for early stage development. Once the drug candidate has advanced to late stage development, the idea could be then sublicensed to a pharmaceutical partner with regulatory experience, an established infrastructure and market presence in the relevant therapeutic area.

In certain cases, it might be ‘best’ if the early and late-stage development partner were one in the same (figure 6, path B). As fully-integrated organizations, pharma companies have capabilities across the entire value chain. To the extent that pharma companies have historically focused on small molecule development, they may be the best partner for early development of small molecules and having a biotech as an intermediary might not be productive. In certain cases, biotechnology companies have forward-integrated to establish in-house late stage development capabilities as well as a commercial infrastructure. Genentech, for example, historically dependent on Roche for commercialization, has since established a large commercial presence in oncology. A biotech partner for early and late-stage development might also be ideal for indications that are historically not a focus for large pharma. Genzyme, for example, has established itself as a leader in the development and commercialization of treatments for rare genetic disorders.
2.3 **Biotech as innovator**

With the rise of the biotech industry three decades ago, pharma companies realized that they could, in effect, outsource much of their discovery and early stage development to biotechnology companies (figure 6, path C). Biotech companies have long been considered to be both innovative and entrepreneurial. In many cases, biotechs have evolved from single platform technology companies, providing contract screening and discovery services to third parties, into discovery and development companies focused on developing internally generated drug candidates. A robust technology platform could function as an important source of new discoveries and could be used to fill early stage biotech pipelines which could then be licensed to pharma as they enter late-stage clinical development. Biotechs have also benefited from this type of arrangement, as pharma capital has become an important source of funding, and relationships with leading pharma companies often serve as validation to the biotech company’s approach.

2.4 **In-house discovery and development**

Prior to passage of legislation favorable to universities and the establishment of a viable biotechnology industry, the majority of pharmaceutical drug discovery and development likely took place internally within a pharmaceutical company (figure 6, path D). Clearly, pharma companies and certain biotech companies have the capacity and capabilities to develop and commercialize internally discovered molecules. Pharma and large biotech companies may have significant scale in screening capabilities which could act as important sources of innovation. Internal discovery might also be preferred as drug candidates could be screened and optimized as part of a comprehensive development plan. In addition, the ability to keep
discovery as an internal process may be a competitive advantage, as drug candidates could be moved along the discovery and early development process without attracting competitors.

2.5 A traditional path to drug development?

Edwards, Murray and Yu studied the sharing of ideas that occurs between universities, biotech and pharma companies by analyzing deal structures from 1975 to 2000 to determine how value is distributed among the different parties (Edwards et al, 2003). The authors discovered that the increasing numbers of patents filed by universities led to a sharp increase in licensing activity. The authors also noted that although biotech companies are willing to in-license patents from universities, they also tend to sublicense the IP to a more established pharma company as the development moves downstream. In fact, of the top ten selling biotechnology products in 2002, 8 of 10 had been in-licensed from a university, and 10 of 10 had been sublicensed to a pharma company (Edwards et al 2003).

The authors estimated that once all relevant payments are factored in, the total value split between university, biotech and pharma is roughly 7:29:64 (Edwards et al, 2003), which reflects the increasing value contribution as the product moves downstream and changes hands from university to biotech to pharma. A separate study noted that 62.5% of oncology drug candidates being developed by early stage companies (primarily clinical stage biotechnology companies) have some kind of an alliance with a pharmaceutical company at the initiation of a Phase I trial (Scharfstein, 2003). These conclusions suggest that the way to maximize value among these parties is for universities to make the initial discoveries, biotech to take responsibility for early stage development and pharma to control late stage development.

3.1 Discussion of source data

This analysis is based on a database created by Roberts et al and described in a JAMA article titled “Trends in the Risks and Benefits to Patients with Cancer Participating in Phase I Clinical Trials” (Roberts 2004). In the article, the authors examined trends in the rates of treatment-related death, toxicity and objective response in cancer clinical trials in order to identify factors which may contribute to patient outcomes. The authors searched abstracts and journal articles reporting the results of Phase I trials submitted to annual meetings of the American Society of Clinical Oncology (ASCO) from 1991 through 2002. ASCO was chosen as the primary source of abstracts, as the conference is generally considered to be the world’s premier oncology specialty society.

