Well Begun Is Half Done: Initial R&D Competence and Firm Growth

Kyungran Lee S. Katie Moon Seungjoon Oh*

November 29, 2017

*The University of Hong Kong, the University of Colorado, and Peking University HSBC Business School, respectively. Lee can be reached at kyunglee@hku.hk, Moon at katie.moon@colorado.edu, and Oh at sjoonoh@phbs.pku.edu.cn. We thank Sugato Bhattacharyya, Seong Byun, Hursit Selcuk Celil, Lin Chen, Jaewon Choi, Tony Cookson, Diego Garcia, Jayant Kale, Dukhee Lee, Danmo Lin, Stefan Nagel, Uday Rajan, Michael Schill, Brain Waters, Ed Van Wesep, and Jaime Zender for excellent comments and suggestions. We also thank seminar participants at 2017 NFA Meeting, 2017 MFA Meeting, 2017 University of Warwick Frontiers of Finance, 2017 Asian Meetings of the Econometric Society, 2017 CAFM Meeting, 2016 PKU-NUS Annual International Conference on Finance and Economics, 2015 KSA Meeting, Korea University, Northeastern University, Peking University HSBC Business School, the University of Colorado Boulder, Sungkyunkwan University, and Yonsei University.

Well Begun Is Half Done: Initial R&D Competence and Firm Growth

ABSTRACT

We examine the effects of initial R&D competence on innovation strategies and firm growth. Using project-level drug development data for entrepreneurial pharmaceutical firms, we show that R&D performance persists over long periods, suggesting initial conditions largely determine the cross-sectional variation in R&D successes. We find that firms with high initial competence are exploitative in their best segment and exit through an IPO earlier or receive greater amounts of venture capital funding. By contrast, firms with low initial competence are explorative, diversifying into multiple segments. Medicare Part D legislation as an exogenous shock to diversification incentives suggests a likely causal relationship. Our results are consistent with firm initial skills (versus luck) leading to different innovation strategies and growth.

Keywords: Firm Initial Conditions, R&D Competence, Innovation, Diversification, Firm Growth

JEL Classification: G32, G34, O32

1 Introduction

What fundamentally determines how an early-stage firm innovates and grows? The nature versus nurture or innate versus environment debate spans a variety of disciplines. Especially in finance, there has been an extensive literature that supports the conclusion that institutional environments play an important role in determining success of innovation and firm growth.[1] Despite the fact that many aspects of firm characteristics are innately determined, it is less emphasized in the literature how such early-stage firm characteristics influence entrepreneurial outcomes. In this study, we examine how entrepreneurial firms with heterogeneous initial competence grow differentially by analyzing the effects of initial R&D performance on innovation strategies and exit outcomes.

Very few previous studies have explored firm initial conditions. Time-invariant firm characteristics of firms influence, for example, future capital structure (Lemmon, Roberts, and Zender (2008)), public status and acquisition activities subsequent to going public (Maksimovic, Phillips, and Yang (2013)), and differences in productivity between firms and countries (Bloom and Van Reenen (2007) and Bloom, Eifert, Mahajan, McKenzie, and Roberts (2013)). Using survey data on manufacturing firms in India, a more recent work by Ayyagari, Demirgüç-Kunt, and Maksimovic (2015) finds that initial size, productivity, and legal form of a start-up in developing countries are persistent and have greater explanatory power in determining growth than the effects from financial institutions. Our study considers initial R&D competence and its effects on subsequent innovation strategies and growth outcomes. This complements the previous literature on R&D that often concludes that time-varying firm efforts such as R&D expenses or institutional supports primarily determine R&D performance.

Specifically, we highlight that a firm has inherent R&D competence that is one of the initial conditions possessed or able to be acquired at the initiation of business and that continue to affect the firm's performance and potential throughout the life of the firm. R&D competence is likely to be based on a set of firm skills and technologies that include skilled human capital, organizational endowment and culture, and superior technology (Drucker (2007)), and heterogeneous across firms because it is a complex harmonization of individual technologies

[1]Those institutional environments include legal environment (Demirgüç-Kunt and Maksimovic (1998), Rajan and Zingales (1998), Beck, Demirgüç-Kunt, and Maksimovic (2005), Acharya, Baghai, and Subramanian (2014)), banking competition (Cornaggia, Mao, Tian, and Wolfe (2015)), financial market development (Hsu, Tian, and Xu (2014)), product market competition (Aghion, Bloom, Blundell, Griffith, and Howitt (2005)), and policy uncertainty (Bhattacharya, Hsu, Tian, and Xu (2015)).

and production skills (Hamel and Prahalad (1990)).[2] Similarly, previous studies of industry life cycles often assume that firms randomly differ in their product innovation expertise which influences their success at product R&D (Klepper (1996)).[3] We expect that initial R&D competence as a key economic channel drives future R&D performance, and consequent venture capital (VC) funding and ultimate outcomes of firm exits.

We use new detailed project-level drug development data for all privately held firms in the drug industry sector from the BioMedTracker database for the sample period from 1985 to 2014. Our data have a unique advantage to examine how firm initial conditions affect innovation strategies and outcomes. The U.S. Food and Drug Administration (FDA) requires all firms that develop new drugs to follow identical steps from the initiation to the final approvals of their projects, irrespective of their outcomes. Therefore, distinct from widely used patent data (*e.g.,* Griliches, Pakes, and Hall (1988) and Hall, Jaffe, and Trajtenberg (2001)) that register only on innovation projects already succeeded and thus patentable, our new drug development data based on the FDA phase information include ongoing innovation that potentially leads to either a successful or unsuccessful outcome.[4] Our data thus allow us to capture firm-specific R&D competence based on the extent to which initiated drug projects are suspended, remain active or successfully proceed to the next phase of the development process. More specifically, focusing on 2,722 individual new drugs projects by 799 entrepreneurial firms that span 30 years and 21 different disease classifications, we create initial and time-varying (current) annual R&D performance measures that reflect the initial three years and each year, respectively. We effectively consider a firm's initial R&D competence within their very first years in combination with our hand-collected founding year data and also compare main versus non-main segment competence using the specific disease group information.

We begin our analyses by examining heterogeneous R&D competence across firms. We discuss and quantify how much of the observed firm-year R&D performance can be attributable to the time-invariant firm fixed effects and particularly to our initial R&D competence measure. We find that firm R&D performance is highly heterogeneous across firms from the earliest stage

[2]See also Henderson and Cockburn (1994), Hallen (2008), Powell and Sandholtz (2012), and Marquis and Tilcsik (2013) for other studies on sources of firm competence.

[3]For a better understanding of R&D competence, we also examine founder characteristics of entrepreneurial firms and find that related education and experience of the founders are associated with initial competence.

[4]Failed or ongoing projects which may better represent the challenges of innovation projects are not observable in patent data. A recent work by Farre-Mensa, Hegde, and Ljungqvist (2016) overcomes this limitation of patent data by using detailed microdata from the U.S. Patent and Trademark Office (USPTO) on all patent applications filed by startups since 2001. Although their data include both approved or rejected patent applications, whether innovation is always patentable remains as an issue (Moser (2013)).

and also highly persistent over time. Further, R&D competence within the initial three years is significantly associated with overall future performance, indicating that our initial R&D competence measure is one of the important initial conditions of firms in highly innovative industries. Our results are robust to controlling for other initial conditions, such as numbers of projects, competitors, and operating segments, and also an alternative R&D competence measure that takes project-specific or phase-specific difficulties into consideration.

The key challenge in our study related to persistent R&D competence is to identify whether our initial competence measure signifies firm skills or potentially persistent effects of pure luck. Initial R&D success can be random, and the randomly achieved good luck for the first project can be positively correlated with successes of subsequent projects. Alternatively, an initial R&D success can be indeed firm skills, if there is a significantly positive correlation between within-firm project successes.[5] We find evidence that firms with high initial R&D competence perform well in their other projects not only across time but also across different disease groups, consistent with a skill-based interpretation of our results.

We then explore how the persistent R&D competence that we find influences innovation strategies, exploitative versus exploratory.[6] In Maksimovic and Phillips (2002)'s model, firms with relatively low comparative advantage are more likely to diversify and operate in multiple segments. We note that this prediction is likely to be stronger for firms in innovative industries because success in the highly unpredictable environment of innovation depends on generating sufficient variation that at least some will prove to yield desirable results ex-post (McGrath (2001)). Therefore, our prediction regarding the relation between initial R&D competence and innovation strategies is that firms with relatively good initial performance in their main segment are more likely to remain focused and develop drugs within the same disease group because a shift to other industries would incur higher opportunity costs. In contrast, firms with low competence that have consequently been doing poorly in their initial main segment have greater incentives to pursue a high level of exploration by expanding to multiple other disease groups.

By tracking project portfolios of our sample firms across 21 different disease classifications, we find strong evidence that supports the above predictions. Our results show that firms with high initial R&D competence in their main segment are less likely to diversify into other disease

[5]We discuss this point in detail in Section 4.2 based on the mechanism motivated by a theoretical multi-armed bandit problem with dependent arms.

[6]Benner and Tushman (2002) define exploitative innovation as involving improvements to existing components and advancing the existing technological trajectory, and exploratory innovation as involving a shift to a different technological trajectory. See March (1991), Levinthal and March (1993), Benner and Tushman (2002), and Gao, Hsu, and Li (forthcoming) for other studies.

groups. On the other hand, firms with low R&D competence are more likely to expand into multiple other disease groups. This finding, summarized as variance-seeking by firms with low competence, is consistent with the notion that investment in innovation can be viewed as a purchase of real options (Thakor (2013)); under the option-like payoffs of innovation, firms with lower likelihood of in-the-money (*i.e.,* distant from the break-even point) are more likely to take risk in their innovation strategies, as the option value increases with underlying risk.

To better understand the patterns of variance-seeking exploration, we next examine into which industries such firms expand. One prediction regarding destination is that those firms will expand into industries previously proven to have higher success rates to be on the safe side. Alternatively, they may expand into industries with less matured projects by competitors and thus higher failure rates. We find that such firms are more likely to enter industries in which the number of incumbents is large (*i.e.,* more competitive industries), incumbent firms are less successful in developing new drugs yet, and their products tend to be in the earlier stage of the development process. Such industries can be categorized as the industries in the first (fluid) phase of Abernathy and Utterback (1978)'s industry life cycle model, in which technological and market uncertainties prevail and firms are still in pursuit of product innovations.[7] This finding is consistent with firms that have higher failure rates, potentially due to low comparative advantage in their original innovation, adopting more explorative innovation strategies not only by increasing the number of operating industries, but also diversifying into relatively younger phase industries.

We next examine the effects of the initial R&D competence on firm growth including IPO exit and VC funding. Going public is considered an important stage in the growth of a company (Pagano, Panetta, and Zingales (1998)), and also the best exit choice, the most successful firms choosing an IPO over other exit methods (Chemmanur, He, He, and Nandy (2015)). Prior studies also find that firms are more likely to go public when their expected profitability is highest (Pástor, Taylor, and Veronesi (2009), Chemmanur, He, and Nandy (2009), and Chemmanur, He, He, and Nandy (2015)). We, therefore, predict a positive link between initial R&D competence and likelihood of going public which implies successful firm growth. Continued or increased VC funding is another indicator of successful firm growth before exit, because VC investors and other financial intermediaries are viewed as a mechanism for evaluating prospective entrepreneurs and funding the most promising ones (King and Levine (1993) and Ueda

[7]The characteristics of such industries are also consistent with those of the emerging industries in Klepper (1996).

(2004)).[8] Thus, given that R&D competence is persistent, we similarly predict that firms with high initial competence are more likely to receive greater amounts of VC funding.

Recalling that exploration is consistent with a value-enhancing strategy under the option-like payoffs of innovation, we further examine whether a poorly performing firm's pursuit of an explorative innovation strategy through diversification also affects subsequent firm growth. Although we predict exploration to increase the likelihood of IPO exit or continued venture capital funding, how to test this effect is unclear, given our finding of a significant relationship between initial competence and exploration as well. To disentangle the direct effect of initial R&D competence on firm growth from the potential indirect effect of initial R&D competence on firm growth through diversification, we particularly introduce a mediation model that employs diversification as an indirect channel. We hypothesize that initial R&D competence in a firm's main segment influences diversification which, in turn, influences the firm's growth in concert with the direct relation of R&D competence itself to firm growth.

Our results show that both direct and indirect effects of R&D competence on growth are significant. We first find that firms with high R&D competence at their earliest stage are more likely to exit through an IPO earlier or receive more venture capital funding than firms with lower competence (direct effect). Second, we find that firms with relatively low initial R&D competence tend to diversify into other disease groups, thereby also increasing the likelihood to successfully exit through an IPO at the end or receive more venture capital funding before exit (indirect effect). We further find that the direct effect is much greater than the indirect effect.

We recognize that diversification decisions are endogenous. Thus, we instrument for diversification decisions using Medicare Part D legislation. Medicare Part D is a US federal government subsidized program of prescription drug insurance for Medicare beneficiaries aged 65 and over. The program can affect the incentives of drug manufacturing firms, especially those performing poorly in their original main industries, to explore specific new industries related to Medicare Part D with anticipation of a positive demand shock in such industries. We find the predicted result that firms with lower R&D competence increase the likelihood of going public or obtaining additional venture capital funding by diversifying into the industries covered by the

[8]Unlike previous studies that focus on the role of VC investment in promoting innovation and growth for early-stage entrepreneurial firms (see, for example, Hellmann and Puri (2000), Chemmanur, Krishnan, and Nandy (2011), Puri and Zarutskie (2012), Bernstein, Giroud, and Townsend (2016)), we highlight the role of VC funding as a key milestone of entrepreneurial firms' successful growth. This is less likely subject to a reverse causality issue because most of the firms in our sample receive their VC funding after the initial three years.

Medicare Part D program, whereas firms with higher competence are un- (or less) affected by the shock remaining focused on comparative advantages in their respective industries.

Our paper first contributes to a growing literature on the effects of firm initial conditions.[9] Our study is the first to show that firm R&D competence is inherently given and highly persistent over time. This further highlights the importance of innate characteristics over firm efforts or institutional supports, especially in highly innovative industries. Using the unique data on developments and outcomes of highly innovative projects, we provide evidence that initial R&D competence is a key economic channel that drives future R&D performance and consequent firm growth outcomes.[10] Moreover, by distinguishing between a skill-based and a luck-based explanation for persistent R&D performance, our results shed light on what initial conditions can potentially represent.

Second, our paper adds to the literature on determinants of innovation strategies by providing the first evidence on the effects of initial firm characteristics on explorative versus exploitative strategies. The existing literature makes numerous important contributions to identifying determinants of innovation strategies including managerial myopia in learning (Levinthal and March (1993)), variance seeking (McGrath (2001)), dynamic environment as characterized, for example, by fluctuation in demand and supply (Sidhu, Commandeur, and Volberda (2007)), tolerance for early failure (Manso (2011)), and form of equity financing or public market listing (Ferreira, Manso, and Silva (2014) and Gao, Hsu, and Li (forthcoming)) among others. Especially, Ferreira, Manso, and Silva (2014) find that the incentives of public firms are biased towards exploitative innovation, while those of private firms towards explorative innovation. Our paper, distinctively going back to the beginning of a firm, shows that the innovation incentives of startup firms vary between exploitative versus explorative according to their initial conditions concerning R&D competence. Particularly, our finding that firms adopt explorative innovation strategies not only by increasing the number of operating industries, but also diversifying into relatively younger and riskier industries extends the notion of variance seeking behavior in highly innovative industries.

Lastly, our paper contributes to the diversification literature. We find new economic effects of diversification, in which the indirect effect of R&D competence inheres, on successful firm

[9]See for example Lemmon, Roberts, and Zender (2008), Maksimovic, Phillips, and Yang (2013), and Ayyagari, Demirgüç-Kunt, and Maksimovic (2015) for the literature.

[10]We acknowledge that our data do have a limitation in that, being only for firms that develop new drugs, the sample size is considerably smaller than what can be drawn from patent data. Notwithstanding the smaller sample size consequent to focusing on a specific industry sector, our findings strongly support our hypothesis and draw important implications for IPO market and VC industry.

growth manifested in IPO exits and VC funding. A large literature finds diversification discounts as value-destroying firm decisions, by focusing on the potential frictions in governance and financial markets (*e.g.,* Lang and Stulz (1994), Berger and Ofek (1995), and Comment and Jarrell (1995)). Another strand of the literature highlights that diversification discounts arise endogenously as a result of value-enhancing firm decisions (*e.g.,* Campa and Kedia (2002), Villalonga (2004), and Maksimovic and Phillips (2002)). Consistent with the latter strand of the literature, we find that explorative innovation through diversification by firms with relatively low R&D competence can be value-enhancing under the option-like payoffs of innovation.

