This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: The Role of Innovation and Entrepreneurship in Economic Growth

Volume Authors/Editors: Michael J. Andrews, Aaron Chatterji, Josh Lerner, and Scott Stern, editors

Volume Publisher: University of Chicago Press

Volume ISBNs: 978-0-226-81078-2 (cloth), 978-0-226-81064-5 (electronic)

Volume URL: https://www.nber.org/books-and-chapters/role-innovation-andentrepreneurship-economic-growth

Conference Date: January 7-8, 2020

Publication Date: Februrary 2022

Chapter Title: Venture Capital-Led Entrepreneurship in Health Care

Chapter Author(s): Amitabh Chandra, Cirrus Foroughi, Lauren Mostrom

Chapter URL:

https://www.nber.org/books-and-chapters/role-innovation-andentrepreneurship-economic-growth/venture-capital-led-entrep reneurship-health-care

Chapter pages in book: p. 475 – 498

Venture Capital–Led Entrepreneurship in Health Care

Amitabh Chandra, Cirrus Foroughi, and Lauren Mostrom

10.1 Introduction

Even though venture capital (VC) funds are generally raised for a limited period of time (usually 10 years) and only account for about \$450 billion in assets (compared to several trillion in private equity and \$43 trillion in public equity), VCs hold disproportionate influence over financing innovation in all sectors of the economy. Lerner and Nanda (2020) note that among nonfinancial firms that issued an IPO between 1995 and 2018, 47 percent were backed by a VC fund. Of those firms that were still public at the end of 2019, the firms originally backed by VC made up 76.2 percent of the total market capitalization and were responsible for 88.6 percent of total R&D expenditure.

As of 2014, about 37 percent of the healthcare industry's market capitalization was backed by VC, making it the third-most VC-backed industry, behind electronics and software (Gornall and Strebulaev 2015). In 2017, health-care spending in the US made up 17.1 percent of GDP (\$3.3 trillion), far exceeding the shares of other developed nations, and per capita health-

Amitabh Chandra is the Henry and Allison McCance Professor of Business Administration at Harvard Business School, the Ethel Zimmerman Wiener Professor of Public Policy and Director of Health Policy Research at the Harvard Kennedy School of Government, and a research associate of the National Bureau of Economic Research.

Cirrus Foroughi is an associate at Analysis Group.

Lauren Mostrom is a research associate at Harvard Business School.

We thank Heidi Williams for helpful comments, and Scott Stern and the editors for very helpful suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research. For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see https://www.nber.org/books-and-chapters/role-innovation-and-entrepreneurship -economic-growth/venture-capital-led-entrepreneurship-health-care.

care spending has nearly doubled since 1995 (Nunn, Parsons, and Shambaugh 2020). This increase in spending has been partially attributed to Baumol's "cost disease," the phenomenon that service sectors face persistently rising costs and limited opportunities for productivity improvements, and partially to innovations that improve health-care quality but do not reduce costs (Baumol and Bowen 1965; Sheiner and Malinovskaya 2016). Since the future quality and affordability of patient care depends on new medical innovations, the way VC investments shape the treatments, technologies, and delivery systems that come to market are particularly salient in this industry. Put more starkly, if VC investments in a therapeutic area are small relative to the social value of these potential treatments (e.g., a transformational medicine for Alzheimer's or a novel telemedicine platform), then society would forgo the benefits of these discoveries, because commercialization of these ideas depends on the decisions of venture capitalists. These concerns are amplified by the observation that venture capitalists ought to be insulated from public markets given their long time-horizons, but there is some evidence that the quality of research conducted by VC-backed early-stage companies is of lower quality during recessions, which would introduce cyclicality in quality (Howell et al. 2020).

Our model of VC-backed investments in health care is the same as R&D investments outside health care, but with augmented risk and return parameters. Venture capitalists will make R&D investments if the expected net present value (NPV) of a project, discounted at the appropriate cost-ofcapital, is positive. Expected NPV will depend on expectations about future revenues—prices and quantities, the risk of failure, and the cost of R&D. These cash flows will be discounted by a cost-of-capital that depends on the correlation between project returns and overall market returns (beta will be low for projects with high scientific and regulatory risk and high for health-care products that are directly sold to consumers, because consumer's willingness to pay is cyclical with the rest of the market). It is important not to conflate uncertainty with the cost of capital—the latter will be lower for pharmaceutical investments than for other investments, but it should be noted that the pharmaceutical investments may still have a low expected NPV because of scientific and regulatory uncertainty.

In this simple model, venture capitalists will want to prioritize projects that are able to move from an idea to a commercialized product in a short time, and this force will—*ceteris paribus*—discourage investments in early-stage pharmaceuticals and biotechnology firms because of the fundamentally long arc of science and regulatory review, and the additional uncertainty stemming from policy uncertainty given government's large role as a purchaser and regulator of health-care services. These additional sources of uncertainty are not present in product markets that have a direct-to-consumer channel for sales, and venture capitalists will seek to overcome these forces by seeking a larger ownership stake in early-stage companies that are particularly risky.

Another concern with the quality of VC-backed investments in health care is that the short time between discovery and commercialization may privilege a set of ideas that are economically viable but less connected to the burden of disease or to the social value of these discoveries. These potential mismatches may be smaller outside the health-care sector, where the regulatory burdens are lower and direct-to-consumer selling is the dominant sales channel. One strategy that reduces the time pressure and uncertainty of innovations is the use of multiple funding rounds in a startup's life, which benefits investors and the entrepreneurs alike by enabling investors to choose the amount of competition and regulatory risk they are exposed to. Multiple funding rounds are likely to be more important for health-care investments, but for this reason, we might expect funding for VC R&D in health care to go to earlier stage companies relative to non-health-care investments, because venture capitalists will have a particular comparative advantage in making sense of these early-stage investments. Another approach used by some specialized venture capitalists to mitigate these challenges, especially in the biotech and pharmaceuticals industries, is to create new startups inside "VCfoundries," to reduce information asymmetry between the venture capitalist and the entrepreneur.

