Knowledge Complementarity between Seed Investors and Startup Founders: Lessons from Accelerators

Simone Santamaria
National University Singapore

Stefano Breschi
Università Bocconi

Abstract

Which entrepreneurs benefit most from accelerator programs? In this paper, we argue that the value-added by startup accelerators depends critically on whether the knowledge these supporting institutions provide is complementary to or a substitute for that of the founding entrepreneurs themselves. As the main value-adding activity of accelerators is the paced and intense business training offered to participants, we hypothesize that such intervention is either redundant or not very beneficial if the entrepreneurial team already has a business education background. Conversely, accelerator programs are particularly efficient channels through which early-stage entrepreneurs, equipped with strong technological competence but lacking business knowledge, can access this valuable complementary resource. We test our ideas using a matched sample of 956 entrepreneurial teams, combining information from Crunchbase and LinkedIn.

Keywords: accelerators, high-tech startups, knowledge complementarity

1 INTRODUCTION

The emergence of startup accelerators as new players in the entrepreneurial ecosystem has attracted the attention of strategy and entrepreneurship scholars (Drover *et al.*, 2017; Hochberg, 2016). By offering educational and mentorship programs to startup founders, these organizations are often thought to facilitate entrepreneurial entry, growth, and success (Hochberg, 2016). While the empirical literature consistently identifies training and mentorship as accelerators' main value-added channels, there is mixed evidence on whether accelerators are systematically beneficial for startup founders (Hallen, Cohen & Bingham, 2020). The mixed empirical findings have prompted a search recently for contingencies that can help explain this observed heterogeneity (Cohen *et al.*, 2019b; Cohen, Bingham & Hallen, 2019a). In this paper, we theoretically argue and empirically document how the value-added provided by accelerators critically depends on the resource configuration of the accelerated ventures, and on how they match with the resources provided by these seed investors. More specifically, in the context of high-tech ventures, accelerator participation is a particularly efficient channel through which early-stage entrepreneurs, equipped with strong technological competences but lacking business knowledge, can access this valuable complementary resource. Conversely, the startup does not benefit much from accelerators when the entrepreneurial team already has a balanced mix of technological and business knowledge. In that case, it should prioritize raising funds from purely financial investors to speed up the go-to-market process.

Recent empirical research on startup accelerators has convincingly established the important role played by training and mentorship in affecting startup success (Gonzalez-Uribe & Leatherbee, 2018; Hallen *et al.*, 2020; Yu, 2020). However, the inconsistent results linking accelerator programs with startup success have raised some questions about the role of accelerators in the entrepreneurial ecosystem (Hallen, Bingham & Cohen, 2014; Smith & Hannigan, 2015). Specifically, there seems to be a lack of consensus in the extant literature on the type of ventures that are most likely to benefit from accelerator participation. If the benefits of accelerators' training are independent of the founders' pre-entry experience, the economic outcome of such programs would be maximized by selecting primarily entrepreneurs with greater capabilities and resources (Hallen *et al.*, 2020). But, a different view suggests that accelerator programs are most effective when providing specific complementary resources to individuals who are lacking them, and thus they should target the entrepreneurs most in need (Lyons & Zhang, 2018). Interestingly, the empirical evidence collected so far does not provide any definitive answer to this debate. Going beyond recent literature that broadly measures entrepreneurs' resources and capabilities in the form of previous entrepreneurial experience (Hallen *et al.*, 2020; Lyons & Zhang, 2018), this study explores in more depth how the knowledge possessed by startup founders complements or substitutes accelerators' training effect. Our findings show that, in the context of high-tech ventures, accelerators are effective only when the knowledge and resources they provide are complementary to those of the entrepreneurs. Building on this idea, we develop a theoretical framework on how the knowledge composition of the entrepreneurial team should determine the selection of the optimal seed investor for the startup.

We test our ideas using a sample of 956 startups. Data on founding team background were gathered by combining information from both Crunchbase and LinkedIn. In our empirical design, we use a matched sample of startups with similar team characteristics, background, and human capital. We compare the performance of new ventures that raised first-round funding from purely financial investors with those that went through an acceleration program, controlling for the financial resources obtained in the first funding round. Our results show that accelerators have a strong value-added effect only for startups whose founding team has a specialized technological background and no one on the team has a business education. For other startups, accelerators do not have any significant impact on their performance. Further, we observe a clear substitution effect of accelerator training on teams with at least one member with a business background. We complement this analysis with a qualitative survey targeting a subsample of accelerated companies and find consistent evidence of the theorized mechanism.

Our study contributes to the ongoing research on entrepreneurial skills (Chatterji *et al.*, 2019; Colombo & Grilli, 2005; Hsieh, Parker & van Praag, 2017; Lazear, 2004; Lechmann & Schnabel, 2014) and startup accelerators (Hallen *et al.*, 2014, 2020; Lyons & Zhang, 2018; Smith & Hannigan, 2015; Yu, 2020), demonstrating the importance of considering resource complementarity between different entrepreneurial actors (Cassiman & Valentini, 2016; Makri, Hitt & Lane, 2009). Our results suggest that accelerators provide important complementary knowledge to a specific type of entrepreneur—one with strong technological ideas but lacking the business knowledge to implement them. This type of knowledge is mostly redundant and marginally beneficial for entrepreneurs already possessing it via a business education. The findings of this paper suggest that accelerators act mostly as providers of complementary (business) resources rather than launching pads for the best ventures. Finally, our empirical approach has some advantages compared to other similar studies. First, our detailed data about founders' background allow us to explore in more depth whether the type of knowledge possessed by startup founders complements or substitutes accelerators' training. Second, we look at a large sample of startups and accelerators, instead of focusing solely on one or two programs (Lyons & Zhang, 2018). Comparing accelerators with competing seed investors also helps us develop a more comprehensive framework on how early-stage entrepreneurs should select their seed investors, highlighting the key trade-offs between different options.

2 LITERATURE AND THEORY

2.1 Startup Accelerators

Recent years have seen the emergence of new players in the entrepreneurial ecosystem that have dramatically changed how people become entrepreneurs and start their businesses, especially in the high-tech sector. Crowdfunding platforms, co-working spaces, corporate venture capital programs, hackathons, seed incubators, and accelerators (Drover *et al.*, 2017) are all altering the entrepreneurial landscape. Among the most important actors in this new ecosystem, startup accelerators gained prominence among policy-makers, practitioners, and academic scholars. Today, the number of accelerators is estimated to be more than 3,000, and growing rapidly (Cohen, 2013; Cohen & Hochberg, 2014). The attractiveness of accelerators for startups is not so much related to the financial resources they bring, which are usually very limited, as their value-added activities like mentorship and training programs. The focus on learning and knowledge acquisition is what distinguishes accelerators from other seed investors in the entrepreneurial ecosystem. Techstars and Y Combinator, two of the most prominent accelerators, describe themselves as startup schools providing "unfettered access to several mentors to help guide people through the strategy, implementation, funding, marketing and legal obstacles every startup faces" (Hallen *et al.*, 2020). While some angel investors and venture capitalists (VCs) may provide similar resources, they usually lack the formal structure and planned activities common in accelerators (Huang & Pearce, 2015).

Early academic studies on accelerators focused on the identification and quantification of accelerators' value-added. Hallen *et al.* (2014) tested the efficacy of top programs (Y Combinator and Techstars) in accelerating ventures. They found a positive effect in shortening the time necessary to reach key milestones, and increased fundraising and web traffic for top accelerators, but not for less popular ones. Smith and Hannigan (2015) found that participation in an accelerator program increases the likelihood of VC financing or exit either by acquisition or quitting. More recent studies shifted the focus to understanding how accelerators add value to startups. As expected, most of these studies point to training and mentorship as accelerators' most valuable activities. Hallen *et al.* (2020) were able to identify learning as the key channel through which accelerators benefit startups, ruling out other possible mechanisms like signaling (Spence, 1973), drawing on a mixed empirical methods approach on a sample of ventures accepted and "almost accepted" to a set of top accelerators. Specifically, the authors labelled the inter-organizational learning that takes place in accelerators as broad, intensive, and paced (BIP) consultation. This training has a relevant impact on accelerated ventures and can shape their early strategic decisions (Cohen *et al.*, 2019a). These results are also supported by Gonzalez-Uribe & Leatherbee (2018), who investigated a similar question in the context of Start-Up Chile, an ecosystem accelerator. Using a regression discontinuity design, they discovered that training and mentoring bundled with basic services like funding and co-working space can significantly increase new venture performance. In contrast, they find no evidence that basic services alone affect startup performance. Finally, Yu (2020) provides interesting perspective on how the entrepreneurial training provided by accelerators benefits startup founders. Using a matched sample of accelerator and non-accelerator startups, she discovered that the main benefit of accelerators is the speed at which the uncertainty around a startup business idea is resolved, helping entrepreneurs reject bad ideas early on. Thus, through accelerator feedback effects, accelerated companies close down earlier and more often and raise less money conditional on closing, but appear to be more efficient investments compared with non-accelerated companies.

In summary, while the empirical literature consistently identified training and mentorship as the main value-added channels of accelerators, there is mixed evidence on whether accelerators' training is systematically beneficial for startups. These mixed findings prompted a search for contingencies to help explain this observed heterogeneity. Cohen *et al.* (2019b) focus their attention on the design choices of accelerator programs. While accelerators have core defining features, there is also significant variation among them. They document descriptive correlations between some design elements like the type of program sponsors or the training organization and the performance of the startups that attend these programs. Similarly, Cohen *et al.* (2019a) analyze how three key design choices made by accelerators, namely (1) whether to space out or concentrate consultations, (2) whether to foster privacy or transparency between peer ventures, and (3) whether to tailor or standardize the program, affect venture performance.

Another important set of contingencies that can help explain the observed heterogeneity in accelerators' effectiveness lie on the participants' side. Indeed, it is reasonable to hypothesize that accelerator participation does not benefit all entrepreneurs equally and thus it is important to understand the characteristics of the individuals who benefit the most. Answering this question is also important for understanding the role of accelerators in an entrepreneurial ecosystem and better informing governments and private corporations interested in using the accelerator model as a catalyst for innovation and entrepreneurship. If accelerators' training has a positive and genuine effect on individuals' entrepreneurial skills, and has the potential to increase the quality of any accelerated venture, independent of the pre-entry background of founders, the economic impact of such programs would be maximized by selecting primarily entrepreneurs with greater capabilities and resources (Hallen et al., 2020). Alternatively, accelerators might be more effective when providing specific complementary resources to individuals lacking them. In this case, such programs should be directed at entrepreneurs most in need. Both views are reasonable from a theoretical point of view and have some empirical evidence backing them. In their detailed study of eight different accelerators, Hallen et al. (2020) documented how a venture's learning is effective and largely independent of its founding team's pre-entry experience. In contrast, Lyons & Zhang (2018) were the first to introduce the idea of entrepreneurship programs as providers of complementary resources to founders who are lacking them. In their empirical study, they discovered that entrepreneurship training programs are not effective for individuals who already have resources and capabilities in entrepreneurship, as measured by prior entrepreneurial experience. Our study builds on these recent findings and contributes to solving this apparent contradiction by unpacking the type of knowledge resources startup teams are endowed with. Going beyond the effect of previous entrepreneurial experience (Hallen *et al.*, 2020; Lyons & Zhang, 2018), we explore in more depth how the type of knowledge possessed by startup founders complements or substitutes accelerators' training.

2.3 The Optimal Resource Configuration for High-Tech Startups

A long-standing theory in entrepreneurship research suggests that entrepreneurs should have a balanced set of skills in order to perform well (Hsieh *et al.*, 2017; Lazear, 2004; Lechmann & Schnabel, 2014). Being multi-skilled is rewarded because starting a new company requires the combination of many different resources, such as physical and financial capital, people, and ideas. Founders are required to have deep knowledge of an advanced technical field as well as knowledge of how to manage people, raise financing, build a network of suppliers, and other business-related tasks (Colombo & Grilli, 2005). The relevance of having balanced technological and business skills persists if we move the unit of analysis from the founder to the founding team (Eisenhardt & Schoonhoven, 1990; Hmieleski & Ensley, 2007).

We can easily formalize the intuition that the optimal resource configuration for high-tech entrepreneurs is a balanced mix of technological and business resources. Specifically, the value of a business idea can be represented as the product of business value and risk. According to practitioners, the two most prominent risks entrepreneurs face are market risk and technology risk (Blank, 2009). Market risk is the concern that the company will find enough customers before running out of funding. Technology risk asks if the appropriate technology is in place to bring the startup idea to market successfully. Both of these concerns are vital for an entrepreneurial venture, and both have the potential to cause a young business to fail (Blank, 2009). Thus, we can represent the expected value of an entrepreneurial idea as:

$$E(V) = V \cdot t(x) \cdot m(y) \tag{1}$$

The parameter V represents the value of the idea, while $t(.)$ and $m(.)$ are the associated probabilities of realization with $0 < t(x) < 1$ and $0 < m(y) < 1$. The first probability is an inverse measure of technological risk. The second probability is an inverse measure of market risk. Entrepreneurs have some initial technological and business resources, x and y respectively, that can reduce venture risk. Business resources y reduce market risk while technological resources x limit technology risk (Graham, 2005). Finally, it seems reasonable to assume that $t(.)$ and $m(.)$ have decreasing returns. Indeed, it is impossible to completely eliminate venture risk regardless of the founder's resources. Thus, for $\alpha, \beta < 1$, we write the risk functions as: $t(x) = x^{\alpha}$ and $m(y) = y^{\beta}$. When the parameters α and β have similar values, we can derive:

Proposition 1. *Startups founded by entrepreneurial teams with a balanced mix of technological (x) and business (y) resources outperform startups founded by teams with an unbalanced mix of them.*

The initial amount of entrepreneurs' resources depends on their education and skills, as well as on the resources provided by co-founders. Below, we argue that some seed investors in the entrepreneurial ecosystem– namely startup accelerator programs– can help founders acquire some of these "knowledge" resources.

2.4 Accelerators as Providers of Complementary Resources

We hypothesize that while people generally acquire business knowledge through formal education or experience, accelerators can provide an alternative route to it. As stated above, the main value-adding activity of accelerators is the paced and intense training provided to participants, paired with the continuous interactions with mentors, peers, and potential customers (Hallen *et al.*, 2020). Although these programs are relatively shorter and less detailed than a university program in business, they can be a reasonable substitute to formal education for entrepreneurs equipped with strong technical knowledge but lacking in business understanding.