Our paper proceeds as follows. In Section 2, we discuss our data and variables. Section 3 presents summary statistics. In Section 4, we discuss R&D competence more in detail regarding its persistence and also by contrasting skill versus luck. Section 5 discusses our empirical models and presents results for R&D competence, diversification, and firm growth. Section 6 concludes.

2 Data and Variable Construction

2.1 Drug Project Development Data

We obtain our project-level drug development data for all publicly and privately held firms in the drug industry sector from the BioMedTracker database. BioMedTracker is a real-time database that identifies biotech and pharmaceutical investment opportunities by assessing drug pipelines and future catalysts. The database tracks drug impact events from the 1950s using multiple sources including the FDA approval database, news articles, press releases, company filings to the Securities Exchange Commission (SEC), medical conferences, conference calls, direct communication with companies, and the ClinicalTrials.gov database. Although the FDA provides comprehensive information on the approved drugs including the approval date, the FDA approval database does not disclose in-process information for each ongoing project. Differently, BioMedTracker provides information on all drug pipelines which contains the detailed development phase and the outcome of each project.

The drug development is closely related to the FDA requirements. It is mainly separated into the pre-clinical research on micro-organisms and animals, and the clinical trials (phase 1, phase 2, and phase 3) on humans. In the pre-clinical stages, new compounds are identified in laboratories and companies perform safety tests for phase 1. Then, an Investigational New Drug (IND) application should be submitted to the FDA with the information on the effect of

the active ingredients and toxicities. Once the IND is approved, the development moves to the clinical phases. The clinical phases include three steps - phase 1, phase 2, and phase 3. In phase 1, safety and dosing issues are reviewed with healthy volunteers. In phase 2, the effectiveness of a drug is tested with a small number of people who have a certain disease or condition. Phase 3 conducts large-scale trials for safety and effectiveness of a drug with several hundred to 3,000 people. At the end of phase 3, a New Drug Application (NDA) or Biologic License Application (BLA) should be submitted for the FDA's review and final approval. Finally, the FDA thoroughly reviews all of the submitted data with an NDA or BLA and considers approving a new drug for marketing.

Our sample primarily comes from these project-level new drug development data for the sample period from 1985 to 2014 according to the availability of the BioMedTracker data. We note that the data coverage before 2000 in the BioMedTracker database may not be complete. This is because one of the important sources for the phase information of the database is the ClinicalTrials.gov website, but the website was created and publicly available from February 2000. Our results are robust to the further refinement of the sample period from 2000 to 2014. Our final sample size is 3,851 annual firm observations with 799 unique private firms.[11] Our sample is a subset of firms in the biotech and pharmaceutical industries that only focus on new drug discovery. The relevant SIC codes for firms in our sample are 2834 and 2836. Our sample also excludes firms that develop, manufacture or sell generic drug products. Although our sample is only for firms with new drug development, the number of unique private firms in our sample reaches approximately 65% of the number of private firms in prior studies that use patent data.[12]

2.2 R&D Competence Measure

From the BioMedTracker database, we primarily obtain the information on the status of drug development phase including the date of an advance to the next phase or suspension in the middle of the development. We measure firm-specific R&D competence based on whether each

[11]The number of unique firms in the BioMedTracker database during our sample period is 1,582. Among them, 533 are identified as public firms in their first observations. Among 1,049 private entities, we drop universities, hospitals, and firms whose ages are greater than 10 in their first observations. We also drop 19 firms with reverse mergers during the sample period.

[12]For example, Gao, Hsu, and Li (forthcoming) also consider private innovative firms using the patent data and the CapitalIQ database for all industries. Their sample contains 1,221 unique private firms that stay private (829) and also go public (392) during the sample period from 1997 to 2008. The total number of firms in our sample for the period after 1997 is about 65% of the number of firms in their sample.

project within a firm in a given year is suspended, stays, or advances to the next phase.[13] Distinct from the widely used patent data, our new drug development data includes not only successful projects but also all failed and ongoing projects.[14] Using this unique advantage of the data, we construct our measure of R&D competence that reflects the numbers of events for the advance to the next phase and the suspension of a project. More specifically, we assign -1 for the suspension event, 0 for no change in phase, and +1 for the advance event, and then scale the sum of these scores by the total number of projects in the pipeline in a given year. Thus our competence measure ranges from -1 to +1. It is important to note that a suspension or an advance to the next phase in new drug development is less likely to be affected by rival firms and thus the firm-specific success rates potentially capture each firm's competence.

This R&D competence measure, how frequently a firm advances, carries on or suspends its drug development projects, reflects the firm's potential to grow in the drug industry sector. Whether a firm successfully develops a new drug and finally gains an FDA approval to market the drug is vital to a firm's success and continuation. In the middle of clinical trial phases, however, trials can be suspended for a number of reasons. Firms may voluntarily suspend or terminate their clinical trials at any time when they believe that their trial results do not show the expected effectiveness. Furthermore, if the trial results are not successful, firms cannot proceed to the next trial phase. Also, regulatory agencies can order a suspension if they believe that the clinical trials are not being conducted in accordance with applicable regulatory requirements or the trials present an unacceptable safety risk to participants including significant side effects. Suspended projects can be resumed with a new clinical trial design or a modification of the existing trial design when the issues are in the experimental design, but we note from the data that resuming a suspended project is not a common event.

According to a study on the cost of developing a new drug by the Tufts Center for the Study of Drug Development, the clinical phase transition probabilities from phase 1 to phase 2 is 59.52%, phase 2 to phase 3 is 35.52%, phase 3 to an FDA submission is 61.95%, and

[13]The identification of suspension events became more accurate after the passage of the FDA Amendments Act of 2007 (FDAAA). Section 801 of FDAAA requires firms to submit the results of clinical trials within 12 months after the completion date.

[14]The NBER patent database (Hall, Jaffe, and Trajtenberg (2001)) provides patent filing and citation information that is widely used as an innovation measure in the literature. We note, however, that more matured firms are likely to patent their ideas. We match startup firms in our sample to firms in the NBER patent data for 1976-2006. Only 114 firms out of 360 firms (32%) in our sample with founding years prior to 2007 are matched to the firms in the patent data. Also, the correlation coefficient between our competence measure and the number of patents or citations is negative but insignificant. In particular, no correlation coefficients with the patent characteristic variables exceed 10% in absolute magnitude, demonstrating that the information contained in our competence measure is unique relative to the information in the patent data.

an FDA submission to final approval is 90.35%. Overall, the clinical approval success rate is only 11.83% and historically the rates have declined significantly.[15] Drug development is a long process from synthesis to an approval and with 16 to 31 months of the average phase transition time. The estimated total cost associated with new drug development is $2.9 billion, and the cost is also increasing over time. Due to the increases in the costs, continuing the drug development without suspensions and proceeding to the next phase quickly is critical to the success and continuation of a startup firm in the drug industry sector. The average number of projects that a firm in our sample has is only 2.2, and the median is 1 (the detailed summary statistics are presented in Table 2). This also indicates how a firm can quickly advance its new drug development without any suspension can be a relevant measure of firm R&D competence.

Our R&D pipeline data spans 30 years and 21 different disease classifications, which allows us to effectively consider a firm's initial R&D competence within their very first years using our founding year data, and also main versus non-main segment competence using the specific disease group information. Based on the fact that the average phase transition time is 16 to 31 months, we use the average of the first three non-missing R&D competence values to construct a firm's initial competence. We also consider the average of the first five non-missing R&D competence values, and our results are robust to using the five-year window. We define a firm's major segment (disease group) based on where the firm has the largest number of projects. Then we construct a firm's main R&D competence using the average R&D competence of the firm's main segment. Throughout this paper, we focus on a firm's initial R&D competence in its main disease group. Figure 1 shows distributions of the initial main R&D competence based on the first three and five year observations. The variable ranges from -1 to +1 by construction and roughly shows a bell shape distribution that is highly concentrated at zero.

[Insert Figure 1 Here]

2.3 Potential Concerns in Our Competence Measure

In this section, we address potential concerns regarding our initial competence measure. One possible concern is that firms may have different risk-taking preferences and thus some firms could initially start with multiple projects as random experimentations simply anticipating luck, which does not necessarily indicate true firm competence. Inconsistent with this hypothesis, firms in our sample maintain only 2.2 projects on average as we later show in the summary

[15] See DiMasi, Grabowski, and Hansen (2016) or the study by the Tuft Center available at `http://csdd.tufts.edu/files/uploads/Tufts_CSDD_briefing_on_RD_cost_study_-_Nov_18,_2014..pdf`

statistics. The low number of total projects is likely due to the high costs of developing a new drug estimated as approximately $3 billion. This may preclude the concern based on the view that entrepreneurship is fundamentally about random experimentation.

Our measure of R&D competence includes suspended projects when we count the total number of projects in the pipeline for a firm in a given year for normalization because there is a non-zero possibility that a suspended project can be resumed. However, this causes a concern that we may count permanently suspended projects that will have no further change in phase as on-going projects. Therefore, in one of our alternative measures, we drop all suspended projects when counting the total number of projects in the subsequent years after they are initially identified as suspended. Our findings are robust to this alternative measure.

There can be a concern that firms endogenously choose the difficulty of their project across disease groups and this choice may be correlated with firm competence. For example, Hay, Thomas, Craighead, Economides, and Rosenthal (2014) document that the probability of success in new drug development widely varies across disease groups with oncology drug development being more likely to fail compared to non-oncology drug development. If so, more skilled firms are more likely to engage in difficult drug development such as oncology projects. We address this concern by including 21 different disease group fixed effects in our analyses. In unreported results, we also find that our results are robust to using a more stringent grouping based on 227 different indications.

A related concern is that firms endogenously choose more difficult but more profitable projects as the expected payoff of a project depends on both the probability of success and the payoff conditional on success. Because the BioMedTracker database does not provide actual or expected future sales data on each drug, we acknowledge that we cannot completely eliminate this concern. We instead proxy the expected profitability of a new drug with actual profits of selling drugs with the same indication by public firms that have available financial data. More specifically, project profitability is measured by the actual growth in profits (EBITDA) divided by sales using public firms' financial data for the subsequent years after their projects with the same indication advance to the final stage of drug development (phase 5: FDA-approved for marketing). Then we examine whether our competence measure is correlated with the profitability. We find there is no correlation between our measure and the profitability.

For our main measure of competence, we assume that failure rates across phases are identical as 50% and thus assign -1 for a suspension event and +1 for an advance event regardless of the

phases. However, the probability of success varies across phases with the transition probability from phase 2 to phase 3 being the lowest. To address this concern, instead of using -1 and +1 we consider another weighting scheme that takes account of phase-specific difficulties. For this, we use the phase-specific success rate estimates in DiMasi, Hansen, and Grabowski (2003). For example, when the success rate from phase 2 to phase 3 is 36%, we assign -0.72 (=-36/50) for a suspension event, 0 for no change, and +1.28 (=64/50) for an advance event. Our results are robust to using any set of phase-specific success rates in the prior studies.

Also, we note that a single drug can be effective for multiple indications. For example, Epeius Biotechnologies Corporation in our sample developed a drug named *Rexin-G*, and it was initially tested for multiple indications including Breast cancer, Pancreatic cancer, and Sarcoma. Although most drugs have a lead indication which is their primary treatment focus, we consider the distinction between the drug and its indication and define a project at the drug and indication level. Therefore, in the previous example, Epeius Biotechnologies Corporation has three different projects at the same time under our definition of a project. Our results are robust to keeping only one lead indication for one drug instead of considering all possible indications that one drug is tested for.

Another possible concern is a change in the coverage of the BioMedTracker data, as they have expanded with the launch of the ClinicalTrials.gov website that is publicly available from February 2000. Therefore, the initial three or five-year data used to compute the initial competence can be random three or five years in the middle of life for some firms that are newly added with the expansion of the coverage in 2000. To address this concern, we drop firms whose ages are greater than 10 in their first observations for any reasons.[16]

Taken together, robustness tests using multiple alternative R&D competence measures suggest that it is unlikely that our main results critically depend on our construction of a measure of R&D competence, although we acknowledge that we cannot completely eliminate all concerns mainly due to the data limitations.

2.4 Other Measures

We also importantly consider a firm's diversification, as we interpret diversification to multiple disease groups as one of the important representations of explorative innovation strategies. Our diversification measure is based on how a firm creates its project portfolio across different

[16]Our results are robust to using different age cutoffs for the first observations.

disease groups. It is the number of all different disease groups where a firm has new drug projects in a given year. Alternatively, we also measure a firm's diversification by considering relative project shares of each disease group. It is calculated as one minus the sum of squared project share of each disease group in a given year. The project share of each disease group is the number of projects in the disease group divided by the number of total projects over all disease groups. For example, a firm with one project in each of two different disease groups has diversification measure of 0.5 $(1 - (0.5)^2 - (0.5)^2 = 0.5)$.

Our industry classification is based on the 21 different disease classifications that the BioMedTracker database identifies. The main industry of a firm is set as an industry where the firm has the largest number of projects in a given year. If there are more than one industries with the largest number of projects, we randomly select one industry. Table 1 provides the list of our 21 industry groups. Oncology is the most common disease area in our sample (30.43%), and then Neurology (14.62%), Cardiovascular (8.18%), and Endocrine (7.50%) follow.

[Insert Table 1 Here]

We consider two different measures that potentially capture overall firm growth - IPOs and venture capital funding. Going public is considered as an important stage in the growth of a firm and the best exit choice where the most successful startup firms choose to have.[17] Also, VC funding is an important indicator of growth. VC funding is an important external financing that is provided by venture capitalists to seed early-stage startup companies generally with a novel technology or business model. The previous VC literature suggests that VC funding increases the performance of a firm and thus the likelihood to exit private (*e.g.,* Lerner (1995) and Kortum and Lerner (2000)). Our VC funding data comes from the VentureXpert provided by the SDC database. We consider either a firm-level dummy variable of VC funding that equals one if the firm has ever received funding from VC investors or Log(1+VC funding), defined as the log of one plus the amount of VC funding in million dollars. Our IPO data comes from the SDC New Issues database. We construct a firm-year level dummy variable of an IPO event that equals one if a firm goes public in a given year.

We control other relevant firm-specific characteristics that may potentially affect firm R&D competence, diversification and firm growth. Those control variables include the number of phase 0 projects, the composition of R&D projects (early versus matured), the presence of partner organizations, and the firm age. We classify early- or matured-stage projects based on

[17] In our later analysis, we consider both IPOs and acquisitions as successful exits of startup firms. Among 799 unique firms, 48 firms have exits via acquisitions. Among 48, 17 (35%) are post-IPO acquisitions.

their phase information. The early-stage includes phase 1, phase 2, and phase 3 projects, and the matured-stage includes phase 4 (NDA/BLA) and phase 5 (Approved). The composition of R&D projects potentially captures growth opportunities of a firm. In the drug industry sector especially, firms often team up with so-called partners or sponsors to be supported technically, financially or with regulatory expertise. Therefore, the existence of partner organizations can affect the success rate of R&D projects. The BioMedTracker database provides the names of partner organizations for each project if exists. We create a dummy variable for the presence of partner organizations for each project and then calculate the fraction of the partnered projects over all projects. We expect that firms with more partnered projects are likely to show better performance and thus more likely to grow successfully. Lastly, firm age may capture any differences in the development stage based on tenure. These firm-specific control variables help to mitigate remaining concerns with our R&D competence measure. For example, remaining concerns include possibilities that firms with a small number of total projects likely show higher initial competence, and that firms may delay entering markets until they have successfully matured projects and thus higher initial competence.

We also control for the effects of the industry characteristics including the total number of firms with new drug development projects in a given industry (*i.e.,* main disease group), the industry's overall failure rate, and the industry's composition of R&D projects alongside industry fixed effects. We consider the number of all firms with new drug development in each industry as one of the proxies for R&D competition. We also consider the overall failure rate within an industry, which potentially indicates the difficulty of developing new drugs for the specific disease group. Lastly, the industry's composition of R&D projects is the fraction of matured projects within an industry. Although our main focus is on private firms, these three industry variables are based on our extended sample that considers all privately and publicly held companies that intend to market their drugs in U.S. This is more accurate way of calculating industry conditions than using either private or public firm sample. The Appendix describes all our variables more in detail.

3 Summary Statistics

Table 2 presents summary statistics of the variables used in our analyses. The sample consists of 3,851 firm-year observations of 799 private firms in the drug industry sector during our sample period from 1985 to 2014. The main focus of our analyses is on diversification and firm

growth determined by the initial R&D competence of a firm. Our diversification measure with the number of different industries is 1.38 on average, and the median firm in our sample is not diversified. 14.2 % of our sample firms went public at the end. 52.2 % of firm-year observations have VC funding, and the average amount of VC funding is 3.812 million dollars.