Roberts et al opted to develop a proprietary database based on meeting abstracts in order to avoid the possibility of introducing selection bias in published trials. The authors searched indexes of the published conference proceedings, as well as individual abstracts to identify relevant studies. This initial search yielded 2,460 Phase I and Phase I/II abstracts, which were subsequently reviewed. As described in figure 7, of the original 2,460 abstracts, a total of 2,247 abstracts were excluded from the analysis. 1,749 abstracts were excluded for having a Phase II component, the use of radiation, or indicating the testing of multiple agents. An additional 165 abstracts were excluded because the agents were already approved by the FDA. Of the remaining 546 studies, 304 were excluded because a journal publication was not identified. Finally, 29 additional abstracts were excluded because they were studied in hematologic cancers, or the trial was without therapeutic intent.
The authors analyzed the remaining 213 abstracts using observers to independently extract information including trial design, types of tumors treated, clinical outcomes and others. In aggregate, 149 unique experimental agents were tested, approximately half of which (47%) were targeted or biological agents. Based on their analyses, the authors concluded that the level of risk experienced by cancer patients who participated in Phase I trials improved over the 12-year study period. Furthermore, since the rate of toxic death decreased more quickly than have objective response rates, the ratio of risk benefit appears to also have improved over time. The authors speculate that this finding may be due the targeted nature of novel cancer drugs, in addition to an increased focus on the safety of clinical research over this period.

Figure 7: Trial Flow Used in Identifying Studies for Detailed Analysis (Roberts et al. 2004, p. 2133)
3.2 Database development and rationale

The focus of this thesis will be to analyze a sample set of novel drug candidates to identify elements of drug discovery and early-stage development which may have an impact on clinical trial success. In order to facilitate comparison among different drug candidates, the sample set was limited to a single therapeutic area. Oncology was selected as the focus therapeutic area, as cancer is the disease class with the largest number of ongoing clinical trials during the last decade (Scharfstein 2004). The analyses described in this thesis utilize the database described above as the initial source database, which was made available by Drs. Thomas Roberts, Stan Finkelstein and their colleagues. This specific database was selected as it provides a robust sample set of drugs that entered early stage clinical development over a decade. As the database includes all Phase I and Phase I/II clinical trial abstracts presented at ASCO, all molecules entering clinical development can be tracked regardless of downstream clinical trial outcome, which in effect reduces the risk of selection bias.

Starting with the complete set of 2,460 Phase I and Phase I/II studies, certain data points were excluded from this analysis, although in many cases the screening criteria for identifying studies for further analysis differed from Roberts. The order in which the trials were excluded from the data set may also result in discrepancies from the Roberts trial flow figure. Of the complete set, 1354 trials were excluded for testing more than a single agent or using radiation therapy (Phase I/II trials were not excluded to the extent that a unique molecule was being tested), 201 trials were excluded for testing an agent already approved by the FDA, and 68 trials were excluded for studying pediatric, hematologic or miscellaneous cancers. Drugs that were approved at the time of the abstract were determined by comparing the abstract dates in the
Roberts database and the approval dates of all currently approved drugs on the FDA’s Center for Drug Evaluation and Research (CDER) website (FDA 2005). According to the CDER website, 253 drugs are currently approved for treatment in various cancer indications. However, many drugs are approved for more than one indication – of the 253 approvals, 133 are unique chemical or biological entities.

As described above, Roberts et al used individual Phase I clinical trial abstracts as the basic unit of measurement in their study. As this thesis intends to analyze the development paths of molecules as they move from academia to industry, the original dataset was adjusted to enable the use of unique molecules as the basic unit of measurement. Since many drugs undergo more than one trial, each trial in the Roberts database was analyzed to identify the number of unique molecules present in the data set. In addition, this analysis will focus on the development of novel molecules, as opposed to studies focused on re-formulations or drug delivery technologies encompassing previously approved molecules. Identical or similar drugs were identified using trade names, investigational names and/or chemical names. Similar drugs were identified by tracking all known alternate names, or "synonyms", for a given molecule and grouping the trials of similar drugs together and assigning each group a unique identifier and using the generic name to identify groups of similar molecules. For the purposes of this analysis, “unique” molecules are defined as new chemical entities or biologics that have not been previously approved by the FDA. Abstracts testing NCEs or biologics that can be considered to be reformulations or enhanced delivery of an existing drug were excluded when possible. In situations where numerous abstracts were found to test similar or identical small molecules or biologics, only the abstract with the earliest initiation date was included as a “unique” molecule. Of the 837 potentially appropriate Phase I and Phase I/II trials, 156 abstracts were excluded as the specific molecule being tested was not possible to determine. Of the remaining abstracts, 364 unique molecules were identified and used for detailed analysis.
Of these, an additional 54 were excluded as limited or no information was found using commercial databases. The remaining 310 molecules were the basis for the analysis described below.

