Venture Capital Led Entrepreneurship in Health Care

Amitabh Chandra, Cirrus Foroughi and Lauren Mostrom

Harvard Business School, Harvard Kennedy School and the NBER
Harvard Business School
Harvard Business School

January 29th, 2021

Despite the fact that 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 note that among non-financial firms that issued an IPO between 1995 and 2018, 47% were backed by a VC fund[1]. Of those firms that were still public at the end of 2019, the firms originally backed by VC made up 76.2% of the total market capitalization, and were responsible for 88.6% of total R&D expenditure.

As of 2014, about 37% of the healthcare industry's market capitalization was backed by VC, making it the third-most VC-backed industry, behind electronics and software[2]. In 2017, health care spending in the US made up 17.1% of GDP ($ 3.3 trillion), far exceeding the shares of other developed nations, and per capita health care spending has nearly doubled since 1995[3]. 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[4]. 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 (say 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 VCs. These concerns are amplified by the observation that VCs ought to be insulated from public markets given their long time-horizons, but there is some

[1] Lerner, J. and R. Nanda. Venture Capital's Role in Financing Innovation: What We Know and How Much We Still Need to Learn. *Journal of Economic Perspectives* 34(3) (2020): 237-261.
[2] Gornall, W. and I. A. Strebulaev. The Economic Impact of Venture Capital: Evidence from Public Companies. Stanford GSB Research Paper No. 15-55 (2015), Available at SSRN: https://ssrn.com/abstract=2681841
[3] Nunn, R., J. Parsons, and J. Shambaugh. A Dozen Facts about the US Health-Care System. Brookings Institution: The Hamilton Project, Economic Facts March (2020).
[4] Sheiner, L. and A. Malinovskaya. Measuring Productivity in Healthcare: An Analysis of the Literature. Brookings Institution: Hutchins Center, May (2016); Baumol, W. J. and W. G. Bowen, On the Performing Arts: The Anatomy of their Economic Problems. *American Economic Review* 55(1/2) (1965): 495-502

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[5].

Our model of VC backed investments in healthcare is the same as R&D investments outside of healthcare, but with augmented risk and return parameters. VCs will make R&D investments if the expected NPV of a project, discounted at the appropriate cost-of-capital 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.

Within this simple model, VCs will want to prioritize projects that are able to move from an idea to a commercialized product within a short time, and this force will—*ceteris paribus*--discourage investments in early-stage pharmaceuticals and biotechnology firms because of 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 VCs 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 healthcare 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 the social value of these discoveries. These potential mismatches may be smaller outside of the healthcare sector where the regulatory burdens are lower and direct-to-consumer selling are the dominant sales channel. One strategy that reduces the time pressure and uncertainty of innovations is the use of multiple funding

[5] Howell, S., J. Lerner, R. Nanda, and R. Townsend. Financial Distancing: How Venture Capital Follows the Economy Down and Curtails Innovation. NBER Working Paper 27150 (2020).

rounds in a start-up'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 healthcare investments but for this reason, we might expect funding for VC R&D in healthcare to go to *earlier* stage companies relative to non-healthcare investments because VCs will have a particular comparative advantage in making sense of these early-stage investments. Another approach used by some specialized VCs to mitigate these challenges, especially in the biotech and pharmaceuticals industries, is to create new start-ups inside "VC-foundries", to reduce information asymmetry between the VC and the entrepreneur.

With these motivations, we describe trends in life cycle of innovation in the healthcare sector, starting with a dataset from Preqin on VC deals in the healthcare sector, in order to develop a more detailed picture of early-stage innovation in healthcare. We find that VC funding in the healthcare 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% of all money was invested in firms working on drugs, another 20% was invested in firms working on a project related to medical devices, and 20% was given to firms working on health care delivery. We also find enormous geographic concentration of healthcare deals which motivates us to explore the 'valley of death' hypothesis (the idea that many useful inventions are not explored because VCs 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 emphasize the need to evaluate it more carefully.

This fact in turn 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 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 a number of additional 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. Second, there is some evidence, albeit directional evidence, 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. 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 privately funded research. This finding raises a number of questions of whether the government should rethink how it allocates money to projects.

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 to the investment deals themselves, in order to develop a fuller understanding of which research ideas and developments are determined a priori to be the most commercializable by VCs. The Venture Capital Deals dataset 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 will be referring to the amounts of money invested by VCs in young companies, not the amounts transacted between VCs and their limited partners (LPs). In order to compare these deal volumes across time, all deal values were converted to 2020 USD using the Consumer Price Index for All Urban Consumers.

The Preqin database has been assembled based on voluntary reporting from general partners (GPs) and LPs of venture funds and public filings from pension funds. This dataset has an advantage of transparency, since GPs are able and willing to submit corrections for inaccurate information about their funds. One potential bias, however, is that it misses certain high

performing VCs, such as Sequoia and Accel, due to the way it collects information[6]. Despite these limitations, Preqin data have been used in recent scholarship to conduct various analyses of performance[7]. 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.

Within the healthcare sector, deals are divided by Preqin into one of 7 industries: Biotechnology, Biopolymers, Healthcare, Healthcare IT, Healthcare Specialists, Medical Devices & Equipment, and Pharmaceuticals. We simplified this to arrive at 5 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 subindustries within Healthcare include Diagnostics, Laboratories, and Hospitals, and the top subindustries 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 combined Series C and D, Venture Debt, Add-On, and Growth Capital into "Expansion;" finally we combined 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.

[6] For a more complete discussion of the Preqin data compared to similar sources, see Kaplan, S. N. and J. Lerner. 2016. Venture Capital Data: Opportunities and Challenges. NBER Working Paper 22500 (2016).
[7] E.g. Gompers, P. and S. Q. Wang. And the Children Shall Lead: Gender Diversity and Performance in Venture Capital. NBER Working Paper 23454 (2017); Harris, R. S., T. Jenkinson, and S. N. Kaplan. Private Equity Performance: What Do We Know? Journal of Finance 69(5) (2014): 1851-1882; Korteweg, A. and S. Nagel. Risk-Adjusting Returns to Venture Capital. Journal of Finance 71(3) (2016): 1437-1470.