[Insert Table 2 Here]

Table 2 shows that firm overall R&D competence in our sample is -0.064 and -0.063 if we only consider main segments. The negative R&D competence indicates that the number of projects suspended is larger than the number of projects advanced to the next phase on average. This is consistent with the difficulties in developing a new drug. We also note that firms are more likely to stay in the same phase without frequent events of suspension or advance as the distribution of R&D competence is highly concentrated at zero. The initial competence measure (the average of the first three initial observations) for main segments is 0.011 on average with a standard deviation of 0.235. Firm age is on average 7.3 years, thereby indicating our sample well captures startup firms in the drug industry sector. As we discuss in Section 2.3 regarding concerns with our R&D competence measure, the summary statistics show that firms maintain 2.2 projects on average. It is unlikely that our measure merely captures the extent to which entrepreneurial firms engage in a large number of random experimentations. Approximately 9% of projects are in the matured stage, and 40% have partner firms. The average project profitability, proxied by public peer firms' EBITDA growth divided by sales in the year after their projects with the same indication advance to the final stage of drug development, is 27% and 10% when it is industry-adjusted by subtracting the industry median.

All our industry variables are based on firms' main segments. The average number of competing firms in a firm's main industry that include both privately- and publicly-held companies is approximately 50. The average industry-wide failure rate is 21.7% with a standard deviation of 10.4%. In an untabulated result, Orthopedics, Gastroenterology, and Ophthalmology industries' suspension rates are lowest as approximately 15 %, while suspension rates in Psychiatry, Cardiovascular, and Respiratory industries are highest as 25%. The average fraction of matured projects at the industry level is 23.3% with a standard deviation of 18.0%.

Next, in Table 3, we compare our variables between firms that have high versus low initial main competence in Panel A, and between firms that are diversified versus focused in their operations in Panel B. In Panel A, the sample consists of 3,851 firm-year observations of 799 firms during the sample period of 1985-2014, while in Panel B the sample consists of 3,091 firm-year observations of 646 firms that have low initial main competence. Firms with low

initial main competence are the firms that have nonpositive initial main segment competence. A firm is diversified if the firm develops new drugs over multiple disease groups, and focused otherwise.

[Insert Table 3 Here]

In Table 3, Panel A shows that nearly all our variables are significantly different between the two groups. First, firms with high initial main competence are more likely to go public or get VC funding during our sample period. 20.7 % of firms with high main competence go public through IPO, whereas only 12.6 % of firms with low main competence go public. VC funding amounts are also significantly greater for the high competence group. 59.2% of firm-years within high main competence group get VC funding, whereas only 50.5% of firm-years within low main competence group get VC funding. The log amount of VC funding is 0.56 (1.75 million dollars) for the high group and 0.42 (1.52 million dollars) for the low group. High competence firms have the greater number of total projects but less matured-stage projects. High main segment competence firms are less likely to have projects jointly with partners. It is worth noting that expected project profitability and its industry-adjusted variable are not different between high and low competence groups. This indicates that there is no evidence in our data that some firms select more difficult projects with greater potential profits and it mechanically leads to low competence for those firms, one of the previously discussed concerns regarding our R&D competence measure.

In Panel B of Table 3, we compare between diversified and focused firms within the subset of low main competence firms. Firms in the drug industry sector are in general focused because roughly 77% of firm-year observations in our sample period are in the focused status. The table shows that the diversified firms are also more likely to go public and get VC funding during our sample period. Approximately 4.4% more firms in the diversified group go public than firms in the focused group. VC funding amounts are also significantly larger for the diversified group. Diversified firms are about 8.2% more likely to get VC funding. It is also worth noting that diversified firms in low main competence group have both greater numbers of total projects and more matured-stage projects compared to the focused firms. Overall results indicate that there is heterogeneity in innovation strategies and subsequent growth patterns even within the group of firms with low initial main competence.

[Insert Figure 2 Here]

Figure 2 graphically confirms our findings in Table 3. The figure illustrates the time trends of the percentage firms of going public and the amount of VC funding over firm age. In Panel

(a) and (b), the sample is split by the initial competence of main disease group, and in Panel (c) and (d) by diversification. Panel (a) shows that firms with high initial main competence are more likely to go public in the first ten years and earlier than the firms with low initial competence in general. Firms with high initial main competence are the most likely to go public at the age of 5 and 6, but the likelihood stays relatively stable for the firms with low initial competence. Panel (b) shows that firms with high initial main competence continue to get the greater amount of VC funding at least for the first ten years. VC funding amount is especially greater for the high competence group in the early years of the firm life. The VC funding amounts for the low competence group are smaller compared to the high competence group and do not fluctuate as much. The gap in the VC funding amount between the two groups becomes smaller as firms get older. It is partly because high competence firms tend to go public earlier and drop out from the sample.

Panels (c) and (d) present analogous results for focused and diversified firms within the subset of firms with low initial main competence. The figures show that, for the subset of firms with low initial main competence, diversified firms are more likely to go public and get the greater amount of VC funding on average. In Panel (c), focused firms are more likely to go public in the first two years. It is potentially because the relatively better firms within the group of low initial main competence are likely to stay focused than other firms. Also, in Panel (d), the converging pattern of VC funding at the later stage is likely associated with the fact that relatively better diversified firms with higher competence tend to go public earlier and thus drop out from the sample.

4 R&D Competence

4.1 Persistence

This section examines whether the initial R&D competence is persistent over time and the R&D performance at the beginning of a firm's life significantly determines the firm's subsequent on-going R&D performance. Our predictions and following tests are to show whether we can consider R&D competence as one of the firm initial conditions that continue to affect the firm's performance and potential throughout the life of the firm.

[Insert Figure 3 Here]

Figure 3 presents the evolution of firms' overall R&D competence over firm age. We split

our sample by high versus low initial main R&D competence. Main competence is the R&D competence of the firm's main segment, and we take the average of the first three non-missing values of main competence for the initial main R&D competence. High main competence is the group of the firms with positive initial main competence, and low main competence is for the group of the firms with non-positive initial main competence. We find that the firm overall R&D competence does not fluctuate much in both high and low main competence groups. The firms in the high initial main competence group maintain their R&D competence positive. Also, the firms in the low initial main competence group stay low all the time during the period that spans more than ten years of firm life.

[Insert Table 4 Here]

Table 4 confirms the persistency of R&D competence that we find in Figure 3 using various regression models. Specifically, we estimate the following specification:

$$\begin{aligned} \textit{Main R\&D competence}_{i,t} = \alpha + \beta_1 \textit{Initial main R\&D competence}_i + \beta_2 \textit{R\&D characteristics}_{i,t} \\ + \beta_3 \textit{Other firm characteristics}_{i,t} + \beta_4 \textit{Industry characteristics}_{i,t} + \alpha_t + \epsilon_{i,t}, \end{aligned}$$

where α_t capture the year fixed effects. Before estimating the model, we first begin by showing the effect of firm and year fixed effects in predicting current R&D competence. Column (1) of the table shows that firm fixed effects alone explain 30.0% of current main R&D competence. This supports that in R&D competence there exists a permanent component and the permanent component is firm-specific nature. Column (2) shows that combination of firm and year fixed effects explains 37.5% of current R&D competence, indicating that the explanatory power of firm fixed effects is significantly greater than that of year fixed effects.

Then, we compare explanatory powers of the variables that potentially affect R&D competence. We first regress the current main R&D competence on the initial main R&D competence (our main variable of interest) with year fixed effects. Because the initial main R&D competence is the firm level variable, we exclude firm fixed effects. In column (3), we find that initial main R&D competence alone significantly explains 15.2% of current R&D competence. In columns (4) to (6), we compare explanatory powers between the initial main R&D competence and each of the following three other initial condition variables: Initial log(number of projects), Initial diversification, and Initial log(number of competitors). All these initial condition variables are the averages of the first three non-missing values.[18] Both the R-square and the adjusted R-

[18]The correlation coefficients between initial main R&D competence and Initial log(number of projects), Initial diversification, and Initial log(number of competitors) are -0.121, -0.046, and -0.061, respectively.

square in each column from (4) to (6) do not or minimally increase relative to that in column (3), as we additionally include each of the other initial condition variables in the regression. The results in columns (4) to (6) show that the initial main competence has the greatest power to explain current R&D competence among all initial condition variables we consider.

In column (7), as in the above model, we take account of all other variables that have potential impacts on current R&D competence. These variables include the number of phase 0 projects, firm age, percentage of matured projects, percentage of projects that have outside partners, number of competitors that also develop new drugs in the same industry, industry failure rate, and industry percentage of matured projects. These control variables are categorized into the following three groups: The R&D characteristics include the number of phase 0 projects, percentage of matured projects, and percentage of projects with partners. Other firm characteristics include firm age. Lastly, industry R&D characteristics include the number of firms in the same industry as a measure of competition, industry failure rate, and industry percentage of matured projects. We find in column (7) that initial main competence significantly explains the current main competence even after controlling for other variables. The estimated coefficient of the initial main competence is 0.738, suggesting a 0.173 (=0.235*0.738) increase in main competence with a one standard deviation increase in initial main competence. This increase is economically large translating into a 44.5 percentage change from the unconditional mean of main competence. The results also show that the newly added independent variables explain the variation of current R&D competence, but the increase in the adjusted R-square is just 4% relative to that in column (3).[19]

Next, we perform a variance decomposition of current main competence to quantify the explanatory power of existing determinants that include the initial main competence. Table 5 presents the variance decomposition results for the same model specifications as in Table 4.

[Insert Table 5 Here]

We report the fraction of the total variances for each variable in each specification, which is the partial sum of squares for the effect of the variable scaled by the aggregate partial sum of squares across all effects in the specification. The reported fractions in each column sum to one and assess the extent to which the total variation in R&D competence is attributable to the variations of particular variables in the model. For example, in column (2), 87.7% sum of squares in the model is attributable to the firm fixed effects, and the remaining 12.3% is to the

[19]In Appendix Table A.1, we consider alternative measures of R&D competence to address potential concerns discussed in Section 2.3. We find that our results are robust to using these alternative measures.

year fixed effects. When we replace the firm fixed effects with our main variable of interest, initial main competence, the attribution of the variable becomes even greater as 91.3%. In the comparisons to other initial condition variables throughout columns (4) to (6), the estimated attributions of the initial main competence remain the largest as approximately 90%. In column (7), we find that firm age and percentage of matured projects explain 10.7% and 4.2% of the variation in current competence, but the effects are not comparable to the effect of initial main competence. Overall, we find from both regression and variance decomposition analyses that initial R&D competence persists over time and shows statistically significant and positive relation to the firm's future competence in the long periods of time.

4.2 Skill versus Luck

Next, we take a deeper consideration of the notion of firm competence, especially whether firm competence is indeed firm skills compared to luck. One might argue that firm initial competence can be random, and the randomly achieved good luck for the first project might result in an ultimate firm success. We discuss this argument based on the mechanism motivated by a bandit problem and present evidence that initial competence is more likely to be firm skills.

In probability theory, a bandit problem is a setting in which a gambler sequentially pulls arms of a slot machine to maximize his/her expected total reward. The reward distribution of each arm is generally assumed to be independent and identically distributed ($i.i.d$) from a known distribution with unknown parameters. A basic formulation is Bernoulli k-armed bandit model, where the reward of each arm is one with probability p_k and zero otherwise. One of the important results of the model is the "stay-on-a winner" principle in Berry (1972). This principle states that if an arm proves to be successful at a stage, selecting the arm again is optimal at the following stage and the ultimate outcome is therefore path-dependent.

Bernoulli k-armed bandit problem can be translated into our context where a startup firm in our sample randomly selects one initial project (arm) among available k projects, and the distribution of each project's success rate determines the firm's switching behavior among projects and ultimate success. The standard solution of Bernoulli k-armed bandit problem, the stay-on-a-winner principle also applies to our problem. If a firm selects a new drug project with a higher chance of a good outcome, the firm will stay with the project and persistently show high competence. Furthermore, the firm is also more likely to exit successfully with an IPO or receive greater amounts of VC funding. Hence, the ultimate success of the firm is path-dependent and

significantly relies on the initial random selection of a project (*i.e.,* luck) not firm skills.

However, in the classical bandit problem, the assumption of independent arms is critical. When the rewards of arms are correlated at a different degree for each player, the final total rewards also vary across players. In Bernoulli k-armed bandit problem with *dependent* arms, the total reward is always greater when the correlation between arms is higher (Pandey, Chakrabarti, and Agarwal (2007)). The interpretation of this result in our setting is that if firms are heterogeneous in the correlation between the available projects, a firm with higher correlation (*i.e.,* skills) will always perform better in any stage than a firm with low correlation. We empirically test for this prediction using within-firm project performance correlations for a given time. Table 6 presents the results.

[Insert Table 6 Here]

The dependent variable is performance of one project within a firm for a given year, an indicator variable that equals one if a firm's project advances to the next phase, minus one if the project is suspended, and zero otherwise. Peer project performance is a variable of interest, which is the average of all other projects' performance for the same firm in the given year. We examine whether the correlation between project performance within a firm is significant and positive and whether it is greater for high competence firms than low competence firms.

In column (1) of the table, we find that performance of a project is significantly associated with other projects' performance within the same firm in the same year. The average correlation between the projects within a firm is estimated as 13.9% and significant at the 1% level. In column (2), we additionally include initial main competence and its interaction term with peer project performance, to examine whether the within-firm project correlation increases or decreases with firm initial competence. We find that firms with higher initial competence show greater correlation for within-firm projects performance, as the interaction term is significantly positive at the 1% level. We also consider a discrete version of initial main competence in column (3) and find the similar result that higher competence is significantly related to the higher correlation between the project performance. Overall, the results in this table suggest that firms with high initial competence are more likely to perform better almost always due to the reason related to performance correlation within their own projects (*i.e.,* firm skills). Hence, our results are not merely driven by a luck-based explanation.

4.3 Firm Skills and Entrepreneur Characteristics

To better understand the source of heterogeneity in firm initial skills, we further investigate associations between entrepreneurial characteristics and our measure of initial competence. We manually collect available characteristics on whether a founder has a Ph.D. degree, an academic major related to drug development such as Biology, Chemistry, and Medicine, and previous entrepreneurial or executive experience, whether the founder is a female or an inventor, whether a firm is based on a research laboratory in a university or hospital, and whether the firm has multiple founders. Table 7 shows the summary statistics for these entrepreneurial characteristics and their correlation coefficients with our competence.

[Insert Table 7 Here]

Panel A presents summary statistics on founder and other related entrepreneurial characteristics of the firms in our sample with available data. 78% and 80% of founders in our sample, respectively, have a doctoral degree and a major related to drug development for their higher education, for example, Biology, Chemistry, or Medicine. 34% of the founders are inventors with patent records, and 8% are female founders. 51% and 68% have previous entrepreneurial and executive experience which may represent the skills in successfully running a startup firm and a firm in general, respectively. We find that 38% of the firms in our sample start off with a research laboratory at a university or hospital and 32% of them have more than one founder.

Panel B of the table displays Pearson correlation coefficient between our initial main competence measure and these entrepreneurial characteristics. With the estimated correlations, we examine which entrepreneurial characteristics are positively or negatively associated with the initial firm skills measured within its first three years. We find that the initial competence of a firm is positively related to the founder's education and research activity as shown by the positive correlation coefficients of a doctoral degree, related major, and inventor. A founder's previous experience as an executive is positively related to the initial competence of the founder's firm, while the association with experience as a previous entrepreneur or a female founder is estimated negatively. Also, when a firm originally starts off with a research laboratory at a university or hospital, the initial competence is shown to be higher. In contrast, having multiple founders is negatively related to the firm initial competence. Although this correlation analysis provides limited information, it helps reify what can be initial skills for the innovative early-stage firms.

5 Empirical Methodology and Results

The key question is whether firm initial conditions, especially in R&D competence, explain the differential patterns of innovation strategies through diversification, and firm growth through an IPO exit or VC funding. We employ multiple different empirical methodologies to explore the direct and indirect relations between R&D competence and diversification, and firm growth.

First, we examine the effect of the initial R&D competence on diversification. We predict that higher opportunity costs prevent firms with high R&D competence in their main segment from diversifying into other industries. This prediction is associated with exploitative (explorative) innovation strategies by firms with higher (lower) competence. Second, we simultaneously estimate the effects of initial R&D competence on diversification and firm growth, and also the effect of diversification on firm growth. Figure 4 illustrates this empirical design that uses a simultaneous mediation model with diversification as a mediator enables us to disentangle the direct effect of R&D competence and the indirect effect of R&D competence through the diversification channel on firm growth.[20] Lastly, we address potential endogeneity concerns between diversification decisions and firm growth, using the Medicare Part D legislation in 2003 as a shock to diversification incentives in firms with low R&D competence.