Figure 8: Dataset adjustment to identify unique molecules

Scharfstein and his colleagues used the same source data, the Roberts database, for their analyses. They determined that of the original dataset, 1,180 abstracts met their search criteria (single agent, non-pediatric indication, not FDA approved). The 1,180 abstracts describe 377 unique agents (as compared to the 364 identified for this thesis). Using Thomson's Investigational Drug Database (IDdb) and PJB Publications' PharmaProjects to identify the trial sponsor, they determined that public companies accounted for 62.3% of molecules, private companies 27.6% and universities and government agencies 10.1%. Their analysis centered on the 235 Phase I trials undertaken by public firms in their data set (Scharfstein et al 2004). The study does not separate the data by biotech or pharma sponsor. Rather, it compares the
probability of mature and early-stage firms (using company revenue as a proxy for early stage and mature firm classification) advancing drug candidates from Phase I to Phase II, and from Phase II to Phase III. Although the aggregate data is not broken down separately, the authors conclude that molecule type (small molecule or biologic) or NIH-sponsorship did not have a statistically significant impact on the probability of a drug candidate from Phase I to Phase II (the focus of their analysis).

As the Scharfstein paper describes, the early-stage firms in the study sample tend to have less promising late-stage clinical results (Phase II or Phase III) and are less likely to receive FDA approval. On the basis that the molecules tested by both early-stage and mature firms are similar along their dimensions of novelty, the authors conclude that this result is due primarily to an agency problem between shareholders and managers of single product early stage firms, as they are less likely to abandon development of their only viable drug candidates (Scharfstein 2004). In their study, novelty was calculated using the pharmacological description of each drug candidate in the PharmaProjects R&D database. For each pharmacological description, drug candidates with the same description were ranked chronologically, with the first candidate being the most novel and the last the least novel within a certain category. In the data set used for the Scharfstein study, early stage firms had an average novelty rank of 22, while mature firms had a rank of 24.6, a statistically insignificant difference. In other words, there was little difference in the novelty of the drug candidates being developed by early stage and mature firms. It should be noted, however, that this measurement of novelty appears to group candidates into fairly broad categories (i.e., DNA antagonist) and not necessarily by molecular target or mechanism of action.

While Sharfstein et al concluded that novelty was not predictive of clinical trial outcome, the use of novelty alone may be misleading in terms of the ‘quality’ of the molecule entering
Phase I clinical development. Despite the fact that there may be similarities in how novel the drugs were, there may be other factors inherent in the discovery and early development of a preclinical candidate that may contribute to the clinical trial outcome of a given molecule. In order to develop a more complete picture of the ‘quality’ of drug candidates entering clinical development, this analysis attempted to describe the amount of scrutiny and scientific validation a molecule had undergone prior to entering Phase I clinical trials.

3.3 Measure of validation

For the purposes of this analysis, validation was defined as the amount of time that a molecule underwent research/preclinical development prior to entering Phase I clinical trials. A proxy for this time period was the time estimated from the filing date of the original patent describing the key discovery of the molecule to the abstract date of the describing the Phase I clinical trial. In other words, the analysis assumes that the time from the initial patent filing to initiation of clinical trials was used to research and test the molecule. The filing date of a patent is defined by the United States Patent and Trademark Office as “the date of receipt in the Office of an application which includes (1) a specification containing a description and, if the application is a nonprovisional application, at least one claim, and (2) any required drawings” (USPTO). The filing date is essentially the first notice the USPTO receives of a patent application. In this case, it is used to approximate the date of invention, as it is likely to reflect the earliest recognition of the discovery of a molecule with therapeutic, and therefore commercial, potential. Of course, this is an imperfect measurement, as other factors certainly exist that would contribute to the length of time that elapses from a patent filing to the initiation of clinical trials for a particular molecule, including access to limited resources, development strategy for a sponsoring company and competitive factors.
Patent filings associated with a specific molecule were identified using the *Thomson Pharma* database. Information about the patent with the earliest filing date was recorded, including the filing date, assignee, patent number and inventors. The patent assignee is referred to as the “source” of a molecule of the remainder of the discussion. In many cases, patent information was not available for a molecule, and those data points were not used for this analysis. Of the 310 molecules in this sample, patent information was available for 180 molecules (58%). Missing patent data across molecule type is nearly identical, as 57% of biologics and 58% of small molecules had patent information associated with the molecule. Lack of patent data, however, may have biased the sample against molecules entering clinical trials that are sponsored by academia. While patent data was available for 61% and 60% of molecules sponsored by biotech and pharma, respectively, the figure for molecules sponsored by academia was only 39%. Occurrences where the initial patent filing date to Phase I initiation was greater than 12 years were excluded from this analysis, as it was assumed that the identified patent was either not a key patent or other factors had impacted the decision to move the molecule into Phase I testing.