The case of Y Combinator, one the first and most successful accelerators[1], provides a good example of this. The company played an important role in legitimizing the accelerator model by backing extremely successful companies like Dropbox, Airbnb, and Twitch. Founded in 2005 by Paul Graham and located in Silicon Valley, Y Combinator's original goal was to help technically skilled hackers coming from the best local universities start their digital businesses. In a 2007 blog post, Graham lamented that few smart hackers decided to start their own businesses or rarely succeeded at it: *"The big mystery to me is: why don't more people start startups? The great majority of programmers still go straight from college to cubicle, and stay there"* (Graham, 2007). The main reason for the low number of hackers among startup founders, he wrote, was a lack of business understanding, leading technologically competent people to focus on the wrong

[1] According to the Seed Accelerator Rankings Project, Y Combinator is the most successful startup accelerator based on various metrics of performance (http://seedrankings.com/2017-rankings.html#home).

ideas and fail frequently: *"Venture capitalists have a list of danger signs to watch out for. Near the top is the company run by techno-weenies who are obsessed with solving interesting technical problems, instead of making users happy. In a startup, you're not just trying to solve problems. You're trying to solve problems that users care about"* (Graham, 2004, p. 105). Co-founders with business training were often necessary to help hackers overcome these biases and develop a viable business model: *"One expert on 'entrepreneurship' told me that any startup had to include business people, because only they could focus on what customers wanted (…) 80% of MIT spinoffs succeed provided they have at least one management person in the team at the start. The business person represents the 'voice of the customer' and that's what keeps the engineers and product development on track"* (Graham, 2005). According to Graham, however, developing this "voice of the customer" was something hackers could do themselves if provided with the right business training: "*The hard part about figuring out what customers want is figuring out that you need to figure it out. But that's something you can learn quickly (…) A hacker who has learned what to make, and not just how to make, is extraordinarily powerful"* (Graham, 2005).

The business model, structure, and training program of Y Combinator are very similar to those of other very influential accelerators like Techstars or 500 Startups, suggesting that the above principles are commonly shared across different accelerators. The idea that accelerators can provide complementary business knowledge to entrepreneurs who lack it is also supported by the qualitative evidence collected by Hallen *et al.* (2020) in their detailed study of eight different accelerators. Specifically, one participant stated: "*We were certainly nerds that can code, but we didn't know a lot about product and customer development, and that was immensely helpful"* (Hallen *et al.*, 2020, p. 395). Some participants openly suggested that accelerators' training is mainly beneficial for people lacking a formal business education: "*[Before Accelerator H], I didn't know what a business was or how to pitch something. I didn't know any of that stuff. I* *think if I had done an MBA I wouldn't have learned as much as I learned at (accelerator)"* (Hallen *et al.*, 2020, p. 396).

2.5 Choosing the Optimal Seed Investor

We can formalize the above ideas to develop a more general framework for how early-stage entrepreneurs should select their seed investors. As stated above, some investors provide knowledge in addition to monetary resources that can be useful in the startup development process. Thus, the choice of the optimal investor depends on the value-added created by the entrepreneur-investor match. This value-added, in turn, depends on the entrepreneur's resource configuration. In contrast to previous studies, our proposed framework analyzes the decision to enroll in a startup acceleration program versus raising funds from purely financial investors as an alternative option. This issue is particularly relevant as entrepreneurs frequently compare alternative seed investors in their decision to secure first-round funding, and might not be fully aware of the different trade-offs.

Let us assume that bringing an idea to market is a costly process requiring the entrepreneur to pay a sunk cost $F > 0$. The entrepreneur has the opportunity to choose an investor to help in the launch process. The entrepreneur can go to either an accelerator or a purely financial investor to seek financial support. Purely financial investors, like wealthy individuals or business angels, can cover a fraction $a > 0$ of the sunk cost F, while accelerators only cover a fraction $b > 0$, with $b < a$. However, as documented in the previous section, accelerators provide value in the form of additional "business resources," denoted here with the variable z. Thus, entrepreneurs can effectively trade money for additional business resources, thus reducing market risk[2]. The entrepreneur opts for the accelerator if:

$$Vx^{\alpha}(y+z)^{\beta} - Vx^{\alpha}y^{\beta} > (a-b)F \tag{2}$$

Equation (2) shows that it is more convenient to choose the accelerator when the value added activity—the left side of the equation—is greater than the loss in financial resources—the right side of the equation. Note that the accelerator's value-added function is decreasing in y (*business resources*), but increasing in x *(technological resources).* Thus, we can conclude:

Proposition 2. *The entrepreneurial teams who benefit most from acceleration programs are those high in technological resources (x) but low in business resources (y).*

Proposition 3. *There is a substitution effect between the resources provided by the acceleration programs and the business resources (y) owned by the entrepreneurial team.*

In summary, our theoretical framework provides insight into how entrepreneurs should choose their seed investor. If the entrepreneurial team already has a balanced mix of technological and business resources, the startup should prioritize raising funds from purely financial investors like business angels or venture capitalists, who traditionally could provide more financial resources than accelerators, and speed up the go-to-market process. Conversely, ventures operated by entrepreneurial teams high in technological resources but lacking business resources should consider enrolling in an accelerator program.

[2] Our framework can easily be extended to a case in which the main effect of accelerators' training is not a direct reduction in market risk but simply a quality signal, as suggested by Yu (2020). Note that we abstract from the cost of capital in our stylized framework and assume it is always convenient for the entrepreneur to raise external funding, as it is a way to launch the company without risking his or her own money.

3. EMPIRICAL ANALYSIS

Our empirical analysis follows a two-pronged approach using both quantitative and qualitative data. First, we exploit a large dataset of startups and information on the educational background of their founders as a proxy for their knowledge resources (i.e., business v. technological) to arrive at an estimate of the differential impact of accelerators. Second, we use primary data collected through an anonymous survey to provide additional evidence of the theoretical mechanism. In the remainder of this section, we discuss the secondary data sources, empirical methodology, and econometric results. We then present findings from our qualitative survey. Finally, we carry out robustness checks.

3.1. Startups and Investors Data

The major source of secondary evidence we used to test our propositions is the Crunchbase (CB) data set[3]. CB collects detailed information on recently founded startups, mostly in high-tech sectors, and their investors. It has rapidly become a point of reference for professionals seeking to invest in new ventures. This data source offers several advantages over more commonly used alternatives, such as VentureXpert, and it is increasingly used in academic research (Lyons & Zhang, 2018; Ter Wal *et al.*, 2016; Yu, 2020). First, CB coverage is sufficiently exhaustive across major developed and emerging economies, and it has more early-stage transaction data compared to other similar databases. Second, existing company accounts are generally not cancelled, which reduces problems related to survival bias. Third, CB also contains granular information on the name, gender, job title, education, and employment history of a large number

[3] Data have been obtained through the academic Crunchbase API, https://data.crunchbase.com/docs.

of startup founders. Moreover, this information can be further complemented through public LinkedIn profiles, which are available for a significant number of founders.

From CB, we extracted all startups founded after January 1, 2004[4]. Within this subset, we identified all ventures that received their first investment round from an accelerator[5]. This sample represents our *Accelerator* group. Following the same methodology, we selected a sample of control startups, i.e. ventures similar in many respects to the ones that received the treatment (acceleration), but that experienced a "purely financial" early-stage investor like micro-venture capital, business angel, or wealthy individual. We describe in detail the matching procedure in Section 3.6 below. This sample represents our *Control* group. Because our research question focuses on early-stage startups and investors, we limit our analysis to those investors that provided "seed funding" defined as US$150,000 or less in the first funding round. This threshold is selected to match the average amount of money invested by accelerators with that of other "purely financial" early-stage investors[6]. It is worth noting that our results are robust to using different thresholds to define "seed funding"[7].

3.2. Founding Team Data

We combined the biographical information on startup founders reported in CB with the curricula vitae obtained from public LinkedIn profiles. We provide in the Online Appendix a detailed description of the methodology used to detect startup founders and to code their biographical

[4] The first known accelerator, Y Combinator, was launched in 2005 and was quickly followed by TechStars in 2006.
[5] CB classifies investors into 22 different *types*, e.g. venture capital, corporate venture capital, funding platform, and so on. Our initial sample includes startups funded by investors classified in CB either as accelerators or as incubators because practitioners often use the two labels interchangeably. Moreover, due to the crowdsourced nature of CB, some investors that label themselves accelerators would be considered incubators by scholars, while others that refer to themselves as incubators would be labelled as accelerators (Cohen, 2013).
[6] According to our raw data, accelerators disbursed on average approximately US$40,000 from 2008 to 2013. After that year, however, the average amount of seed funding provided by accelerators increased considerably, reaching nearly US$90,000 in 2018. More information about this trend is available in the Online Appendix.
[7] Results are robust with US$50,000, US$100,000, and US$200,000 as thresholds.

information, specifically their educational background. As we discuss there, augmenting information on founders reported in CB with the one available from LinkedIn leads to almost complete coverage of the number of startup founders (Retterath & Braun, 2020)[8]. It is worth highlighting that the pre-matching average number of founders per startup retrieved by our methodology (i.e., 2.1, Table A1, Appendix 1) is similar to the one reported by Hallen *et al.* (2020), which is based on proprietary, highly confidential primary data provided directly by a set of accelerators. For each startup founder, we classified educational attainment and we coded as STEM any educational qualification belonging to any of three fields: (i) natural sciences, mathematics, and statistics, (ii) information and communication technologies, and (iii) engineering, manufacturing, and construction. Similarly, we coded as business any educational qualification belonging to the field of business administration and law. We combined different text-analysis and machine learning techniques, in addition to human judgement, to classify the self-reported degrees on CB and LinkedIn into the aforementioned categories. A detailed description of this process is available in the Online Appendix.

3.3. Dependent Variables and Econometric Specifications

Given the absence of sales and employment data, we use two alternative measures that have been widely adopted in the literature as proxies for the successful performance of new ventures (e.g. Gompers, 1995). First, we use a continuous variable represented by the (logged) total amount (in USD) raised by the company in all funding rounds (*Total funding amount*). Second, we use a discrete variable capturing whether the company collected enough funding to be in the top 50%,

[8] To identify founders we relied exclusively on information from CB and LinkedIn. For a variety of reasons, we decided not to use startup websites to collect this information. First, this approach is prone to selection bias as the websites of failed startups are often shut down. Second, startups may decide to drop the name of some founders (or co-founders) from the website due to conflicts or for other strategic reasons. As we detail in the Online Appendix, we observed several such cases in our review of startup websites.

25%, or 10% of the distribution in terms of total funding (*Top 50%, Top 25%, Top 10%*)[9]. We estimated linear regression models (OLS) for both dependent variables to simplify the interpretation of coefficients. Results of the limited dependent variable regressions are similar with a Logit or Probit specification. It is worth noting that our empirical design is unaffected by survival bias as our sample still contains failed companies[10].

3.4. Independent Variables

Consistent with our theoretical framework, we have two main independent variables of interest. The first relates to the investor type in the first funding round. This is a dummy variable (*Accelerator*) that takes the value one if the company went through an accelerator, and zero otherwise. As noted above, our control group consists of companies that raised first-round funding from purely financial investors like micro-venture capital, business angels, or wealthy individuals. Indeed, the goal of our paper is to develop a framework for how early-stage entrepreneurs should select their seed investors, highlighting the key trade-offs between different options. The second variable relates instead to the type of knowledge resources (i.e., business v. technological) possessed by the entrepreneurial teams. We use the educational background of the company founders as the best available proxy. In the context of our research, educational background has several advantages over other measures like job experience. First, formal education is the most often used measure of human capital as it molds individuals' problem-solving skills and how they approach creative thinking (Mumford *et al.*, 2010). Accordingly, how

[9] Another measure often used is the occurrence of a successful exit either through IPO or acquisition. In this regard, it is worth noting that our sample includes mostly startups founded in recent years. The median foundation date is 2012. Most of the firms in our sample simply did not have time to reach a maturity stage to be acquired or to go public. For these reasons, we opted not to include this measure in our context.

[10] We decided not to use survival as a performance measure. Paradoxically, low-growth startups tend to survive longer and have a lower probability of failure than high-growth startups (Arora & Nandkumar, 2011).

founders think of and execute their business plan is strongly affected by their educational background[11]. Second, while it is relatively uncontroversial to classify the education of people as either technological or business, it is much less so for job experience. This is often self-reported and hard to classify into categories such as technology or business without many simplifying and arbitrary assumptions. Finally, while our regressions include controls for prior entrepreneurial experience, we believe that such a measure is not a complete substitute for business education[12]. Indeed, to the extent that founders managed their past ventures through the *lenses* of their inherited cognitive skills, previous entrepreneurship experience does not necessarily lead to a change in an individual's set of competences.

Leveraging information on the educational background of founders, we constructed binary variables capturing the characteristics of the founding team, consolidated as follows. Ventures that have either a mix of business and STEM founders or have founders with a mixed background were labelled as mixed tech-business (*Mixed tech-business*). Ventures in which founders have only STEM degrees or STEM degrees together with other degrees that are not related to business were labelled as pure tech companies (*Pure tech*). Finally, a residual category comprises the remaining ventures (*Other education*). This latter category includes ventures in which all people have a business background or/and other degrees that are not related to business or STEM. The three categories *Mixed tech-business*, *Pure tech*, and *Other education* are mutually exclusive and exhaustive, i.e. each venture in our sample falls in one and only one category. Finally, we constructed another variable, which takes value one for ventures with at least one team member

[11] This effect is amplified in our context as the average number of years between the time at which founders received their degree and the time at which they founded the venture is only 5.4 years.
[12] In this regard, some studies have looked at whether one or more founders had previous entrepreneurial experience as a way to measure the ex-ante resources and capabilities in entrepreneurship (Hallen *et al.*, 2020; Lyons & Zhang, 2018).

with a business background (*One business*), and zero otherwise. Note that this latter variable can include ventures from both *Mixed tech-business* and *Other education* groups.

3.5. Control Variables

In addition to the above main variables, we control for a number of additional covariates. First, for each venture, we recorded the year of foundation (*Foundation year*) and the amount of funding raised in the first funding deal (*Amount first round*). Second, we computed the number of founders at the company foundation (*Founding team size*). From the gender of founders, we also created a variable capturing the share of women within the founding team (*Share females*). In addition to the type of educational background, we used CB information together with public LinkedIn profiles to collect data about the educational attainment of our founders as indicators of their human capital. We created three binary variables capturing whether the venture has, respectively, at least one team member with a PhD (*PHD*), Master in Business Administration (*MBA*), or Master of Science (*MSC*). We also have a dummy variable identifying ventures started by a founder with academic experience (*Academic Founder*) or a founder who graduated from a top university (*Top university*). Regarding the former variable, we checked whether prior to founding the focal venture any of the founders reported employment at a university as a professor, research assistant, lecturer, or postdoc. To measure the prestige of the universities from which founders received their degrees, we exploited the QS World University Rankings (2012)[13]. *Top university* takes the value one if a founder received a degree from one of the top 50 universities according to the QS ranking, and zero otherwise. Finally, for each founder we coded whether prior to starting the enterprise they had already founded other startups. On the basis of

[13] https://www.topuniversities.com/. This is one of the most widely used university rankings in the world.

this information, we created two binary variables capturing the presence on the founding team of at least one founder with experience in a startup founding prior to the focal one (*Serial founder*), and at least one founder with self-employment experience before the focal startup (*Self-employee founder*)[14]. To account for the fact that founders may acquire skills and knowledge of how to manage a startup via direct experience or by working with specific employers, we followed Hallen *et al.* (2020) and created two other variables: *Work experience* and *Top employer.* The former captures the amount of work experience and is measured as the average number of years between the time at which founders received their degree (excluding MBA, which is typically attained after entry into the labor market) and the time at which they founded the venture[15]. The latter captures the prominence of the prior employer in terms of spawned startups. *Top employer* takes the value one if any founder worked for the top 50 employers before starting their own venture, and zero otherwise[16]. We used industry category tags provided by CB to classify our startups' activity sector. Because companies can simultaneously have multiple tags, we performed a Latent Class Analysis (LCA) to assign each startup to a unique sector (*Sector fixed effects*). The methodology is detailed in the Online Appendix. Finally, we rely on CB data to identify the geographical location of our companies. As the startups in the sample are scattered across a large number of countries, we grouped them into eleven broader geographical areas

---

[14] Regarding the variable *Serial founder*, we checked whether prior to founding the focal venture any of the founders reported employment with a job title as founder, co-founder, or entrepreneur. The variable *Self-employee* founder was built in a similar way, by looking at whether prior to founding the focal venture any of the founders reported employment as self-employee, freelance, or independent employee.