[Insert Figure 4 Here]

5.1 Initial R&D Competence and Diversification

In this section, we examine whether a firm's R&D competence in its main segment affects the firm's decision to diversify into other industries. We assess how initial or current main R&D competence affects diversification. We specifically estimate the following specification:

$$\begin{aligned} &\textit{Diversification}_{i,t} = \alpha + \beta_1 \textit{Main R\&D competence}_{i,t} + \beta_2 \textit{R\&D characteristics}_{i,t} \\ &+\beta_3 \textit{Other firm characteristics}_{i,t} + \beta_4 \textit{Industry characteristics}_{i,t} + \alpha_i + \alpha_t + \epsilon_{i,t}, \end{aligned}$$

where *Main R&D competence* is either the initial main R&D competence or the current main R&D competence based on specifications. *Diversification* is either the number of all different disease groups where a firm has new drug projects or one minus the sum of squared project share of each disease group in a given year. α_i and α_t capture the industry fixed effects and year fixed effects, respectively. We regress our time-varying firm diversification measure on the

[20] Mediation analysis has frequently been used in social sciences. See for example Pearl (2012) in the statistics literature and Heckman, Pinto, and Savelyev (2013) and Heckman and Pinto (2015) in the economics literature.

variables that potentially affect the diversification decision including the main R&D competence (our main variable of interest), firm age, number of phase 0 projects, percentage of matured projects, percentage of projects that have outside partners, indicator of VC financing, number of competitors in the same industry, industry failure rate, and industry percentage of matured projects. All control variables are one year lagged.

Table 8 presents the results from the above regression specification. In columns (1) and (2), we use the diversification measure that is the total number of all disease groups where the firm has projects as a dependent variable. In columns (3) and (4), we use the other diversification measure calculated by one minus the sum of squared project share of each disease group in a given year as a dependent variable. Columns (1) and (3) examine the effect of the initial main R&D competence (*Initial main competence*), and columns (2) and (4) examine the effect of the current main competence (*Main competence*).

[Insert Table 8 Here]

From columns (1) and (3), we find that firms that have high initial competence in their main segment are less likely to diversify into other industries. The results are statistically and economically significant. A one standard deviation decrease in the initial R&D competence in a firm's main segment results in an increase in diversification by five percentage point (0.235*-0.206) as in column (1), and by one percentage point (0.235*-0.0399) as in column (3). This is consistent with our prediction that firms with high initial competence in their initial main segment are more likely to stay focused due to the high opportunity costs of entering into other industries. In columns (2) and (4), we also find the similar results using the current main R&D competence instead, but the magnitudes of the coefficients are smaller than those of the initial main competence in columns (1) and (3).

Table 8 also shows how other firm R&D characteristics and industry characteristics relate to diversification decisions. First, firm R&D progresses as shown by the number of phase 0 projects and the percentage of matured projects are positively associated with diversification decisions. Firm age is also positively associated with diversification, indicating that firms start with a single segment in their early years and tend to diversify into multiple other industries as they grow older. The coefficients of the percentage of projects with partners and VC backing indicator are also positive, implying that firms with more operational and financial resources are more likely to diversify. At the industry level, competition measured by the total number of competing firms and the percentage of matured projects in a firm's main industry are positively related to diversification decisions.

5.2 Diversification Patterns

Next, we further examine which industries firms with relatively low initial competence in their original main industry expand their business into. Table 9 presents summary statistics on characteristics of the destination industries where the diversifying firms newly enter. We find that 165 firms among 799 firms in our sample expand to other disease groups, with 220 unique diversification events. We examine the following relevant industry characteristics to better understand their diversification patterns. This is important to shed light on the fate of the unsuccessful startup firms in innovative industries, which has not been developed in the literature.

[Insert Table 9 Here]

We define the following indicator variables that equal to one if the industry average is above the median value for all industries. Those variables include industry R&D competence (High competence), project suspension (High suspend rate) and advance rates (High advance rate), number of products (More products), percentage of matured projects (More matured product), percentage of phase 0 product (More phase 0 product), competition (High Competition), number of industry competitors (More incumbents), and project profitability (High profitability).

We find from the table that firms tend to enter into the industries with higher suspension rates, lower advance rates, more total products, more phase 0 products, higher competition, but less matured. All these patterns are significant at the 1% level. However, we do not find evidence that diversifying firms enter into industries with average higher competence or historically higher profitability. These stylized facts are consistent with unsuccessful startup firms in their initial main industries newly starting their businesses in other emerging industries. The high suspension rates, more products, and more products in the early stages are the characteristics of the first phase (fluid phase) industries in the industry life cycles models such as the models by Abernathy and Utterback (1978) and Klepper (1996).

5.3 Initial R&D Competence, Diversification and Firm Growth: Simultaneous Analysis

Because initial R&D competence has twofold effects on firm growth, which are the direct effect and also the indirect effect through diversification, we simultaneously analyze both effects using a system of linear equations. In the system of equations, the two dependent variables are *Diversification* and a measure of firm growth. We consider either *Going public* or *VC funding* as a measure of firm growth. The dependent variable of the first equation, *Diversification* is

continuous and the second dependent variable, *Going public* is a dummy variable (*VC funding* is continuous). To precisely incorporate the potential correlation in the residuals between the two equations, we estimate seemingly unrelated regressions (SUR). Specifically, we estimate the following specifications:

$$\begin{aligned} Diversification_{i,t} = v + \beta_1 Initial\ main\ competence_{i,t} + \beta_2 R\&D\ characteristics_{i,t} \\ +\beta_3 Other\ firm\ characteristics_{i,t} + \beta_4 Industry\ characteristics_{i,t} + v_i + v_t + \epsilon_{i,t} \end{aligned}$$

$$\begin{aligned} Going\ public_{i,t} \text{ or } VC\ funding_{i,t} = w + \gamma_1 Initial\ main\ competence_{i,t} + \gamma_2 Diversification_{i,t} \\ +\gamma_3 R\&D\ characteristics_{i,t} + \gamma_4 Other\ Firm\ characteristics_{i,t} \\ +\gamma_5 Industry\ characteristics_{i,t} + w_i + w_t + \eta_{i,t}, \end{aligned}$$

where *Diversification* is the firm-year level diversification measure calculated by the total number of different disease groups where the firm has projects in a given year, *Going public* is one if the firm went public via an IPO in a given year, and *VC funding* is Log (1+VC funding amount) in a given year. The same set of control variables as in Table 8 is included, except the VC backing indicator when *VC funding* is used for a measure of firm growth in the system. As before, all control variables are one year lagged.

Table 10 reports the effects of initial main competence on either going public or getting VC funding directly and indirectly through firm diversification using a simultaneous mediation model. The results in columns (1) and (3) show that the diversification intensity declines with the initial main competence. The coefficients of *Initial main competence* is negative and statistically significant at the 1%. This is consistent with the result in Table 8, implying that low initial competence in a firm's main segment drives the firm to diversify into other industries.

[Insert Table 10 Here]

In column (2), we find that both initial main competence and diversification increase the likelihood of going public. The initial R&D competence in the main segment alone significantly increases the probability of going public. This direct effect of the initial R&D competence on the likelihood of going public is also economically significant. The economic direct effect of the initial competence is comparable to those of the VC backing indicator and the percentage of joint projects with partners.[21] The results show that firm inherent R&D competence, which is highly persistent over time, is a fundamental component of comparative advantages in highly

[21] In unreported results, we find that the current main competence, not initial competence also has a significant and positive effect of increasing the likelihood of going public.

innovative industries that drive exits through IPOs at the end.

The result shows not only the direct effect of the initial R&D competence but also the indirect effect of the initial R&D competence through diversification as a potential mediator. We find that a firm's low R&D competence in its main segment increases the firm's incentive to diversify its portfolio of new drug projects into multiple disease groups, and then the diversification increases the likelihood of going public. This indirect effect of the initial R&D competence on the likelihood of going public is statistically significant at the 5% level, but economically much weaker than the direct effect. The estimated coefficients in columns (1) and (2) imply that, regarding the likelihood of going public, the direct effect (0.0328) is roughly fourteen times greater than the indirect effect (-0.206*0.0115).

Regarding VC funding, we find in column (4) that only the indirect effect of the initial R&D competence through the diversification channel is statistically significant at the 5% level. The direct effect shown in the coefficient of initial R&D competence is positive but not statistically significant. The result of this specification implies that the initial R&D competence in a firm's main segment affects the likelihood of receiving more VC funding mainly through the diversification channel.

Overall, the results from our simultaneous analysis are consistent with the predictions that firms with higher R&D competence initially in their main segment are more likely to go public, whereas firms with lower R&D competence initially in their main segment increase the likelihood of going public or getting more VC funding by diversifying into multiple different industries. We further stress-test our results by focusing only on between-effects in the cross-sections of our sample. We consider a refined sample that consists of only the final year observations for each firm. For a firm that eventually goes public, the IPO event takes place in the final year. For other firms, we select observations in the last available year. Then, we analogously run the simultaneous mediation model as in Table 10. The results are available in Appendix Table A.2. We confirm that our results are robust in the refined cross-sectional sample.

5.4 Simultaneous Instrumental Variable Analysis with the Medicare Part D Legislation

We recognize that the two main variables of the system of equations - diversification and firm growth (the mediator and the outcome) are potentially endogenous. For example, managerial risk aversion may lead to diversification decision and the successful exit through an IPO as well.

When the mediator is endogenous in mediation analysis, the estimated indirect effect on the outcome is not causal. In such case, an exogenous treatment such as using instruments inducing an exogenous variation in the mediator is suggested in the mediation analysis literature (see for example Imai, Keele, Tingley, and Yamamoto (2011)). Following this methodology, we consider an exogenous shock that is likely to affect diversification incentives and firm growth only through the diversification channel to address these potential endogeneity concerns. Particularly, we use the passage of the Medicare Part D legislation (also called the Medicare prescription drug benefit) in 2003 as an exogenous shock to the diversification incentives.

Medicare Part D is a government program to subsidize the costs of prescription drugs and prescription drug insurance premiums for Medicare beneficiaries. Before Medicare Part D, Medicare program covered the medicine only associated with physician services. Part D significantly expanded drug coverage covered by Medicare. In particular, under Part D the drug coverage must include all drugs within the following indication classes: antipsychotics, antidepressants, anticonvulsants, immunosuppressants, cancer, and HIV/AIDS. Therefore, only specific firms with lack of the project pipelines in such indication classes should be materially affected by the shock regarding their increased incentives to expand into the relevant disease groups. Although firms anecdotally anticipated large excess profits as a result of the new Part D program, there was substantial uncertainty about prices, and thus profits related to Part D covered drugs. Previous studies on Medicare Part D show that the passage and the implementation of the program are associated with significantly lower prescription drug prices (Duggan and Morton (2010) and Duggan and Morton (2011)), but increases in pharmaceutical R&D for therapeutic classes that have higher Medicare market shares (Blume-Kohout and Sood (2013)).

The shifted demand for prescription drugs but their lower prices potentially due to increased competitions and government intervention create a suitable setting for variance-seeking exploration by firms that have been doing poorly in their original disease groups. Therefore, firms without any projects in the Part D drug classes before the legislation are more likely to have large incentives to expand their project portfolios into the Part D drug classes in pursuit of an explorative innovation strategy. In contrast, the diversification incentives are not as large in firms that have had active projects in the Part D drug classes before the legislation or firms that have been doing successfully in their original disease groups outside the Part D drug classes due to the opportunity costs of expansion. Based on these differential effects of the Medicare Part D on diversification incentives, we identify the treated and control groups as follows: Firms without any projects in the Part D drug classes before the legislation are included in the treated

group, while firms with active projects in the Part D drug classes before the legislation are included in the control group. The Part D shock should significantly affect the diversification incentives of firms in the treated group, but not (or less) affect those of firms in the control group. Furthermore, we expect that within the treated group the increased diversification incentives following the shock will be more pronounced for firms that show low initial competence in their original disease groups.

There may be a concern related to the exclusion restriction of our instruments that the Medicare Part D can have a direct positive effect on firm growth not through the diversification channel. However, this concern is not valid with our empirical design that groups firms with and without active projects in the Part D drug classes separately. For firms in the treated group, it is impossible to undergo the direct influence of the Part D on IPO and VC funding unless they diversify into the new industries supported by the Part D program. In contrast, it is obvious that for firms in the control group that already have active projects in the Part D drug classes, the Part D shock affects their IPO and VC funding directly. However, this channel is irrelevant to the exclusion restriction of our treated-group based instruments, because the instruments only concern firms in the treated group that have no active projects in the Part D drug classes. Furthermore, the increased likelihood of going public or receiving greater amounts of VC funding for firms in the control group through this channel leads to a result working against our founding. Overall, although no one set of instruments will assuage all concerns, the Medicare Part D shock we consider is highly likely to satisfy the exclusion restriction requirement, as the effect of Medicare Part D on firm growth comes only through the diversification channel by which firms without any project in the affected Part D drug classes can be exposed to the shock.

We reconsider the model specification in Table 10 with one key change: the introduction of instrumental variables for diversification using the treated group, the post-Medicare Part D, and low initial main R&D competence indicators and their interaction terms. Our continuous use of the system of equations with SUR is to examine both the direct effect of initial R&D competence on firm growth and the indirect effect of it through diversification. Separately, the use of the instrumental variables for diversification is to take into account the endogeneity between diversification and firm growth in the systems of equations. Specifically, we estimate

the following specification with a simultaneous instrumental variable approach:

$$\begin{aligned} Diversification_{i,t} = v + \beta_1 Treated_i \times Post_t \times Low\ competence_i \\ \beta_2 Treated_i \times Post_t + \beta_3 Treated_i \times Low\ competence_i + \beta_4 Post_t \times Low\ competence_i \\ +\beta_5 Treated_i + \beta_6 Post_t + \beta_7 Low\ competence_i \\ +\beta_8 R\&D\ characteristics_{i,t} + \beta_9 Other\ firm\ characteristics_{i,t} \\ +\beta_{10} Industry\ characteristics_{i,t} + v_i + v_t + \epsilon_{i,t}, \end{aligned}$$

$$\begin{aligned} Going\ public_{i,t} \text{ or } VC\ funding_{i,t} = w + \gamma_1 Low\ competence_i + \gamma_2 Diversification_{i,t} \\ +\gamma_3 R\&D\ characteristics_{i,t} + \gamma_4 Other\ Firm\ characteristics_{i,t} \\ +\gamma_5 Industry\ characteristics_{i,t} + w_i + w_t + \eta_{i,t}, \end{aligned}$$

where *Treated* is one if a firm does not have any projects in the required Part D drug classes before the legislation, *Post* is one after the passage of the Medicare Part D legislation in 2003, and *Low competence* is an indicator variable for a firm with nonpositive initial main competence. *Post* alone is subsumed by year fixed effects. *Treated*, *Post*, *Low competence*, and their interaction terms are the set of our instruments with the triple interaction term having the most important economic relevance to diversification.

[Insert Table 11 Here]

Table 11 reports the estimation results of the system of equations. We find that firms with lower competence in the treated group indeed increase diversification after the passage of the Medicare Part D legislation, consistent with our prediction. The coefficients of the triple interaction terms in the second row are significantly positive at the 1% level for both columns (1) and (3)–which can be thought as equivalents to first-stage regressions in non-simultaneous instrumental variable approaches. The economic impact of the estimated triple interaction terms (1.132 and 1.125) implies that following the Medicare Part D shock the treated firms with low initial competence entered into roughly one more disease group. The implications of other firm and industry characteristics on diversification are comparable to those in Table 10.

In column (2), regarding IPO exits, we find that both the initial R&D competence (*Low competence* in this specification) and the instrumented diversification significantly increase the likelihood of going public. As before, the economic significance of the direct effect of the initial R&D competence on the likelihood of going public is estimated as comparable to those of the VC backing indicator and the percentage of joint projects with partners. Also, the result shows

that the effect of diversification, even after our control of potential endogeneity concerns using the Medicare Part D shock, is significant and positive at the 5% level. We find that a firm with low R&D competence in its main segment that is not covered by the Medicare Part D program has higher incentives to diversify into the new industries supported by the Medicare Part D program, thereby being able to increase the likelihood of going public.[22]

Regarding VC funding, we also find consistent results. In column (4), the result shows that both coefficients for the initial R&D competence (*Low competence*) and instrumented diversification are positive and significant. This indicates that the diversification channel, as well as the initial R&D competence, increases the likelihood of getting more VC funding. For the former channel, we especially address potential endogeneity concerns between diversification and VC funding by using the Medicare Part D shock as an instrument for diversification.