In this sample set, the period of validation varied significantly across all molecules. The average period between the first patent filing date associated with a molecule and the initiation of Phase I clinical trials was approximately 5.7 years, with a standard deviation of 3.1. Although only a proxy, this suggests that, on average, molecules in this sample set will spend just under 6 years in research and preclinical development. According to PhRMA estimates, an average drug candidate molecule will spend approximately 3.5 years in preclinical development (PhRMA, 2006), a figure which does not necessarily include preliminary research prior to identifying a potential drug candidate. Patents may be filed as soon as an invention can be claimed, which could potentially predate the identification of a specific preclinical candidate. Precisely what is covered in a patent is also likely to determine how early in the research and
development process a patent application is filed. According to Murray et al, a patent can cover any useful, non-obvious and novel application including a newly discovered natural product, therapeutic proteins, drug targets, diagnostics and isolated DNA sequence (Murray et al, 2005).

3.4 Molecule Type

Each molecule was identified as either a small molecule or a biologic. For the purposes of this analysis, biologics included monoclonal antibodies, conjugated antibodies, peptides, vaccines and other natural products. Of the 310 unique molecules identified by this initial screen, approximately 73% of the sample consisted of small molecule drug candidates and the remaining 27% of the sample were labeled biologics. The proportion of biologics represented in this sample set varies from the sample set used in the Roberts study, which estimated that biologics account for 47% of molecules entering Phase I. This discrepancy could be due to a bias in the sample used for this analysis – of the 156 abstracts that were excluded for this analysis due to a lack of information regarding the specific molecule being tested, 91% were biologics. Although the overall split in this sample is 73% small molecules and 27% biologics, the relative contribution of biologics has increased over the course of the decade. In 1991, biologics accounted for only 22% of all molecules entering clinical development. By 2000, biologics comprised approximately 43% of all oncology drugs entering clinical development. As the promise of targeted therapeutics is realized, biologics are expected to represent an increasing proportion of novel drug candidates, a trend that may be further driven by improvements in manufacturing processes for protein-based therapeutics.
3.5 **Source, sponsor and licensor classification**

In this analysis, the source of a unique drug is the assignee found on the patent with the earliest filing date in the *Thomson Pharma* database. The source is considered to be the inventor, or original source of innovation, for a given molecule. The sponsor is defined as the entity responsible for the Phase clinical trial development. Sponsors are usually companies or academic research centers. When available, the sponsor for a specific molecule, as identified in the Phase I abstract, was taken from the original Roberts database. If the sponsor name for a molecule was not available in the Roberts database, as was the case for approximately 36% of molecules, the sponsor was identified using the *Thomson Pharma* database. This may have potentially introduced a source of error if the definition of sponsor used by *Thomson Pharma* differed from Roberts’ definition. In cases where the sponsor information was available in both the Roberts database and the *Thomson Pharma* database, the sponsor information matched in 78% of cases. Based on the *Thomson Pharma* drug profile, the sponsor was assumed to be
the entity that had ownership or control of the molecule as of the date of the Phase I abstract. In cases where there were numerous Phase I abstracts in the Roberts database for a single drug, only the earliest abstract was included in the analysis.

License agreements were also available from the Thomson Pharma database. License agreements described as co-marketing, co-promotion, strategic alliance, co-development, joint venture, exclusive rights and sub-license agreements were included. As this analysis is focused on the development path of the molecule, license agreements for marketing right to foreign markets (Europe, Japan, etc.) were not included. In certain cases, the company that was listed in Thomson Pharma had licensed the drug to a second party prior to the date of the Phase I abstract. If the date of the licensing agreement was prior to the initiation of the Phase I trial, the licensee was considered the sponsor in cases where the sponsor information was not available in the Roberts database. The name of the licensor and date of the agreement were recorded. Each originator, sponsor and licensor was classified as an academic institution, biotech or pharma company.


4.1 Sources and sponsors of innovation

Based on the measures described above, of the 180 molecules with available patent information, pharmaceutical companies were the source of discovery for approximately 52% of molecules, with biotechnology companies and academia accounting for 24% each. As in the original 310 molecule sample, the split between biologics and small molecules was 27% and
73%, respectively. When sorted according to molecule type (small molecule or biologic), the contributions of each of the players differed significantly. Pharma was responsible for 66% of small molecule discoveries, but only 15% of biologic discoveries. Biotech was the source of 42% of biologic molecules and only 17% of small molecules. Academia accounted for 43% of biologic discoveries and 17% of small molecules. While this overall finding contradicts the null hypothesis that academia is the source for the majority of novel drugs, academia seems to play a larger role in discovery of novel biologics than small molecules.

The sponsor of a molecule entering early clinical development (Phase I or Phase I/II) appears to be dependent on the source of the molecule.