[15] For the few founders that did not graduate, we followed again Hallen *et al.* (2020) and took the year they started full-time employment or, otherwise, the year they turned 22.

[16] As expected, the list of employers spawning more startups include well-known corporations, such as Microsoft, Google, Apple, and so on.

characterized by relative economic and cultural homogeneity (*Geo area fixed effects*). For the United States, we grouped startups at the state level[17].

3.6. Matching

Because the selection of the investor type (e.g. the decision to apply to an accelerator) does not occur randomly, startups that participate in accelerators are likely to be different on average from those that do not participate. To address this potential selection bias, we used matching techniques that paired each accelerator startup with the most similar startups that do not receive treatment based on pre-accelerator characteristics. This approach relies on the assumption that startups that are similar in their observable characteristics are also on average similar in their unobservable characteristics, including—crucially—those driving the selection into the accelerator. Although this assumption cannot be formally tested, we attempted to mitigate this issue by matching startups along a rich set of pre-treatment characteristics. To obtain a comparable sample of ventures, we used the Coarsened Exact Matching (CEM) approach (Conti & Graham, 2020; Iacus, King & Porro, 2011). This process guarantees a good degree of covariate balance while keeping the sample size large enough[18]. In our specific case, we matched our startups based on foundation year, location, sector, and all our variables measuring founding team human capital and educational background[19]. Table 1 reports the descriptive statistics for the sample of companies resulting from the matching[20]. In the last column of the table, we report the

[17] It is worth reporting that our results are the same if we use the CB category tags to identify sectors without any further modification. Similarly, the level of aggregation of the geographical variable is selected to ease reporting but does not meaningfully affect our main findings.

[18] The advantage of CEM is that rather than matching observations on specific covariate values, it coarsens the support of the joint distribution of the covariates into a finite number of strata, and then matches a treated observation if and only if a control observation can be found in the same stratum.

[19] Because CEM relies on arbitrary cut-off points to balance continuous variables, we use primarily categorical variables in the matching procedure. We control for the continuous variables (e.g. initial funding) in the main regression.

[20] CEM weights are not included in the table. Results are similar if CEM weights are included.

standardized mean difference for each variable[21]. A common rule of thumb is to consider the two groups balanced if the standardized mean difference is below 0.10 (Linden & Samuels, 2013). The descriptive statistics of the pre-matching sample are available in Appendix 1 (Table A1). Table 2 reports the distribution of startups by first-round investor type, while Tables 3 and 4 report, respectively, the sector and geographical distribution of ventures, separately for the *Accelerator* and *Control* groups. Appendix 1 (Table A2) reports the distribution across startups in U.S. states.

--- Insert Tables 1, 2, 3 and 4 Here ---

The CEM matching produced good results in balancing the team-level variables between groups. We observe no substantial differences (i.e., standardized mean difference below 0.1) in terms of *Founding team size,* educational attainment (*MSC, PHD, MBA,* and *Academic founder*), past entrepreneurial experience (*Self-employee founder* and *Serial founder*), years of work experience (*Work experience*), or gender composition (*Share females*). The two groups have the same share of founders who graduated from a top university (*Top university*) or worked for a prominent company (*Top employer*). Finally, the shares of *Pure tech*, *Mixed tech-business*, and *Other education* ventures are evenly distributed between the two groups. There is a slightly larger share of *Pure tech* ventures and a slightly lower share of *One business* ventures in the *Accelerator* group. This small imbalance is consistent with our theoretical framework.

There are, however, differences between the two groups in terms of some venture-level covariates. Specifically, we observe a difference between the two groups in the average amount of funding received in the first round, US$48,300 for the *Accelerator* group versus US$68,200

[21] The standardized mean difference for a given variable j is defined as $SMD_j = \left(\bar{X}_{jT} - \bar{X}_{jC}\right) / \sqrt{{SD_{jT}^2 + SD_{jC}^2} / 2}$, where $\bar{X}_{jT}$ and SD_{jT} are, respectively, the mean and the standard deviation of the covariate in the treatment group.

for the *Control* group. Consistent with the findings of previous studies (Yu, 2020) and our theoretical framework, ventures going through an acceleration program generally receive a smaller amount of financing compared to startups selected by angel investors or other seed investors, even when explicitly selecting only "seed investments" below US$150,000. As we detail in the theory section, this does not necessarily reflect a different unobservable quality between the two types of companies. Rather, it is a consequence of how the intervention of accelerators is designed, i.e. access to training and mentorship in exchange for a smaller amount of financial resources. Even though the difference between groups is quite small in economic terms, the imbalance forces us to control for the amount provided by the seed investor in our regressions. Nonetheless, to address any potential bias resulting from unobserved quality differences between the *Accelerator* and *Control* groups, in the Robustness Checks section we replicate our results using an alternative matching approach relying on the exact amount of money raised in the first round as a proxy for company "quality". Our results are consistent in both approaches.

Looking at Tables 3 and 4, we can see that our samples are quite balanced in terms of sectorial and geographical distribution of startups. Accelerated startups are, however, slightly overrepresented in the U.S. and South America. Conversely, Control startups are overrepresented in Western Europe. We believe this small imbalance is partially explained by the presence of important accelerators like Y Combinator, 500 Startups, TechStars, or Startup Chile, on the American continent. In contrast, accelerators are relatively rarer in the European startup ecosystem, which still relies on more traditional forms of financing. Just to be on the safe side, we control for sector and geographical region fixed effects in our regressions.

3.7. Regression Results

We begin by investigating the interaction between *Mixed tech-business* startups and Accelerator support. *Pure tech* and *Other education* ventures act as a baseline. Table 5 reports our main results. Columns (1), (2), and (3) display the results when the dependent variable is *Total Funding* and progressively add the control variables. Columns (4), (5), and (6) display the results when the dependent variables are discrete percentiles of *Total Funding*.

--- Insert Table 5 Here ---

The results in the table support Proposition 1. A balanced mix of technology and business competences (*Mixed tech-business*) is the best resource configuration for high-tech entrepreneurs. Estimates in columns (1), (2), and (3) show that having a team of both STEM and business people increases overall funding. According to model (3), having a team of both STEM and business people increases total funding by 58% (SE = 0.22, p = .08). Similarly, estimates in columns (4), (5), and (6) show that having a team of both STEM and business people increases the probability of being in the top of the total funding distribution. Specifically, *Mixed Tech-Business* startups are 20 percentage points more likely to be in the top 50% (SE = 0.06, p = .01), 13 percentage points more likely to be in the top 25% (SE = 0.05, p = .07), and 4 percentage points more likely to be in the top 10% of the funding distribution (SE = 0.03, p = .22).

The *Accelerator* coefficient is generally negative or not statistically significant. The most relevant coefficient in the regression tables is, however, the interaction between *Accelerator* and *Mixed tech-business*. The coefficient is negative and statistically significant in all the specifications. According to the model with all controls, *Mixed tech-business* startups that join an accelerator raise 80% less funding (SE = 0.30, p = .07) than similar startups that engaged with purely financial investors. We obtain similar results when we change the dependent variable in columns (4), (5), and (6). Specifically, *Mixed tech-business* startups that join an accelerator are 26 percentage points less likely to be in the top 50% (SE = 0.08, *p* = .02), 15 percentage points less likely to be in the top 25% (SE = 0.07, *p* = .02), and 10 percentage points less likely to be in the top 10% (SE = 0.04, *p* = .02). These results provide initial support for Proposition 2.

We now investigate the interaction between *Pure tech* startups and Accelerator support. *Mixed tech-business* and *Other education* ventures act now as a baseline. Table 6 reports our main results.

--- Insert Table 6 Here ---

The results in Table 6 are the other side of the coin of Table 5. All models clearly indicate that *Pure tech* teams systematically underperform. Estimates in columns (1), (2), and (3) show that the lack of business people on a team of tech individuals reduces the startup's overall funding. According to model (3) *Pure tech* teams raise 43% less funding (SE = 0.17, *p* = .01). Similarly, estimates in columns (4) and (5) show that having a team of *Pure tech* people reduces the probability of being in the top 50% or 25% of the total funding distribution. Specifically, *Pure tech* startups are 12 percentage points less likely to be in the top 50% (SE = 0.05, *p* = .01) and 7 percentage points less likely to be in the top 25% (SE = 0.04, *p* = .07). Results are not statistically significant at the top 10% of the funding distribution.

Similar to the previous table, the *Accelerator* baseline coefficient is generally negative or not statistically significant. However, our estimates show a statistically significant and economically large positive impact of accelerators on *Pure tech* ventures. According to the model with all the controls, *Pure tech* startups that join an accelerator raise 20% more funding than similar startups that engaged with purely financial investors[22]. Interestingly, going through an early-stage

[22] The accelerator net effect for *Pure tech* ventures is -0.21 (*Accelerator* coefficient, SE = 0.14, *p* = .14) + 0.41 (interaction, SE = 0.14, *p* = .14) = 0.20.

accelerator program helps *Pure tech* ventures filling their performance gap with respect to other ventures[23] (*Mixed tech-business* and *Other education)*. Results are similar but slightly less significant if we change the dependent variable. Specifically, *Pure tech* startups that join an accelerator are 15 percentage points more likely to be in the top 50% (SE = 0.06, p = .02) and 6 percentage points more likely to be in the top 25% (SE = 0.05, p = 0.21). Results are not statistically significant at the top 10% of the funding distribution.

Finally, in the last table we test more explicitly whether accelerator programs complement or substitute the business knowledge owned by the entrepreneurial team. Thus, we use the variable *One business* as the main independent variable. The baseline group consists of ventures with teams lacking any business background. Our main results are reported in Table 7.

--- Insert Table 7 Here ---

According to the estimates in columns (1), (2), and (3), having a founder with business knowledge on the team increases startup performance by between 50% to 60%. We find a similar relationship when we change the dependent variable in columns (4), (5), and (6). Also in this case, the empirical results show a baseline *Accelerator* effect that is not statistically significant. However, looking at the interaction, we see a strong substitution effect between *Accelerator* and having a person with a business background (*One business*). Specifically, the interaction coefficient is -0.52 (SE = 0.23, p = .02) in model 3, -0.15 (SE = 0.06, p = .02) in model 4, and -0.09 (SE = 0.05, p = .08) in model 5. Results are not statistically significant at the top 10% of the funding distribution.

In summary, the combined results of our regressions provide strong empirical support for our propositions. The best resource configuration for entrepreneurs is a mixed balance of

[23] -0.43 (*Pure tech* coefficient, SE = 0.17, p = .01) + 0.41 (interaction, SE = 0.14, p = .14) = 0.02

technological and business resources (Proposition 1). Thus, entrepreneurs with only technological resources tend to underperform. However, the seed investors in the ecosystem can have a relevant impact on addressing the resource gap between entrepreneurial teams. Specifically, accelerators' training can help pure tech teams close their performance gap (Proposition 2). Such training, however, is marginally valuable or even counterproductive (a waste of time for the startup) when the entrepreneurial team is equipped with sufficient business resources and knowledge (Proposition 3). Our results are consistent across different econometric specifications except when the dependent variable is the top 10% of the funding distribution, suggesting that while our theory is good at determining the average effect, it does not accurately predict outliers.

3.8. Qualitative Evidence: Survey Data

To corroborate our findings, we performed a second analysis using data gathered through an anonymous survey. The survey was developed using Qualtrics, directed to all companies in our dataset, distributed via email, contained 12 questions in total, and took no more than three minutes to complete. The survey was intentionally short to avoid biased answers affected by interviewee fatigue and did not mention our research question. The entire survey is available in Appendix 2. Our final sample contained 236 unique responses, 56% of them from startups that are accelerated and 44% from startups (control group) that raised seed funding from purely financial investors[24]. The breakdown of control startups by type of seed investor shows that 30% were backed by VCs, 28% by business angels, 27% by individuals, and 15% by other purely financial investors. Based on these measures, the sample appears to be representative and comparable to the one used in the main analysis. There is no evidence of a response bias related

[24] Because the structure of our questionnaire does not allow respondents to differentiate their answers based on ventures, we excluded entrepreneurs managing more startups.

to investor type. At the end of the survey, we asked entrepreneurs in the control group to explain why they did not consider enrolling in an accelerator program. Reassuringly, the vast majority of responses (43%) reported an “exogenous reason”: they were not aware of an accelerator or accelerators were not available in their city. Interestingly, 30% of the cases reported a collaboration with another seed investor as the major reason. This evidence is consistent with the idea that entrepreneurs frequently compare alternative seed investors and that important trade-offs exist. Finally, 17% suggested they did not need any help from acceleration programs. Only, 10% were rejected applicants.

Following the procedure outlined in the main analysis, we identified *Pure tech* teams and *One business* teams (i.e., ventures with at least one team member with a business background). The share of *Pure tech* and *One business* teams in the survey are in line with the descriptive statistics in the pre-matched sample (see Appendix 1), suggesting that a response bias based on educational background is very unlikely. In addition to the main independent variables, we control for entrepreneur’s gender (*Female* dummy variable), *Age bracket* (<18, 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, >85), and *Education level* (1. no college education, 2. college education, 3. Bachelor or equivalent, 4. Master or equivalent, 5. PhD). The most important question in the survey is the rating of seed investor impact on startup performance using a seven-point Likert scale. The variable *Overall investor impact* reports entrepreneurs’ answers to this question. As a follow-up, we asked entrepreneurs to rate *how* the seed investor helped their startup. This question was divided into different items encompassing all the different aspects of startup launch—from *Fundraising* to *Training on business issues*. As in the previous question, entrepreneurs were asked to rate impact using a seven-point Likert scale. All the different survey items as well as summary statistics for all variables are reported in Table 8.

--- Insert Table 8 Here ---

The survey responses are consistent with the theory outlined in this study. Entrepreneurs recognize that accelerators provide more value-adding activities than purely financial seed investors like VCs or business angels (*Overall investor impact*). Specifically, accelerators offer more valuable feedback on the business model (*Business model validation*), idea (*Feedback on the idea*), customer development (*Customer development*), and pitching (*Pitching the idea).* Overall, as hypothesized, they provide superior training on business-related issues (*Training on business issues*) and mentoring (*Networking with mentors*). Conversely, VCs and business angels contribute with more financial resources (*Fundraising*). The picture becomes more interesting when we break down the impact of accelerator value-adding activities based on team background. As expected, *Pure tech* entrepreneurs report a stronger positive impact of the acceleration program on their startup (*Overall investor impact).* The main theoretical arguments of the paper find additional support if we look at how accelerators add value for entrepreneurs. As shown in Table 8, *Pure tech* entrepreneurs report similar scores to other entrepreneurs in most questions related to value-adding activities, except three—*Training on business issues* (p = .02), *Pitching the idea* (p = .06), and *Networking with mentors* (p = .08). Conversely, *we do not observe* such differences between *Pure tech* and non-*Pure tech* entrepreneurs in the control group.