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:

i. VC funding in the healthcare sector has grown more slowly and been directed at *earlier stage* firms than VC funding in other sectors.

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

iii. American firms dominate the VC deals in the Preqin dataset 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 within the US, and are joined by New York City on the investment side and San Diego on the innovation side. We explore this pattern of allocation in the last section of our paper for it is consistent with VC not knowing about health care innovation from other cities.

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

Healthcare deals are more likely to be made to Series A and B firms (35.3%) than to Expansion firms (29.1%), while firms in other sectors were more likely to receive VC funding during their Expansion stage (34.1%) rather than in Series A and B (27.9%). R&D in health care, and particularly in biopharmaceuticals, has higher risks earlier in the life-cycle of companies and VCs play an important role in allocating capital to such projects (see Table 1). Note that the funding stage of the deal is unavailable for about a quarter of VC deals, but the overwhelming majority of deals with a known funding stage occurred either in the Series A and B stage of a firm's life or the Expansion stage.

The pattern of VC investments, among healthcare deals and non-healthcare 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 2). The fact that the trends are very similar across industries suggests that the explanation is not healthcare specific. With the data available to us we are not able to explore relationship between deal-size and changes in the availability of capital from sources other than VCs.

(ii) Funding to younger companies is overwhelmingly directed towards the development of drugs; this is in fact even truer of grant money from the NIH than investments by VCs. Among VC deals, 60% of all money transacted was invested in firms working on drugs, another 20% was invested in firms working on a project related to medical devices, and 20% was given to firms working on a project related to health care delivery (see Table 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 benefitting 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 to firms in the Series A and B stages (see Figure 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[8]. This is one reason why 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 dominated by contributions to pharmaceuticals firms (see Figure 4). In total, 76.4% of NIH grant money given to start-ups over the past 20 years has supported the development of drugs, another 13% has supported the development of medical devices, and only 10.5% supported investments in health care delivery and infrastructure. Since

[8] 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).

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 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.

(iii) American firms dominate the VC deals in the Preqin dataset on both sides of the transactions, with the Bay Area and the Greater Boston Area serving as hubs for healthcare innovation and investment within the US. In the Preqin dataset, 57% of VC investments in the healthcare sector come from American VCs, and the top ten investing countries contribute 88% of the money invested (see Table 3). Of the money invested by American VCs, 50% was originated from VCs in the Bay Area, New York City, and the Greater Boston Area alone, with the top ten MSAs contributing 62% together. The recipient firms of these investments are even more concentrated at the country level, but slightly less concentrated at the MSA level within the US. American firms receive fully 72.9% of the money accounted for in the Preqin deals dataset, and the top ten MSAs received 55% 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% of the money invested in American firms over the last 20 years.

We find that the top investing MSAs carry diversified portfolios across all five industries within healthcare (see Figure 5, top panel). In contrast, Boston and San Diego show a clear specialization in drugs development, with an overwhelming proportion of their VC backed portfolio firms focused on either biotechnology or pharmaceuticals (see Figure 5, bottom panel). The Bay Area, on the other hand, has well diversified innovations as well as investments. While at the MSA level investment portfolios appear to be well diversified, at least within the healthcare sector, individual VCs appear to focus on one or two industries in particular (namely biotechnology and pharmaceuticals). This may be an indication these VCs, 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.

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. On the other hand, some commentators have wondered if this geographic concentration is a consequence of a phenomenon known as the "valley of death"[9], whereby early stage ventures often fail before commercialization, often as a result of venture capitalists' potential preference for innovation local to them. 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 innovation 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 United States Patent Office's PatentsView-- 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[10]. Data was 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

[9] Hudson, John, and Hanan F. Khazragui. "Into the valley of death: research to innovation." *Drug discovery today* 18.13-14 (2013): 610-613.

[10] Baruffaldi, Stefano H., and Markus Simeth. "Patents and knowledge diffusion: The effect of early disclosure." *Research Policy* 49.4 (2020): 103927.
Melero, Eduardo, Neus Palomeras, and David Wehrheim. "The effect of patent protection on inventor mobility." *Management Science* (2020).

lends itself to economic analysis[11]. 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. In order 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[12]. This data was 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 dataset of all United States biotech and pharmaceutical drug patenting that is used for all analyses to follow.

The first fact that we present is that innovation in the United States, as measured by patenting, is incredibly concentrated across geographies, and that this concentration increases for pharmaceutical and biotech patenting, respectively. Figure 6 below plots Gini curves for patenting within select industries across MSAs in the United States. While Pharmaceutical patenting is roughly as concentrated as patenting in Chemicals, Biotech uniquely stands out as the most concentrated industry in the sense of the origin of biotech patents. The specific geographies driving this concentration can be visualized in Figure 7 below, a heat map of pharmaceutical patenting activity across the continental United States. The specific geographies that account for the top-10 clusters are visualized in Figure 8 which reports their contribution to total patenting in pharma 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 2000-2015.

[11] Hall, B. H., A. B. Jaffe, and M. Trajtenberg. The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Paper 8498 (2001).
[12] Shapefiles are publicly accessible at https://catalog.data.gov/

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 one-year lag of logged patenting in those two same industries. Results from this exercise can be found in Table 4, where column one includes year fixed effects and column two includes year and MSA (location) fixed effects. Column one 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 venture capital 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 one and two-year lags of patents as instruments to correct for idiosyncratic noise in current-year patent rates and 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 a fixed across cities but not responsive to changes in the level of patenting (for example, a 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 active future research, as the consequences of undiscovered innovation are not just a cost to investors, but a cost to society. In on-going 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 place 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).