The transition that took place from the discovery (source) to early clinical development (sponsor) is summarized in figure 10 below. Pharmaceutical companies were most likely to retain sponsorship of internally discovered molecules, with approximately 89% of molecules with a pharma source maintaining a pharma sponsor at initiation of clinical trials. Molecules with a biotech company identified as the source of discovery were also highly likely to maintain a biotech sponsor at initiation of clinical trials, with 79% of molecules with a biotech source are sponsored by biotech. On the other hand, approximately 86% of molecules sourced within academia are transitioned to industry at the initiation of clinical development. Of all molecules sourced by academia, biotech companies sponsor 48% of molecules, while pharma sponsors 39%. This finding appears to contradict the null hypothesis that discoveries sourced in academia primarily feed biotech. Not only do a large proportion of novel drugs originate in industry (as opposed to academia), but of those that transition from academia to industry, nearly 45% are transferred to pharma (as opposed to biotech).
4.2 Validation

The source of a patent has the potential to impact the timing of patent filing. Academia, pharma and biotech may have very different patent strategies that could affect the amount of time between a filing and the initiation of Phase I clinical trials. Academic institutions may have technology licensing offices that encourage early filings to protect intellectual property without restricting a scientist’s ability to submit publications to scientific journals. Pharma companies may choose to file patent applications later in the discovery process to enable them to keep novel discoveries more closely guarded against potential competitors, and to maximize years of patent life following commercialization. Biotechs, on the other hand, may rush to file patents to increase pipeline visibility for investors, and initiate clinical trials to meet milestones. In terms of molecule type, biologics might be expected to have longer validation periods, due to their relative complexity.
Despite the potential for these factors to impact the timing if patent filings, there does not appear to be a statistically significant difference in validation according to source, sponsor or molecule type. As mentioned above, the period of validation across all 180 molecules with available patent data was 5.7 years (SD=3.1). The period of time between patent filing and the Phase I abstract year differs when the entity that filed the initial patent is considered. Novel molecules originating in biotechnology companies had the shortest average period between patent filing and the Phase I abstract date with 5.3 years (SD=2.7). Pharmaceutical companies, on average, had a period of validation of 5.4 (SD=3.2) while molecules originating in academia had an average period of 6.6 years (SD=3.3). When the molecules are grouped according to the entity that sponsored the initiation of Phase I clinical trials, the average validation period is 5.3 (SD=3.73) for academic sponsors, 6.0 (SD=3.1) for biotech sponsors and 5.55 (SD=3.1) for pharma sponsors. When the data is analyzed for differences in molecule type, biologics and small molecules have average patent to Phase I times of 6.0 (SD=3.3) and 5.6 (SD=3.1) years respectively. A t-test for the validation by molecule data shows that the differences in the mean values between the two samples is not statistically significant (P(t)=.419>>.05). The differences in the average validation period between academia and industry, as well as the differences between small molecules and biologics, are not statistically significant.
4.3 Biologics and small molecules

Given that biologics have become an increasingly significant presence in the oncology drug development pipeline over the last decade, the source of these molecules could be an important factor in establishing a better understanding of potential development paths in oncology. As with the measure of validation, the assignee of a key patent for a molecule was used as a proxy for the inventor (source) of a novel molecule. The patent assignee was recorded for the earliest patent filing in the Thomson Pharma database associated with a given
molecule and classified as academia, biotech or pharma. Of the biologics in the sample with an identified patent, 42% were sourced by biotechnology companies, 43% by research or academic institutions and the remaining 15% by pharmaceutical companies. This suggests that a significant portion of biologics were discovered by biotechnology companies, as opposed to originating in academia and then being transferred or licensed into a biotech company. Small molecules, on the other hand, were discovered primarily by pharmaceutical companies. Pharmaceutical companies hold 66% of small molecule patents in the sample, while universities and research centers and biotechnology companies account for 17% each. These findings suggest that pharma is largely responsible for small molecule discovery, while academia and biotech each account for approximately over 40% of biotech discoveries.

![Source of biologics and small molecules](image)

**Figure 12: Source of discovery for biologics and small molecules**

A similar picture emerges when the entity that sponsored a drug candidate's Phase I clinical trial is considered. As described above, a drug sponsor is defined by the FDA as the entity or individual that is responsible for adhering to all FDA guidelines regarding clinical development. While the division of labor that occurs between academic centers and industry
(and eventually between biotech and pharma) will be analyzed and discussed below, it is important to note here that of the 310 unique molecules in the sample, a significant majority of biologics (60%) are sponsored by biotechnology companies and 67% of small molecules are sponsored by pharmaceutical companies. When only the 180 molecules with patent information are considered, biologics sponsorship shifts further to biotechs (73%), while small molecule sponsorship moves to pharma (72%). As mentioned above, this change could be due to a bias since molecules with academic sponsors entering Phase I were less likely to have associated patent information.