With these motivations, we describe trends in the life cycle of innovation in the health-care sector, starting with a data set from the capital market company Preqin on VC deals in the health-care sector, to develop a more detailed picture of early-stage innovation in health care. We find that VC funding in the health-care sector has grown more slowly and been directed at earlier stage firms than VC funding in other sectors, which suggests that other sectors offer more economically attractive projects. Among VC investments, 60 percent of all money was invested in firms working on drugs, another 20 percent was invested in firms working on a project related to medical devices, and 20 percent was given to firms working on health-care delivery. We also find enormous geographic concentration of health-care deals, which motivates us to explore the "valley of death" hypothesis (the idea that many useful inventions are not explored, because venture capitalists may not know about them). We explore the relationship between patenting and VC funding at the level of cities and find some support for this hypothesis, but we emphasize the need to evaluate it more carefully.

This fact in turn motivates us to consider another way of looking at earlystage entrepreneurship in health care: publications in medical journals and the relative roles of private and public funding of different areas of research in health care (basic science, devices, pharmaceuticals, delivery). Such an analysis would not be possible in other industries, where publication is not a prerequisite for commercialization, but the science and research–intensive nature of innovation in health care means that we can use publications and sources of funding to measure the direction of pre-investment research in health care.

Using publications to measure the development of ideas reveals some addi-

tional facts. Two Metropolitan Statistical Areas (MSAs)—Boston and San Francisco—account for a disproportionate share of basic science research, translational research, and clinical research—which may be why they receive the plurality of VC investments. Some evidence, albeit directional evidence, suggests that the National Institutes of Health (NIH) reduces market failure by allocating relatively more dollars for basic science research (research that is fundamental biology and chemistry and not linked to a particular drug or disease) than industry allocates. While we cannot answer the question of whether the NIH should do more or less of this, the allocation that we find is a necessary condition for allocative efficiency of public dollars. Pushing in the other direction is our finding that when it comes to translational research (research that is directly linked to a disease), NIH funded research does not look different than privately funded research-for example, the NIH funded research projects are just as likely to study cancer over infectious diseases as are privately funded research projects. This finding raises questions of whether the government should rethink how it allocates money to projects.

10.2 Venture Capital Deals

We obtain data on VC deals in healthcare from Preqin. Though most existing literature focuses on Preqin's performance data, we focus our attention on the investment deals themselves to develop a fuller understanding of which research ideas and developments are determined a priori to be the most commercializable by venture capitalists. The Venture Capital Deals data set from Preqin includes not only investments by VC funds and angel investors, but also grants from foundations and government agencies (namely, the NIH), which we analyze separately. When we refer to volumes of VC investments or deals throughout this chapter, we are referring to the amounts of money invested by venture capitalists in young companies, not the amounts transacted between venture capitalists and their limited partners. To compare these deal volumes across time, all deal values were converted to 2020 US dollars using the Consumer Price Index for All Urban Consumers.

The Preqin database has been assembled based on voluntary reporting from general partners and limited partners of venture funds and public filings from pension funds. This data set has the advantage of transparency, since general partners are able and willing to submit corrections for inaccurate information about their funds. One potential bias, however, is that the database misses certain high performing VCs, such as Sequoia and Accel, due to the way it collects information.¹ Despite these limitations, Preqin data have been used in recent scholarship to conduct various analyses of perfor-

^{1.} For a more complete discussion of the Preqin data compared to similar sources, see Kaplan and Lerner (2016).

mance.² Finally, we feel confident in the suitability of the Preqin data for our analysis, because some of the primary concerns with Preqin data—such as survivorship bias, slow updates, and spotty coverage of cash flow data impact performance data in Preqin but do not affect the reporting of deals.

In the health care sector, deals are divided by Preqin into one of seven industries: Biotechnology, Biopolymers, Healthcare, Healthcare IT, Healthcare Specialists, Medical Devices & Equipment, and Pharmaceuticals. We simplified this to arrive at five industry groups used in later tables and figures. Given the importance of VCs to biopharmaceutical innovation, we spend some time on this topic and sometimes refer to the Pharmaceuticals and Biotechnology industries collectively as "Drugs." Similarly, we will sometimes refer to Healthcare and Healthcare IT collectively as "Healthcare Delivery," because the most common sub-industries in Healthcare include Diagnostics, Laboratories, and Hospitals, and the top sub-industries of Healthcare IT include Medical Software, Communication Platforms, Diagnostics, and Laboratories. These groupings will facilitate comparison between VC deals and academic publications in these areas.

We also discuss the funding stages at which these deals were made. For ease of explanation, we combine Series E, F, G, H, I, and J, PIPE, Mergers, Pre-IPO, and Secondary Stock Purchases into "Late Stage;" then we combine Series C and D, Venture Debt, Add-On, and Growth Capital into "Expansion;" finally, we combine Series A with Series B and Seed with Angel. The grants from foundations and government agencies mentioned above were tagged as such using the funding stage variable, so the point in the firm's life cycle at which it received the grant is unobservable in this context.

10.3 Three Facts about R&D in Health Care

Our analysis of deal making in this industry reveals three stylized facts about VC involvement in the healthcare sector over the past two decades:

1. VC funding in the health-care sector has grown more slowly and has been directed at earlier stage firms than VC funding in other sectors.