Predicting future innovations through research publications

This above exploration of the potential mismatch between VC investments and patents motivates us to consider another way of looking at early-stage entrepreneurship in healthcare:

publications in medical journals and the relative roles of private and public funding 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're measuring ideas not dollars, and it is possible that there is not a monotonic relationship between research papers and research dollars. Yet, research papers may provide a superior prediction of future innovation than research dollars 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 PMIDs 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 publications such as dissertations, meeting abstracts, lectures, editorials, and newspaper articles. Second we restricted based on the source of the funding, which can be identified within 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 federal government agency other than the NIH, any state or local government within 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[13]. For this reason we cannot identify particular big private funding sources of academic publications in the health field, because they are not

[13] 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."

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 (up to 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 1 for a major review article on CRISPR technology. Our grouping of publications are 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[14]. Publications included in these counts under "Drugs" include papers about prescription drugs, generic drugs, placebos, drug combinations, biotechnology, and related

[14] These groups of publications based on MeSH topics are not mutually exclusive by default, since a publication can discuss many topics. In order 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 in the next section, 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 dataset 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 healthcare 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.

terms, but not 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 Phenomena" 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 described by Dana-Farber Cancer Institute[15]. 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 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 within the body and chemical compounds, even drugs, but not in the context of treating any disease in particular.

[15] https://blog.dana-farber.org/insight/2017/12/basic-clinical-translational-research-whats-difference/

We define "translational science" as research that connects the findings of basic science to particular 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 group.

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 (see the three panels of Figure 9) for 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 down-stream patents to the quality of upstream basic and translational science research.

One challenge with basic science research is that it is hard for the researcher to expropriate the full social value of their research—because the research may be quite removed from a disease or therapeutic. Recognizing this market-failure is one reason that governments finance basic science research. One 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 health care delivery than private actors. One implication of this is that NIH funded research should skew more towards basic science than clinical science that privately funded research. Figure 10 provides some initial evidence of allocative efficiency as measured by the share of NIH funded work that is basic science relative to clinical or translational.

Conclusion

A variety of pull and push forces influence innovation in health care, with VCs playing an extremely important role in marshalling 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 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 healthcare 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 privately funded translational research when it should not be. This deserves further exploration for it should not be the case that the NIH follows industry in the desire to show commercial benefit: public investments should not be a substitute for private investments, rather 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 the economy as a whole. An additional factor at play in the future of the healthcare 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.

TABLE 1: VC Deals by Funding Stage

	Healthcare Deals				Non-Healthcare 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%

TABLE 2: VC Deals by Industry

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%
Healthcare	4,182	13.10%	$ 43,394.6	11.37%
Healthcare IT	4,322	13.54%	$ 33,801.4	8.86%
Total	31,918	100.0%	$ 381,656.7	100.0%

TABLE 3: Geographic Dispersion of VC Investments

A. Top Recipients of VC Funds

Rank	Top Countries	Number of Deals	Percent of Deals	Deal Volume ($Millions)	Percent of Deal Vol.	Top US MSAs	Number of Deals	Percent of US Deals	Deal Volume ($Millions)	Percent of US Deal Vol.
1	US	18,725	61.30%	$ 241,515.9	72.30%	Bay Area	3,048	16.38%	$ 54,202.3	22.62%
2	China	3,542	11.59%	$ 31,282.5	9.36%	Boston	1,454	7.81%	$ 31,042.3	12.95%
3	UK	1,580	5.17%	$ 13,188.2	3.95%	San Diego	1,086	5.84%	$ 16,912.0	7.06%
4	Canada	784	2.57%	$ 5,723.7	1.71%	New York	621	3.34%	$ 8,132.8	3.39%
5	Germany	715	2.34%	$ 5,631.5	1.69%	Los Angeles	445	2.39%	$ 4,652.2	1.94%
6	Switzerland	439	1.44%	$ 5,617.0	1.68%	Seattle	425	2.28%	$ 4,643.4	1.94%
7	France	635	2.08%	$ 5,293.6	1.58%	Chicago	133	0.71%	$ 2,761.2	1.15%
8	Israel	608	1.99%	$ 5,022.5	1.50%	Boulder	162	0.87%	$ 2,711.2	1.13%
9	Ireland	175	0.57%	$ 2,581.8	0.77%	Durham	198	1.06%	$ 2,396.5	1.00%
10	India	587	1.92%	$ 2,451.6	0.73%	Lexington, MA	138	0.74%	$ 2,141.4	0.89%
	Top 10	27,790	90.97%	$ 318,308.34	95.29%	Top 10	7,710	41.43%	$ 129,595.19	54.08%

B. Top Investors of VC Funds

Rank	Top Countries	Number of Deals	Percent of Deals	Deal Volume ($Millions)	Percent of Deal Vol.	Top US MSAs	Number of Deals	Percent of US Deals	Deal Volume ($Millions)	Percent of US Deal Vol.
1	US	13,717	44.65%	$ 160,512.32	56.17%	Bay Area	7,727	22.03%	$ 47,635.6	23.35%
2	China	3,540	11.52%	$ 26,297.17	9.20%	New York	3,458	9.86%	$ 31,772.7	15.58%
3	UK	2,438	7.94%	$ 19,617.73	6.87%	Boston	3,413	9.73%	$ 23,026.1	11.29%
4	Switzerland	1,075	3.50%	$ 10,917.39	3.82%	Chicago	812	2.32%	$ 5,798.4	2.84%
5	Hong Kong SAR	480	1.56%	$ 6,814.13	2.38%	Washington, DC	586	1.67%	$ 4,469.3	2.19%
6	Germany	1,116	3.63%	$ 6,490.22	2.27%	Houston	305	0.87%	$ 3,156.3	1.55%
7	France	1,022	3.33%	$ 6,041.80	2.11%	Los Angeles	571	1.63%	$ 3,117.0	1.53%
8	Canada	843	2.74%	$ 5,536.95	1.94%	Seattle	536	1.53%	$ 3,067.0	1.50%
9	Singapore	389	1.27%	$ 4,870.63	1.70%	Princeton	767	2.19%	$ 2,722.7	1.33%
10	Japan	669	2.18%	$ 4,689.15	1.64%	Baltimore	252	0.72%	$ 1,964.2	0.96%
	Top 10	25,289	82.31%	$ 251,787.49	88.12%	Top 10	18,427	52.55%	$ 126,729.20	62.13%

TABLE 4: OLS and IV Estimates of VC Investment Elasticity

	(1) OLS	(2) OLS	(3) IV	(4) IV
Outcome Variable	Log Pharma VC dollars			
Log Patents	0.885*	0.149	1.071*	0.147
	(0.118)	(0.135)	(0.146)	(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

* p<0.10, ** p<0.05, *** p<0.01

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.