Figure 13: Sponsor into Phase I clinical development for biologics and small molecules

As expected, the biggest change from source (discovery) to sponsor (early clinical development) appears to occur in the academic segment. Although academic centers account for nearly 25% of novel discoveries of all molecules, they sponsor only 10% of all novel drugs entering clinical development, or 4% of the 180 molecules with patent information available. Of course, this change can be explained by the fact that academic centers are more likely to license their discoveries to industry partners.
4.4 A snapshot of the oncology drug development landscape

The Roberts database spans a period of over ten years (1991-2002), which enabled a retrospective look to see how these new discoveries and ideas that emerged over the course of a decade fared through the drug development process. 310 molecules were tracked for the purposes of this analysis. Of these molecules, approximately 16% have been launched and 12% have reached Phase III or are already in registration, while 41% have failed or have been discontinued. Overall, these rates were in line with the results Scharfstein et al reported in their study sample (Scharfstein, 2004). Based on the attrition rates that were discussed above, these rates are also in line for the risk generally associated with oncology drug development. The fact that 16% of drug candidates in this sample have been launched, approximates the 5% success rate for cancer clinical development from first-in-man registration reported by Kola and colleagues (Kola et al, 2004). Of course, as drug candidates that are currently in the pipeline are launched, the final total percentage of commercial products from this sample may be higher than would have been predicted by historical trends, which may reflect the impact of targeted therapies showing greater efficacy in pivotal trials.
The summarized data can also be broken down further to show development status for biologics and small molecules. Despite the fact that biologics (especially monoclonal antibodies) are generally considered to be more specific or targeted therapies, small molecules and biologics have similar development profiles. Both had similar rates of failure or discontinuation, approximately 39% for biologics and 42% for small molecules. Launch rates to date were also similar for small molecules (17%) and biologics (13%), as was the distribution of molecules in the different phases of clinical development.
Of the 180 molecules with patent information, 44 were sourced by academia, 43 by biotechs and 93 by pharma. The launch rate was similar across the three players, although academia had a slightly higher launch rate: academia (18%), biotech (14%) and pharma (13%). The failure and discontinuation rates between academia and industry however were different. Biotechs had a 44% failure or discontinuation rate, while pharma’s rate was 34% and 20% for academia. Although the sample sizes are small in this subset analysis are small, this could potentially suggest that academia is somehow producing higher ‘quality’ molecules that are not captured by the validation measure used in this thesis.
4.5 Division of labor – a map of the oncology drug development landscape

In order to develop a better understanding of where new ideas are generated and how they flow through the development process, it is important to look at the entirety of the process. The tables below track the transition novel molecules undergo from discovery (source) to Phase I initiation (sponsor), and into late clinical development (1st alliance). This analysis includes only the 180 molecules with patent information, as the goal is to develop a better understanding as to where new ideas are generated, which entities sponsor early development of these new ideas.
and who is ultimately responsible for late stage development and commercialization. It is important to note that this analysis does not account for the current development status of the molecules. In other words, molecules that failed early would not have had the opportunity to advance to the stage where an alliance may have been considered (and would therefore appear in the “no alliance” box in the following maps). As mentioned above, the first license was determined by recording the first commercial license in the Thomson Pharma database that occurred after the initiation of Phase I (per the abstract date in the Roberts database). Also, the license agreements were not reviewed. Special situations, such as return of product rights following a clinical trial failure, were not excluded and may have skewed the analysis.

Below is a map of oncology drug candidates that were sourced by academia. Academia appears to be an approximately equal source of small molecules and biologics. Of the 44 total molecules originating in academia, 48% were biologics and 52% were small molecules. As mentioned above, the majority of molecules (86%) are transferred from academia to industry at the initiation of Phase I. Of those that were transferred to biotech, 71% were biologics, while 71% of molecules transferred to pharma were small molecules. Of the candidates sponsored by biotech, 5 were subsequently partnered with pharma, while two were partnered to another biotech, and did not partner the remaining 12 candidates. Pharma was more likely to retain the molecules it sponsored, or to partner with another pharma company.
As mentioned previously, industry is significantly more likely to sponsor its own clinical candidates. In the case of biotech, nearly 80% of molecules that were sourced by biotech stayed in the hands of a biotech company. While biotechs are the source for approximately 42% of all biologics, they discover biologics and small molecules in almost equal numbers. That is, biotechs and academia source the majority of all biotechs, while biotechs contributions to small molecule discovery is dwarfed by pharma. They were also nearly as likely to sponsor small molecules as biotechs for Phase I clinical trials. This suggests that biotechs in oncology are not necessarily Of those that were sponsored by biotech, the majority (23 of 34 molecules) were never partnered, while the remainder was split nearly evenly between pharma (5 molecules) and biotech (4 molecules). It is interesting to note that pharma only sponsored molecules originating in biotech in 5 cases, 4 of which were small molecules, none of which were subsequently partnered.
It was interesting to note that the pharma industry produces as many novel discoveries as the biotech industry and all of academia (93 for pharma vs. 87 for biotech and academia). Although this seems to contradict the conventional view that biotech and academia are the primary sources of innovation, this result can possibly be explained by pharma's large chemical screening capabilities, a resource that most biotechs and academic centers do not have.

