Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Spatial Organization of Firms and Location Choices Through the Value Chain

Juan Alcácer, Mercedes Delgado

To cite this article:
Juan Alcácer, Mercedes Delgado (2016) Spatial Organization of Firms and Location Choices Through the Value Chain. Management Science 62(11):3213-3234. http://dx.doi.org/10.1287/mnsc.2015.2308

MANAGEMENT SCIENCE
Vol. 62, No. 11, November 2016, pp. 3213–3234
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

https://doi.org/10.1287/mnsc.2015.2308

Spatial Organization of Firms and Location Choices Through the Value Chain

Juan Alcácer
Harvard Business School, Boston, Massachusetts 02163, jalcacer@hbs.edu

Mercedes Delgado
MIT Sloan School of Management, Cambridge, Massachusetts 02142; and Institute for Strategy and Competitiveness, Harvard Business School, Boston, Massachusetts 02163, delgadom@mit.edu

We explore the impact of geographically bounded, intrafirm linkages (internal agglomerations) and geographically bounded, interfirm linkages (external agglomerations) on firms' location strategies. Using data from the Census Bureau's Longitudinal Business Database, we analyze the locations of new establishments of biopharmaceutical firms in the United States from 1993 to 2005. We consider all activities in the value chain and allow location choices to vary by research and development, manufacturing, and sales. Our findings suggest that internal agglomerations have a positive impact on location. The effects of internal agglomerations vary by activity, and they arise both within an activity (e.g., among plants) and across activities (e.g., between sales and manufacturing). Our results also suggest that previous estimates of the effect of external agglomerations may be overestimated because the existing literature abstracted from internal agglomerations.

Keywords: location choices; agglomeration economies; internal agglomerations; value chain; spatial organization
History: Received June 6, 2013; accepted December 20, 2014, by Bruno Cassiman, business strategy. Published online in *Articles in Advance* January 27, 2016.

1. Introduction

The determinants of firms' location choices are the subject of a large body of research spanning multiple disciplinary fields. Most research in economics (Hanson 2001, Delgado et al. 2010) and strategy (Shaver and Flyer 2000, Alcácer and Chung 2014) focuses on elements of the external environment, specifically agglomeration economies, as drivers of firm location choices. Taken as a whole, this research suggests that geographic proximity *between firms* boosts productivity and therefore creates incentives for collocation.

Much less is known about the spatial organization of firms and the role of *same-firm* geographic proximity as a driver of location choices. Neglecting the role of internal drivers in firms' location decisions is a particularly surprising omission in the strategy literature, which recognizes that links between activities across the value chain are important levers for developing competitive advantage (Porter 1996) and for innovating (Cohen and Levinthal 1990). Understanding same-firm geographical proximity is also important because its effect on location choices is likely to be present in any multiunit firm, which has been the most common organizational form for firms since the 1940s (Chandler 1977). Geographic space is an important dimension of a firm's organizational structure that may affect performance and thus is paramount to understand the internal and external drivers of the spatial organization of firms.

To improve our understanding of the drivers of firms' spatial organization, this study investigates two related questions: *Does the location of a firm's existing facilities affect its subsequent location choices? If so, does that effect vary by activity in the value chain?* We argue that internal agglomerations—geographically bounded, intrafirm linkages that positively impact performance—do exist and that they prompt firms to collocate activities across the value chain. More specifically, we suggest that the three external agglomeration economies identified by Marshall (1920)—access to knowledge spillovers, specialized labor, and specialized suppliers—are as likely to boost firm performance when they occur within firms as when they occur across firms. Internal agglomerations also arise because geographical proximity enhances firms' ability to control (Giroud 2013, Kalnins and Lafontaine 2013) and coordinate (Chandler 1977, Henderson and Ono 2008) activities across the value chain, which results in better firm performance. Additionally, we suggest that these internal agglomerations vary by activity.

The presence of internal agglomerations does not preclude the effect of external ones. In fact, we

advocate for a comprehensive framework where *both* internal and external agglomerations drive location choices. We conceptualize internal and external agglomerations as separate forces that either complement or oppose each other depending on firms' previous locations. For firms located in clusters (i.e., geographic concentrations of related economic activity), internal and external agglomerations may work in the same direction. For firms not located in clusters, external agglomerations act as centrifugal forces that may drive firms to disperse their activities geographically, and internal agglomerations as centripetal forces that may drive within-firm collocation. Regardless of the direction these forces push firm locations, failing to consider internal agglomerations leads to an omitted variable problem, one that may well have biased previous work estimating external agglomeration's effect on location choice. This possible bias prompts a related research question: *Does the effect of external agglomerations on location choice vary when internal agglomerations are properly accounted for?*

We explore these questions using data from the Longitudinal Business Database of the Census Bureau. Our data set encompasses U.S. location choices for new establishments of biopharmaceutical firms for distinct activities of the value chain (research and development (R&D), production, and sales) from 1993 to 2005. Our empirical approach is to infer the existence of internal agglomerations through location decisions rather than by measuring them directly. Specifically, our core empirical analysis estimates conditional logit models predicting the location of new biopharmaceutical establishments as a function of both internal and external agglomerations.