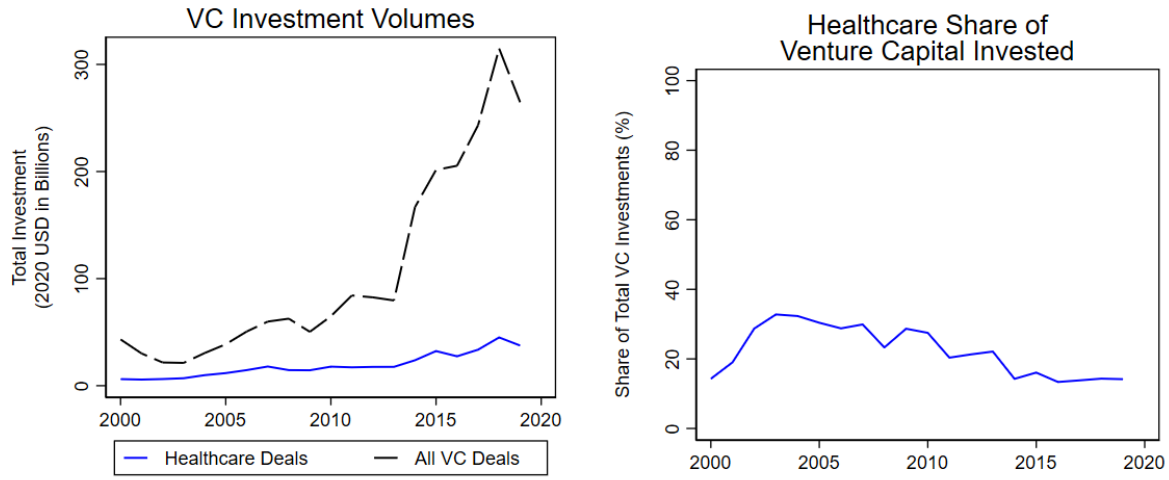


Figure 1. VC Deals in Healthcare as a Share of All VC Deals

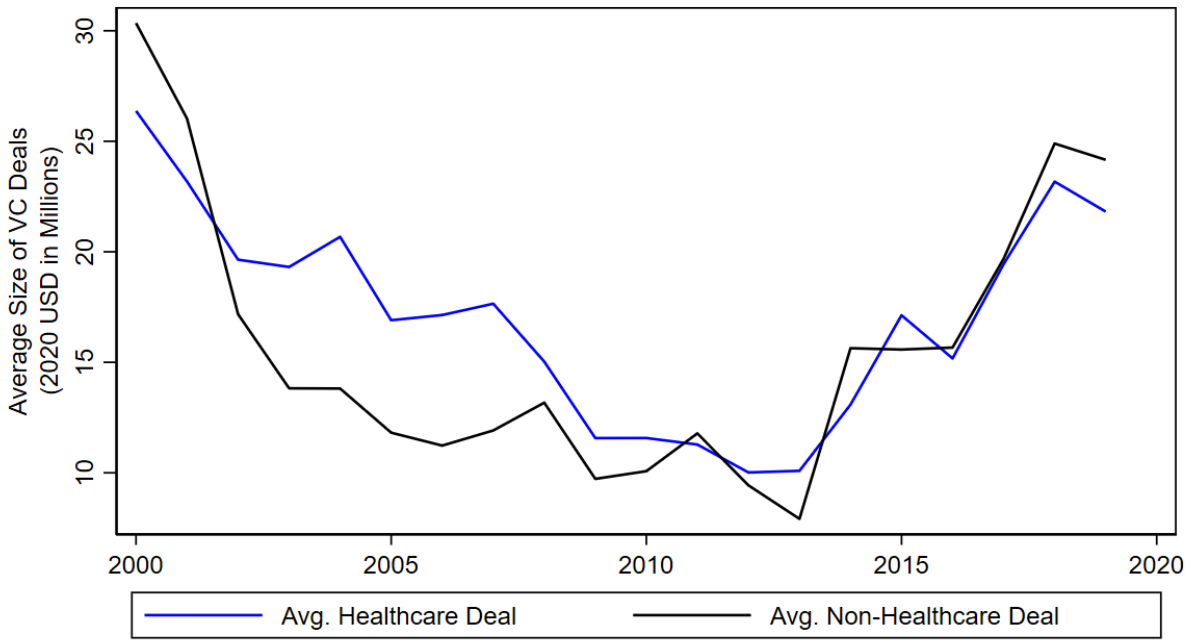


Figure 2. Average Deal Sizes, Healthcare and Non-Healthcare