Pharmas are overwhelmingly focused on the novel discovery of small molecules (92% of total). Not surprisingly, the majority of molecules that are sourced by pharma are also sponsored by pharma (89%). Of those that are sponsored by pharma, the majority (86%) are either never partnered or partnered with another pharma company. Although the sample size is small, it is interesting to note that only two of seven biologics sourced by pharma are sponsored by biotechs (and one is subsequently partnered with pharma).
4.6 Potential paths to development

As discussed in Section 2, there are many potential paths a drug candidate can take prior to reaching the marketplace. Four potential development paths are highlighted below. Based on the maps above, each occurrence of a specific path was recorded. The “traditional” path, where molecules are sourced in academia, transition to biotech and are sub-licensed to pharma (path A) was only observed in seven cases. In 33 cases, molecules originated in academia and were transitioned to industry and remained either in biotech or pharma (path B). The situation where a molecule originated in biotech and ultimately partnered with pharma occurred in 24 cases (path C). The most common path was in-house development by a pharma or biotech company, which occurred in 85 cases. Looking at path A and B combined suggests that less approximately 20% of molecules in the sample set originated in academia, while nearly 60% originated in industry (paths C and D). The remaining 20% followed a path not highlighted in figure 20.
The molecule type appears to have some affect on the path of development, however. While the path that small molecules take closely mirrors the results above, biologics, which make up approximately 20% of all molecules in this analysis, are more evenly split between different paths. 26% of all biologics take path A and 28% take path B, which suggests that over 50% of the biologics that move down the path to development originated in academia. Of the remainder, 24% of biologics took path C and 22% path D.

The development status of each molecule was sorted according to the development path. Paths A and C appeared to have the highest rate of launches (21% an 14%, respectively). Although the sample sizes are relatively small, on the surface this appears to support
Scharfstein's finding that involvement by a pharma company is correlated to more favorable outcomes.

![Pie charts showing development status by path]

**Figure 21: Current development status by development path**

5. Discussion and Conclusions

5.1 Summary of findings

In this thesis, I have used the database provided by Drs. Finkelstein and Roberts to analyze how novel drug discoveries flow through the drug development process and, more
specifically, how these molecules change hands between academia and industry on their path to commercialization. I chose oncology as the case study for this study, as it is a major focus for both research and industry and, as the largest therapeutic area in terms of number of ongoing clinical trials, would provide a more complete picture of the dynamics of the marketplace for novel discoveries. While previous work has focused on the how organizational structures impact the processes that drive development and commercialization decisions between mature and early stage companies, (Scharfstein et al 2004), this thesis focused on the nature of the ideas and discoveries entering clinical development. In particular, I have considered the impact of the sources of novel discoveries, the role of Phase I sponsors, the type of molecules entering clinical trials, their novelty and validation. An important consideration of the analyses was to attempt to understand the connection between these factors and the path that the molecule ultimately takes.

In Section 1, I examined the need for new drug approvals, particularly in oncology. Since cancer has been a focus of increasing research expenditure and remains an area with significant unmet needs, it is important to understand the process of innovation. The US government, and particularly the NCI, plays a central role in the drive for innovation through R&D funding. Universities, in turn, have established relationships with industry to enable the transfer of intellectual property. Biotech and pharma companies, of course, focus on the development and eventual commercialization of drug candidates. The conventional view has been that the common path for novel drugs is discovery in academia, early development in risk-taking biotechnology companies, and finally late stage development and commercialization. The first question is, of course, does this perception reflect reality? If not, what does the ‘typical’ development path for a drug candidate look like? Are there specific elements of a drug’s profile that determines the path it will take?
The primary analysis consisted of constructing a database of novel molecules entering clinical development for cancer indications. This database was based on previous work conducted by Roberts et al and additional data was acquired through the NCBI and Thomson Pharma databases. The database was filtered down from the 2460 Phase I and Phase I/II abstracts included in the Roberts database to 364 unique and novel oncology molecules, of which 310 were used for analysis.

Based on this dataset, it appears that while the majority of drugs entering clinical development for cancer have historically been small molecules, this is changing as biologics have begun to show promise as targeted approaches to treatment. The data further supported the view that biologics are primarily developed by biotechnology companies, although of all molecules discovered by biotech, nearly 50% are small molecules. In line with the conventional view, small molecules continue to be the focus of pharma companies.