The economic significance of both direct and indirect effects of initial competence is large. The economic interpretation of the direct effect is that the IPO likelihood and the VC funding amounts reduce by 60 percentage and 30 percentage of unconditional means, respectively, if a firm has low initial competence. For the indirect effect, firms with low initial competence are more likely to have one additional disease group than firms with high initial competence, and such an increase in the number of disease groups leads to 38 percentage and 11 percentage increases with respect to the unconditional means of the IPO likelihood and the VC funding amounts, respectively.[23]

As our instrument has a time series component, we consider a set of tests showing the timing of the Medicare Part D shock more directly and also falsification tests examining whether firms in treated and control groups behave divergently in a placebo year. The results are presented in Appendix Table A.4. We run the analogous test as in columns (1) and (3) in Table 11 in a non-simultaneous setting in column (1) and confirm that the effect of the triple interaction term on diversification is positive and significant. We next examine the timing of the Medicare Part D effects with the two-year windows for pre- and post-Medicare periods in column (2) and

[22]In Appendix Table A.3, we consider both IPOs and acquisitions as successful exits of startup firms and run the analogous IPO regressions as in columns (1) and (2) of Table 10 and Table 11. We find that our results are robust of considering both exit choices.

[23]The unconditional means of the IPO likelihood and the VC funding are 0.032 and 0.447 (=$3.812 million) in Table 2. The coefficients for *Low competence* in columns (2) and (4) translate into 60 percentage (=0.0192/0.032) and 30 percentage (=0.130/0.447) decreases with respect to the unconditional means. Also, the coefficients for *Low competence* in columns (1) and (3) translate into 1.132 and 1.125 more disease groups, and such increases in the number of disease groups lead to 38 percentage (=1.132*0.0108/0.032) and 11 percentage (=1.125*0.0439/0.447) increases with respect to the unconditional means of the IPO likelihood and the VC funding amounts, respectively.

the five annual windows around the Medicare shock in column (3). We find that the effect of the Part D shock on diversification for firms with lower competence in the treated group only manifests in the post two year period but not in the pre-period. Also, the effect is strongest right after the shock in 2003. In column (4), we consider a place event in 2010 and run a falsification test. We find that the triple interaction term with the placebo year is insignificant, supporting that our results cannot be explained by non-parallel trends between the treated and control groups in non-event years. Our results are robust to considering other placebo years.

As robustness, we consider the same test using two alternative measures of R&D competence that we discuss in Section 2.3 regarding concerns with our R&D competence measure. First, we drop all suspended projects when counting the total number of projects in the subsequent years after they are initially identified as suspended. Second, we consider a conditional success probability weighted R&D competence to reflect phase-specific difficulties of transition based on success rates estimated in DiMasi, Hansen, and Grabowski (2003). Our findings are robust to these alternative measures of R&D competence and the regression results are available in Appendix Table A.5 and A.6, respectively. Also, Povel, Sertsios, Kosova, and Kumar (2016) find that there is a significant and persistent entry year effect, more specifically whether it is during industry booms or glooms, on the performance of hotel investment. We thus consider entry year fixed effects instead of year fixed effects and run the analogous test. The results are available in Appendix Table A.7.[24]

Overall, our results are consistent with our predictions that firms with higher R&D competence initially in their main segment are more likely to go public or get more VC funding, whereas firms with lower R&D competence increase the likelihood of such successful growth by exploring other industries, especially where positive demand shocks are expected, for example in the Medicare Part D shock.

6 Conclusion

We examine private firms' diversification and growth in instances in which initial conditions in R&D competence exert strong effects on both simultaneously. Our study focuses on heterogeneous levels of R&D competence inherent at a firm's inception or acquired in the earliest stages of its life. Our detailed project-level drug development data that spans 30 years and

[24]Our results are also robust to dropping firm years where all projects are identified as suspended in a given year and thereafter. These firms are considered as firms that no longer operate.

21 disease groups in the drug industry sector enables us to measure innate R&D competence overall and in a firm's main segment by calculating the suspension and success rates of new drug development projects.

We find that R&D competence within its first three years persists throughout a firm's lifetime. We further examine how this persistent R&D competence affects innovation strategies, as measured by diversification into multiple disease groups and firm growth through IPO exit and venture capital funding. We find that firms with high initial R&D competence in their main segment are more likely to focus on their best segment and grow faster and more successfully than firms with low initial competence. Specifically, they exit through IPOs earlier or receive greater amounts of venture capital funding. By contrast, firms with low initial R&D performance in their main segment tend to diversify into other disease groups, thereby also increasing the likelihood of going public at the end or securing more venture capital funding.

Using a simultaneous mediation model, we show that both the direct effect of initial R&D competence on firm growth and the indirect effect, whereby the mediation effect of diversification exists, are significant, but the direct effect is much greater than the indirect effect. We confirm the mediation effect of diversification on firm growth to likely be causal, using the Medicare Part D legislation as an exogenous shock to diversification incentives. Our results add new evidence to the literature that initial conditions in the earliest stage of firm life largely explain variations in corporate diversification and growth patterns.

Appendix

Variable Definitions	
Going public	is the firm-year dummy variable that is one if the firm goes public via an IPO in a given year and zero otherwise.
Exit	is the firm-year dummy variable that is one if the firm either goes public via an IPO or is acquired in a given year, and zero otherwise.
VC funding	is the log of one plus the amount of VC funding in million dollars in a given year.
Diversification w/ number of industries	is the total number of different disease groups where the firm has projects.
Diversification w/ project shares	is one minus the sum of squared project share of each disease group in a given year. The project share is the number of projects in each disease group divided by the number of total projects over all disease groups.
R&D competence	is the number of advancing events to the next phase minus the number of suspension events divided by the total number of projects in the firm's new drug development pipeline in a given year.
Main competence	is R&D competence of the firm's main segment. A firm's main segment is the disease group where the firm has the largest number of projects.
Initial main competence	is the average of the first three non-missing values of R&D competence of the firm's main segment.
Matured phase projects	is the drug projects in the pipeline in phase 4 and 5.
Number of projects	is the firm's total number of projects in the pipeline in all phases in a given year.
% of matured projects	is the percentage of matured projects in the firm's pipeline in a given year.
% of projects with partner	is the percentage of the projects in the firm's pipeline that have partners in a given year.
Log(1+firm age)	is the log of one plus the firm's founding years. Founding years for IPO firms are from Jay Ritter's IPO data web site. We thank Jay Ritter for kindly providing us founding year data. Other private firms' founding years are hand-collected.
VC backed	is firm-level dummy variable of venture capital funding that equals one if a firm has ever received funding from venture capital investors.
Log(number of projects)	is the log of one plus the firm's total number of projects in the pipeline in all phases in a given year.
Log(number of phase 0 projects)	is the log of one plus the firm's phase 0 in the pipeline in a given year.
Log(number of competitors)	is the log of the total number of both private and public firms with new drug development in each industry in a given year.
Ind failure rate	is the industry average of firm failure rate in a given year.
Ind % matured projects	is the industry average of firm % matured projects in a given year.
Initial log(number of projects)	is the average of the first three non-missing values of *Log(number of projects)*.

Initial diversification	is the average of the first three non-missing values of *Diversification w/ number of industries*.
Initial log(number of competitors)	is the average of the first three non-missing values of *Log(number of competitors)*.
Project profitability	is the growth in profits (EBITDA) divided by sales in the year after projects with the same indication advance to the final stage of drug development (phase 5: FDA-approved for marketing) using public firms' financial data.
Ind-adj project profitability	is the industry median adjusted project profitability.
FDA Phase Information	
Phase 0	The drug or treatment is in pre-clinical stage, or Investigational New Drug (IND) application is submitted to the FDA.
Phase 1	Researchers test the drug or treatment in a small group of healthy volunteers for the first time to evaluate its safety, determine a safe dosage range, and identify side effects.
Phase 2	The drug or treatment is given to a small group of people who have a certain disease or condition to see if it is effective and to further evaluate its safety.
Phase 3	The drug or treatment is given to large groups of people from several hundreds to 3,000 to confirm its effectiveness, monitor side effects, compare it to commonly used treatments, and collect information that will allow the drug or treatment to be used safely.
Phase 4	Clinical studies are done. New Drug Application (NDA) or Biologic License Application (BLA) is submitted for the FDA review process.
Phase 5	The FDA has approved the drug or treatment for marketing in the United States.

References

Abernathy, William J, and James M Utterback, 1978, Patterns of industrial innovation, *Technology Review* 64, 254–228.

Acharya, Viral V, Ramin P Baghai, and Krishnamurthy V Subramanian, 2014, Wrongful discharge laws and innovation, *Review of Financial Studies* 27, 301–346.

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt, 2005, Competition and innovation: An inverted-U relationship, *The Quarterly Journal of Economics* 120, 701–728.

Ayyagari, Meghana, Asli Demirgüç-Kunt, and Vojislav Maksimovic, 2015, What determines entrepreneurial outcomes in emerging markets? The role of initial conditions, *World Bank Policy Research Working Paper*.

Beck, Thorsten, ASLI Demirgüç-Kunt, and Vojislav Maksimovic, 2005, Financial and legal constraints to growth: Does firm size matter?, *The Journal of Finance* 60, 137–177.

Benner, Mary J, and Michael Tushman, 2002, Process management and technological innovation: A longitudinal study of the photography and paint industries, *Administrative Science Quarterly* 47, 676–707.

Berger, Philip G, and Eli Ofek, 1995, Diversification's effect on firm value, *Journal of Financial Economics* 37, 39–65.

Bernstein, Shai, Xavier Giroud, and Richard R Townsend, 2016, The impact of venture capital monitoring, *The Journal of Finance* 71, 1591–1622.

Berry, Donald A, 1972, A bernoulli two-armed bandit, *The Annals of Mathematical Statistics* pp. 871–897.

Bhattacharya, Utpal, Po-Hsuan Hsu, Xuan Tian, and Yan Xu, 2015, What affects innovation more: Policy or policy uncertainty?, *Working Paper*.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, 2013, Does management matter? Evidence from India, *The Quarterly Journal of Economics* 128, 1–51.

Bloom, Nicholas, and John Van Reenen, 2007, Measuring and explaining management practices across firms and countries, *The Quarterly Journal of Economics* 122, 1351–1408.

Blume-Kohout, Margaret E, and Neeraj Sood, 2013, Market size and innovation: Effects of Medicare Part D on pharmaceutical research and development, *Journal of Public Economics* 97, 327–336.

Campa, Jose Manuel, and Simi Kedia, 2002, Explaining the diversification discount, *The Journal of Finance* 57, 1731–1762.

Chemmanur, Thomas J, Jie He, Shan He, and Debarshi K Nandy, 2015, The exit choices of entrepreneurial firms, in *Working Paper*.

Chemmanur, Thomas J, Shan He, and Debarshi K Nandy, 2009, The going-public decision and the product market, *Review of Financial Studies* p. hhp098.

Chemmanur, Thomas J, Karthik Krishnan, and Debarshi K Nandy, 2011, How does venture capital financing improve efficiency in private firms? A look beneath the surface, *The Review of Financial Studies* 24, 4037–4090.

Comment, Robert, and Gregg A Jarrell, 1995, Corporate focus and stock returns, *Journal of Financial Economics* 37, 67–87.

Cornaggia, Jess, Yifei Mao, Xuan Tian, and Brian Wolfe, 2015, Does banking competition affect innovation?, *Journal of Financial Economics* 115, 189–209.

Demirgüç-Kunt, Asli, and Vojislav Maksimovic, 1998, Law, finance, and firm growth, *The Journal of Finance* 53, 2107–2137.

DiMasi, Joseph A, Henry G Grabowski, and Ronald W Hansen, 2016, Innovation in the pharmaceutical industry: New estimates of R&D costs, *Journal of Health Economics* 47, 20–33.

DiMasi, Joseph A, Ronald W Hansen, and Henry G Grabowski, 2003, The price of innovation: New estimates of drug development costs, *Journal of Health Economics* 22, 151–185.

Drucker, Peter Ferdinand, 2007, *Management Challenges for the 21st Century* (Routledge).

Duggan, Mark G, and Fiona Scott Morton, 2010, The effect of Medicare Part D on pharmaceutical prices and utilization, *The American Economic Review* 100, 590–607.

——— , 2011, The medium-term impact of Medicare Part D on pharmaceutical prices, *The American Economic Review* 101, 387–392.

Farre-Mensa, Joan, Deepak Hegde, and Alexander Ljungqvist, 2016, The bright side of patents, *NBER Working paper*.

Ferreira, Daniel, Gustavo Manso, and André C Silva, 2014, Incentives to innovate and the decision to go public or private, *Review of Financial Studies* 27, 256–300.

Gao, Huasheng, Po-Hsuan Hsu, and Kai Li, forthcoming, Innovation strategy of private firms, *Journal of Financial and Quantitative Analysis*.

Griliches, Zvi, Ariel Pakes, and Bronwyn H Hall, 1988, The value of patents as indicators of inventive activity, *NBER Working paper*.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg, 2001, The NBER patent citation data file: Lessons, insights and methodological tools, Discussion paper, National Bureau of Economic Research.

Hallen, Benjamin L, 2008, The causes and consequences of the initial network positions of new organizations: From whom do entrepreneurs receive investments?, *Administrative Science Quarterly* 53, 685–718.

Hamel, Gary, and Coimbatore K Prahalad, 1990, The core competence of the corporation, *Harvard Business Review* 68, 79–91.

Hay, Michael, David W Thomas, John L Craighead, Celia Economides, and Jesse Rosenthal, 2014, Clinical development success rates for investigational drugs, *Nature biotechnology* 32, 40–51.

Heckman, James, Rodrigo Pinto, and Peter Savelyev, 2013, Understanding the mechanisms through which an influential early childhood program boosted adult outcomes, *The American economic review* 103, 2052–2086.

Heckman, James J, and Rodrigo Pinto, 2015, Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs, *Econometric reviews* 34, 6–31.

Hellmann, Thomas, and Manju Puri, 2000, The interaction between product market and financing strategy: The role of venture capital, *The Review of Financial Studies* 13, 959–984.

Henderson, Rebecca, and Iain Cockburn, 1994, Measuring competence? Exploring firm effects in pharmaceutical research, *Strategic Management Journal* 15, 63–84.

Hsu, Po-Hsuan, Xuan Tian, and Yan Xu, 2014, Financial development and innovation: Cross-country evidence, *Journal of Financial Economics* 112, 116–135.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto, 2011, Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies, *American Political Science Review* 105, 765–789.

King, Robert G, and Ross Levine, 1993, Finance, entrepreneurship and growth, *Journal of Monetary Economics* 32, 513–542.

Klepper, Steven, 1996, Entry, exit, growth, and innovation over the product life cycle, *American Economic Review* pp. 562–583.

Kortum, Samuel, and Josh Lerner, 2000, Assessing the contribution of venture capital to innovation, *RAND Journal of Economics* pp. 674–692.

Lang, Larry HP, and Rene M Stulz, 1994, Tobin's Q, corporate diversification, and firm performance, *Journal of Political Economy* 102, 1248–1280.

Lemmon, Michael L, Michael R Roberts, and Jaime F Zender, 2008, Back to the beginning: Persistence and the cross-section of corporate capital structure, *Journal of Finance* 63, 1575–1608.

Lerner, Josh, 1995, Venture capitalists and the oversight of private firms, *Journal of Finance* 50, 301–318.

Levinthal, Daniel A, and James G March, 1993, The myopia of learning, *Strategic Management Journal* 14, 95–112.

Maksimovic, Vojislav, and Gordon Phillips, 2002, Do conglomerate firms allocate resources inefficiently across industries? Theory and evidence, *Journal of Finance* pp. 721–767.

——— , and Liu Yang, 2013, Private and public merger waves, *Journal of Finance* 68, 2177–2217.

Manso, Gustavo, 2011, Motivating innovation, *Journal of Finance* 66, 1823–1860.

March, James G, 1991, Exploration and exploitation in organizational learning, *Organization Science* 2, 71–87.

Marquis, Christopher, and András Tilcsik, 2013, Imprinting: Toward a multilevel theory, *The Academy of Management Annals* 7, 195–245.

McGrath, Rita Gunther, 2001, Exploratory learning, innovative capacity, and managerial oversight, *Academy of Management Journal* 44, 118–131.

Moser, Petra, 2013, Patents and innovation: Evidence from economic history, *The Journal of Economic Perspectives* 27, 23–44.

Pagano, Marco, Fabio Panetta, and Luigi Zingales, 1998, Why do companies go public? An empirical analysis, *Journal of Finance* 53, 27–64.

Pandey, Sandeep, Deepayan Chakrabarti, and Deepak Agarwal, 2007, Multi-armed bandit problems with dependent arms, in *Proceedings of the 24th International Conference on Machine Learning* pp. 721–728. ACM.

Pástor, L'uboš, Lucian A Taylor, and Pietro Veronesi, 2009, Entrepreneurial learning, the IPO decision, and the post-IPO drop in firm profitability, *Review of Financial Studies* 22, 3005–3046.

Pearl, Judea, 2012, The mediation formula: A guide to the assessment of causal pathways in non-linear models, in *Causality: Statistical Perspectives and Applications*.