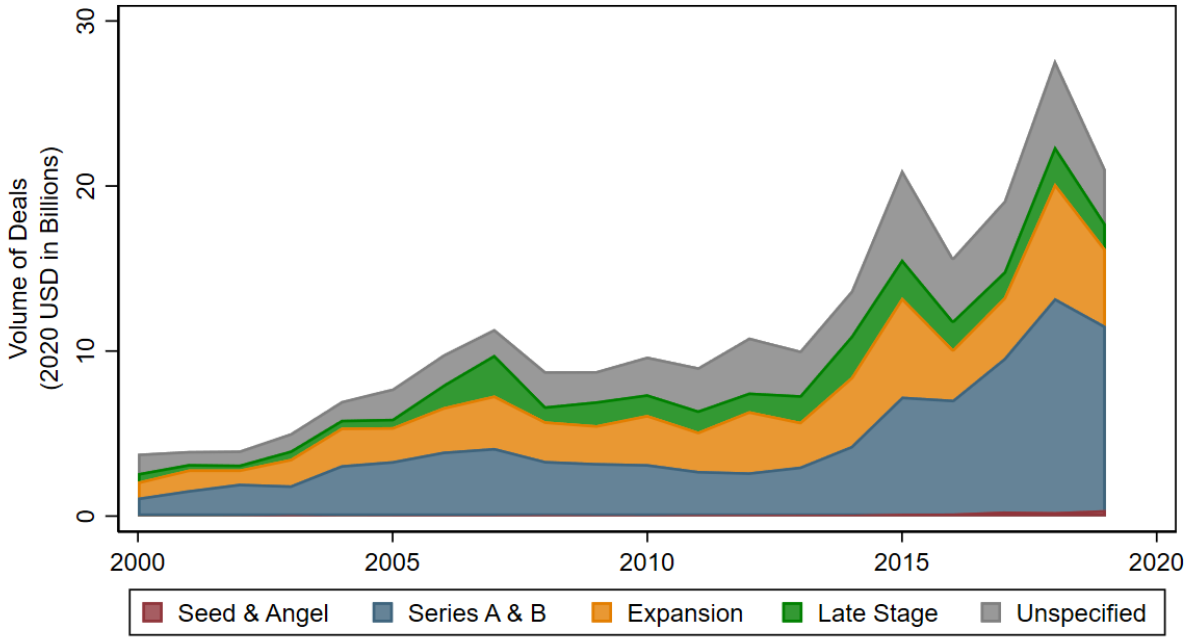


Figure 3. Funding Stage Breakdown of Drugs Investments

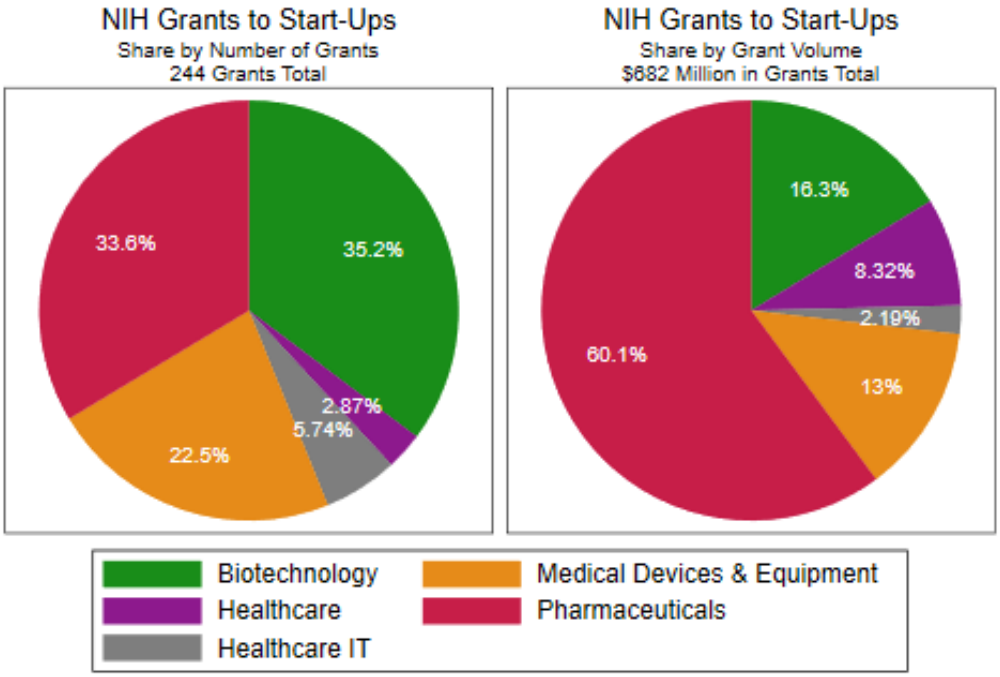


Figure 4. Industry Breakdown of NIH Grants