We run two OLS regressions with the survey data to test the statistical significance of our key results. In Table 9, we test how *Pure tech* and *One business* evaluate the overall efficacy of accelerator programs. Thus, we focus only on accelerated companies and use *Overall investor impact* as a dependent variable. As expected, *Pure tech* entrepreneurs report a stronger positive impact of the acceleration program (0.57 Likert scale point, SE = 0.33, p = .09) while teams with at least one business founder (*One business*) report a lower impact (-0.70 Likert scale point, SE = 0.31, p = .03). In line with our theory, running the same regression in the control group of non-accelerated companies does not provide any statistically significant results. Table 10 reports regression results using *Training on business issues* as a dependent variable. The results show a strong substitution effect of acceleration training on *One business* teams (-0.56 Likert scale point, SE = 0.34, p = .10) while reporting a strong positive effect of accelerator training on *Pure tech* teams (0.70 Likert scale point, SE = 0.36, p = .06). In this case, too, running the same regression in the control group of non-accelerated companies does not provide any statistically significant result.

--- Insert Tables 9 and 10 Here ---

4. ROBUSTNESS CHECKS

In this section, we further test the robustness of our main analysis. First, we replicate our results using exact matching on the first funding amount collected by the venture as an indicator of quality. Second, we test the validity of our results focusing only on the top accelerators. Finally, we replicate our results in the pre-matched sample.

4.3. Exact Matching on First Funding Amount

Ventures going through an acceleration program generally receive a significantly smaller amount of financing compared to startups selected by angel and other institutional investors (Yu, 2020). Instead of relying on several different covariates related to the human capital of the startup team, we now adopt a different approach and match accelerated and non-accelerated startups by the exact amount of funds raised in the first round. This latter variable can indeed be considered a simple and comprehensive measure of startup “quality” or “potential”. The results, reported in Appendix 1 (Table A3), are largely consistent with the main analysis.

4.4. Top Accelerators

In the second robustness check, we replicate the results of the secondary data analysis focusing only on the top accelerators. We built the syndication network, where a pair of investors is linked by a tie if they co-invested in the same venture, and computed the eigenvector centrality index (Bonacich, 1987) to identify the most prominent accelerators. For this robustness check, we restricted our analysis to the startups (and related control ventures) that participated in the top 100 accelerators based on their eigenvector centrality scores[25]. Results are reported in Appendix 1 (Table A4). All the interaction coefficients show the expected signs. It is worth reporting that as we restrict the sample to only the top accelerators, the baseline *Accelerator* coefficient becomes larger and the interaction effects weaker. These findings suggest that startups can benefit from the very top accelerators through channels other than training and knowledge acquisition (e.g. reputation or signaling). These results are consistent with Hallen *et al.* (2014), who found a positive effect of top accelerators on startup performance, independent of teams' educational background, but no effect for less popular ones.

4.5. Pre-matched Sample

We ensure that our results are generalizable in the wider population by replicating the empirical analysis on the pre-matched sample, controlling for all the available covariates. The results, reported in Appendix 1 (Table A5), are largely consistent with the main analysis.

[25] The top accelerators in terms of eigenvector centrality score feature some of the most well-known and -reputed organizations. The top three accelerators in the ranking are 500 Startups, Y Combinator, and Techstars.

5. CONCLUSION AND DISCUSSION

Our study shows how accelerators can act as providers of complementary knowledge to specialized tech entrepreneurs who lack the necessary business knowledge to launch a company. Interestingly, our results show that the value-added activity of accelerators is negligible in the case of entrepreneurial teams already possessing these resources in the form of business education. Indeed, there is a clear substitution effect between accelerator training and business education. This study contributes to the literature on the background of successful entrepreneurs (Hsieh *et al.*, 2017; Lazear, 2004; Lechmann & Schnabel, 2014) and the literature on accelerators (Hallen *et al.*, 2014, 2020; Hataway, 2016; Smith & Hannigan, 2015; Yu, 2020), highlighting the importance of considering complementarities between entrepreneurial team resources and other supporting institutions like seed investors. Our results provide evidence consistent with the view that accelerators are beneficial for a specific category of early-stage entrepreneurs who lack business knowledge and capabilities (Lyons & Zhang, 2018). Even though these institutions play an important role in an entrepreneurial ecosystem, their training programs are far from unconditionally effective. By comparing accelerators with competing seed investors like business angels or venture capitalists we provide a more comprehensive framework for how early-stage entrepreneurs should select their seed investors, highlighting the key trade-offs between different options. It is worth noting that our theoretical framework produces good results in anticipating the average effects, but it is weaker in predicting outliers and the impact of the very top accelerators.

It is important to acknowledge the main limitations of this study. First, our empirical analysis is unable to identify strict causality because of the sorting process between entrepreneurs and seed investors. Entrepreneurs endogenously select their seed investors based on their resources.

Conversely, seed investors try to select entrepreneurs with the highest potential. Nevertheless, any sorting effect favoring or penalizing accelerators is partially accounted for by the double interaction in our models. Indeed, even assuming that startups selecting an accelerator have systematically lower/higher potential, this effect should be reflected homogenously between entrepreneurs with different backgrounds. As well, our theoretical framework explicitly models a selection effect between actors aimed at maximizing economic outcome. Such a selection effect does not contradict our empirical results. Second, our operationalization of performance is far from perfect. While common in the literature to use variables such as total funding (Gompers, 1995), we do not have information about startup revenue or revenue growth, which might be better measures for startup success. Finally, we use educational background to assess the type of knowledge/resources that our entrepreneurs have. Needless to say, this approach is imperfect as entrepreneurs can acquire knowledge/resources through alternative channels that are more difficult to observe.

References

Arora A. & Nandkumar A. (2011). Cash-out or flameout! Opportunity cost and entrepreneurial strategy: Theory, and evidence from the information security industry. *Management Science 57*(10), 1844–1860.

Blank S. (2009). Market Risk and Technology Risk. *Stanford eCorner*. Available at: https://ecorner.stanford.edu/in-brief/market-risk-and-technology-risk/ [13 January 2020].

Bonacich P. (1987). Power and Centrality: A Family of Measures. *American Journal of Sociology*. University of Chicago Press *92*(5), 1170–1182.

Cassiman B. & Valentini G. (2016). Open innovation: Are inbound and outbound knowledge flows really complementary? *Strategic Management Journal 37*(6), 1034–1046.

Chatterji A., Delecourt S., Hasan S. & Koning R. (2019). When does advice impact startup performance? *Strategic Management Journal 40*(3), 331–356.

Cohen S. L. (2013). What Do Accelerators Do? Insights from Incubators and Angels. *Innovations: Technology, Governance, Globalization 8*(3–4), 19–25.

Cohen S. L., Bingham C. B. & Hallen B. L. (2019)(a). The Role of Accelerator Designs in Mitigating Bounded Rationality in New Ventures. *Administrative Science Quarterly 64*(4), 810–854.

Cohen S. L., Fehder D. C., Hochberg Y. V. & Murray F. (2019)(b). The design of startup accelerators. *Research Policy 48*(7), 1781–1797.

Cohen S. L. & Hochberg Y. V. (2014). *Accelerating Startups: The Seed Accelerator Phenomenon. SSRN Electronic Journal*. Available at: http://dx.doi.org/10.2139/ssrn.2418000.

Colombo M. G. & Grilli L. (2005). Founders’ human capital and the growth of new technology-based firms: A competence-based view. *Research Policy 34*(6), 795–816.

Conti A. & Graham S. J. H. (2020). Valuable choices: Prominent venture capitalists’ influence on startup CEO replacements. *Management Science 66*(3), 1325–1350.

Drover W., Busenitz L., Matusik S., Townsend D., Anglin A. & Dushnitsky G. (2017). A Review and Road Map of Entrepreneurial Equity Financing Research: Venture Capital, Corporate Venture Capital, Angel Investment, Crowdfunding, and Accelerators. *Journal of Management 43*(6), 1820–1853.

Eisenhardt K. M. & Schoonhoven C. B. (1990). Organizational Growth: Linking Founding Team, Strategy, Environment, and Growth Among U.S. Semiconductor Ventures, 1978-1988. *Administrative Science Quarterly 35*(3), 504.

Gompers P. A. (1995). Optimal Investment, Monitoring, and the Staging of Venture Capital. *The Journal of Finance 50*(5), 1461.

Gonzalez-Uribe J. & Leatherbee M. (2018), April 1. The effects of business accelerators on venture performance: Evidence from start-Up Chile. *Review of Financial Studies*. Oxford University Press.

Graham P. (2004). *Hackers & Painters: Big Ideas from the Computer Age*. O’Reilly & Associates, Inc.: Sebastopol.

Graham P. (2005). Why Smart People Have Bad Ideas. Available at: http://paulgraham.com/bronze.html?viewfullsite=1 [6 May 2020].

Graham P. (2007). Why to Not Not Start a Startup. Available at: http://www.paulgraham.com/notnot.html?viewfullsite=1 [6 May 2020].

Hallen B. L., Bingham C. B. & Cohen S. L. (2014). Do accelerators accelerate? A study of venture accelerators as a path to success. In *Academy of Management Proceedings*. Academy of Management: 12955.

Hallen B. L., Cohen S. L. & Bingham C. B. (2020). Do Accelerators Work? If So, How? *Organization Science 31*(2), 378–414.

Hataway I. (2016). Accelerating growth: Startup accelerator programs in the United States. *The Brookings Institution*. Available at: https://www.brookings.edu/research/accelerating-growth-startup-accelerator-programs-in-the-united-states/ [13 January 2020].

Hmieleski K. M. & Ensley M. D. (2007). A contextual examination of new venture performance: entrepreneur leadership behavior, top management team heterogeneity, and environmental dynamism. *Journal of Organizational Behavior 28*(7), 865–889.

Hochberg Y. V. (2016). Accelerating entrepreneurs and ecosystems: The seed accelerator model. *Innovation Policy and the Economy 16*(1), 25–51.

Huang L. & Pearce J. L. (2015). Managing the Unknowable. *Administrative Science Quarterly 60*(4), 634–670.

Iacus S. M., King G. & Porro G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association 106*(493), 345–361.

Lazear E. P. (2004). Balanced skills and entrepreneurship. *The American Economic Review 94*(2), 208–211.

Lechmann D. S. J. & Schnabel C. (2014). Are the self-employed really jacks-of-all-trades? Testing the assumptions and implications of Lazear’s theory of entrepreneurship with German data. *Small Business Economics 42*(1), 59–76.

Linden A. & Samuels S. J. (2013). Using balance statistics to determine the optimal number of controls in matching studies. *Journal of Evaluation in Clinical Practice 19*(5), 968–975.

Lyons E. & Zhang L. (2018). Who does (not) benefit from entrepreneurship programs? *Strategic Management Journal 39*(1), 85–112.

Makri M., Hitt M. A. & Lane P. J. (2009). Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions. *Strategic Management Journal 31*(6), 602–628.

Mumford M. D., Antes A. L., Caughron J. J., Connelly S. & Beeler C. (2010). Cross-field differences in creative problem-solving skills: A comparison of health, biological, and social sciences. *Creativity Research Journal 22*(1), 14–26.

Retterath A. & Braun R. (2020). Benchmarking Venture Capital Databases. *SSRN Electronic Journal*. Elsevier BV. Available at: https://papers.ssrn.com/abstract=3706108.

Smith S. W. & Hannigan T. J. (2015). Swinging for the fences: How do top accelerators impact the trajectories of new ventures? *DRUID Working Paper*.

Spence M. (1973). Job Market Signaling. *The Quarterly Journal of Economics 87*(3), 355–374.

Stuetzer M., Obschonka M. & Schmitt-Rodermund E. (2013). Balanced skills among nascent entrepreneurs. *Small Business Economics 41*(1), 93–114.

Ter Wal A. L. J., Alexy O., Block J. & Sandner P. G. (2016). The Best of Both Worlds: The Benefits of Open-specialized and Closed-diverse Syndication Networks for New Ventures' Success. *Administrative Science Quarterly 61*(3), 393–432.

Yu S. (2020). How do accelerators impact the performance of high-technology ventures? *Management Science 66*(2), 530–552.

TABLE 1 Descriptive statistics, matched sample

	Accelerator group		Control group		
	Mean	SD	Mean	SD	Std. mean diff.
Venture level variables					
Total funding amount (in '000s USD)	5106.16	76432.11	1346.96	6715.57	0.06
Accelerator	1.00	0.00	0.00	0.00	-
Foundation year	2013.54	2.61	2012.91	2.88	0.23
Amount first round (log)	10.45	0.87	10.91	0.78	-0.51
Amount first round (in '000s USD)	48.34	38.59	68.19	38.47	-0.55
Team level variables					
Pure tech	0.42	0.49	0.37	0.48	0.10
Mixed tech-business	0.15	0.36	0.18	0.39	-0.09
Other education	0.43	0.50	0.45	0.50	-0.03
One business	0.31	0.46	0.37	0.48	-0.13
Founding team size	1.55	0.66	1.55	0.68	-0.01
Share females	0.03	0.16	0.04	0.16	-0.00
Serial founder	0.47	0.50	0.49	0.50	-0.05
Academic founder	0.03	0.16	0.03	0.16	-0.01
Self-employee founder	0.01	0.10	0.01	0.09	0.01
MSC	0.41	0.49	0.42	0.49	-0.03
PHD	0.07	0.25	0.07	0.25	-0.00
MBA	0.06	0.24	0.06	0.23	0.02
Work experience	5.37	4.88	5.44	4.93	-0.01
Top employer	0.02	0.15	0.03	0.16	-0.02
Top university	0.03	0.18	0.03	0.18	-0.01
Observations	505		451		

Notes: SD, standard deviation. Std. mean diff., Standardized mean difference.

TABLE 2 Distribution of startups by first round investor, matched sample

Investor type	Accelerator group	Control group
Accelerator	1	-
Venture Capital	-	0.37
Individuals	-	0.33
Micro Venture Capital	-	0.11
Government Office	-	0.07
Funding Platform	-	0.04
Angel Group	-	0.03
Co-working Space	-	0.02
University	-	0.01
Other	-	0.02
Total	1	1

Notes: The table reports the distribution of all startups in the matched sample by type of investor in the first funding round, separately for the treated and control groups. *Other* includes private equity firms, family firm offices, investment banks, fund of funds, technology transfer offices, and non-equity funding.

TABLE 3 Distribution of startups by sector of activity, matched sample

Sector	Accelerator group	Control group
Commerce	0.40	0.36
Software	0.24	0.24
Media & entertainment	0.10	0.12
Hardware	0.07	0.07
Mobile apps	0.04	0.04
Data Analytics	0.03	0.04
Fintech	0.04	0.03
Biotech	0.02	0.03
Sales & marketing	0.02	0.03
Green tech & energy	0.02	0.02
Internet services	0.01	0.01
Total	1	1

Notes: Sectors of activity are based on the industry tags reported in the CB database. See the Online Appendix for a discussion of the methodology adopted.