There was no statistically significant difference in validation according to molecule type, source or sponsor. In terms of outcomes, molecule type and source also do not appear to have a differentiating impact. Looking at the current status of the drug molecules, I calculated that approximately 16% of the drug candidates in the data set have reached commercial launch, which is directly in line with published success rates (although the number will likely increase as pipeline candidates continue to move through the process). Although the sample size is small, it appears that involvement with a pharma company may have a favorable impact on launch rates, which would also support Scharfstein’s findings (Scharfstein, 2004).

In terms of the division of labor between academia, biotech and pharma, the data suggests that pharma and biotech are more likely to sponsor molecules that are internally discovered. Academia, on the other hand, is more likely to license molecules to industry prior to
initiation of Phase I trials. Although academia has limited resources with which to pursue clinical development, the molecules that are sponsored by academia entering Phase I are possibly funded by SBIR grants and other sources of funding that support early stage feasibility and Phase II studies, particular in cancer studies where might be smaller and of shorter duration. Another interesting result is that, pharma companies are responsible for the majority of novel discoveries entering Phase I clinical development in oncology (approximately equal to the combined output of biotech and academia).

5.2 An ideal path?

Based on the results, there is no clear path for the development of drug candidates. Biologics are more likely to emerge through academia or biotech, while pharmaceutical companies depend heavily on internal discovery programs. While it is clear that the majority of biologic drug candidates do not necessarily move through the traditional path, it appears that molecules that travel this path may have marginally higher success rates. This also needs to be considered within the context of Edwards et al's findings that 9 of the top 10 biologics have originated in academia, licensed by biotechs and are then sublicensed by pharma. It is possible that the development path is not what made these molecules successful, but rather certain aspects of the molecules informed which development path it would take. That is, perhaps these molecules showed early promise, which attracted the best biotech partners who began development of a promising molecule in a large indication, which in turn attracted the large pharma partners.
5.3 Limitations

Although the analysis yielded several interesting results, it is important to note that this approach has several limitations that may have affected the results. First, although the use of abstracts was a starting point was included to reduce the risk of selection bias, the results may be biased to the extent that there is limited information on drugs that failed early in the development process. Additionally, the use of patent filings for measurement of validation and discoveries is a proxy that is subject to errors. Licensing information used for the purposes of determining the existence of alliances was based on information found in the Thomson Pharma database. Each individual license agreement was not reviewed and therefore the inclusion of the alliance may not reflect the reality of the agreement between the companies. Finally, it is important to note that each drug molecule has its own 'story' and is the product of a unique set of circumstances which may have significantly impacted the course of development in ways that are unlikely to be reflected in the available data.

5.4 Areas for future research

This thesis provides a first step towards developing a better understanding of how ideas and discoveries travel through the drug development process. Of course, drug development is an incredibly complex process and further analysis will certainly be required. Among these, statistical models can be used to determine the impact of each of the factors described above. Second, as discussed above, the measurements used in this analysis are proxies for discovery and validation. A clear understanding of the preclinical status of a molecule is needed in order to better quantify the ‘quality’ of the product entering clinical development. Other measures of validation could potentially include a review of animal data and assessment of available animal
models used to study preclinical candidates. To the extent that these factors can be quantified directly, it would significantly impact our ability assess preclinical candidates. Ultimately, an ideal goal would be to develop a model that would use these and other factors to help predict probability of success for drug candidates.

It would certainly be useful to analyze the impact that targeted therapeutics, where the specific biological or genetic target is known, have had on attrition rates. Pharmacogenomics and novel diagnostics will enable better screening for clinical trial subjects to improve attrition rates in this area. In addition it would be interesting to study how external parties impact the interactions between academia and industry, specifically the role that venture capital plays in portfolio biotechnology companies’ decisions to move into clinical development. Finally, the interactions between academia and industry in different therapeutic areas would need to be examined. Oncology is a unique therapeutic area and, therefore, the findings of this study are not likely to translate over to other therapeutic areas.
6. Sources and Bibliography


National Cancer Institute website (www.cancer.gov).

National Institutes of Health website (www.nih.gov).


Powell, W. “Learning from collaboration: Knowledge and networks in the biotechnology and pharmaceutical industries” California Management Review 40: (3) 228-263. 1998


Saxenian, A. Regional advantage: culture and competition in Silicon Valley and Route 128. Cambridge, MA: Harvard University Press. 1994

Therapeutic Categories Outlook, SG Cowen & Co., New York, NY. October 2005


Surveillance, Epidemiology, and End Results Program. Cancer Facts & the War on Cancer. 2002


US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, Health, United States, 2005 With Chartbook on Trends in the Health of Americans. Hyattsville, MD: NCHS, 2005