Povel, Paul, Giorgo Sertsios, Renata Kosova, and Praveen Kumar, 2016, Boom and gloom, *The Journal of Finance*.

Powell, Walter W, and Kurt W Sandholtz, 2012, Amphibious entrepreneurs and the emergence of organizational forms, *Strategic Entrepreneurship Journal* 6, 94–115.

Puri, Manju, and Rebecca Zarutskie, 2012, On the life cycle dynamics of venture-capital-and non-venture-capital-financed firms, *The Journal of Finance* 67, 2247–2293.

Rajan, Raghuram G, and Luigi Zingales, 1998, Financial dependence and growth, *The American Economic Review* 88, 559–586.

Sidhu, Jatinder S, Harry R Commandeur, and Henk W Volberda, 2007, The multifaceted nature of exploration and exploitation: Value of supply, demand, and spatial search for innovation, *Organization Science* 18, 20–38.

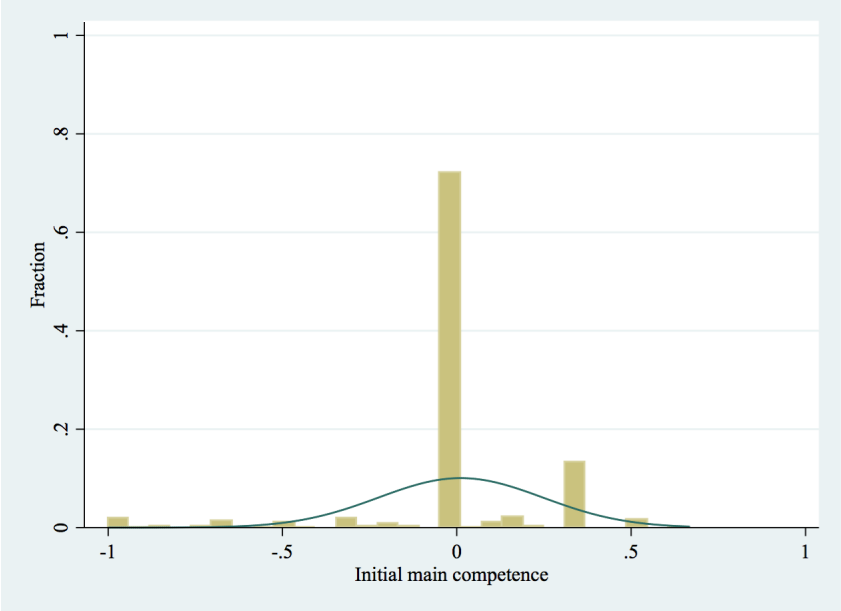
Thakor, Anjan, 2013, Innovation and growth: What do we know?, *World Scientific*.

Ueda, Masako, 2004, Banks versus venture capital: Project evaluation, screening, and expropriation, *The Journal of Finance* 59, 601–621.

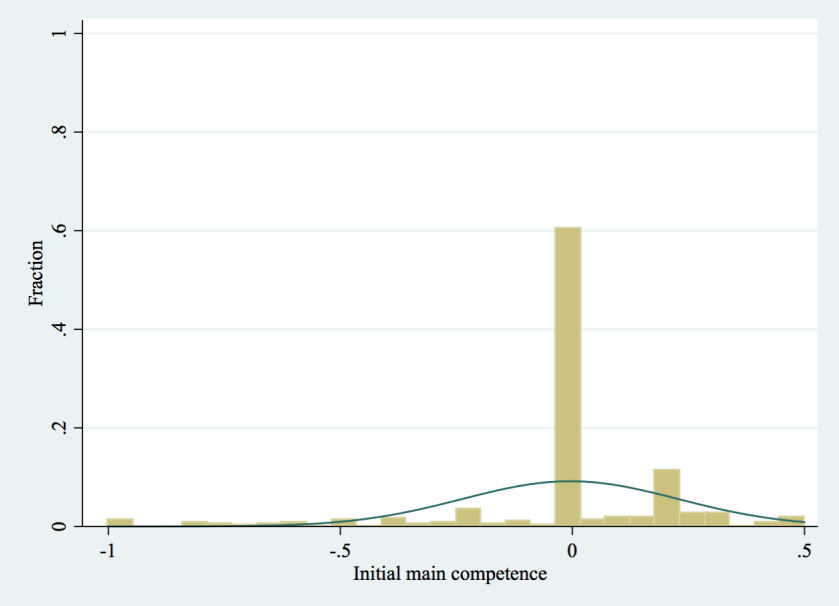
Villalonga, Belen, 2004, Does diversification cause the "diversification discount"?, *Financial Management* pp. 5–27.

Figure 1: Histograms of Initial Main R&D Competence

The figures illustrate the distributions of initial R&D competence of a firm's main segment. We define a firm's main segment based on the disease group where the firm has the largest number of projects. *Initial main competence* is the average of the first three non-missing values of R&D competence for the firm's main segment in figure (a) and the average of the first five non-missing values in figure (b). 222 firms (28%) and 316 firms (40%) out of 799 sample firms have non-zero initial main competence in figure (a) and figure (b), respectively.



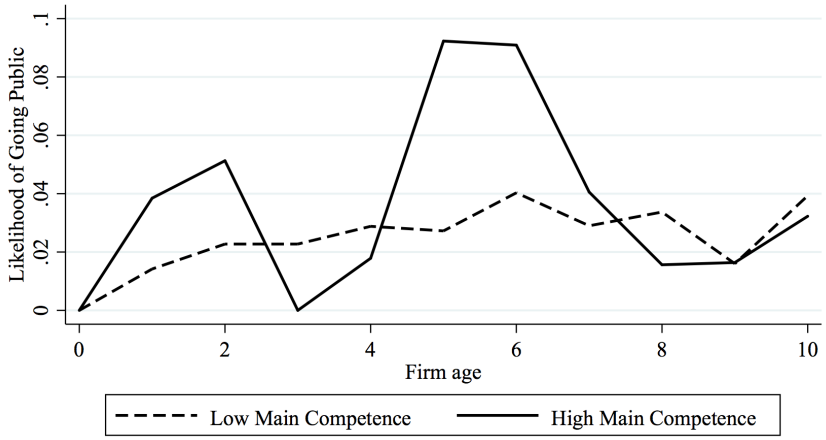
(a) The first 3 year main competence



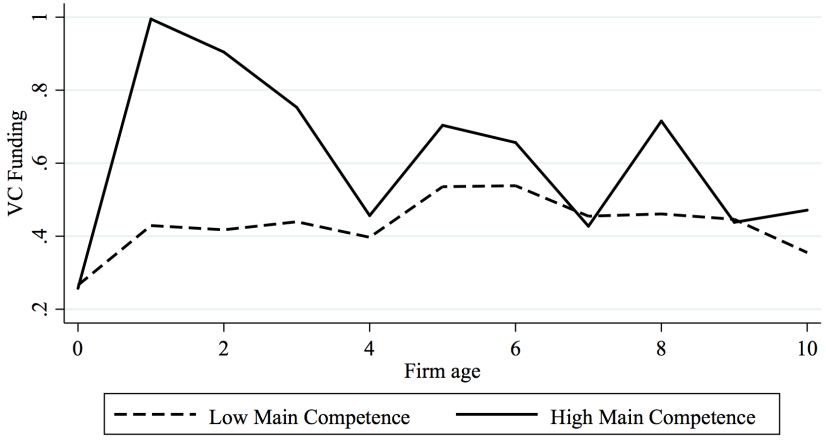
(b) The first 5 year main competence

Figure 2: Time Trends of Competence, Diversification, and Firm Growth

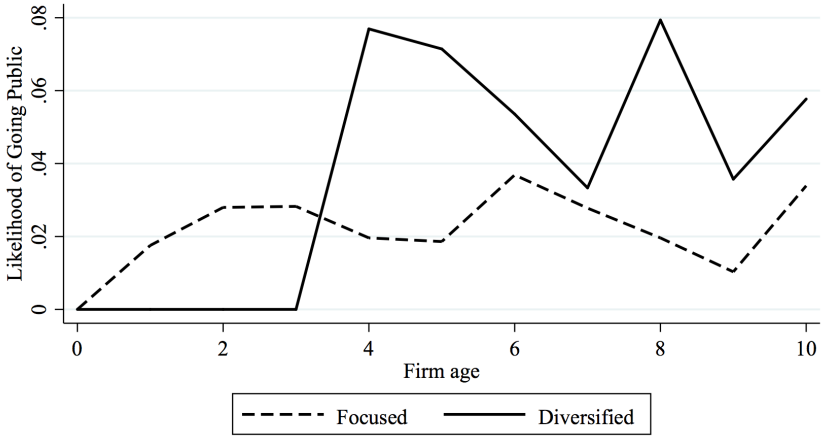
The figures display the time trends of percentage firms of going public and VC funding over firm age. VC funding is the log of one plus the amount of VC funding in million dollars. In Panels (a) and (b), the sample is split by the initial competence of firm's main segment, and in Panels (c) and (d) by diversification. *Initial main competence* is the average of the first three non-missing values of R&D Competence of the firm's main segment. *Initial main competence* is high (solid) if it is positive, and low (dash) otherwise. Panels (c) and (d) are for focused (dash) versus diversified (solid) firms within the subset of firms with low initial main competence. A firm is diversified if the firm develops new drugs over multiple disease groups, and focused otherwise.



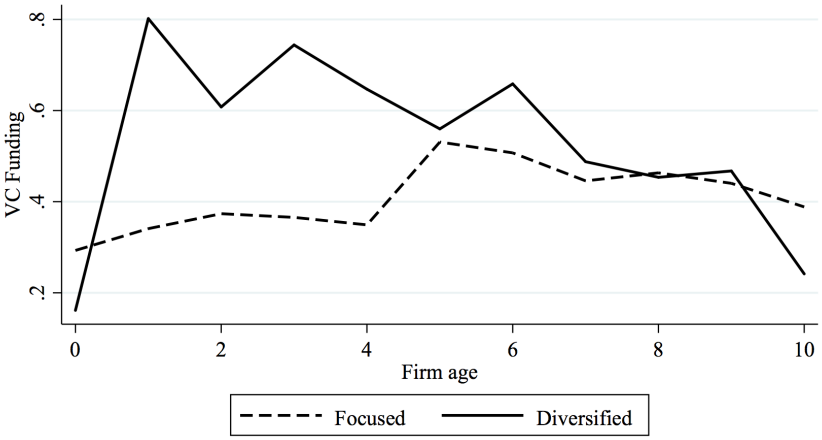
(a) Competence and IPO



(b) Competence and VC Funding



(c) Diversification and IPO



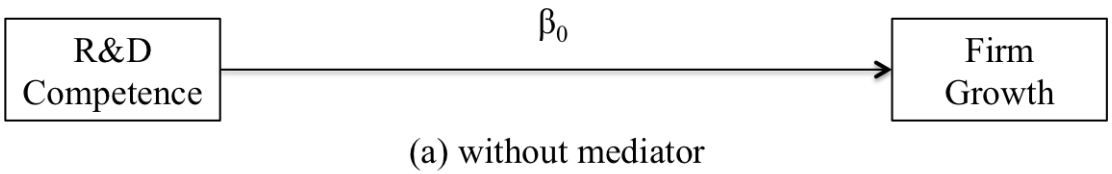
(d) Diversification and VC Funding

Figure 3: Persistency of R&D Competence

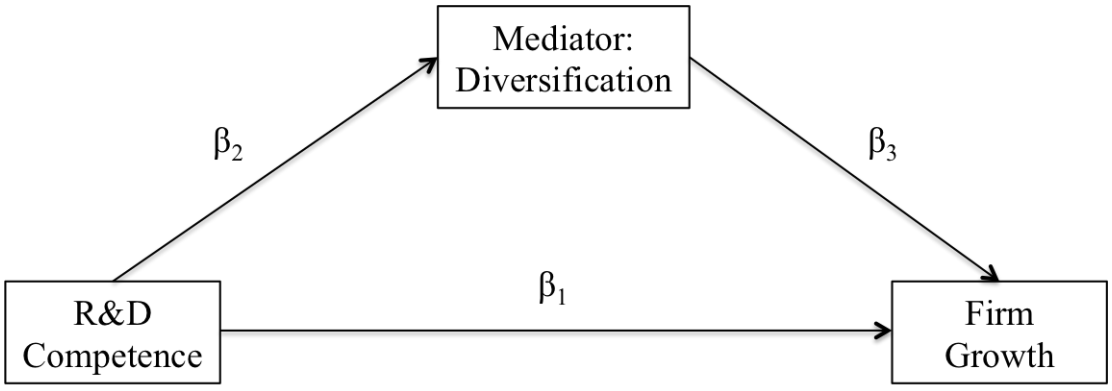
The figure illustrates the persistence of overall R&D competence. *Main competence* is R&D competence of the firm's main segment. *Initial main competence* is the average of the first three non-missing values of R&D competence of the firm's main segment. *Initial main competence* is high (solid) if it is positive, and low (dash) otherwise.

Figure 4: Mediation Model

The figures illustrate our model with one mediator, diversification. Figure (a) displays the direct effect of R&D competence on going public decisions (β_0) without the mediator. Figure (b) displays the effects of R&D competence both directly (β_1) and indirectly through the mediator, diversification ($\beta_2 \times \beta_3$).



(a) without mediator



(b) with mediator

Table 1: Industry Classification

The table presents our industry classification based on disease codes from the BioMedTracker database. Our sample comprises of all private firms in the drug industries that are required to report to the U.S. Food and Drug Administration (FDA). The sample consists of 3,851 firm-year observations of 799 firms during the sample period of 1985-2014.

Disease code	Disease group	Observations	Percent
1	Allergy	9	0.23
2	Autoimmune/Immunology	232	6.02
3	Cardiovascular	315	8.18
4	Dermatology	19	0.49
5	Ear, Nose, Throat/Dental	25	0.65
6	Endocrine	289	7.50
7	Gastroenterology	59	1.53
8	Hematology	113	2.93
9	Infectious	381	9.89
10	Metabolic	120	3.12
11	Neurology	563	14.62
12	Obstetrics/Gynecology	18	0.47
13	Oncology	1,172	30.43
14	Ophthalmology	136	3.53
15	Orthopedics	3	0.08
16	Psychiatry	119	3.09
17	Renal	30	0.78
18	Respiratory	151	3.92
19	Rheumatology	19	0.49
20	Urology	60	1.56
21	Not Specified	18	0.47
Total		3,851	100

Table 2: Summary Statistics

The table presents summary statistics for private firms in our sample from the drug industries that are required to report to the U.S. Food and Drug Administration (FDA). The sample consists of 3,851 firm year observations of 799 firms during the sample period of 1985-2014. All variable definitions are in Appendix.