2. VC funding to young and innovative companies is overwhelmingly directed to the development of drugs; this is in fact even truer of grant money from the NIH than of investments by venture capitalists.

3. American firms dominate the VC deals in the Preqin data set on both sides of the transaction, but on the innovation side in particular; the Bay Area and the Greater Boston Area are hubs of both health-care innovation and investment in the US. They are joined by New York City on the investment side and San Diego on the innovation side. We explore this pattern of

^{2.} For example, Gompers and Wang (2017); Harris, Jenkinson, and Kaplan (2014); Korteweg and Nagel (2016).

A. VC Investment Volumes



B. Healthcare Share of Venture Capital Invested



Fig. 10.1 VC deals in health care as a share of all VC deals

allocation in the last section of our chapter, because it is consistent with VC not knowing about health-care innovation from other cities.

1. VC funding in the health-care sector has grown more slowly and has been directed at slightly less mature firms than VC funding in other sectors. VC deals in the health-care sector have been increasing fairly steadily over the past 20 years, but they have not grown as rapidly as VC deals in other sectors in the past decade. As a consequence, the health-care share of all VC deals has steadily declined over this period, from 33 percent in 2003 to just 14 percent in 2019 (see figure 10.1). If investments in health care commanded supranormal returns—perhaps because of the guarantee of high drug prices from future launches—then this would not be case.

		Healt	h-care deals		Non-health-care deals			
Funding stage	Number of deals	Percent of deals	Deal volume (\$ millions)	Percent of deal vol.	Number of deals	Percent of deals	Deal volume (\$ millions)	Percent of deal vol.
Seed & angel	3,971	13.00	4,521.43	1.20	42,231	26.57	41,445.84	2.39
Series A & B	9,163	30.00	133,585.76	35.31	42,786	26.92	457,008.58	26.30
Expansion	6,058	19.83	110,100.96	29.11	23,432	14.74	611,918.10	35.22
Late stage	1,271	4.16	38,430.69	10.16	2,868	1.80	215,477.93	12.40
Unspecified	10,085	33.01	91,638.13	24.23	47,614	29.96	411,768.08	23.70
Total	30,548	100.0	378,277.0	100.0	158,931	100.0	1,737,618.5	100.0

VC Deals by funding stage

Table 10.1



Fig. 10.2 Average deal sizes, health care and non-health care

Health-care deals are more likely to be made to Series A and B firms (35.3 percent) than to Expansion firms (29.1 percent), while firms in other sectors were more likely to receive VC funding during their Expansion stage (34.1 percent) rather than in Series A and B (27.9 percent). R&D in health care, and particularly in biopharmaceuticals, has higher risks earlier in the life cycle of companies, and venture capitalists play an important role in allocating capital to such projects (see table 10.1). Note that the funding stage of the deal is unavailable for about a quarter of VC deals, but nearly all deals with a known funding stage occurred either in the Series A and B stage of a firm's life or in the Expansion stage.

The pattern of VC investments, among health-care deals and non-health-care deals alike, has also been characterized by shrinking average deal sizes between 2000 and 2013, and a gradual return since 2013 to the average deal sizes as they were in 2000 (see figure 10.2). The fact that the trends are very

Primary industry	Number of deals	Percent of deals	Deal volume (\$millions)	Percent of deal vol.
Biotechnology	8,500	26.63	\$118,086.2	30.94
Pharmaceuticals	6,412	20.09	\$112,512.2	29.48
Medical devices & equipment	8,502	26.64	\$73,862.4	19.35
Health care	4,182	13.10	\$43,394.6	11.37
Health-care IT	4,322	13.54	\$33,801.4	8.86
Total	31,918	100.0	\$381,656.7	100.0

Table 10.2	VC deals by industry
------------	----------------------

similar across industries suggests that the explanation is not health-care specific. With the data available to us, we are not able to explore the relationship between deal size and changes in the availability of capital from sources other than VCs.

2. Funding to young and innovative companies is overwhelmingly directed to the development of drugs; this is in fact even truer for grant money from the NIH than of investments by VCs. Among VC deals, 60 percent of all money transacted was invested in firms working on drugs, another 20 percent was invested in firms working on a project related to medical devices, and 20 percent was given to firms working on a project related to health-care delivery (see table 10.2). If innovations in the drugs and devices industries allow VCs to capture the value of their investments more than innovations in health-care delivery (perhaps because health-care delivery investments have positive externalities, because of benefiting government payers or network externalities, that are difficult to fully capture), it would be intuitive to expect that private sources of funding favor drugs and devices. Over about the past 7 years, we see the most consistent growth in VC investments in the area of biopharmaceuticals is to firms in the Series A and B stages (see figure 10.3). Expansion investments appear to exhibit less consistent growth, and instead show occasional spikes driven by large deals, most clearly occurring in 2015 and 2018.³ This is one reason we will spend some time considering VC investments in biotech and pharmaceutical companies.

Interestingly, we find that NIH grants to companies are over-represented in the area of pharmaceuticals—the NIH portfolio of grants is domi-

3. The spike in Drugs Expansion investments in 2015 was primarily driven by Horizon Pharma's fundraising efforts to acquire Hyperion Therapeutics and Crealta Holdings, primarily for their orphan drugs (see https://www.nytimes.com/2015/03/31/business/dealbook/horizon-pharma-offers-to-acquire-hyperion-therapeutics-for-1-1-billion.html and https://www.chicagotribune.com/business/ct-horizon-buys-small-drugmaker-1212-biz-20151211 -story.html). The spike in 2018 was driven by a large joint venture undertaken by Novartis and Aduro Biotech in the field of immuno-oncology (see https://www.novartis.com/news/media -releases/novartis-accelerates-cancer-immunotherapy-efforts-aduro-biotech-alliance-and -launch-new-immuno-oncology-research-group).