TABLE 4 Distribution of startups by location, matched sample

Sector	Accelerator group	Control group
Africa/Middle East	0.04	0.04
Asia	0.05	0.05
Australia and New Zealand	0.02	0.03
Canada	0.03	0.03
Eastern Europe	0.06	0.07
United Kingdom	0.07	0.07
India	0.04	0.04
Israel	0.00	0.00
South America	0.08	0.05
United States	0.25	0.21
Western (Continental) Europe	0.36	0.41
Total	1	1

Notes: For the U.S., a breakdown of startups across states is provided in Appendix 1.

TABLE 5 Impact of accelerators on Mixed tech-business ventures

	Total funding (1)	Total funding (2)	Total funding (3)	Top 50% total funding (4)	Top 25% total funding (5)	Top 10% total funding (6)
Accelerator	-0.292	0.112	0.069	0.067	-0.021	0.029
	(0.027)	(0.376)	(0.572)	(0.045)	(0.452)	(0.106)
Mixed tech-business	0.744	0.781	0.587	0.203	0.135	0.039
	(0.001)	(0.000)	(0.008)	(0.001)	(0.007)	(0.223)
Accelerator × Mixed tech-business	-0.864	-0.932	-0.805	-0.258	-0.155	-0.098
	(0.009)	(0.002)	(0.007)	(0.002)	(0.021)	(0.024)
Amount first round (log)		0.826	0.785	0.179	0.051	0.012
		(0.000)	(0.000)	(0.000)	(0.001)	(0.241)
Founding team size			0.360	0.094	0.049	0.032
			(0.000)	(0.001)	(0.032)	(0.032)
Share females			-0.390	-0.045	-0.071	-0.005
			(0.275)	(0.648)	(0.373)	(0.920)
Serial founder			0.092	-0.035	0.009	0.019
			(0.470)	(0.313)	(0.741)	(0.295)
Academic founder			-0.107	-0.047	0.044	-0.014
			(0.783)	(0.656)	(0.612)	(0.801)
Self-employee founder			1.773	0.322	0.422	0.103
			(0.001)	(0.033)	(0.001)	(0.204)
MSC			-0.187	-0.096	-0.019	0.016
			(0.191)	(0.013)	(0.549)	(0.444)
PHD			0.951	0.188	0.166	0.081
			(0.001)	(0.017)	(0.010)	(0.053)
MBA			0.176	0.096	-0.000	-0.005
			(0.495)	(0.174)	(0.999)	(0.893)
Work experience			-0.011	0.003	-0.003	-0.003
			(0.371)	(0.340)	(0.345)	(0.103)
Top employer			-0.440	-0.022	-0.049	-0.088
			(0.263)	(0.840)	(0.581)	(0.123)
Top university			0.476	0.216	0.069	-0.003
			(0.164)	(0.021)	(0.364)	(0.946)
Constant	11.894	2.863	3.482	-1.481	-0.218	-0.121
	(0.000)	(0.000)	(0.000)	(0.000)	(0.223)	(0.298)
Geo area fixed effects	No	No	Yes	Yes	Yes	Yes
Sector fixed effects	No	No	Yes	Yes	Yes	Yes
Foundation year fixed effects	No	No	Yes	Yes	Yes	Yes
Observations	956	956	956	956	956	956
R-squared	0.024	0.158	0.254	0.198	0.134	0.088

Notes: In columns (1)-(3) the dependent variable is the log of the total funding amount. *p*-values in parentheses. CEM weights are included in the regression.

TABLE 6 Impact of accelerators on Pure tech ventures

	Total funding (1)	Total funding (2)	Total funding (3)	Top 50% total funding (4)	Top 25% total funding (5)	Top 10% total funding (6)
Accelerator	-0.597	-0.233	-0.211	-0.031	-0.069	0.010
	(0.000)	(0.111)	(0.137)	(0.426)	(0.030)	(0.644)
Pure tech	-0.408	-0.282	-0.435	-0.123	-0.070	-0.016
	(0.027)	(0.102)	(0.012)	(0.009)	(0.069)	(0.516)
Accelerator × Pure tech	0.452	0.498	0.415	0.150	0.063	0.011
	(0.071)	(0.033)	(0.067)	(0.016)	(0.215)	(0.729)
Amount first round (log)		0.824	0.776	0.177	0.049	0.011
		(0.000)	(0.000)	(0.000)	(0.001)	(0.263)
Founding team size			0.377	0.101	0.055	0.029
			(0.000)	(0.000)	(0.012)	(0.045)
Share females			-0.412	-0.047	-0.075	-0.010
			(0.250)	(0.630)	(0.353)	(0.852)
Serial founder			0.114	-0.027	0.015	0.022
			(0.371)	(0.441)	(0.610)	(0.245)
Academic founder			-0.061	-0.031	0.056	-0.013
			(0.876)	(0.771)	(0.522)	(0.821)
Self-employee founder			1.808	0.336	0.427	0.111
			(0.001)	(0.027)	(0.001)	(0.170)
MSC			-0.125	-0.079	-0.005	0.017
			(0.377)	(0.043)	(0.880)	(0.423)
PHD			1.029	0.200	0.179	0.089
			(0.000)	(0.013)	(0.007)	(0.037)
MBA			0.110	0.089	-0.007	-0.014
			(0.674)	(0.214)	(0.907)	(0.720)
Work experience			-0.013	0.003	-0.003	-0.003
			(0.291)	(0.422)	(0.276)	(0.089)
Top employer			-0.405	-0.012	-0.040	-0.089
			(0.304)	(0.914)	(0.649)	(0.120)
Top university			0.486	0.218	0.070	0.000
			(0.155)	(0.020)	(0.361)	(0.997)
Constant	12.164	3.119	3.796	-1.398	-0.163	-0.099
	(0.000)	(0.000)	(0.000)	(0.000)	(0.366)	(0.400)
Geo area fixed effects	No	No	Yes	Yes	Yes	Yes
Sector fixed effects	No	No	Yes	Yes	Yes	Yes
Foundation year fixed effects	No	No	Yes	Yes	Yes	Yes
Observations	956	956	956	956	956	956
R-squared	0.019	0.150	0.252	0.194	0.130	0.083

Notes: In columns (1)-(3) the dependent variable is the log of the total funding amount. *p*-values in parentheses. CEM weights are included in the regression.

TABLE 7 Impact of accelerators on ventures with a business person on the team

	Total funding (1)	Total funding (2)	Total funding (3)	Top 50% total funding (4)	Top 25% total funding (5)	Top 10% total funding (6)
Accelerator	-0.191	0.197	0.133	0.080	-0.012	0.026
	(0.203)	(0.171)	(0.338)	(0.035)	(0.694)	(0.192)
One business	0.609	0.547	0.477	0.140	0.089	0.033
	(0.001)	(0.001)	(0.006)	(0.003)	(0.022)	(0.186)
Accelerator × One business	-0.636	-0.649	-0.520	-0.148	-0.090	-0.034
	(0.012)	(0.006)	(0.024)	(0.019)	(0.080)	(0.310)
Amount first round (log)		0.817	0.780	0.177	0.050	0.011
		(0.000)	(0.000)	(0.000)	(0.001)	(0.254)
Founding team size			0.359	0.096	0.052	0.027
			(0.000)	(0.000)	(0.020)	(0.058)
Share females			-0.395	-0.046	-0.071	-0.008
			(0.269)	(0.640)	(0.374)	(0.874)
Serial founder			0.092	-0.034	0.010	0.020
			(0.472)	(0.330)	(0.724)	(0.289)
Academic founder			-0.081	-0.037	0.052	-0.014
			(0.835)	(0.731)	(0.550)	(0.802)
Self-employee founder			1.765	0.320	0.419	0.107
			(0.001)	(0.035)	(0.001)	(0.184)
MSC			-0.198	-0.097	-0.019	0.012
			(0.166)	(0.013)	(0.562)	(0.568)
PHD			0.970	0.193	0.169	0.086
			(0.001)	(0.015)	(0.009)	(0.041)
MBA			0.095	0.074	-0.013	-0.017
			(0.722)	(0.310)	(0.832)	(0.660)
Work experience			-0.011	0.003	-0.003	-0.003
			(0.385)	(0.334)	(0.352)	(0.106)
Top employer			-0.428	-0.016	-0.044	-0.090
			(0.276)	(0.880)	(0.618)	(0.116)
Top university			0.501	0.223	0.073	0.001
			(0.142)	(0.017)	(0.341)	(0.982)
Constant	11.785	2.884	3.453	-1.486	-0.221	-0.117
	(0.000)	(0.000)	(0.000)	(0.000)	(0.216)	(0.316)
Geo area fixed effects	No	No	Yes	Yes	Yes	Yes
Sector fixed effects	No	No	Yes	Yes	Yes	Yes
Foundation year fixed effects	No	No	Yes	Yes	Yes	Yes
Observations	956	956	956	956	956	956
R-squared	0.025	0.156	0.254	0.195	0.132	0.085

Notes: In columns (1)-(3) the dependent variable is the log of the total funding amount. *p*-values in parentheses. CEM weights are included in the regression.

TABLE 8 Impact of accelerators, descriptive statistics (survey data)

	Entire Sample		Accelerator		Control	
	Accelerator	Control	Pure tech	Non-Pure tech	Pure tech	Non-Pure tech
Overall investor impact	5.15	4.30	5.43	5.01	4.30	4.29
	(1.69)	(2.08)	(1.62)	(1.76)	(1.97)	(2.14)
Fundraising	4.61	6.08	4.88	4.46	6.23	6.02
	(2.25)	(1.69)	(2.22)	(2.30)	(1.63)	(1.72)
Business model validation	4.07	3.16	4.18	4.01	3.00	3.23
	(1.89)	(1.93)	(1.85)	(1.91)	(2.21)	(1.82)
Feedback on the idea	4.15	3.13	4.38	4.03	3.19	3.10
	(1.83)	(1.92)	(1.83)	(1.83)	(2.11)	(1.86)
Customer development	3.92	2.81	4.00	3.89	2.73	2.93
	(1.90)	(1.85)	(1.86)	(1.95)	(1.90)	(1.85)
Advice on operations	3.72	3.32	3.90	3.62	2.96	3.43
	(1.84)	(1.89)	(1.71)	(1.90)	(1.92)	(1.87)
Training on business issues	3.44	2.73	3.97	3.16	2.6	2.78
	(1.86)	(1.73)	(1.60)	(1.93)	(1.80)	(1.71)
Training on technical issues	2.26	1.94	2.40	2.19	1.92	1.95
	(1.55)	(1.29)	(1.52)	(1.55)	(1.28)	(1.31)
Pitching the idea	5.21	3.07	5.59	5.01	3.00	3.10
	(1.65)	(1.83)	(1.41)	(1.77)	(1.95)	(1.79)
Technology development	2.70	2.32	2.63	2.73	1.96	2.46
	(1.54)	(1.94)	(1.27)	(1.67)	(1.31)	(1.71)
Team building	2.92	2.74	3.04	2.85	2.30	2.92
	(1.69)	(1.62)	(1.66)	(1.68)	(1.46)	(1.71)
Networking with mentors	5.19	3.00	5.54	5.00	3.23	2.90
	(1.65)	(1.93)	(1.31)	(1.80)	(2.02)	(1.90)
Access to physical space	4.60	2.25	4.75	4.60	2.07	2.32
	(2.13)	(1.78)	(1.23)	(1.11)	(2.52)	(1.89)
Pure tech	0.33	0.25	1	0	1	0
	(0.48)	(0.46)	(0)	(0)	(0)	(0)
One business	0.37	0.39	0	0.55	0	0.52
	(0.48)	(0.49)	(0)	(0.49)	(0)	(0.49)
Female	0.13	0.06	0.11	0.13	0.08	0.05
	(0.33)	(0.23)	(0.30)	(0.35)	(0.24)	(0.23)
Age bracket	3.3	3.4	3.02	3.45	3.30	3.43
	(0.87)	(0.71)	(0.85)	(0.91)	(0.74)	(0.69)
Education level	3.7	3.7	3.88	3.58	4.07	3.67
	(0.94)	(0.90)	(0.84)	(0.98)	(1.03)	(0.80)
Observations	133	103	44	89	26	77

Notes: The table reports the mean and the standard deviation (in parenthesis) of all variables. Variable scores are reported using a Likert scale from 1 to 7.

TABLE 9 Impact of accelerators on Pure Tech and One business ventures (survey data)

	Overall investor impact (Accelerator)	Overall investor impact (Accelerator)	Overall investor impact (Accelerator)	Overall investor impact (Accelerator)
Pure tech	0.420	0.573		
	(0.184)	(0.093)		
One business			-0.711	-0.685
			(0.020)	(0.031)
Female		0.018		-0.082
		(0.968)		(0.855)
Age bracket		0.127		0.071
		(0.501)		(0.696)
Education level		-0.244		-0.135
		(0.167)		(0.429)
Constant	5.012	5.418	5.425	5.685
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	129	126	129	126
R-squared	0.014	0.032	0.042	0.046

Notes: The table includes only accelerated ventures and report estimates of OLS regressions. *p*-values in parentheses.

TABLE 10 Accelerator training as a substitute of business knowledge (survey data)

	Training on business issues (Accelerator)	Training on business issues (Accelerator)	Training on business issues (Accelerator)	Training on business issues (Accelerator)
Pure tech	0.809	0.706		
	(0.020)	(0.057)		
One business			-0.565	-0.567
			(0.097)	(0.103)
Female		0.198		0.095
		(0.688)		(0.847)
Age bracket		-0.183		-0.268
		(0.373)		(0.179)
Education level		0.094		0.212
		(0.623)		(0.258)
Constant	3.169	3.438	3.667	3.760
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	127	126	127	126
R-squared	0.043	0.051	0.022	0.043

Notes: The table includes only accelerated ventures and report estimates of OLS regressions. *p*-values in parentheses.

Appendix 1 - Additional tables

TABLE A1 Pre-matching sample, descriptive statistics

	Accelerator group		Control group		
	Mean	SD	Mean	SD	Std. mean diff.[a]
Venture level variables					
Total funding amount (in '000s USD)	5463.42	84274.03	3249.81	32026.02	0.03
Accelerator	1.00	0.00	0.00	0.00	
Foundation year	2013.59	2.43	2013.08	2.85	0.19
Amount first round (log)	10.60	0.83	10.93	0.82	-0.40
Amount first round (in '000s USD)	54.45	40.97	71.29	41.40	-0.41
Team level variables					
Pure tech	0.38	0.49	0.33	0.47	0.10
Mixed tech-business	0.28	0.45	0.25	0.43	0.06
Other education	0.34	0.37	0.42	0.40	-0.06
One business	0.45	0.50	0.47	0.50	-0.04
Founding team size	2.13	1.18	1.89	1.02	0.22
Share females	0.16	0.31	0.13	0.29	0.11
Serial founder	0.60	0.49	0.55	0.50	0.10
Academic founder	0.19	0.39	0.15	0.36	0.10
Self-employee founder	0.05	0.23	0.04	0.19	0.07
MSC	0.48	0.50	0.45	0.50	0.05
PHD	0.15	0.35	0.16	0.37	-0.04
MBA	0.19	0.39	0.19	0.40	-0.01
Work experience	6.61	6.00	8.15	7.04	-0.24
Top employer	0.19	0.39	0.13	0.33	0.17
Top university	0.17	0.37	0.13	0.34	0.11
Observations	4221		2598		

Notes: The table reports the descriptive statistics for the sample of accelerated firms and the sample of potential control ventures before applying CEM matching. For the selection of the two samples, please refer to the text and for additional details to the Online Appendix. [a] See Linden & Samuels (2013) for the definition. SD, standard deviation. Std. mean diff., Standardized mean difference.