We find strong evidence of same-firm collocation of biopharmaceutical activities. For example, our results suggest that a 1% increase in same-firm biopharmaceutical employment in a location leads to a 0.36% increase in the probability of the firm choosing that location. This finding, although present in all activities, varies by activity in the value chain: it is larger for R&D and manufacturing than for sales. Same-firm collocation occurs both within an activity (e.g., among plants) and across activities (e.g., between sales and manufacturing). Our results document unexpected patterns in the data, such as asymmetries of internal agglomeration between sales and manufacturing, whose explanation requires further conceptual development. We also find that external agglomerations positively influence location choices, but their estimated effects decline with the inclusion of the variables that proxy for internal agglomerations. This suggests that previous estimates of the effect of external agglomerations actually may overstate their influence, because internal agglomerations were not sufficiently modeled.

Our main findings persist even after controlling for alternative explanations that may drive same-firm collocation, such as sociological mechanisms, strategic location behavior, the initial location of a firm, and different sources of firm heterogeneity. Moreover, survival models suggest that establishments opened in locations where a firm was not previously present are more likely to fail, providing additional evidence that internal agglomerations may be behind same-firm collocation. Taken together, our analysis suggests that internal agglomerations are, in fact, an important driver of location choices that has been overlooked in the literature.

We examine internal agglomerations taking into account the spatial organization of *all* activities in the value chain of a firm. Failing to consider all activities leads to an omitted variable problem—pervasive in the literature—in which the estimated collocation levels between, say, sales and manufacturing may be biased by failing to control for the existing locations of another activity, such as R&D. Empirically, our paper is the first, to our knowledge, to systematically explore collocation decisions for all firm activities rather than for subsets of activities. Conceptually, we offer a comprehensive framework for location choices that encompasses both internal and external drivers—a framework that, according to our results, better explains actual location choices. This new framework encompasses a diverse array of literatures, which we curated and synthesized to identify five primary sources of internal agglomerations and the theoretical mechanisms on which they are based.

2. What Drives Location Choices?

Most research on firms' location choices is based on external drivers. This approach, although fruitful, offers a limited portrait of actual location choices by omitting or minimizing the role of internal agglomerations (geographically bounded, intrafirm linkages). Internal agglomerations are likely to play an important role in the location choices of multiunit firms, the most common organizational form since the 1940s (Chandler 1977). Multiunit firms are bound together through intrafirm linkages that impact performance (Porter 1996). To the extent that these interdependencies are weakened by distance, firms are likely to consider their existing physical footprint as they decide where to expand.

More recent studies on external drivers implicitly recognize the existence of internal agglomerations by taking steps to rule them out of empirical tests. For example, Alcácer and Chung (2007) focused on first-time entrants into the United States because "incumbent firms have prior investments that may affect subsequent location choices and create dependence

among observations by the same firm" (p. 761). Other studies have focused on how external drivers influence the location of start-ups (e.g., among others, Glaeser and Kerr 2009 and Delgado et al. 2010), a choice that will not be affected by internal agglomerations. Our focus here is to study the location decisions of existing firms as a function of both external drivers and internal agglomerations. In the rest of this section, we explain the factors that contribute to internal agglomerations (in §2.1) and the external factors that influence firm location choices (in §2.2).

2.1. Internal Agglomerations: Bringing Firm Activities Together?

Internal agglomerations arise because geographical proximity between within-firm activities improves firm performance. Previous research suggests diverse mechanisms of internal agglomerations, some of which are analogous to those identified by Marshall (1920) for external agglomeration economies: improved access to knowledge spillovers, specialized labor, and specialized suppliers.

For example, Cohen and Levinthal (1990) highlighted that the continuous exchange of information among activities, such as R&D and manufacturing, allows firms to develop the absorptive capacity needed to acquire external knowledge and innovate on it. To the extent that physical proximity increases communication and knowledge sharing, collocating activities will allow for productivity gains. Indeed, knowledge spillovers between R&D and manufacturing have received special interest in the literature. Tecu (2013) used patent data as an output measurement to find that R&D in chemicals was 2.5 times more productive in locations where manufacturing was also present (after controlling for external drivers of innovation such as proximity to universities and peer firms). Examining the reverse effect in the case of R&D and manufacturing, Adams and Jaffe (1996) demonstrated that the positive effect of parent-firm R&D on plant productivity diminishes with geographic distance. In both cases, within-firm knowledge spillovers were cited as the internal agglomeration mechanism behind superior performance. Despite these insights, few studies have examined how these types of spillover might work simultaneously.