Fig. 10.3 Funding stage breakdown of drug investments

nated by contributions to pharmaceuticals firms (see figure 10.4). In total, 76.4 percent of NIH grant money given to startups over the past 20 years has supported the development of drugs, another 13 percent has supported the development of medical devices, and only 10.5 percent supported investments in health-care delivery and infrastructure. Since this money was given in the form of grants, with no claim to future earnings or repayment, these grants should have been allocated based on expected future social good rather than on profitability. One justification for this allocation would be if NIH granted these funds for studying treatments for diseases that primarily affect communities unlikely to be able to pay high drug prices, or for treatments that were just below the threshold for economic viability. Evaluating this claim is beyond the scope of our analysis, but it would be important to know whether NIH grants to early-stage companies induce socially valuable innovation, or whether they are a substitute for private investments.

3. American firms dominate the VC deals in the Preqin data set on both sides of the transactions, with the Bay Area and the Greater Boston Area serving as hubs for health-care innovation and investment in the US. In the Preqin data set, 57 percent of VC investments in the health-care sector come from American VCs, and the top 10 investing countries contribute 88 percent of the money invested (see table 10.3). Of the money invested by American venture capitalists, 50 percent was originated from venture capitalists in the Bay Area, New York City, and the Greater Boston Area alone, with the top 10 MSAs contributing 62 percent together. The recipient firms of these investments are even more concentrated at the country level, but slightly less concentrated at the MSA level in the US. American firms receive fully 72.9 percent of the money accounted for in the Preqin deals data set, and A. NIH Grants to Startups, Share by Number of Grants, 244 Grants Total



B. NIH Grants to Startups, Share by Grant Volume, \$682 Million in Grants Total



Fig. 10.4 Industry breakdown of NIH grants

the top 10 MSAs received 55 percent of that money. The top three MSAs in terms of received investments were the Bay Area, the Greater Boston Area, and San Diego, which together received 44 percent of the money invested in American firms over the past 20 years.

We find that the top investing MSAs carry diversified portfolios across all five industries in health care (see figure 10.5, top panel). In contrast, Boston and San Diego show a clear specialization in drug development, with an overwhelming proportion of their VC-backed portfolio firms focused on either biotechnology or pharmaceuticals (see figure 10.5, bottom panel). In contrast, the Bay Area has well diversified innovations as well as investments. While at the MSA level investment portfolios appear to be well diversified,

Table 1	0.3 Geogral	phic dispersio	n of VC inve	stments						
Rank	Top countries	Number of deals	Percent of deals	Deal volume (\$millions)	Percent of deal vol.	Top US MSAs	Number of deals	Percent of US deals	Deal volume (Smillions)	Percent of US deal vol.
	Top commo		6 11771 10					a mon	(monito)	
				A. $Top i$	ecipients of	f VC funds				
1	SU	19,538	61.21	\$278,026.80	72.85	Bay Area	3,267	16.82	\$63,607.90	23.05
7	China	3,543	11.10	\$33,178.60	8.69	Boston	1,675	8.62	\$37,376.80	13.55
e	UK	1,714	5.37	\$14,961.30	3.92	San Diego	1,131	5.82	\$19,815.70	7.18
4	Germany	767	2.40	\$6,607.60	1.73	New York	649	3.34	\$8,916.10	3.23
5	Canada	805	2.52	\$6,529.50	1.71	Los Angeles	457	2.35	\$5,392.70	1.95
9	Switzerland	480	1.50	\$6,519.80	1.71	Seattle	452	2.33	\$5,388.80	1.95
7	France	672	2.11	\$6,174.30	1.62	Boulder	167	0.86	\$3,015.20	1.09
8	Israel	635	1.99	\$5,767.80	1.51	Durham	210	1.08	\$2,970.70	1.08
6	Ireland	190	0.60	\$2,950.90	0.77	Chicago	137	0.71	\$2,937.00	1.06
10	India	592	1.85	\$2,712.80	0.71	Austin	250	1.29	\$2,305.30	0.84
	Top 10	28,936	90.66	\$363,429.34	95.22	Top 10	8,395	43.23	\$151,726.33	54.99
				B. Top i	investors of	VC funds				
1	NS	14,464	45.27	\$185,778.77	56.90	Bay Area	7,778	21.77	\$54,680.50	23.30
7	China	3,541	11.08	\$27,929.07	8.55	New York	3,494	9.78	\$35,910.60	15.30
Э	UK	2,562	8.02	\$22,642.82	6.93	Boston	3,450	9.66	\$26,255.90	11.19
4	Switzerland	1,097	3.43	\$12,557.25	3.85	Chicago	812	2.27	\$6,529.90	2.78
5	Germany	1,146	3.59	\$7,562.86	2.32	Washington, DC	984	2.75	\$6,076.70	2.59
9	Hong Kong SAR	480	1.50	\$7,260.34	2.22	Houston	305	0.85	\$3,630.50	1.55
7	France	1,047	3.28	\$6,918.16	2.12	Seattle	557	1.56	\$3,554.10	1.51
8	Canada	863	2.70	\$6,483.07	1.99	Los Angeles	574	1.61	\$3,454.00	1.47
6	Singapore	393	1.23	\$5,406.41	1.66	Princeton	767	2.15	\$3,332.10	1.42
10	Japan	671	2.10	\$5,234.35	1.60	San Diego	458	1.28	\$2,238.10	0.95
	Top 10	26,264	82.21	\$287,773.11	88.13	Top 10	19,179	53.67	\$145,662.34	62.06





B. Top 10 Investing MSAs



Fig. 10.5 Geographic dispersion and industry specialization of MSAs

at least in the health-care sector, individual VCs appear to focus on one or two industries in particular (namely, biotechnology and pharmaceuticals). This may be an indication that these venture capitalists, while the nature of their work requires some degree of idiosyncratic risk, may be carrying more idiosyncratic risk than necessary. We highlight this observation as a suggestion for further research using other sources.