	Mean	Std. Dev	Min	Median	Max	Obs.
Diversification Decisions and Firm Growth						
Diversification w/ number of industries	1.380	0.874	1.000	1.000	9.000	3851
Diversification w/ project shares	0.123	0.229	0.000	0.000	0.821	3851
Going public (firm)	0.142	0.349	0.000	0.000	1.000	3851
Going public (firm year)	0.032	0.177	0.000	0.000	1.000	3851
VC backed	0.522	0.500	0.000	1.000	1.000	3851
VC fund ($MM)	3.812	15.413	0.000	0.000	471.070	3851
Log(1+VC fund) or VC funding	0.447	1.076	0.000	0.000	6.157	3851
Firm Characteristics						
R&D competence	-0.064	0.381	-1.000	0.000	1.000	3851
Main competence	-0.063	0.390	-1.000	0.000	1.000	3851
Initial main competence	0.011	0.235	-1.000	0.000	0.667	3851
Firm age	7.318	4.447	0.000	7.000	36.000	3851
Log(1+firm age)	1.946	0.650	0.000	2.079	3.611	3851
Number of projects	2.218	2.144	1.000	1.000	23.000	3851
Number of phase 0 projects	0.234	0.684	0.000	0.000	7.000	3851
Number of phase 1 projects	0.483	0.837	0.000	0.000	10.000	3851
Number of phase 2 projects	0.781	1.139	0.000	0.000	13.000	3851
Number of phase 3 projects	0.145	0.414	0.000	0.000	4.000	3851
Number of phase 4 projects	0.018	0.155	0.000	0.000	3.000	3851
Number of phase 5 projects	0.239	1.047	0.000	0.000	12.000	3851
% matured projects	0.087	0.270	0.000	0.000	1.000	3851
% projects with partner	0.396	0.466	0.000	0.000	1.000	3851
Project profitability	0.269	1.109	-2.222	0.168	18.029	1961
Ind-adj project profitability	0.106	1.109	-2.370	0.000	17.881	1961
Industry Characteristics						
Number of competitors	47.294	17.253	2.000	47.000	71.000	3851
Log(number of competitors)	5.277	0.856	1.792	5.493	6.366	3851
Ind failure rate	0.217	0.104	0.000	0.241	0.414	3851
Ind % matured projects	0.233	0.180	0.056	0.200	1.000	3851

Table 3: Comparison in Main Segment Competence and Divesification

The table presents summary statistics for the firms that have high vs low initial main competence (Panel A) and that are diversified vs. focused (Panel B) among our private firm sample in the drug industries. In Panel A, the sample consists of 3,851 firm-year observations of 799 firms during the sample period of 1985-2014. In Panel B, the sample consists of 3,091 firm-year observations of 646 firms from the subset of firms with low initial main competence. Firm initial main competence is low when it is nonpositive. A firm is diversified if the firm develops new drugs over multiple disease groups, and focused otherwise. All variable definitions are in Appendix. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: High vs. Low Main Segment Competence

	High Main Competence		Low Main Competence		
	Mean	Median	Mean	Median	Mean Difference
VC backed	0.592	1.000	0.505	1.000	0.087***
VC funding	0.560	0.000	0.420	0.000	0.140***
Going public (firm)	0.207	0.000	0.126	0.000	0.080***
Going public (firm year)	0.049	0.000	0.028	0.000	0.021***
R&D competence	0.111	0.000	-0.107	0.000	0.217***
Main competence	0.121	0.000	-0.108	0.000	0.229***
Initial main competence	0.306	0.333	-0.061	0.000	0.368***
Diversification	0.128	0.000	0.122	0.000	0.006
Log(1+firm age)	1.994	2.079	1.934	2.079	0.059**
Number of projects	2.462	2.000	2.158	1.000	0.304***
% matured projects	0.064	0.000	0.093	0.000	-0.029***
% projects with partner	0.327	0.000	0.414	0.000	-0.087***
Project profitability	0.290	0.176	0.261	0.165	0.029
Ind-adj project profitability	0.132	0.010	0.097	-0.001	0.035
Log(number of competitors)	5.317	5.501	5.267	5.493	0.050
Ind failure rate	0.220	0.236	0.217	0.242	0.004
Ind % matured projects	0.214	0.200	0.238	0.202	-0.024***
Observations	760		3091		

Panel B: Diversified vs. Focused

	Diversified		Focused		
	Mean	Median	Mean	Median	Mean Difference
VC backed	0.568	1.000	0.486	0.000	0.082***
VC funding	0.471	0.000	0.405	0.000	0.066
Going public (firm)	0.160	0.000	0.116	0.000	0.044***
Going public (firm year)	0.043	0.000	0.024	0.000	0.020***
R&D competence	-0.122	0.000	-0.102	0.000	-0.020
Main competence	-0.130	0.000	-0.102	0.000	-0.028*
Initial main competence	-0.069	0.000	-0.059	0.000	-0.009
Log(1+firm age)	2.053	2.197	1.899	2.079	0.154***
Number of projects	4.143	3.000	1.563	1.000	2.580***
% matured projects	0.152	0.000	0.075	0.000	0.078***
% projects with partner	0.422	0.333	0.411	0.000	0.010
Project profitability	0.197	0.165	0.289	0.160	-0.092
Ind-adj project profitability	0.039	-0.001	0.122	0.000	-0.083
Log(number of competitors)	5.072	5.226	5.326	5.509	-0.254***
Ind failure rate	0.230	0.254	0.212	0.238	0.018***
Ind % martured projects	0.250	0.196	0.234	0.208	0.016**
Observations	713		2378		

Table 4: Persistency of R&D Competence

The table examines the persistence of R&D competence in firm's main disease group. *Main competence* is the R&D competence of the firm's main segment. Initial main competence is the average of the first three non-missing values of R&D competence for the firm's main segment. The table presents coefficient estimates and both R-squares and adjusted R-squares for multiple different model specifications. *t-statistics* (in parenthesis) are robust and adjusted for firm clustering. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Main competence						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Initial main competence			0.762***	0.762***	0.761***	0.761***	0.738***
			(19.19)	(19.22)	(19.18)	(19.27)	(19.10)
Initial log(number of projects)				0.014			
				(0.71)			
Initial diversification					0.012		
					(1.28)		
Initial log(number of competitors)						0.020***	
						(2.91)	
Log(number of phase 0 projects)							0.059***
							(4.31)
Log(1+firm age)							-0.173***
							(-9.23)
% matured projects							0.147***
							(6.25)
% projects with partner							-0.059***
							(-4.02)
Log(number of competitors)							-0.024**
							(-2.55)
Ind failure rate							-0.628***
							(-3.89)
Ind % matured projects							-0.054
							(-0.88)
Observations	3071	3071	3071	3071	3071	3071	3071
R^2	0.457	0.522	0.161	0.161	0.161	0.170	0.206
Adjusted R^2	0.300	0.375	0.152	0.152	0.153	0.162	0.196
Fixed Effects	Firm	Firm, Year	Year	Year	Year	Year	Year

Table 5: Variance Decompositions of R&D Competence

The table presents variance decompositions for the model specifications in Table 4. It reports the fraction of the total variances for each variable in each specification, which is the partial sum of squares for the effect of the variable scaled by the aggregate partial sum of squares across all effects in the specification. The reported fractions in each column sum to one and assess the extent to which the total variation in R&D competence is attributable to the variations of particular variables in the model.

	Main competence						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Firm FE	1.000	0.877					
Year FE		0.123	0.087	0.087	0.086	0.082	0.068
Initial main competence			0.913	0.913	0.911	0.864	0.713
Initial log(number of projects)				0.001			
Initial diversification					0.002		
Initial log(number of competitors)						0.054	
Log(number of phase 0 projects)							0.018
Log(1+firm age)							0.107
% matured projects							0.042
% projects with partner							0.020
Log(number of competitors)							0.015
Ind failure rate							0.017
Ind % matured projects							0.001
Observations	3071	3071	3071	3071	3071	3071	3071
Adjusted R^2	0.300	0.375	0.152	0.152	0.153	0.162	0.196

Table 6: Within-firm Project Performance

The table reports performance correlation between within-firm projects. The dependent variable, *Project performance* equals one if a firm's project advances to the next phase, minus one if the project is suspended, and zero otherwise for a given year. *Peer project performance* is the average of all other projects' performance for the firm in the given year. *Initial main competence* is the average of the first three non-missing values of R&D competence of the firm's main segment. Initial main competence is high if it is positive, and low, otherwise. We define a firm's main segment based on the disease group where the firm has the largest number of projects. Industry and year fixed effects are included. Standard errors (in parenthesis) are robust and adjusted for clustering within firm. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Project performance		
	(1)	(2)	(3)
Peer project performance	0.139***	0.142***	0.116***
	(0.02)	(0.02)	(0.02)
Peer project performance * Initial main competence		0.233***	
		(0.04)	
Initial main competence		0.096***	
		(0.02)	
Peer performance * High initial main competence			0.071**
			(0.03)
High initial main competence			0.044***
			(0.01)
Log(number of phase 0 projects)	-0.013**	-0.012**	-0.010**
	(0.01)	(0.01)	(0.01)
Log(1+firm age)	-0.011**	-0.006	-0.010**
	(0.01)	(0.01)	(0.01)
% matured projects	0.029	0.031*	0.036*
	(0.02)	(0.02)	(0.02)
% projects with partner	-0.002	0.001	0.002
	(0.01)	(0.01)	(0.01)
Log(number of competitors)	-0.098**	-0.085**	-0.096***
	(0.04)	(0.03)	(0.04)
Ind failure rate	-0.039	0.009	-0.009
	(0.13)	(0.13)	(0.13)
Ind % matured projects	0.013	0.009	-0.009
	(0.08)	(0.07)	(0.08)
VC backed	0.006	0.004	0.005
	(0.01)	(0.01)	(0.01)
Observations	8636	8583	8583
Adjusted R^2	0.032	0.038	0.036
Fixed Effects	Ind, Year	Ind, Year	Ind, Year

Table 7: Entrepreneurial Characteristics

The table presents summary statistics for available entrepreneurial characteristics of the firms in our sample (Panel A) and their correlations with *initial main competence* (Panel B). One observation is one firm. The sample consists of 799 entrepreneurial firms in the drug industry sector during the sample period of 1985-2014.

Panel A: Summary Statistics

	Mean	Std. Dev	Min	Median	Max
Doctoral degree	0.781	0.414	0.000	1.000	1.000
Related major (biology, chemistry, or medicine)	0.802	0.399	0.000	1.000	1.000
Inventor	0.335	0.473	0.000	0.000	1.000
Female	0.077	0.267	0.000	0.000	1.000
Previous entrepreneurial experience	0.516	0.500	0.000	1.000	1.000
Previous executive experience	0.684	0.465	0.000	1.000	1.000
Based on a laboratory at a university or a hospital	0.376	0.485	0.000	0.000	1.000
Co-founder	0.318	0.466	0.000	0.000	1.000

Panel B: Pearson Correlation Coefficients

	Initial main competence
Doctoral degree	0.042
Related major (biology, chemistry, or medicine)	0.024
Inventor	0.047
Female	-0.080
Previous entrepreneurial experience	-0.022
Previous executive experience	0.028
Based on a laboratory at a university or a hospital	0.032
Co-founder	-0.076

Table 8: Initial R&D Competence and Diversification

The table examines the relation between diversification and initial main competence as well as time-varying main competence. The sample comprises of all firm-years in our sample period, 1985-2014. In the first two columns, the dependent variable, *Diversification with number of industries* is the total number of all unique disease groups where the firm has projects. In the last two columns, the dependent variable, *Diversification with project shares* is the firm-year level diversification measure calculated by one minus the sum of squared project share of each disease group at the given year. The project share is the number of projects in each disease group divided by the number of total projects over all disease groups. *Main competence* is the R&D competence of the firm's major segment. *Initial main competence* is the average of the first three non-missing values of R&D competence for the firm's main segment. We define a firm's main segment based on the disease group where the firm has the largest number of projects. Industry and year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for industry and year clustering. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification with number of industries		Diversification with project shares	
	(1)	(2)	(3)	(4)
Initial Main competence	-0.206***		-0.0399**	
	(-3.28)		(-2.40)	
Main competence		-0.190***		-0.0347***
		(-5.27)		(-3.98)
Log(number of phase 0 projects)	0.430***	0.439***	0.116***	0.118***
	(6.64)	(6.72)	(9.24)	(9.34)
Log(1+firm age)	0.125***	0.108***	0.0361***	0.0329***
	(6.21)	(5.83)	(6.34)	(6.06)
% matured projects	0.432***	0.447***	0.0986***	0.101***
	(7.90)	(8.15)	(6.63)	(6.79)
% projects with partner	0.108***	0.101***	0.0256***	0.0245***
	(4.37)	(4.13)	(3.56)	(3.48)
Log(number of competitors)	0.604***	0.583***	0.223***	0.219***
	(3.74)	(3.62)	(4.83)	(4.78)
Ind failure rate	0.463	0.379	0.174	0.159
	(1.10)	(0.91)	(1.27)	(1.17)
Ind % matured projects	0.599**	0.561**	0.142*	0.136*
	(2.22)	(2.08)	(1.75)	(1.68)
VC backed	0.0374*	0.0346*	0.0205***	0.0200***
	(1.81)	(1.66)	(3.67)	(3.54)
Observations	3851	3851	3851	3851
Adjusted R^2	0.146	0.150	0.144	0.146
Fixed Effects	Ind, Year	Ind, Year	Ind, Year	Ind, Year

Table 9: Diversification Pattern

The table presents summary statistics for destination industry of diversifying firms in our sample during the sample period of 1985-2014. The sample consists of 220 firm-year observations of 165 firms' diversification patterns. We define the following indicator variables that equal one if the industry average is above the median value for all industries. Variables include industry R&D competence (High competence), project suspension (High suspend rate) and advance rates (High advance rate), number of products (More products), percentage of matured projects (More matured product), percentage of Phase 0 product (More phase 0 product), competition (High Competition), number of industry competitors (More incumbents), and project profitability (High profitability). The table presents the percentages of the firms that enter into the specific destination among all diversified firms, *p-value* and *t-statistics*.

	Percentage	p-value	t-statistics
High competence	49.5%	0.55	-0.13
High suspend rate	96.8%	0.00	39.47
High advance rate	47.3%	0.79	-0.81
More products	81.4%	0.00	11.92
More matured products	1.4%	0.00	-62.06
More phase 0 products	93.2%	0.00	25.35
High competition	99.5%	0.00	109.00
More incumbents	99.5%	0.00	109.00
High profitability	48.9%	0.61	-0.29
Observations		220	

Table 10: Initial Competence, Diversification and Firm Growth

The table reports the effects of initial main competence on (1) Going public and (2) VC funding directly and indirectly through diversification using a simultaneous mediation model. In the first two columns, the dependent variable, *Going Public* is one if the firm went public via an IPO in a given year, and zero otherwise. In the last two columns, the dependent variable, *VC funding* is the log of one plus the VC funding amount in a given year. The mediator variable, *Diversification* is the total number of different disease groups where the firm has projects. *Initial main competence* is the average of the first three non-missing values of R&D competence for the firm's main segment. We define a firm's main segment based on the disease group where the firm has the largest number of projects. Industry and year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for clustering within industry and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification	Going public	Diversification	VC funding
	(1)	(2)	(3)	(4)
Diversification		0.0115**		0.0470**
		(2.30)		(2.19)
Initial main competence	-0.206***	0.0328***	-0.202***	0.0903
	(-3.30)	(2.69)	(-3.25)	(1.29)
Log(number of phase 0 projects)	0.430***	0.00488	0.429***	0.0338
	(6.69)	(0.31)	(6.64)	(0.54)
Log(1+firm age)	0.125***	0.000319	0.129***	-0.0210
	(6.26)	(0.10)	(6.38)	(-0.80)
% matured projects	0.432***	-0.0119	0.427***	-0.397***
	(7.96)	(-0.88)	(7.92)	(-6.11)
% projects with partner	0.108***	0.0131**	0.106***	-0.00947
	(4.41)	(2.50)	(4.36)	(-0.23)
Log(number of competitors)	0.604***	0.0208	0.605***	0.0757
	(3.77)	(0.49)	(3.77)	(0.36)
Ind failure rate	0.463	-0.387***	0.463	-0.450
	(1.11)	(-2.99)	(1.10)	(-0.59)
Ind % matured projects	0.599**	0.176*	0.590**	-0.294
	(2.24)	(1.84)	(2.19)	(-0.60)
VC backed	0.0374*	0.0475***		
	(1.82)	(5.16)		
Observations	3851		3851	
Berndt R^2	0.206		0.186	
Fixed Effects	Ind, Year		Ind, Year	

Table 11: Medicare Part D Shock on Diversification and Firm Growth

The table reports the effects of diversification on (1) Going public and (2) VC funding using an instrumental variable approach with the passage of Medicare Part D legislation. *Treated* is one if a firm does not have any projects in the required Part D drug classes before the legislation. *Post* is one after the passage of the Medicare Part D legislation in 2003. *Low competence* is an indicator variable for a firm with nonpositive initial main competence. In the first two columns, the dependent variable, *Going Public* is one if the firm went public via an IPO in a given year, and zero otherwise. In the last two columns, the dependent variable, *VC funding* is the log of one plus the VC funding amount in a given year. The mediator variable, *Diversification* is the total number of different disease groups where the firm has projects. Industry and year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for clustering within industry and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification	Going public	Diversification	VC funding
	(1)	(2)	(3)	(4)
Diversification		0.0108**		0.0439**
		(2.17)		(2.05)
Treated * Post * Low competence	1.132***		1.125***	
	(2.95)		(2.94)	
Treated * Post	-0.401		-0.390	
	(-1.48)		(-1.45)	
Treated * Low competence	-1.171***		-1.170***	
	(-3.11)		(-3.12)	
Post * Low competence	-0.388		-0.389	
	(-1.26)		(-1.27)	
Treated	0.784**		0.775**	
	(2.55)		(2.52)	
Low competence	0.347	-0.0192**	0.348	-0.130***
	(1.14)	(-2.40)	(1.15)	(-2.60)
Log(number of phase 0 projects)	0.434***	0.00613	0.434***	0.0411
	(6.63)	(0.39)	(6.59)	(0.66)
Log(1+firm age)	0.110***	0.00224	0.114***	-0.0166
	(5.90)	(0.71)	(5.99)	(-0.63)
% matured projects	0.423***	-0.00966	0.419***	-0.388***
	(7.06)	(-0.71)	(7.03)	(-6.00)
% projects with partner	0.130***	0.0124**	0.128***	-0.00589
	(5.11)	(2.42)	(5.08)	(-0.14)
Log(number of competitors)	0.663***	0.0186	0.661***	0.0770
	(3.49)	(0.44)	(3.48)	(0.37)
Ind failure rate	0.0700	-0.391***	0.0651	-0.456
	(0.17)	(-3.03)	(0.16)	(-0.60)
Ind % matured projects	1.010***	0.170*	1.003***	-0.309
	(3.34)	(1.78)	(3.31)	(-0.62)
VC backed	0.0318	0.0474***		
	(1.45)	(5.10)		
Observations	3851		3851	
Berndt R^2	0.213		0.195	
Fixed Effects	Ind, Year		Ind, Year	