TABLE A2 Distribution of U.S. startups across states, matched sample

Investor type	Accelerator group	Control group
California	0.68	0.66
Colorado	0.02	0.03
Illinois	0.01	0.01
Missouri	0.01	0.01
New York	0.16	0.17
Ohio	0.05	0.05
Pennsylvania	0.03	0.03
Tennessee	0.02	0.02
Texas	0.01	0.01
Washington	0.01	0.01
Total	1	1

Notes: The table reports the distribution of the U.S. startups in the matched sample across states, separately for the accelerator and the control groups. For the pre-matching distribution, please refer to the Online Appendix.

TABLE A3 Exact matching on first round funding amount

Panel A. Mean difference between treated and control groups

	Accelerator group		Control group		(Diff)
	Mean	SD	Mean	SD	Std. mean diff.[a]
Amount first round (log)	10.84	0.81	10.84	0.81	0.00
Amount first round (in '000s USD)	66.319	41.042	66.494	41.180	0.00
Observations	1973		1973		

Panel B. OLS regressions

	Total funding (1)	Total funding (2)	Total funding (3)
Accelerator	-0.012	-0.150	0.015
	(0.868)	(0.052)	(0.866)
Mixed tech-business	0.258		
	(0.016)		
Accelerator × Mixed tech-business	-0.317		
	(0.022)		
Pure tech		-0.080	
		(0.402)	
Accelerator × Pure tech		0.158	
		(0.217)	
One business			0.232
			(0.014)
Accelerator × One business			-0.226
			(0.063)
Constant	11.690	11.752	11.647
	(0.000)	(0.000)	(0.000)
Control variables	Yes	Yes	Yes
Geo area fixed effects	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes
Foundation year fixed effects	Yes	Yes	Yes
Observations	3,946	3,946	3,946
R-squared	0.124	0.123	0.124

Notes: [a] See Linden & Samuels (2013) for the definition. SD, standard deviation. Std. mean diff., Standardized mean difference. In columns (1)-(3) the dependent variable is the log of the total funding amount. *p*-values in parentheses.

TABLE A4 Top 100 Accelerators

	Total funding (1)	Total funding (2)	Total funding (3)
Accelerator	0.297	0.096	0.410
	(0.088)	(0.630)	(0.038)
Mixed tech-business	0.690		
	(0.043)		
Accelerator × Mixed tech-business	-0.805		
	(0.082)		
Pure tech		-0.529	
		(0.036)	
Accelerator × Pure tech		0.302	
		(0.382)	
One business			0.724
			(0.005)
Accelerator × One business			-0.620
			(0.074)
Constant	0.848	1.310	0.616
	(0.555)	(0.362)	(0.667)
Control variables	Yes	Yes	Yes
Geo area fixed effects	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes
Foundation year fixed effects	Yes	Yes	Yes
Observations	394	394	394
R-squared	0.348	0.349	0.355

Notes: The table reports OLS estimates for the matched sample restricted to ventures in the top 100 accelerators and related controls. The dependent variable is the log of the total funding amount. *p*-values in parentheses.

TABLE A5 Pre-matched sample

	Total funding (1)	Total funding (2)	Total funding (3)
Accelerator	-0.003	-0.097	0.028
	(0.955)	(0.109)	(0.670)
Mixed tech-business	0.204		
	(0.023)		
Accelerator × Mixed tech-business	-0.148		
	(0.162)		
Pure tech		-0.092	
		(0.255)	
Accelerator × Pure tech		0.158	
		(0.106)	
One business			0.170
			(0.032)
Accelerator × One business			-0.145
			(0.121)
Constant	3.861	3.899	3.820
	(0.000)	(0.000)	(0.000)
Control variables	Yes	Yes	Yes
Geo area fixed effects	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes
Foundation year fixed effects	Yes	Yes	Yes
Observations	6,824	6,824	6,824
R-squared	0.202	0.202	0.202

Notes: The table reports OLS estimates for the pre-matched sample. The dependent variable is the log of the total funding amount. *p*-values in parentheses.

Online Appendix

This Appendix in organized in two parts. Part A is devoted to describing the methodology adopted to collect and codify data used in the empirical analysis of the paper. Part B is devoted to provide further descriptive statistics, tables and figures on the sample of startups examined in this study.

Regarding part A, in section 1, we briefly illustrate the criteria used to select the sample of startups to study. Then, in section 2, we present a detailed account of the steps undertaken to identify startup founders, while in section 3 we describe the way in which their educational background was codified. In section 4, we describe how we coded the classification of startups' sector of activity, while in section 5 we illustrate the way in which we classified their geographical location.

Part A - Methodology

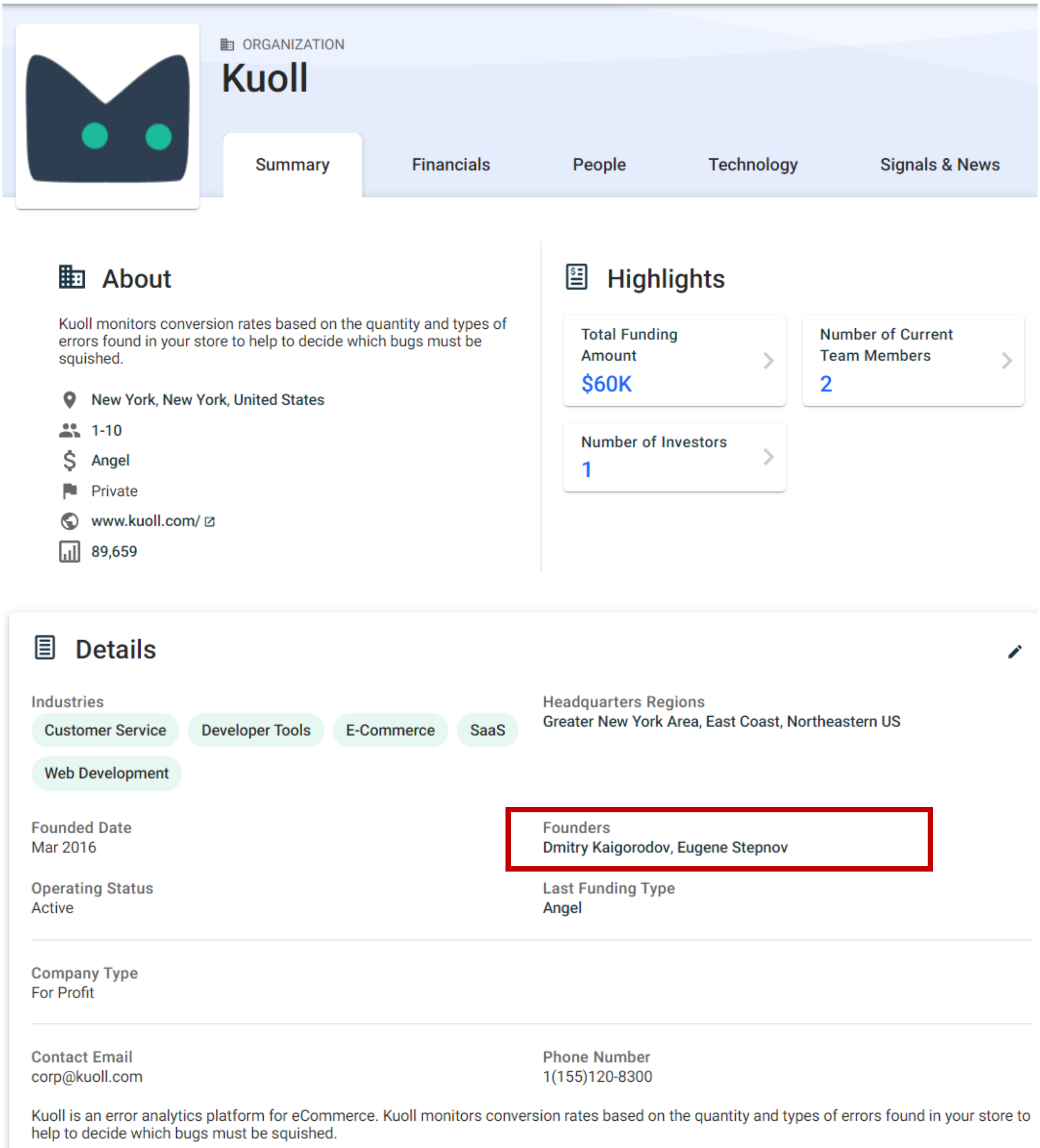
A1. Sample selection

As mentioned in the text, we extracted information from the Crunchbase (hereafter CB) database through the dedicated RESTful API. From the universe of all firms contained in CB, we focused on the startups that received a first funding round between 2004 and 2018 (inclusive). Among them we further selected the startups that in the first funding round received an amount less than or equal to 150 thousand US dollars, excluding those ventures for which the type of investors of the first funding round was not disclosed and those funded by companies or pension funds. These filters led us to identify an initial sample of 12,759 firms. Please see Part B of this appendix for a discussion of the rationale of using a threshold of 150 thousand US dollars for the selection of the sample to study.

A2. Identification of startup founders

For each of the selected startups, we faced the challenge of identifying their founding members. To this purpose, we exploited the information contained in the jobs.csv and people.csv tables retrieved through the RESTful API from the CB database. In particular, we identified all individuals that were affiliated with any of the selected startups (either currently or in the past) and that reported *founder* (or *co-founder*) as job title. We combined this with the information from the CB summary pages of all the ventures in our sample (see Figure 1 for an example).

Figure A1 – Summary page of a startup in the CB website
https://www.crunchbase.com/organization/kuoll
(consulted on May 1st 2021)



Combining these two pieces of information is important for a correct identification of all founders. In the example of Figure 1, one of the two founders of the startup, Dmitry Kaigorodov, reported CEO as the job title at this company, but he appeared among the co-founders in the summary page[1]. It is also worth noting that the information on founders in the summary page of startups is not available among the tables one can retrieve through the CB API. It must be scraped using alternative means for each startup. Hence, any study on startup founders relying only on the tables available from CB and on the reported job titles is likely to underestimate their number, unless autonomous scraping of the summary page is undertaken.

The identification of startup founders through the information contained in CB presents a second challenge. This is related to the fact that not all the *true* founders of a startup necessarily have a CB profile. Consequently, the names, affiliation and job titles of some true founders might be simply missing from the CB database. In order to address this problem, we used information from LinkedIn. More specifically, for each startup in our sample, we collected the public profiles of all employees that have reported affiliation to that startup. Among them, we identified as founders those individuals that in the LinkedIn resume reported *founder* (or *co-founder*) as the job title. This task was performed in two ways. First, the profile of individuals in CB often reports the URLs of various social and professional networks, particularly Facebook, Twitter and LinkedIn. Second, for those founders not reporting the URL of their LinkedIn resume in the CB profile, we searched it manually through the Google search engine.

Of the 12,759 startups in the initial sample, we could find information on founders using the methodology described above for 10,538 of them. Overall, we identified 22,994 *distinct* founders[2]. The distribution of these founders according to the origin of the information is reported in Table A1.

These results are broadly in line with a recent benchmarking exercise carried out by Retterath and Braun (2020). The authors compare information from eight frequently used VC databases, including

[1] It is worth noting that the information on founders in the summary page is not available among the tables one can retrieve through the CB API. It must be scraped using alternative means for each startup.
[2] Note that some of these individuals may have founded more than one startup.

CB, across 339 actual venture capital (VC) financing rounds from 396 investors in 108 different (mostly European) companies. They show that CB (together with Pitchbook) has the best coverage as far as the number (and the educational background) of founders are concerned. More specifically, their findings suggest that of all the *true* founders (as taken from funding contracts and original documentation) CB reports 63% of them. This is smaller than the fraction we find (i.e. 76%). However, it must be also noted that they consider (mostly) European companies and that the sample of firms examined is highly selected. In particular, the startups they examined have received funding from VC, whereas we consider any type of investor, and have received multiple funding rounds from several VC investors, whereas our sample includes mostly startups that have received only one or few funding rounds). Thus, any comparison should be taken with some degree of caution and should be interpreted in the light of these differences.

Table A1. Distribution of founders according to the origin of information

Origin of information	Number of founders	Percentage
(a) Individuals affiliated with a start-up and reporting *founder* or *co-founder* as job title in the CB database downloaded through the RESTful API	17,526	76.2
(b) Individuals affiliated with a start-up that do not report *founder* or *co-founder* as job title in the CB database, but appearing as part of the founding team in the summary webpage of the startup in the CB website	2,968	12.9
(c) Individuals without a CB profile, but reporting to be *founders* (or *co-founders*) of the startup in their LinkedIn resume	2,500	10.9
Total	22,994	100.0

In summary, to identify our founders we relied primarily on the information provided by CB, and used LinkedIn to double-check potentially missing people. We decided not to use the webpages of the startups for two main reasons. First, this approach is prone to a selection bias as the websites of failed startups are likely to be offline or shutdown. As we detailed in the paper, one of the advantages of CB is indeed the fact that failed startups are not deleted, in general, from the database, thereby avoiding any survivorship bias in the analysis. Second, startups might decide to drop the name of some founders (or *co-founder*) from the website due to conflicts or strategic reasons.

Consider the case of YouTern, a startup located in Carson City, Nevada (https://www.youtern.com) as an example. According to the combined information of CB and LinkedIn, the company has two founders: Mark Babbitt and Deb Babbitt. Quite interestingly, the (still active) website of the startup has a section devoted to introducing the company team (https://www.youtern.com/cm/about_us). The webpage reports pictures and links to the Twitter accounts of different team members (though some of these links are not working at the time we consulted the page on May 1st 2021). One of the pictures corresponds to that of Mark Babbit, even though he is not mentioned as one of the founders. However, what is more important to us, none of the persons reported in the page corresponds to Deb Babbit, i.e. the other presumed founder of the company.

Overall, we acknowledge that our sample may have some limitations. Specifically, it may be that some of the *true* founders of startups have not been detected. However, unless one has direct access to the company files and to the persons involved in the founding process, we are quite convinced that it is hardly possible to get a better and more faithful representation of the founding team. Moreover, to the extent that any misrepresentation affects Accelerator and Control startups in a random way, this issue should have a relatively minor impact upon our findings.

A3. Educational background of founders

The CB API allows extracting tables containing information on the educational backgrounds of the people having an account on the platform. However, it is important to point out that the educational background is not available for all the people having a profile on CB. For some individuals, CB only provides basic information on the name, gender, company affiliations and job titles, but no details on their education attainments. In particular, out of all 22,994 founders identified, only 6,940 reported information on the education background in their CB profile (i.e. around 30% of all founders).

For the remaining 16,054 founders, with missing information on the educational background, we exploited LinkedIn data. In particular, for each of the 22,994 founders, we searched the corresponding (public) LinkedIn profile. Out of the 22,994 founders, we could find a corresponding LinkedIn profile for 17,947 (i.e. slightly more than 78%) of them. After this data collection, our sample of 22,994 founder can be classified into four different typologies with respect to the source of information on their educational background. As one can observe from Table A2, for around 30% of all founders in our sample, the only information available to us on their education background was the one reported in the LinkedIn profile. For 23% of all founders, on the other hand, education background was available both from the LinkedIn resume and from the CB profile.