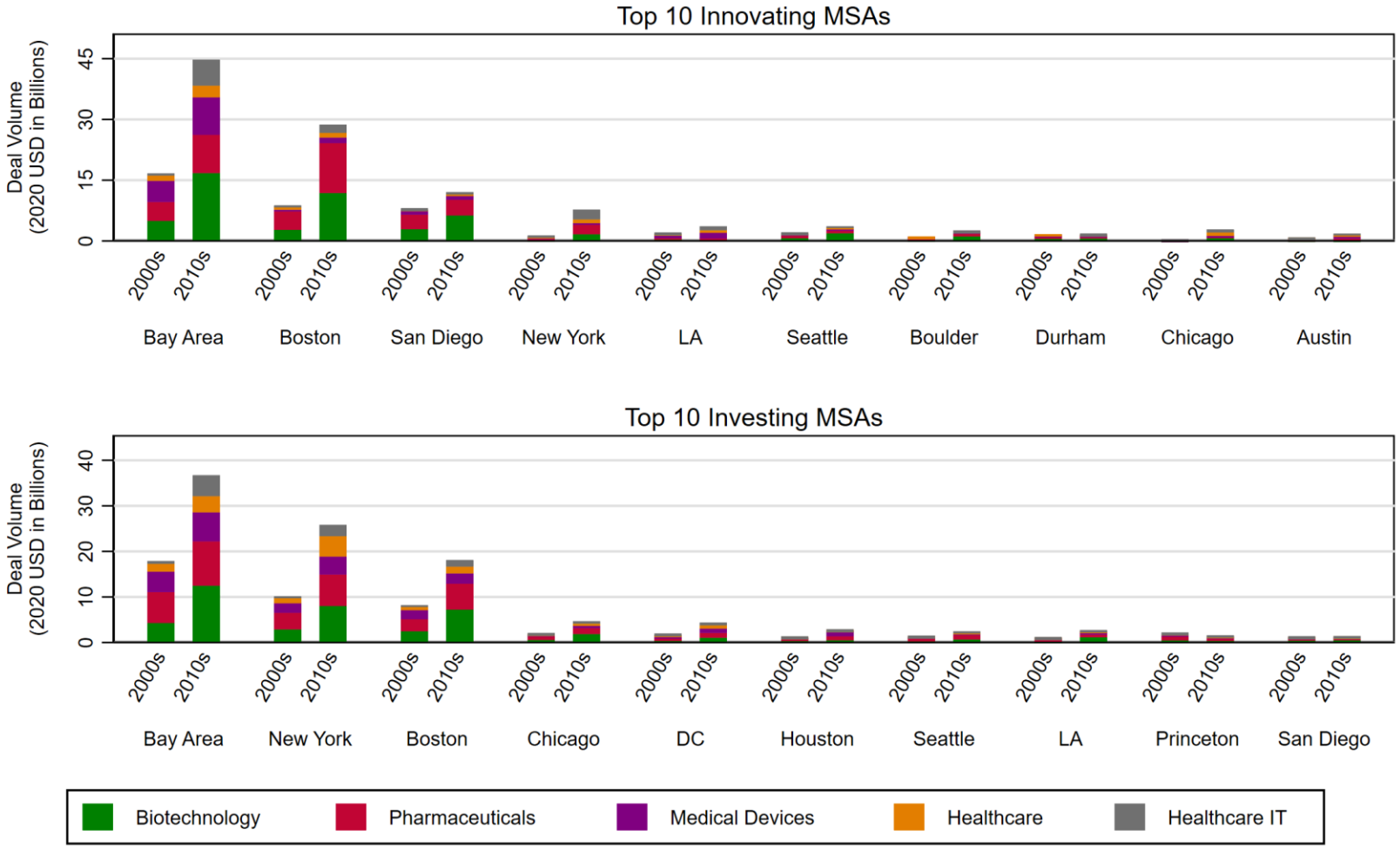


Figure 5. Geographic Dispersion and Industry Specialization of MSAs

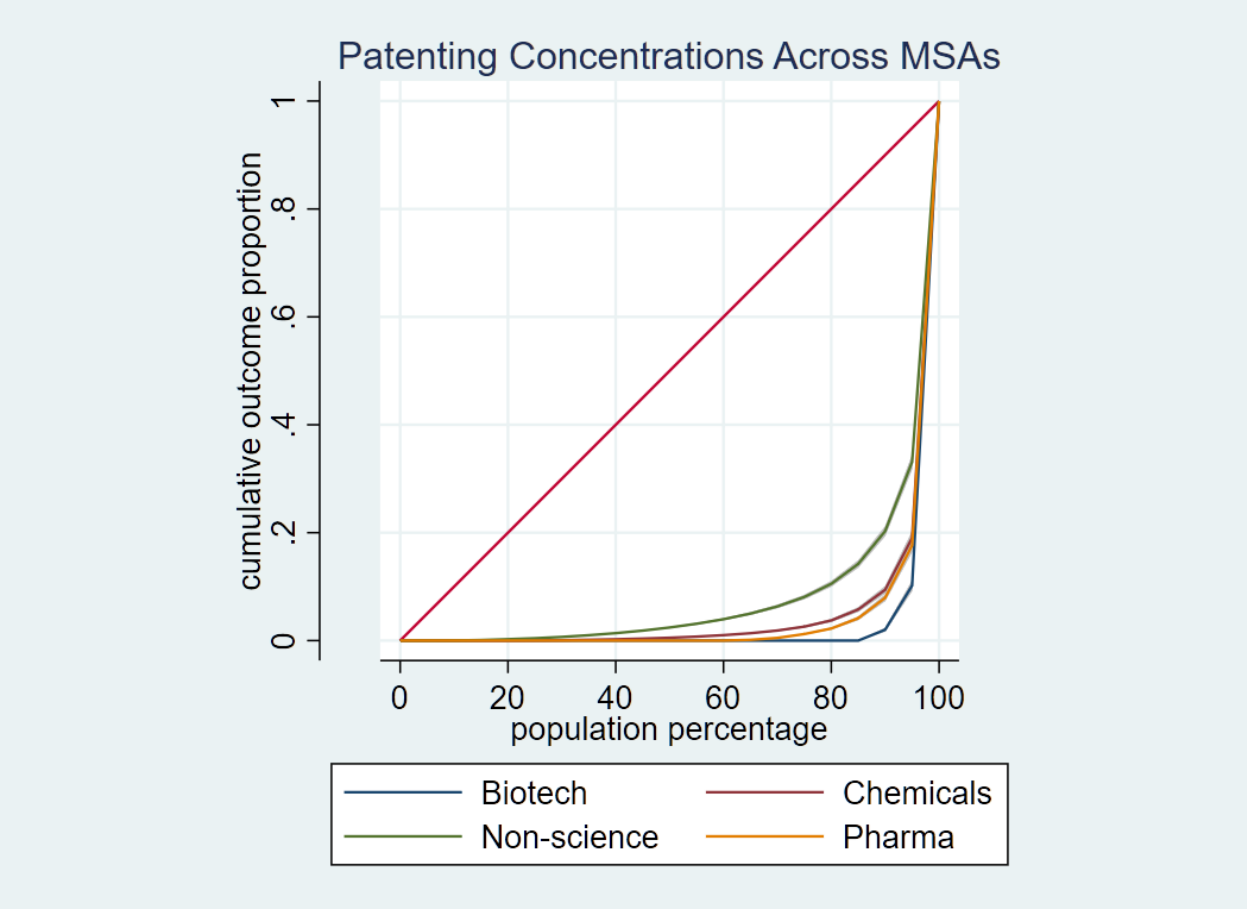


Figure 6: Geographic concentration of patents by industry