10.4 Understanding R&D Clusters in Health Care

The above facts on the flow of VC investments by geography motivated us to ask whether the disproportionate allocations of VC investments to San Francisco and Boston would be explained by a larger number of patents originating from these areas. However, some commentators have wondered whether this geographic concentration is a consequence of a phenomenon known as the "valley of death" (Hudson and Khazragui 2013), whereby early stage ventures often fail before commercialization, often as a result of venture capitalists' potential preference for innovation local to themselves. This preference might be rational—search costs are lower for ideas generated by local inventors, and local entrepreneurs might find it easier to establish a better reputation with venture capitalists. In contrast, it could also be the case that innovation stemming from clusters is simply of higher quality than that stemming from non-clusters.

To distinguish between these hypotheses, we sought to understand whether there was a link between VC dollars and the geographic location of patents. This requires us to subset the analysis to VC investments in the biopharma space, because this area relies on patents for innovation.

We obtained data on patenting comes from the US Patent Office's Patents-View—a modern data initiative organized by the USPTO that uses machine learning methods to disambiguate inventors. PatentsView is widely used in studies that require precise data on the location of inventors (Baruffaldi and Simeth 2020; Melero, Palomeras, and Wehrheim 2020). Data were obtained for all granted patents filed from 2000 until 2015 that were classified as "Drugs and Medical," subcategory "Drugs," using the NBER patent classification system. Introduced by Hall, Jaffe, and Trajtenberg (2001), the NBER patent classification system allows for easy identification of pharmaceutical drugs in this setting and lends itself to economic analysis. For certain analyses, chemicals patents were also identified from the PatentsView data using the NBER classification system. We chose 2015 as a cutoff, because it takes a while for patents to be approved.

Inventor locations in PatentsView are provided as latitudes and longitudes, as well as city-state tuples. To create a more usable form of location that takes into account economic clusters (and combines cities in a sensible way), we aggregate these data to the metropolitan statistical area level using the 2015 Census Bureau Shapefiles in conjunction with QGIS 3.1.0.⁴ These data were then collapsed to the MSA-year level using inventor weights to prevent double-counting of patents with inventors in multiple locations. To use a stylized example, consider a patent filed in 2011 with three inventors, one from the greater New York City region and two from the Boston metro area. The New York–Newark–Jersey City, NY–NJ–PA MSA would be assigned 1/3 of a patent in 2011, while the Boston–Cambridge–Newton, MA–NH MSA would be assigned 2/3 of a patent in 2011. What remains is an MSA-year-level data set of all US biotech and pharmaceutical drug patenting that is used for all subsequent analyses.

The first fact we present is that innovation in the US, as measured by patenting, is incredibly concentrated across geographies, and that this concentration increases for pharmaceutical and biotech patenting, respectively. Figure 10.6 plots Gini curves for patenting in select industries across MSAs in the US. While Pharmaceutical patenting is roughly as concentrated as patenting in Chemicals, Biotech uniquely stands out as the most concen-

^{4.} Shapefiles are publicly accessible at https://catalog.data.gov/.



Fig. 10.6 Geographic concentration of patents by industry



Fig. 10.7 Heat-map of biopharma patenting

trated industry in the sense of the origin of biotech patents. The specific geographies driving this concentration can be visualized in figure 10.7, a heat map of pharmaceutical patenting activity across the continental US. The specific geographies that account for the top 10 clusters are visualized in figure 10.8, which reports their contribution to total patenting in pharma

488



Fig. 10.8 Share of pharma and biotech patents in top-10 patenting metropolitical statistical areas (MSAs), 2000–2015

and biotech, respectively. Boston, New York City, and the Bay Area (defined as San Francisco and San Jose, California) are the largest contributors to overall patenting in pharma and biotech from 2000 to 2015.

As a first step in exploring the potential for a "valley of death" phenomena, we examine whether venture capital dollars flow to areas with increased patenting activity—a sign that venture capitalists are responding to increased innovation in geographies. Formally, we regress the log of yearly venture capital flows in pharmaceuticals and biotech on a 1-year lag of logged patenting in those two industries. Results from this exercise can be found in table 10.4, where column 1 includes year fixed effects, and column 2 includes year and MSA (location) fixed effects. Column 1 shows a statistically significant correlation between last year's patenting activity on this year's venture capital flows. When we add MSA fixed effects (column 2), we find that VC funding is no longer correlated with patenting activity, a sign that funding flows are simply a function of geography itself, wherein venture capitalists favor certain locations over others. Columns 3 and 4 use 1- and 2-year lags of patents as instruments to correct for idiosyncratic noise in current-year patent rates; we find similar results in these estimates. To be clear, the IV approach is only to clean up measurement error in the reporting of patents.