Table A2. Origin of information for founders' education

Origin of information	Number of founders	Percentage
(i) Founders for whom we have information on the educational background both from CB and from LinkedIn	5,180	22.5
(ii) Founders for whom we have information on the educational background only from CB	1,760	7.7
(iii) Founders for whom we have information on the educational background only from LinkedIn	6,931	30.1
(iv) Founders for whom we do not have information on the educational background from any source	9,123	39.7
Total	22,994	100.0

Most importantly, we could not find any information on the education background for about 40% of all the founders in the sample. This happened either because the person did not report that information in the CB profile or because we could not supplement CB with information from the LinkedIn resume. As discussed above, not all founders have a corresponding (public) LinkedIn profile or they have one, but the information reported in it is incomplete, i.e. it does not provide any detail on the education attainments.

Because our empirical analysis crucially relies upon the correct identification of the educational background of *all* the founders in the company, we were forced to drop from our sample the companies for which we were unable to do that. After dropping the startups for which data on the educational background of at least one founder was missing, our sample reduced to 6,867 firms.

For the 13,871 founders of the remaining 6,867 startups, we coded the educational attainments in terms of: a) type and level of degree, b) subject area of degree, c) start and end year of degree. In the next two subsections, we describe the methodology we followed to code the type and level of degree and the subject area.

A3.1 Education level

As far as the coding of the type and level of degree, we adopted as a reference classification the UNESCO International Standard Classification of Education (ISCED) 2011[3] that classifies educational attainments into nine levels. For the purposes of our paper, the relevant levels are (6) Bachelor or equivalent, (7) Master or equivalent and (8) Doctoral or equivalent. In order to code the education level, we implemented a dictionary approach. More specifically, for each relevant education level (e.g. Master) we created a dictionary containing a large number of different (regular) expressions that may correspond to that educational level. For example, some of the expressions used to classify Master education are the following ones:

{'Master': ['MSC', 'MPHIL', 'MSEE', 'MTECH', 'MMATH', etc.]}

Those education levels that we were unable to classify automatically through the ad-hoc built dictionary were reviewed and classified manually. Table A3 reports the distribution of founders according to the highest title attained.

Table A3. Founders by level of education

[3] http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf

Highest title attained	Number of founders	Percentage
Bachelor	5,080	36.6
Master (including MBA)	5,295	38.2
PhD	1,331	9.6
Others	2,165	15.6
Total	13,871	100.0

Note: the category others includes a miscellanea of titles and certifications that escape any standard classification. It also includes primary and secondary education. It also includes titles and certifications reported in the resume and achieved after graduation.

The figures reported in Table A8 are broadly comparable to those in the benchmarking exercise carried out by Retterath and Braun (2020). Founders with a PhD are 9.6% of all founders (as compared to 10.9% in the source cited). Those with a Master are 38.2% (as compared to 46.8% in the source cited), and those with a Bachelor are 36.6% (as compared to 15.2% in the source cited). Quite interestingly, in the study mentioned above by Retterath and Braun (2020), the founders with no information on the education background are 26.5% compared to only 15.6% in our sample. Thus, some of the discrepancies between our figures and theirs (particularly, with respect to the fraction of founders with a Bachelor) are likely due to the higher precision of the data used here, especially with regard to the use of LinkedIn resumes.

A3.2 Education subject areas

As far as the subject area of the degrees is concerned, we adopted machine-learning techniques. In particular, we started by taking the detailed description of the fields of education reported in the ISCED-F 2013 classification, published by the UNESCO[4]. This taxonomy classifies training and education programs into 11 broad areas according to the subject content of the education, plus a residual category (see Table A4):

[4] http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-fields-of-education-and-training-2013-detailed-field-descriptions-2015-en.pdf

Table A4. ISCED-F 2013, Fields of education

00 – Generic programs and qualifications
01 – Education
02 – Arts and humanities
03 – Social sciences, journalism and information
04 – Business, administration and law
05 – Natural sciences, mathematics and statistics
06 – Information and communication technologies
07 – Engineering, manufacturing and construction
08 – Agriculture, forestry, fisheries and veterinary
09 – Health and welfare
10 – Services
99 – Field unknown

For each of these 11 broad fields, the ISCED-F 2013 classification also provides a detailed list of the programs and qualifications whose content is classified in the field. In practice, this can thought of as a dictionary that associates each education class (key) to a list of subject contents (values), as in Figure A2 below.

Figure A2 – Dictionary of education fields and subject contents based on ISCED-F 2013

```
{
 'natural sciences': ['biology', 'genomics', etc.],
 'engineering': ['ceramics', 'electronics materials', etc.],
 'business': ['management science', 'business finance', etc.],
 'humanities': ['classical languages', 'history', etc.],
 'ict': ['computer science', 'informatics', etc.],
 'agriculture': ['fisheries', 'farming', etc.],
 'education': ['didactics', 'teacher training', etc.],
 'health': ['psychiatry', 'physiology', etc.],
 'services': ['catering', 'cosmetology', etc.],
 'social sciences': ['politics', 'ethnology', etc.],
}
```

The classification problem we faced is that the degree titles reported in the CB and LinkedIn resumes do not necessarily appear among any of the subject contents listed in the dictionary available from ISCED-F 2013 and reported in Figure A2. For example, consider the following degree title reported in the CB resume of a founder in our sample:

Variable name	Description	Example
uuid	Unique identifier of degree	829cd8ca-80ac-253a-c85b-587832b83df7
name	Degree name	Ph.D. Bioinformatics @ University of California, Santa Cruz
person_uuid	Unique identifier of person	caf8dbdb-5380-a162-3124-ec93d9d79754
person_name	Person name	Charles Vaske
institution_uuid	Unique identifier of educational institution	f2d37262-3642-64b7-e584-08a04c5698b4
institution_name	Educational institution name	University of California, Santa Cruz
degree_type	Type of degree	Ph.D.
subject	Educational subject	Bioinformatics
started_on	Starting date of education	2003
completed_on	Completion date of education	2009

The subject area of this degree, i.e., *bioinformatics*, is not included in any of the subject content areas of the ISCED-F 2013 dictionary. Thus, our aim was to build an algorithm capable to predict the most likely field of education of this degree and to classify it into any of the 11 broad fields of the ISCED-F 2013 classification, *even though "bioinformatics" does not appear among any of the subject contents listed in that classification*.

To this purpose, we implemented three different supervised machine-learning algorithms for short text categorization, using the ISCED-F 2013 dictionary described above as our *training set*. The three algorithms belong to the family of *word-embedding cosine similarity classifiers*[5]. In what follows, we explain with some more detail the working of these classifiers. Our aim is to provide the broad intuition behind these techniques; hence, the discussion is descriptive, rather than technical.

The first algorithm is based on the Word2Vec model. Word2Vec is one of the most commonly used word-embedding models (Mikolov, Sutskever, et al. 2013; Mikolov, Chen, et al. 2013). In a nutshell, given a text corpus as input, Word2Vec converts (i.e., *embeds*) words into vectors that carry semantic meaning. The idea and the purpose of the Word2Vec algorithm is to group together similar words in an *n*-vector space. The model *learns* the vectors of words from their co-occurrence information, i.e. how frequently they appear together in large text corpora. For this paper, we used

[5] To this purpose, we used the `shorttext` Python library (https://shorttext.readthedocs.io/en).

the Google's pre-trained Word2Vec model, which comprises word vectors for a vocabulary of 3 million words and phrases trained on about 100 billion words from a Google News dataset[6]. The vector length is 300 features (i.e. each word is embedded in vectors of length *n*=300) [7].

The second classification algorithm we implemented is the GloVe model (Global Vectors for Word Representation). The idea behind this algorithm, proposed by Pennington, Socher, and Manning (2014) is similar to Word2Vec. Words are converted into numerical vectors such that similar words cluster together and different words are distant in the vector space. The major difference between these models is that, while Word2vec relies just on local statistics that exploit the *local context* information of words (i.e. words that are occurring close to a certain word), GloVe also incorporates global statistics leveraging information from the entire corpus in order to obtain word vectors. In particular, we used two GloVe pre-trained models. First, a GloVe model pre-trained on Wikipedia and Gigaword, which comprises word vectors for a vocabulary of 400 thousand words and phrases trained on about 6 billion tokens. Second, a GloVe model pre-trained on *Common Crawl* (an open repository of web crawl data), which comprises word vectors for a vocabulary of 1.9 million words and phrases trained on about 42 billion tokens. In both cases, the vector length for word embedding is 300 features (i.e. each word is embedded in vectors of length *n*=300).

Using the pre-trained models briefly described above, we computed the cosine similarity between the degree title to classify (e.g. bioinformatics) and the subject contents for each of the 11 fields of education reported in Table A4. Thus, in the case of *bioinformatics* (i.e. the example used before), the three models yield the following output:

[6] The Google pre-trained model can be downloaded from https://code.google.com/archive/p/word2vec/

[7] As each word is translated (i.e. embedded) into a numerical vector (of size *n*=300), one can use these vectors to perform various operations, such as solving analogies among words, find the similarity between two words and so on. For example, consider again *bioinformatics*. Using the Word2Vec model pre-trained on the Google news data set, the top five most similar words in terms of vector cosine similarity are: [('genomics', 0.72), ('proteomics', 0.71), ('computational_biology', 0.71), ('informatics', 0.71), ('computational_chemistry', 0.69]. Please note that *words* in the Google pre-trained Word2Vec model comprise also misspelled words and paired words such as computational_biology.

Education field	Word2Vec model pre-trained on Google News Cosine similarity	Glove model pre-trained on Wikipedia Cosine similarity	Glove model pre-trained on Common Crawl Cosine similarity
'agriculture'	0.323195	0.174183056	0.280048788
'arts_humanities'	0.324961	0.109794274	0.242027849
'business'	0.300321	0.050448243	0.260863513
'education'	0.264906	0.084001936	0.273193419
'engineering'	0.405223	0.167354792	0.296524078
'generic'	0.260116	0.086451061	0.272801191
'health'	0.41679	0.267133027	0.380260646
'ict'	0.491649	0.291579843	0.397773743
'natural_sciences'	**0.712261**	**0.641396701**	**0.734733105**
'services'	0.260142	0.008069998	0.188106
'social_sciences'	0.406736	0.233450994	0.366639167
'unknown'	0.158428	0.108978309	0.156550601

The degree title was assigned to the field for which the cosine similarity takes the highest value. In this specific example, all models would assign *bioinformatics* to the field of natural sciences.

In addition to the three word-embedding models discussed above, we also implemented the so-called Soft Jaccard Score, which is a measure of the edit distance between two sets of tokens (Russ et al. 2016). This measure is based on word spellings. To illustrate its computation, two other metrics have to be discussed: the Damerau-Levenshtein distance and the longest common prefix.

The Damerau-Levenshtein distance is a string metric for measuring the edit distance (i.e. the dissimilarity) between two strings. It computes the minimum number of operations (such as inserting, deleting or replacing a character, transposing two adjacent characters and so on) that one has to apply to convert one string to another string. For example, if one applies this metric to the pair of words *bioinformatics* and *biology*

```
damerau_levenshtein('bioinformatics', ' biology')
```

the result is 7, as this is the minimum number of operations that need to be done to convert one word into the other.

The longest common prefix finds the length of common prefix of two words. For example, if one applies this metric to the pair of words *bioinformatics* and *biology*

```
longest_common_prefix('bioinformatics', ' biology')
```

the result is 3 (i.e. the length of the string 'bio').

Using the two metrics defined above, the similarity between two words is defined as the larger of the following:

$$s = 1 - \frac{DL\ distance}{\max[len(word1), len(word2)]} = 1 - \frac{7}{14} = 0.5$$

and

$$t = \frac{longest\ common\ prefix}{\max[len(word1), len(word2)]} = \frac{3}{14}$$

Finally, the Soft Jaccard Score is defined (similarly to the regular Jaccard coefficient) as the ratio between the (soft) intersection between the two sets of tokens (in this example, this is equal to 0.5) and the union (in this example, this is equal to 1.5). Hence, in the example given above the Soft Jaccard Score is equal to 1/3.

For each degree title, we computed the value of the Soft Jaccard Score against all subject content areas included in each education field. For each education field, then, we retained the maximum value of this score. For example, in the case of bioinformatics, the output of this operation is the following:

Education field	Subject content area	Soft Jaccard Score
'agriculture'	forestry	0.4
'arts_humanities'	ethics	0.647059
'business_law'	typing	0.473684
'education'	didactics	0.473684
'engineering'	robotics	0.555556
'generic'	co-operation	0.4
'health'	anatomy	0.4
'ict'	**informatics**	**0.866667**
'natural_sciences'	geoinformatics	0.75
'services'	gymnastics	0.473684
'social_sciences'	civics	0.555556
'unknown'	field unknown	0.272727

Among all education fields, finally, we classified the degree title in the education field with the highest Soft Jaccard Score overall. In the case of *bioinformatics*, this degree title was classified as 'ict' as this is the field registering the highest value of the Jaccard score.

After applying the four methods described above, we collected the results and compared the classifications produced by each of them. For example, in the case of bioinformatics, we would obtain a vector of the following kind:

Degree title	Education field predicted by Word2Vec (trained on Google News)	Education field predicted by GloVe (trained on Wikipedia)	Education field predicted by GloVe (trained on Common Crawl)	Education field predicted by Soft Jaccard Score
bioinformatics	natural sciences	natural sciences	natural sciences	ict

Given this output, we took a rather conservative (and high precision) approach. As long as all four methods yielded the same prediction, we retained that prediction and we classified the degree title in the corresponding education field. On the other hand, if at least one of the four methods yielded a prediction that was discordant from the others, we manually checked the degree title and imputed it to any of the 11 ISCED education fields according to our own evaluation, based on the documentation available from ISCED-F 2013 and information collected from web. For example, in the case of bioinformatics, we decided to classify this title in the field of natural sciences after consulting various websites.

It is also worth noting that we only classified education titles equivalent to BSC or higher (i.e. MSC, PhD and postgraduate diplomas). In other words, we did not classify any degree title at the secondary or post-secondary education level.

Table A5 below reports the distribution of the 13,878 founders by field of education. In the same table, we also report the number and percentage of founders in STEM, Business and Other fields. Please note that the percentage column does not sum up to 100 (and the column for the number of founders does not sum up to the total number of founders), as the same individual may have degree titles classified in different education fields (e.g. a MSC in engineering and an MBA).

Table A5. Founders by field of tertiary education

Education field	Number of founders	Percentage of all founders
00 – Generic programs and qualifications	195	1.4
01 – Education	139	1.0
02 – Arts and humanities	1080	7.8
03 – Social sciences, journalism and information	1675	12.1
04 – Business, administration and law	4206	30.3
05 – Natural sciences, mathematics and statistics	1350	9.7
06 – Information and communication technologies	2961	21.3
07 – Engineering, manufacturing and construction	3204	23.1
08 – Agriculture, forestry, fisheries and veterinary	36	0.3
09 – Health and welfare	339	2.4
10 – Services	159	1.1
STEM (05, 06, 07)	6677	48.1
Business (04)	4206	30.3
Others	3249	23.4

Note: the percentage column does not sum up to 100 as the same individual may have degree titles classified in different education fields (e.g. a MSC in engineering and an MBA, or BSC in engineering and MSC in computer science).