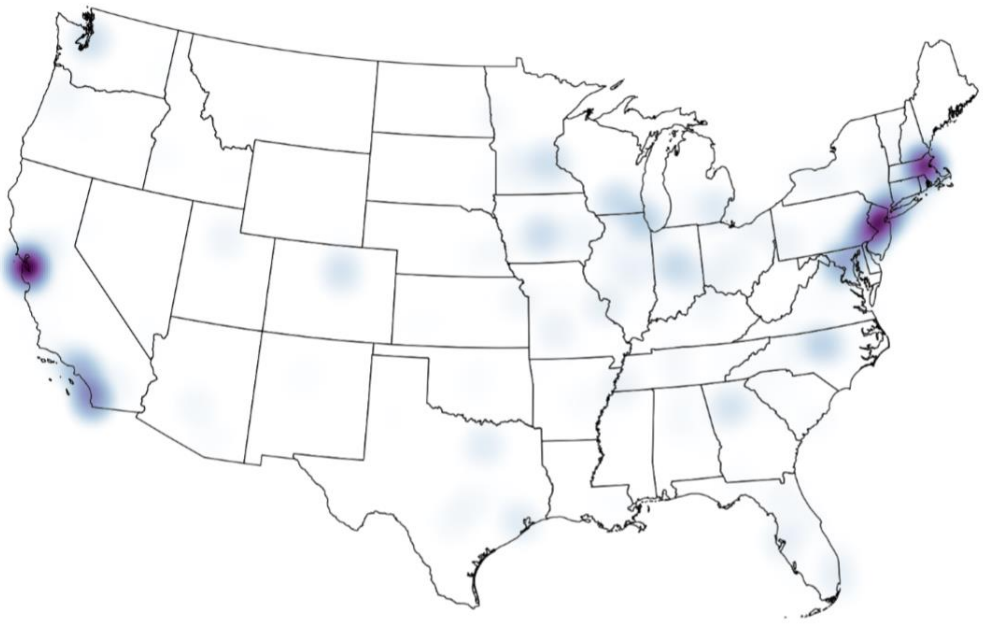


Figure 7: Heat-map of bio-pharma patenting

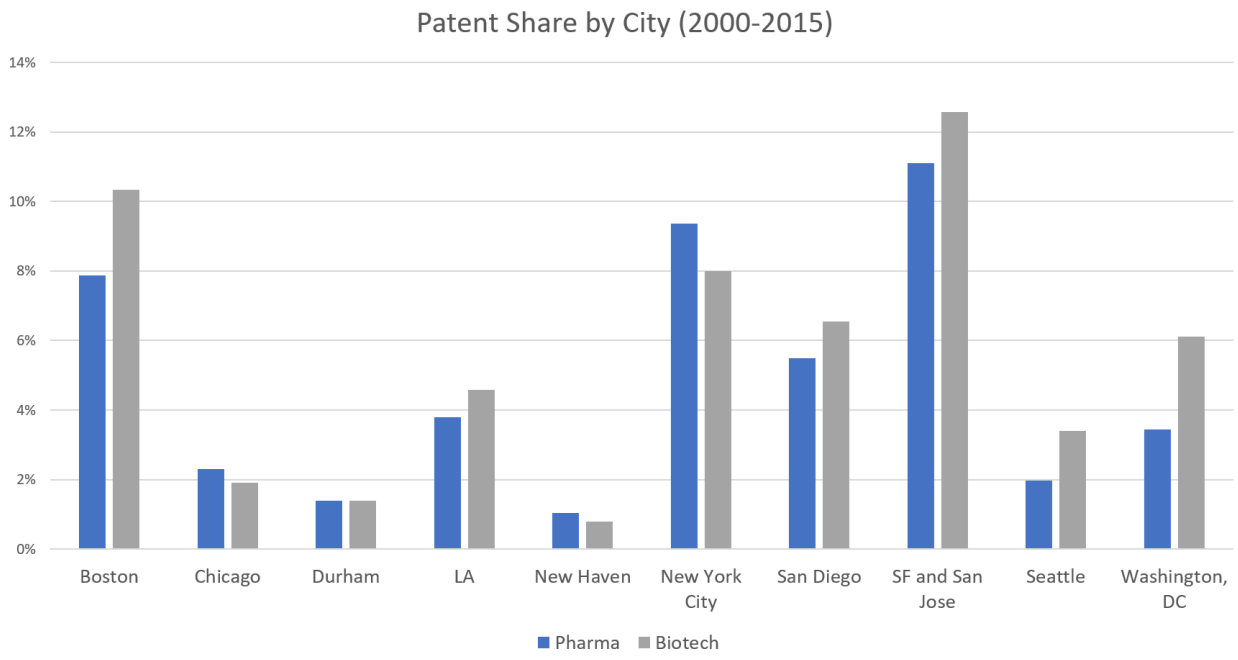
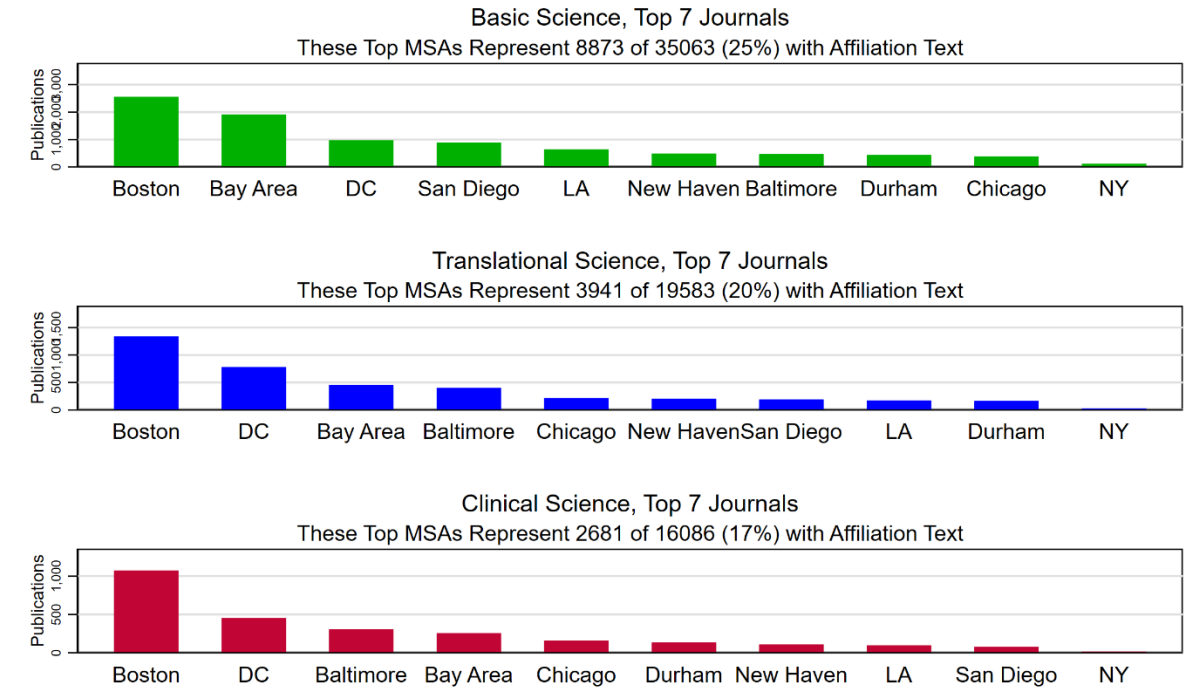