These results provide some preliminary evidence that the valley of death phenomenon does exist in this market as it pertains to VC funding flows as a response to innovation. We label our evidence as preliminary, because we cannot reject a model where the quality of patents is fixed across cities but not responsive to changes in the level of patenting (for example, a

	Log pharma VC dollars			
Outcome variable	OLS (1)	OLS (2)	IV (3)	IV (4)
Log patents	0.885* (0.118)	0.149 (0.135)	1.071* (0.146)	0.147 (0.853)
Controls:	· · · ·		· · · ·	· · · · ·
Year	Yes	Yes	Yes	Yes
MSA	No	Yes	Yes	Yes
Observations	989	989	563	563
Adjusted R-squared	0.356	0.649	0.371	-0.197

Table 10.4 OLS and IV estimates of VC investment elasticity

Notes: Standard errors appear in parenthesis and are clustered at the MSA level. Observations are at the MSA-year level. Controls are indicated above and include year and MSA fixed effects. Columns (1) and (2) present OLS estimates, and columns (3) and (4) present IV estimates, where log patents are instrumented by one and two year lags of log patents. *p < 0.10, **p < 0.05, ***p < 0.01.

model where patents originating in MSAs like Boston and San Francisco are systematically better than patents from other cities). Evaluating this possibility is a potential area of future research, as the consequences of undiscovered innovation are not just a cost to investors but also a cost to society. In ongoing work, we are using a more formal economic model to assess the presence of the valley of death, by examining one implication of this model—that patients from places that do not receive VC dollars are more successful when funded (which would mean that the marginal project from a city that received fewer VC dollars is better than the marginal project that was funded in Boston or San Francisco).

10.5 Predicting Future Innovations through Research Publications

The above exploration of the potential mismatch between VC investments and patents motivates us to consider another way of looking at early-stage entrepreneurship in health care: publications in medical journals and the relative roles of private and public funding in different areas of research in health care (basic science, devices, pharmaceuticals, delivery). Such an analysis would not be possible in other industries, where publication is not a prerequisite for commercialization. But the science-heavy nature of innovation in new medicines, devices, and the emphasis on clinical trials—not only for regulatory review but also as a standard for evidence—means that we can use publications as another measure of R&D in health care. One challenge with using publications as a measure of research activity is that we are measuring ideas, not dollars, and it is possible that no monotonic relationship exists between research papers and research dollars. Yet research papers may provide a superior prediction of future innovation than research dollars do, because they reflect the size of research support and the realization of that support.

Our data on peer-reviewed research publications were obtained by using the PubMed API to query based on publication characteristics, scrape the unique PubMed Identifiers that identify the papers that fit our criteria, and then group and count those publications by various dimensions. First, we restricted these counts to include only journal articles and clinical trials in phase 2 or 3, published between 1980 and 2019. This step dropped such publications as dissertations, meeting abstracts, lectures, editorials, and newspaper articles. Second, we restricted the counts based on the source of the funding, which can be identified in PubMed based on a combination of publication types and grant codes. PubMed has been tagging publications that received NIH funding by including the string "NIH" in the grant code field since 1980 and has been assigning the paper to the publication types "Research Support, N.I.H., Intramural" or "Research Support, N.I.H., Extramural" since 2005. We included in our counts of NIH publications all those that were tagged in one or both of these ways. We identified a publication as "privately funded" if it did not receive any funding from the NIH, any US government agency other than the NIH, any state or local government in the US, any foreign government, or one of the foundations listed explicitly in PubMed's supporting documentation (including Alzheimer's Association, Susan G. Komen, Wellcome Trust, and 54 others).⁵ For this reason, we cannot identify particular big private funding sources of academic publications in the health field, because they are not identified as such in the database. Third, we sometimes restricted publication counts based on the journal in which the publication appeared, in an attempt to adjust for the quality of the paper. For this purpose, we used British Medical Journal (BMJ), Cell, Journal of the American Medical Association (JAMA), Lancet, Nature, New England Journal of Medicine (NEJM), and Science.

Finally, we classified publications on the basis of the content of each publication, as identified by the Medical Subject Heading (MeSH) classifiers assigned to it. Publications have many MeSH classifiers assigned to them (more than 50 in some cases) by scientists who read the paper and determine what key terms define the topics discussed in it. Some of these classifiers for each publication are marked with an asterisk to denote it as a "major MeSH topic," which is reserved for about 10 or fewer MeSH classifiers that define the most central topics or ideas discussed in the publication. An example of PubMed's display of a paper and its MeSH classifiers, with major topics denoted by asterisks, is included as appendix A for a major review article on

^{5.} The list of foundations explicitly excluded from the counts of privately funded publications in PubMed is available at https://www.nlm.nih.gov/bsd/grant_acronym.html under "Other United States Funding Organizations" and "Non-US Funding Agencies/Organizations."

CRISPR technology. Our grouping of publications is assembled using only those MeSH classifiers marked as major topics.

The two other sets of publication groupings based on content we will be discussing are intended to more closely align with our analysis of VC deals. The first of these groups separates publications about drugs, medical devices, surgery and surgical techniques, health-care delivery, and other forms of science and treatment such as non-drug therapies, non-drug chemicals, and biological phenomena.⁶ Publications included in these counts under "Drugs" include papers about prescription drugs, generic drugs, placebos, drug combinations, biotechnology, and related terms, but not about illegal drugs, cannabis, or substance abuse. Publications coded as "Medical Devices" are about devices used both internally and externally, including atmosphere exposure chambers, catheters, diagnostic equipment, tourniquets, and others. Publications coded as "Surgery" include those discussing surgical procedures used either for operating or diagnosing, and structures created inside the body using surgical techniques. Publications coded as "Health Care Delivery" discuss things like administration of health care, access to health care, health-care facilities, disease prevention and outbreak control, and others. Publications coded as "Non-Drug Therapies" include those about various types of treatment that do not involve drugs or medical devices. "Non-Drug Chemicals" include papers about chemical compounds (proteins, amino acids, enzymes, etc.) but not about drugs. In a small set of cases, a publication classified by us as "Non-Drug Chemicals" was in fact about a prescription or over-the-counter drug, but rather than being assigned major MeSH topics related to pharmaceuticals, it was coded only in reference to the chemical composition of the drug (e.g., antihistamine under neurotransmitter agents, or ibuprofen under carboxylic acids). In these cases, the papers were truly about drugs but were misclassified by us as papers about non-drug chemicals. However, these cases are rare, and we do not expect this miscoding to drive our results. Finally, "Biological Phe-