The majority of founders are graduates in STEM fields (48%), followed by Business (30%) and other fields (23%). These figure are somewhat different from those reported in Retterath and Braun (2020). For the sample studied by these two authors, graduates in STEM are only 29% of all founders, whereas 39% of them are graduated in business disciplines. In this respect, it is worth noting again that the samples studied in this paper and the one examined by the two authors mentioned above is quite different. Whereas they consider only startups, which have received multiple funding rounds from VCs, our sample comprises firms that have received one funding round from any type of investor (and that do not necessarily have received other funding).

A4. Sector of activity of startups

Startups in CB are assigned one or more *industry category* tags, which in the CB database are referred to as industry category groups. There are 46 unique tags in our dataset listed in Table A6.

Table A6. CB industry category groups

1	administrative services
2	advertising
3	agriculture and farming
4	apps
5	artificial intelligence
6	biotechnology
7	clothing and apparel
8	commerce and shopping
9	community and lifestyle
10	consumer electronics
11	consumer goods
12	content and publishing
13	data and analytics
14	design
15	education
16	energy
17	events
18	financial services
19	food and beverage
20	gaming
21	government and military
22	hardware
23	health care
24	information technology
25	internet services
26	lending and investments
27	manufacturing
28	media and entertainment
29	messaging and telecommunications
30	mobile
31	music and audio
32	natural resources
33	navigation and mapping
34	payments
35	platforms
36	privacy and security
37	professional services
38	real estate
39	sales and marketing
40	science and engineering
41	software
42	sports
43	sustainability
44	transportation
45	travel and tourism
46	video

Thus, for example, a company can be simultaneously be classified in lending and investments and financial services, depending on the specific type of activity of the startup. In our sample of 6,867 startups, the number of tags per startup goes from a minimum of 1 to a maximum of 14. The average number of tags per company is 2.89. In order to assign each startup to a unique *sector*, we clustered startups using Latent Class Analysis (LCA). LCA is a clustering technique, which is particularly appropriate when the characteristics associated to the subjects to be grouped are categorical. The technique is based on an iterative, maximum likelihood approach. It starts with a random split of subjects (i.e. startups) into a given number of classes and then reclassifies them based on an improvement criterion until convergence is reached (i.e., no further improvement is possible). The split that yielded the best fit in terms of AIC and BIC classified our startups into 12 different clusters. They are reported in Table A7 below.

Table A7. Distribution of startups by sector of activity (LCA analysis)

Sector	Number of firms	Percentage
Biotech & life sciences	298	4.3
Commerce	1837	26.8
Data analytics	413	6.0
Design & fashion	114	1.7
Fintech	485	7.1
Green tech & energy	232	3.4
Hardware	433	6.3
Internet services	195	2.8
Media & entertainment	834	12.1
Mobile apps	389	5.7
Sales & marketing	230	3.3
Software	1365	19.9
Not classified	42	0.6
Total	6867	100.0

The majority of the startups in our sample are in the fields of (e-)Commerce (27%), Software (20%) and Media & Entertainment (12%). Overall, these three fields account for about 59% of all startups in our sample.

Please note that we could not classify 42 startups into any of the 12 clusters as industry tags were missing for these companies in the CB database. Consequently, we dropped these firms from our analysis, which reduced our sample to 6,825 firms.

A5. Geographical areas

The CB database provides the location of startups at the level of country, state and city level. This information is available for all, but a few companies. In our sample, it was missing for just 67 startups.

In those few cases, we manually coded this information based on ancillary information available in the CB database (e.g. the area code associated to the phone number) or through web searches.

Overall, we could find the location for 6,819 of the 6,825 startups in our sample. As the startups in the sample are scattered across a large number of countries, we grouped them into eleven broader geographical areas characterized by relative economic and cultural homogeneity. Table A8 reports the distribution of our startups across these areas.

Table A8. Distribution of startups across broad geographical areas

Geographical area	Number of startups	Percentage
Africa/Middle East	212	3.1
Asia	268	3.9
Australia and New Zealand	167	2.4
Canada	200	2.9
Eastern Europe	331	4.9
United Kingdom	534	7.8
India	233	3.4
Israel	77	1.1
South America	533	7.8
United States	2933	43.0
Western (Continental) Europe	1331	19.5
Total	6819	100

As the United States is the most prominent location, we collected information (and subsequently matched companies) on the different US states (see Table A9).

Table A9. Distribution of US startups by state

State	**Number of startups**	**Percentage**
Alaska	1	0.0
Alabama	7	0.2
Arkansas	8	0.3
Arizona	16	0.5
California	1088	37.1
Colorado	95	3.2
Connecticut	22	0.8
District of Columbia	34	1.2
Delaware	12	0.4
Florida	48	1.6
Georgia	38	1.3
Hawaii	16	0.5
Iowa	13	0.4
Idaho	4	0.1
Illinois	79	2.7
Indiana	18	0.6
Kansas	6	0.2
Kentucky	25	0.9
Louisiana	6	0.2
Massachusetts	159	5.4
Maryland	42	1.4
Maine	4	0.1
Michigan	24	0.8
Minnesota	14	0.5
Missouri	38	1.3
Mississippi	4	0.1
Montana	4	0.1
North Carolina	28	1.0
North Dakota	1	0.0
Nebraska	19	0.6
New Hampshire	5	0.2
New Jersey	19	0.6
New Mexico	3	0.1
Nevada	14	0.5
New York	419	14.3
Ohio	90	3.1
Oklahoma	7	0.2
Oregon	19	0.6
Pennsylvania	131	4.5
Rhode Island	19	0.6
South Carolina	8	0.3
Tennessee	60	2.0
Texas	98	3.3
Utah	18	0.6
Virginia	32	1.1
Virgin Islands	1	0.0
Washington	79	2.7
Wisconsin	32	1.1
Not available	6	0.2
Total	**2933**	**100**

Part B – Further statistics

Out of the 6,819 selected startups (see above section A), 4,221 (about 62%) of them had an accelerator as their first-round investor. Table B1 reports the 50 most important accelerators in terms of funded startups in our sample.

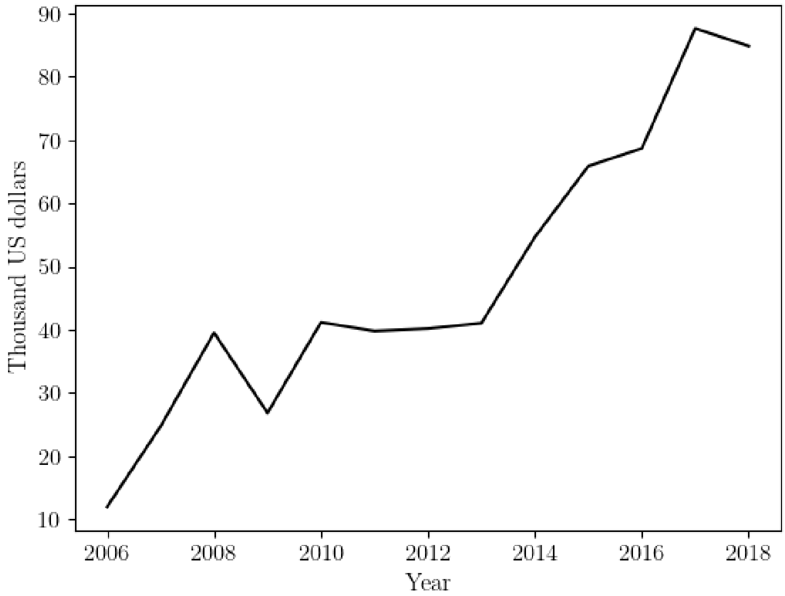
Table B1 – The top 50 accelerators in our sample

Name	**Percentage of all accelerated startups in sample**	**Average amount (first round) '000 USD**	**In seed-db.com**
Y Combinator	13.3	110.8	yes
Start-Up Chile	11.9	40.2	no
Techstars	9.0	91.2	yes
Startupbootcamp	4.3	20.6	yes
500startups	4.0	108.7	yes
Entrepreneurs Roundtable Accelerator	2.7	49.0	yes
AlphaLab	1.8	25.0	yes
Eleven Startup Accelerator	1.6	41.1	yes
Nxtp.labs	1.5	38.8	yes
MassChallenge	1.3	69.4	no
Bethnal Green Ventures	1.3	23.8	yes
Blueprint Health	1.2	20.0	yes
The Alchemist Accelerator	1.0	34.6	yes
TLabs	0.9	30.3	yes
Betaspring	0.9	42.9	yes
GameFounders	0.7	19.2	yes
Propel Capital (Sting Accelerate program)	0.7	35.1	no
Lanzadera Accelerator	0.7	40.7	no
gener8tor	0.7	37.6	yes
Capital Innovators	0.7	80.7	yes
DreamIT Ventures	0.7	29.1	yes
Excelerate Labs (Not Operating)	0.7	58.7	yes
H2 Ventures (H2 Accelerator)	0.6	80.3	no
TURN8 Seed Accelerator	0.6	40.7	no
Axel Springer Plug and Play	0.6	30.4	yes
Plug and Play	0.6	46.2	no
UpTech	0.6	28.2	yes
Acceleprise	0.6	47.9	yes
Seedcamp (Became Seed Fund)	0.6	68.8	yes
Startup Wise Guys	0.6	27.5	yes
Blue Startups	0.5	20.6	yes
Slingshot Accelerator	0.5	26.0	no
StartupYard	0.5	23.2	yes
Entrepreneur First	0.5	47.0	no
The Brandery	0.5	30.0	yes
Boomtown Accelerator	0.5	25.0	no
Collider	0.5	78.4	no
Emerge Venture Lab	0.5	41.8	yes
JumpStartFoundry	0.5	27.7	yes
CanopyBoulder	0.4	23.0	no
FounderFuel	0.4	53.7	yes
Portland Seed Fund	0.4	41.0	yes
StartFast Venture Accelerator	0.4	26.8	yes
SynBio axlr8r	0.4	82.7	yes
AngelPad	0.4	65.1	yes
Flashstarts	0.4	34.9	yes
Launchpad Accelerator	0.4	50.0	no
Start-Up Brasil	0.4	88.9	no
Triangle Startup Factory	0.4	56.4	yes
Accelerator Centre	0.4	31.6	no
Top 50 accelerators	**75.3**	**45.8**	

The top 50 accelerators in our sample feature some of the most well known organizations, including Y Combinator, Techstars, and 500Startups. Overall, they account for 75% of all startups present in our pre-matching sample. Moreover, 35 of them appear in the list of most prominent accelerators compiled by Jed Christiansen and reported in the website https://www.seed-db.com/accelerators.

The average amount of seed funding provided by these organizations is around 46 thousand USD, although there is a quite large variability across accelerators. In addition to variability across accelerators, there is also a significant variability over time. To explore this point, we computed the average amount of first-round seed funding provided by accelerators to the startups in our sample. This is reported in Figure B1.

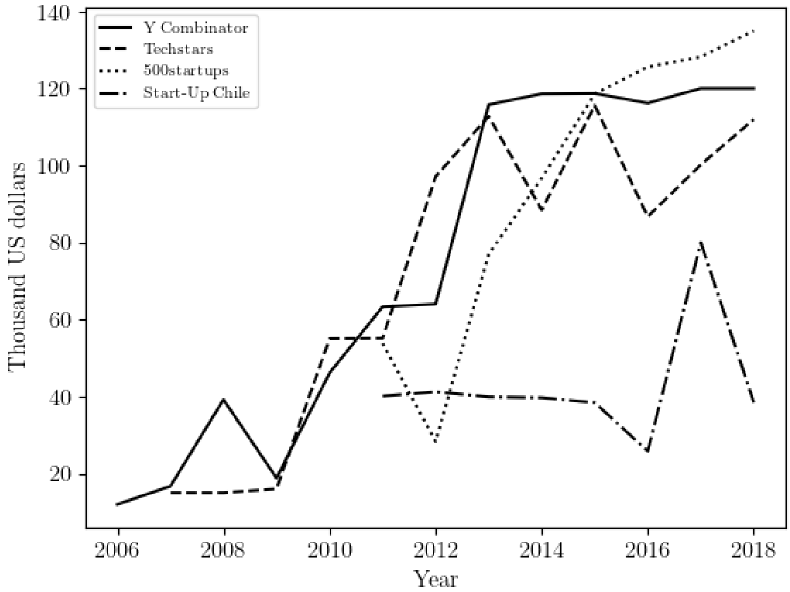
Figure B1. Average amount of first-round seed funding provided by accelerators to startups in our sample



As one can note, after an initial period of adjustment, accelerators disbursed on average around 40 thousand US dollars from 2008 to 2013. After that year, the average amount of seed funding provided by accelerators has increased considerably reaching almost 90 thousand dollars in 2018.

Figure B2 below reports the average amount of seed funding (to startups in our sample) for three of the major US based accelerators and, for comparison, one of the major non-US accelerators, i.e., Startup Chile.

Figure B2 – Average amount of first-round seed funding provided by four major accelerators to startups in our sample



Once again, after an initial period during which the amount funded was comprised between 30 and 60 thousand US dollars, the seed funding provided by the top US accelerators has grown remarkably over the years reaching more than 100 thousand US dollars in 2018. Currently, the website of Y Combinator declares that the amount provided to startups is 125 thousand dollars. On the other hand, the average amount provided by Startup Chile is quite flat, with the exception of a spike in 2017, around 40 thousand US dollars.

As the average amount of seed funding disbursed by accelerators varies both across accelerators and over time, we believed it was very important to control in all our regressions for the initial amount of financial resources raised by startups, in order to account for such heterogeneity. Moreover, since the average amount provided by the top accelerators in our sample exceeded 100 thousand US dollars, we prudentially selected the startups in the control group from among those that obtained less than 150 thousand US dollars from other (i.e., non-accelerator) types of investors. Setting this threshold allowed us comparing ventures that started their development with a reasonably similar amount of financial resources. Moreover, as long as the amount funding raised in the first funding round is a proxy of the quality of the business model and of the founding team, selecting the control sample by retaining only ventures that received less than 150 thousand dollars ensures that we compare startups of a similar ex-ante quality.

References

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." In *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*. International Conference on Learning Representations, ICLR. http://ronan.collobert.com/senna/.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Distributed Representations of Words and Phrases and Their Compositionality." *Advances in Neural Information Processing Systems*, October.

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. 2014. "GloVe: Global Vectors for Word Representation." In *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1532–43. Association for Computational Linguistics (ACL). https://doi.org/10.3115/v1/d14-1162.

Retterath, Andre, and Reiner Braun. 2020. "Benchmarking Venture Capital Databases." *SSRN Electronic Journal*, November.

Russ, Daniel E, Kwan-Yuet Ho, Joanne S Colt, Karla R Armenti, Dalsu Baris, Wong-Ho Chow, Faith Davis, et al. 2016. "Computer-Based Coding of Free-Text Job Descriptions to Efficiently Identify Occupations in Epidemiological Studies." *Occupational and Environmental Medicine* 73 (6): 417–24.