Figure 8: Share of Pharma and Biotech patents in top-10 patenting Metropolitical Statistical Area (MSAs)



Top 7 Journals: BMJ, Cell, JAMA, Lancet, Nature, NEJM, Science

Figure 9. Publications by Stage of Science Research

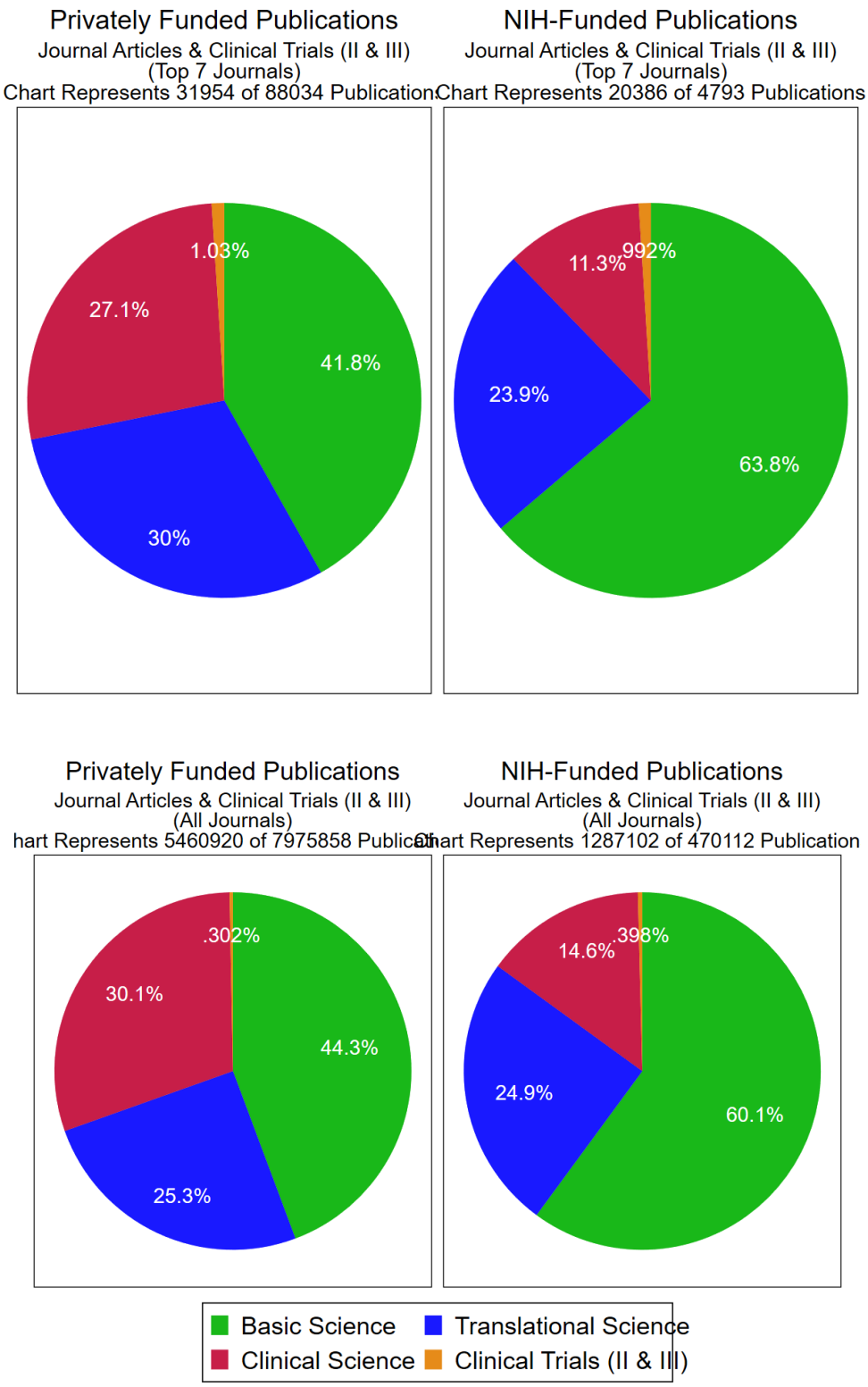


Figure 10. Science Publications by Stage and Funding

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