Table A.1: Persistency of R&D Competence: Robustness

The table examines the robustness of the R&D competence persistence in firm's main disease group. The table displays the results for the last column in Table 4 using alternative measures of R&D competence or different model specifications. In column (1), we drop all suspended projects when counting the total number of projects in the subsequent years after they are initially identified as suspended. In column (2), we use an alternative measure of R&D competence for which we keep one lead indication for each drug when the drug is tested for multiple indications. In column (3), the competence measure is a conditional success probability weighted R&D competence that reflects phase-specific difficulties of transition based on success rates estimated in DiMasi, Hansen, and Grabowski (2003). In column (4), we include entry year fixed effects instead of year fixed effects. The table presents coefficient estimates and both R-squares and adjusted R-squares for multiple different model specifications. *t-statistics* (in parenthesis) are robust and adjusted for firm clustering. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Main competence			
	(1)	(2)	(3)	(4)
Initial main competence	0.660***	0.757***	0.580***	0.727***
	(15.49)	(0.04)	(0.07)	(18.93)
Log(number of phase 0 projects)	0.042***	0.059***	-0.007	0.034**
	(3.16)	(0.02)	(0.00)	(2.38)
Log(1+firm age)	-0.136***	-0.181***	-0.013*	-0.044**
	(-7.67)	(0.02)	(0.01)	(-2.10)
% matured projects	0.106***	0.157***	0.011*	0.209***
	(4.63)	(0.02)	(0.01)	(6.63)
% projects with partner	-0.040***	-0.060***	-0.007	-0.004
	(-2.91)	(0.02)	(0.00)	(-0.29)
Log(number of competitors)	-0.025***	-0.029***	-0.006	-0.012
	(-2.74)	(0.01)	(0.00)	(-1.27)
Ind failure rate	-0.416***	-0.642***	-0.082	-1.164***
	(-2.76)	(0.17)	(0.07)	(-11.82)
Ind % matured projects	0.036	-0.053	0.008	-0.157***
	(0.61)	(0.06)	(0.03)	(-3.03)
Observations	2971	3069	2916	3071
R-squared	0.184	0.207	0.080	0.269
Adjusted R-squared	0.174	0.197	0.068	0.263
Fixed Effects	Year	Year	Year	Entry-Year

Table A.2: Initial Competence, Diversification and Firm Growth: Cross-sectional Analysis

The table reports the effects of initial main competence on (1) Going public and (2) VC funding directly and indirectly through diversification using a simultaneous mediation model. We use a refined sample that consists of only the final year observations for each firm during our sample period of 1985-2014. For a firm that eventually goes public, the IPO event takes place in the final year. For other firms, we select observations in the last available year. In the first two columns, the dependent variable, *Going Public* is one if the firm went public via an IPO during the period, and zero otherwise. In the last two columns, the dependent variable, *VC funding* is the log of one plus the total VC funding amount during the period. The mediator variable, *Diversification* is the total number of different disease groups where the firm has projects at the final observation. *Initial main competence* is the average of the first three non-missing values of R&D competence for the firm's main segment. We define a firm's main segment based on the disease group where the firm has the largest number of projects. Industry fixed effects are included. Standard errors (in parenthesis) are robust and adjusted for clustering within industry and entry-year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification	Going public	Diversification	VC funding
	(1)	(2)	(3)	(4)
Diversification		0.021*		0.251***
		(0.01)		(0.07)
Initial main competence	-0.319*	0.158***	-0.307*	0.309
	(0.18)	(0.04)	(0.18)	(0.25)
Log(number of phase 0 projects)	0.601***	0.005	0.599***	-0.244*
	(0.10)	(0.02)	(0.11)	(0.15)
Log(1+firm age)	0.227***	-0.013	0.237***	0.300**
	(0.06)	(0.02)	(0.06)	(0.12)
% matured projects	0.653***	-0.051	0.636***	-1.005***
	(0.18)	(0.06)	(0.17)	(0.28)
% projects with partner	0.074	0.058**	0.070	0.046
	(0.06)	(0.02)	(0.06)	(0.13)
Log(number of competitors)	-0.554	-1.567***	-0.585	-4.352**
	(0.85)	(0.45)	(0.85)	(1.92)
Ind failure rate	0.995	1.667	0.937	5.071
	(2.54)	(1.41)	(2.54)	(5.84)
Ind % matured projects	-0.575	-3.022***	-0.621	-10.110***
	(1.45)	(0.91)	(1.46)	(3.64)
VC backed	0.074	0.166***		
	(0.07)	(0.02)		
Observations	796		796	
Fixed Effects	Ind, Entry Year		Ind, Entry Year	

Table A.3: Exits through an IPO or an Acquisition

The table reports the effects of initial main competence on the two exit choices including IPOs and acquisitions directly and indirectly through diversification using a simultaneous mediation model. The dependent variable, *Exit* is one if the firm either goes public via an IPO or is acquired in a given year, and zero otherwise. The first two columns report analogous regression results as in the first two columns of Table 10 that replaces the dependent variable *Going public* with *Exit*. The last two columns report analogous regression results as in the first two columns of Table 11 that also replaces the dependent variable *Going public* with *Exit*. The mediator variable, *Diversification* is the total number of different disease groups where the firm has projects. Industry and year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for clustering within industry and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification	Exit	Diversification	Exit
	(1)	(2)	(3)	(4)
Diversification		0.012**		0.012**
		(0.01)		(0.01)
Initial main competence	-0.218***	0.020		
	(0.06)	(0.02)		
Treated * Post * Low competence			1.535***	
			(0.35)	
Treated * Post			-0.407	
			(0.29)	
Treated * Low competence			-1.615***	
			(0.34)	
Post * Low competence			-0.345	
			(0.31)	
Treated			0.852***	
			(0.32)	
Low competence			0.310	-0.017**
			(0.31)	(0.01)
Log(number of phase 0 projects)	0.417***	0.006	0.423***	0.007
	(0.07)	(0.01)	(0.07)	(0.01)
Log(1+firm age)	0.129***	-0.011**	0.122***	-0.010**
	(0.02)	(0.00)	(0.02)	(0.00)
% matured projects	0.407***	-0.008	0.403***	-0.007
	(0.05)	(0.02)	(0.06)	(0.02)
% projects with partner	0.122***	0.014**	0.143***	0.014**
	(0.02)	(0.01)	(0.03)	(0.01)
Log(number of competitors)	0.568***	0.035	0.583***	0.034
	(0.12)	(0.05)	(0.15)	(0.05)
Ind failure rate	0.645	-0.509***	0.043	-0.512***
	(0.44)	(0.19)	(0.39)	(0.19)
Ind % matured projects	0.331	0.217*	0.888***	0.213*
	(0.26)	(0.12)	(0.28)	(0.12)
VC backed	0.019	0.054***	0.009	0.054***
	(0.02)	(0.01)	(0.02)	(0.01)
Observations	3708		3708	
Fixed Effects	Ind, Year		Ind, Year	

Table A.4: The Timing of Medicare Part D and Diversification

The table reports the effects of the passage of Medicare Part D legislation in 2003 on diversification. *Treated* is one if a firm does not have any projects in the required Part D drug classes before the legislation. *Post* is one after the passage of the Medicare Part D legislation in 2003. *Low competence* is an indicator variable for a firm with nonpositive initial main competence. *Pre-Medicare* and *Post-Medicare* are indicator variables for the two years before and after the passage of the Medicare Part D legislation, respectively. *Placebo Post* is one after the placebo event year, which is 2010. Results are robust to different placebo years. Industry and year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for clustering within industry and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification			
	(1)	(2)	(3)	(4)
Treated * Post * Low competence	1.132***			
	(2.92)			
Treated * Pre-Medicare (-2,-1) * Low competence		-0.459		
		(-0.92)		
Treated * Post-Medicare (0, +2) * Low competence		0.344**		
		(1.99)		
Treated * Year 2000 * Low competence			-0.438	
			(-1.01)	
Treated * Year 2002 * Low competence			-0.010	
			(-0.02)	
Treated * Year 2003 * Low competence			0.640*	
			(1.67)	
Treated * Year 2004 * Low competence			0.189	
			(0.65)	
Treated * Year 2005 * Low competence			0.314	
			(1.62)	
Treated * Placebo Post * Low competence				0.145
				(1.00)
Observations	3851	3851	3851	3851
Adjusted R^2	0.152	0.149	0.148	0.150
Fixed Effects	Ind, Year	Ind, Year	Ind, Year	Ind, Year

Table A.5: Alternative Competence Measure: Dropping Suspended Projects

The table reports the effects of diversification on (1) Going public and (2) VC funding using an instrumental variable approach with the passage of Medicare Part D legislation as in Table 11 using an alternative measure of R&D competence. With this R&D competence measure, we drop all suspended projects when counting the total number of projects in the subsequent years after they are initially identified as suspended. *Treated* is one if a firm does not have any projects in the required Part D drug classes before the legislation. *Post* is one after the passage of the Medicare Part D legislation in 2003. *Low competence* is an indicator variable for a firm with nonpositive initial main competence. In the first two columns, the dependent variable, *Going Public* is one if the firm went public via an IPO in a given year, and zero otherwise. In the last two columns, the dependent variable, *VC funding* is the log of one plus the VC funding amount in a given year. The mediator variable, *Diversification* is the total number of different disease groups where the firm has projects. Industry and year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for clustering within industry and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification	Going public	Diversification	VC funding
	(1)	(2)	(3)	(4)
Diversification		0.010**		0.041*
		(0.01)		(0.02)
Treated * Post * Low competence	1.211***		1.206***	
	(0.39)		(0.39)	
Treated * Post	-0.453*		-0.445	
	(0.27)		(0.27)	
Treated * Low competence	-1.247***		-1.247***	
	(0.38)		(0.38)	
Post * Low competence	-0.405		-0.407	
	(0.31)		(0.31)	
Treated	0.827***		0.819***	
	(0.31)		(0.31)	
Low competence	0.360	-0.017**	0.362	-0.117**
	(0.31)	(0.01)	(0.30)	(0.05)
Log(number of phase 0 projects)	0.439***	0.002	0.438***	0.022
	(0.06)	(0.02)	(0.07)	(0.06)
Log(1+firm age)	0.131***	0.003	0.134***	0.005
	(0.02)	(0.00)	(0.02)	(0.03)
% matured projects	0.459***	-0.010	0.455***	-0.397***
	(0.06)	(0.01)	(0.06)	(0.07)
% projects with partner	0.189***	0.007	0.188***	0.007
	(0.03)	(0.00)	(0.03)	(0.03)
Log(number of competitors)	0.739***	0.021	0.737***	-0.015
	(0.19)	(0.04)	(0.19)	(0.21)
Ind failure rate	0.151	-0.395***	0.146	-0.542
	(0.43)	(0.13)	(0.43)	(0.78)
Ind % matured projects	1.102***	0.178*	1.094***	-0.468
	(0.31)	(0.10)	(0.31)	(0.49)
VC backed	0.029	0.049***		
	(0.02)	(0.01)		
Observations	3658		3658	
Fixed Effects	Ind, Year		Ind, Year	

Table A.6: Alternative Competence Measure: Phase-specific Difficulties

The table reports the effects of diversification on (1) Going public and (2) VC funding using an instrumental variable approach with the passage of Medicare Part D legislation as in Table 11 using an alternative measure of R&D competence. This alternative is a conditional success probability weighted R&D competence that reflects phase-specific difficulties of transition based on success rates estimated in DiMasi, Hansen, and Grabowski (2003). *Treated* is one if a firm does not have any projects in the required Part D drug classes before the legislation. *Post* is one after the passage of the Medicare Part D legislation in 2003. *Low competence* is an indicator variable for a firm with nonpositive initial main competence. In the first two columns, the dependent variable, *Going Public* is one if the firm went public via an IPO in a given year, and zero otherwise. In the last two columns, the dependent variable, *VC funding* is the log of one plus the VC funding amount in a given year. The mediator variable, *Diversification* is the total number of different disease groups where the firm has projects. Industry and year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for clustering within industry and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification	Going public	Diversification	VC funding
	(1)	(2)	(3)	(4)
Diversification		0.010**		0.040*
		(0.01)		(0.02)
Treated * Post * Low competence	1.231***		1.226***	
	(0.39)		(0.39)	
Treated * Post	-0.469*		-0.461*	
	(0.27)		(0.27)	
Treated * Low competence	-1.249***		-1.249***	
	(0.38)		(0.38)	
Post * Low competence	-0.423		-0.425	
	(0.31)		(0.31)	
Treated	0.828***		0.821***	
	(0.31)		(0.31)	
Low competence	0.361	-0.020**	0.363	-0.141***
	(0.31)	(0.01)	(0.30)	(0.05)
Log(number of phase 0 projects)	0.439***	0.002	0.439***	0.023
	(0.06)	(0.02)	(0.07)	(0.06)
Log(1+firm age)	0.131***	0.003	0.134***	0.006
	(0.02)	(0.00)	(0.02)	(0.03)
% matured projects	0.460***	-0.009	0.456***	-0.395***
	(0.06)	(0.01)	(0.06)	(0.07)
% projects with partner	0.190***	0.007*	0.189***	0.009
	(0.03)	(0.00)	(0.03)	(0.03)
Log(number of competitors)	0.740***	0.021	0.738***	-0.013
	(0.19)	(0.04)	(0.19)	(0.21)
Ind failure rate	0.146	-0.395***	0.141	-0.542
	(0.43)	(0.13)	(0.43)	(0.78)
Ind % matured projects	1.106***	0.178*	1.099***	-0.470
	(0.31)	(0.10)	(0.31)	(0.49)
VC backed	0.029	0.049***		
	(0.02)	(0.01)		
Observations	3658		3658	
Fixed Effects	Ind, Year		Ind, Year	

Table A.7: Medicare Part D Shock on Diversification and Firm Growth: Entry Year Effects

The table reports the effects of diversification on (1) Going public and (2) VC funding using an instrumental variable approach with the passage of Medicare Part D legislation as in Table 11 with the exception of entry year fixed effects instead of year fixed effects. *Treated* is one if a firm does not have any projects in the required Part D drug classes before the legislation. *Post* is one after the passage of the Medicare Part D legislation in 2003. *Low competence* is an indicator variable for a firm with nonpositive initial main competence. In the first two columns, the dependent variable, *Going Public* is one if the firm went public via an IPO in a given year, and zero otherwise. In the last two columns, the dependent variable, *VC funding* is the log of one plus the VC funding amount in a given year. The mediator variable, *Diversification* is the total number of different disease groups where the firm has projects. Industry and entry year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for clustering within industry and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification	Going public	Diversification	VC funding
	(1)	(2)	(3)	(4)
Diversification		0.012**		0.074***
		(0.01)		(0.02)
Treated * Post * Low competence	0.436**		0.415**	
	(0.20)		(0.20)	
Treated * Post	-0.704***		-0.671***	
	(0.13)		(0.12)	
Treated * Low competence	-0.407**		-0.399**	
	(0.19)		(0.20)	
Post * Low competence	0.238*		0.229	
	(0.14)		(0.14)	
Treated	1.282***		1.254***	
	(0.22)		(0.22)	
Low competence	-0.331**	-0.020***	-0.322**	-0.131***
	(0.14)	(0.01)	(0.14)	(0.05)
Log(number of phase 0 projects)	0.500***	0.010	0.500***	-0.009
	(0.06)	(0.01)	(0.06)	(0.06)
Log(1+firm age)	-0.053***	0.012***	-0.044**	0.053*
	(0.02)	(0.00)	(0.02)	(0.03)
% matured projects	0.274***	0.001	0.262***	-0.354***
	(0.07)	(0.02)	(0.07)	(0.07)
% projects with partner	0.023	0.016***	0.020	0.040
	(0.02)	(0.01)	(0.02)	(0.04)
Log(number of competitors)	0.529***	0.110***	0.518***	-0.048
	(0.08)	(0.03)	(0.08)	(0.15)
Ind failure rate	0.212	-0.127*	0.199	-1.255***
	(0.23)	(0.07)	(0.23)	(0.38)
Ind % matured projects	0.436*	0.317***	0.427*	-0.008
	(0.25)	(0.08)	(0.25)	(0.40)
VC backed	0.076***	0.046***		
	(0.02)	(0.01)		
Observations	3851		3851	
Fixed Effects	Ind, Entry Year		Ind, Entry Year	