6. These groups of publications based on MeSH topics are not mutually exclusive by default, since a publication can discuss many topics. To create groups that were mutually exclusive of one another without dropping a significant number of publications that discuss many aspects of health, these groups had to be prioritized into a hierarchical structure. For the most relevant comparisons between the research in PubMed and the VC deals described below, we first prioritized publications about drugs and medical devices. That is, if a publication was about either drugs or medical devices (or both), it was coded as such, regardless of the publication's other topics. Therefore, the publications in our data set that were coded as surgery or surgical techniques include only papers about surgery and not about drugs or medical devices. This is particularly relevant to understanding the overlap between surgery and medical devices, because some medical devices (e.g., stents) are implanted using surgical techniques. While papers about best practices or new innovations in implanting stents are relevant both to discussions of surgical techniques and medical devices, for the purposes of understanding the pushes and pulls for innovations in the health-care sector, we code them here only as medical devices. This prioritization occurs in a similar way for the other groups as well, with publications coded as health-care delivery being those about health-care delivery and not about drugs, medical devices, or surgery; non-drug therapies, non-drug chemicals, and biological phenomena are similarly defined by excluding the higher priority topics from their corresponding queries.

nomena" include papers about anatomy, organisms, and other biological processes but none of the topics previously mentioned.

In our third method of categorizing PubMed publications based on their content, we mapped the categories of drugs, devices, surgery, non-drug therapies, non-drug chemicals, and biological phenomena directly to designations of basic science, translational science, and clinical science. Our definitions of these three categories and the MeSH classifiers assigned to each were based on the stages of scientific research as it pertains to medical problem-solving as described by the Dana-Farber Cancer Institute.⁷ Here the publications categorized as "basic science" refer to developing knowledge about how body systems and chemical compounds function and interact with one another, and they are defined as publications about biological phenomena, chemicals, and drugs, but not about diseases. Therefore, the publications we group as "basic science" seek to understand the functions and uses of different mechanisms in the body and chemical compounds, even drugs, but not in the context of treating any disease in particular.

We define "translational science" as research that connects the findings of basic science to specific medical issues and challenges. In particular, this group refers to publications about diseases and also chemicals, drugs, biological phenomena, and non-drug therapies, but not surgical or diagnostic techniques. Finally, "clinical science" refers in general to the application of translational science findings to the resolution of medical problems. In particular, we define this group as publications about diseases and also about surgery, diagnostics, anesthesia, analgesia, or medical devices. Basic science, translational science, and clinical science publications counts here include only journal articles that are not clinical studies; that is, they reflect academic research that is still at least one step removed from its implementation in medical problem-solving. Phase 2 and phase 3 clinical trials are included in exhibits as their own separate groups.

With these classifications, we note the extent to which some MSAs appear to be far more specialized in basic science research related to drug discovery at all stages—basic, translational, and clinical (the three panels of figure 10.9 demonstrate the striking degree to which Boston and the Bay area publish papers in basic science, translational science and clinical trials). If basic science research produces better patents because of deeper insights or more novel insights, then the flow of VC dollars to these MSAs may not be surprising, but to truly ascertain this channel, future researchers would have to map quality of downstream patents to the quality of upstream basic and translational science research.

One challenge with basic science research is that it is hard for researchers to expropriate the full social value of their research—because the research may be quite removed from a disease or therapeutic. Recognizing this mar-

 $[\]label{eq:2.1} 7. https://blog.dana-farber.org/insight/2017/12/basic-clinical-translational-research-whats -difference/.$

A. Basic Science, Top 7 Journals. These Top MSAs Represent 8873 of 35063 (25%) with Affiliation Text



B. Translational Science, Top 7 Journals. These Top MSAs Represent 3941 of 19583 (20%) with Affiliation Text



C. Clinical Science, Top 7 Journals These Top MSAs Represent 2681 of 16086 (17%) with Affiliation Text



Fig. 10.9 Publications by stage of science research. *Note:* Top seven journals are: *BMJ, Cell, JAMA, Lancet, Nature, NEJM, Science*

ket failure is one reason that governments finance basic science research. An open question for new innovations in health care is to understand how well governments fill this gap—governments should, for example, be more willing to finance basic science research and research on topics like healthcare delivery than private actors are. One implication is that NIH funded research should skew more toward basic science than it does toward clinical science, in contrast to privately funded research. Figure 10.10 provides A. Privately Funded Publications, Journal Articles & Clinical Trials (II & III),

Top 7 Journals, Chart Represents 31954 of 88034 Publications



B. NIH-Funded Publications, Journal Articles & Clinical Trials (II & III),

Top 7 Journals, Chart Represents 31961 of 36754 Publications



Fig. 10.10 Science publications by stage and funding

some initial evidence of allocative efficiency as measured by the share of NIH funded work that is directed to basic research rather than to clinical or translational science.

10.6 Conclusion

A variety of pull and push forces influence innovation in health care, with venture capitalists playing an extremely important role in marshaling the pull forces that drive innovation. The NIH (or government more generally) is responsible for reducing market failure by subsidizing research that no commercial entity would fund. Our analysis has uncovered several new facts on the operation of these entities. On the VC side, VC dollar allocations have moved away from investment in health care, and a disproportionate share of VC investments—almost 60 percent—are in drugs and devices. While

C. Privately Funded Publications, Journal Articles & Clinical Trials (II & III),

All Journals, Chart Represents 5460920 or 7975858 Publications



D. NIH-Funded Publications, Journal Articles & Clinical Trials (II & III),

All Journals, Chart Represents 1287102 of 470112 Publications



Fig. 10.10 (cont.)

there are many justifications for this allocation, there is also a concern than the VC model will not automatically bring socially valuable innovations to market, because it will emphasize private returns over public ones. Government efforts in health care should therefore try to subsidize socially valuable projects—and while we find that the NIH does do this, we also find that the shape of NIH funded translational research is similar to that of privately funded translational research when it should not be. This deserves further exploration, because the NIH should not follow industry in the desire to show commercial benefit. Public investments should not be a substitute for private investments; instead they should induce complementary investments by the private sector by reducing R&D uncertainty and making these investments more viable for private entities. Another striking fact that deserves further exploration is that three locations account for the plurality of scientific research and also receive the majority of VC investments. Whether this represents higher quality research from these cities, or a "valley of death" is an important question for future research.

The growth prospects of the sector will depend on the answers to these two questions: whether the NIH is overinvesting in clinical and translational research and underinvesting in basic science, and whether the "valley of death" is leading to underinvestment in healthcare innovation outside the investment hubs of Boston and the Bay Area. If these market failures are substantial, remedying them could create sustainable growth in the sector by allocating funding to the marginal invention or research project, which could in turn spur downstream research and innovation. If not, health-care spending will likely continue to grow, but productivity growth in the sector will continue to lag behind that in the economy as a whole. An additional factor at play in the future of the health-care industry will be the lasting impact of the Coronavirus pandemic on health-care delivery infrastructure, particularly in digital health and telehealth resources, for which it is too early to make meaningful projections.

Appendix

Review > Hum Mol Genet. 2014 Sep 15;23(R1):R40-6. doi: 10.1093/hmg/ddu125. Epub 2014 Mar 20.

CRISPR/Cas9 for Genome Editing: Progress, Implications and Challenges

Feng Zhang ¹, Yan Wen ¹, Xiong Guo ²

Affiliations + expand PMID: 24651067 DOI: 10.1093/hmg/ddu125

MeSH terms

- > Animals
- > Clustered Regularly Interspaced Short Palindromic Repeats / genetics*
- > DNA Cleavage
- > Gene Expression Regulation
- > Gene Transfer Techniques
- > Genetic Therapy
- > Genome
- > Genomics / methods*
- > Humans
- > Protein Structure, Tertiary
- > RNA Editing / genetics*
- > Sequence Analysis, DNA

Fig. 10.A.1 Example of PubMed display of a paper with MeSH classifiers

References

- Baruffaldi, Stefano H., and Markus Simeth. 2020. "Patents and Knowledge Diffusion: The Effect of Early Disclosure." *Research Policy* 49(4): 103927.
- Baumol, W. J., and W. G. Bowen. 1965. "On the Performing Arts: The Anatomy of Their Economic Problems." *American Economic Review* 55(1/2): 495–502.
- Gompers, P., and S. Q. Wang. 2017. "And the Children Shall Lead: Gender Diversity and Performance in Venture Capital." NBER Working Paper No. 23454. Cambridge, MA: National Bureau of Economic Research.
- Gornall, W., and I. A. Strebulaev. 2015. "The Economic Impact of Venture Capital: Evidence from Public Companies." Stanford GSB Research Paper No. 15-55. https://ssrn.com/abstract=2681841.
- Hall, B. H., A. B. Jaffe, and M. Trajtenberg. 2001. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." NBER Working Paper No. 8498. Cambridge, MA: National Bureau of Economic Research.
- Harris, R. S., T. Jenkinson, and S. N. Kaplan. 2014. "Private Equity Performance: What Do We Know?" *Journal of Finance* 69(5): 1851–82.
- Howell, S., J. Lerner, R. Nanda, and R. Townsend. 2020. "Financial Distancing: How Venture Capital Follows the Economy Down and Curtails Innovation." NBER Working Paper No. 27150. Cambridge, MA: National Bureau of Economic Research.
- Hudson, John, and Hanan F. Khazragui. 2013. "Into the Valley of Death: Research to Innovation." *Drug Discovery Today* 18(13–14): 610–13.
- Kaplan, S. N., and J. Lerner. 2016. "Venture Capital Data: Opportunities and Challenges." NBER Working Paper No. 22500. Cambridge, MA: National Bureau of Economic Research.
- Korteweg, A., and S. Nagel. 2016. "Risk-Adjusting Returns to Venture Capital." Journal of Finance 71(3): 1437–70.
- Lerner, J., and R. Nanda. 2020. "Venture Capital's Role in Financing Innovation: What We Know and How Much We Still Need to Learn." *Journal of Economic Perspectives* 34(3): 237–61.
- Melero, Eduardo, Neus Palomeras, and David Wehrheim. 2020. "The Effect of Patent Protection on Inventor Mobility." *Management Science* 66(12): 5485–504
- Nunn, R., J. Parsons, and J. Shambaugh. 2020. "A Dozen Facts about the US Health-Care System." Brookings Institution: The Hamilton Project, Economic Facts, March. https://www.brookings.edu/wp-content/uploads/2020/03/HealthCare _Facts_WEB_FINAL.pdf.
- Russell, John. 2015. "Horizon Pharma to Buy Lake Forest Company for \$510 Million." Chicago Tribune, December 11.
- Sheiner, L., and A. Malinovskaya. 2016. *Measuring Productivity in Healthcare: An Analysis of the Literature*. Washington, DC: Brookings Institution, Hutchins Center.