Economies of scale and scope in internal labor markets are another source of internal agglomerations. Fixed costs from investments to attract, retain, and motivate workers, such as day-care facilities, gyms, or cafeterias, are easier to spread over a geographically concentrated labor force. Large-scale operations also enable firms to develop specialized labor that can be shared across activities (e.g., sharing R&D personnel across diverse research projects). Although many relevant studies find evidence of such scope economies in R&D (Helfat 1997, Henderson and Cockburn 1996) and in performance in general (Hamilton et al. 2003, Tate and Yang 2015), they abstract from the spatial organization of firms. We argue that geographic proximity among same-firm units helps firms achieve economies of scope by making labor easier to share or redeploy. For example, Di Minin and Bianchi (2011) found that physical proximity of R&D personnel to intellectual property (IP) lawyers increases the chances that an innovation will be patented, explaining why R&D is homebound in global industries. Proximity of specialized labor will also facilitate personnel rotation across activities, which was identified as a key practice behind the productivity gains of Japanese firms in the 1980s (Clark and Fujimoto 1987, Mansfield 1988).

A third mechanism of internal agglomerations is the access to intermediate inputs. Efficiency reasons for vertical integration are discussed in Grossman and Hart (1986), but there is little empirical evidence on the magnitude and type of these efficiencies. To the extent that firms have designed their operations around flows of intermediate inputs, we expect that collocating units that produce these inputs will decrease transportation and coordination costs and ultimately impact performance.[1]

Other related mechanisms behind internal agglomerations arise because value-chain activities within the firm are governed by organizational hierarchy rather that by market mechanisms. Within hierarchies, activity control and coordination are paramount to performance. In terms of control, Kalnins and Lafontaine (2013) found that the distance to headquarters is associated with lower revenue for establishments in the hotel industry, a finding that they attribute to the difficulty headquarters have controlling and monitoring their establishments from a distance. Similarly, Giroud (2013) argued that better control and monitoring explain his finding that a reduction in travel time between headquarters and plants increases plant investment by approximately 9% and total factor productivity by 1.4%.

Proximity between headquarters and plants has also been associated with the desire to improve performance by improving coordination. Henderson and Ono (2008) found that, after controlling for external drivers such as access to specific services in metropolitan areas, firms prefer to keep headquarters close to their manufacturing bases to facilitate coordination. Coordination as a driver of collocation extends beyond headquarters and plants. Chandler (1977) identified coordination as the fundamental force that

[1] In a recent paper, Atalay et al. (2014) find that vertically diversified firms often do not transfer tangible inputs between their upstream and downstream establishments, but instead they transfer intangible inputs (e.g., intellectual property).

kept different parts of a multiunit firm together and suggested that coordination costs increase with distance. Ketokivi (2006) explored the circumstances under which the collocation of R&D and manufacturing decreases coordination costs.

Despite differences in theoretical frameworks, methods, and the mechanisms proposed, the papers above share a common feature: they each suggest, and find evidence to support, that collocating activities has a positive impact on performance. Therefore, we expect that *firms are more likely to locate new establishments in places where they already have operations.* That collocation would be triggered by firms' attempts to achieve better performance through any of the mechanisms that drive internal agglomerations: knowledge and information flows, development of specialized labor and inputs, control and monitoring, and coordination.

We note that previous research suggests, conceptually and empirically, that the effect of internal agglomerations may change by activity. For example, Kleinbaum et al. (2008) used email frequency as a proxy for coordination needs in a large technology firm and found heterogeneity in information flows, with most information exchanges happening within activity, especially in R&D, and with above-average information exchanges between sales and R&D and between sales and supporting services. Van den Bulte and Moenaert (1998) found that information flows among R&D personnel increased when dispersed R&D personnel were geographically concentrated, suggesting there are informational benefits from R&D collocation.

What accounts for these differences? A plausible explanation is that internal agglomerations vary according to activity. For example, knowledge generated by R&D may be tacit and hard to transfer, making collocation *within* activities more desirable for R&D establishments; coordination across plants may be more important than among sales establishments, etc. Similar arguments can be used to hypothesize that agglomeration economies vary *between* activities. For example, the need to collocate R&D and manufacturing may be greater than the need to collocate R&D and sales because R&D and manufacturing require more coordination and knowledge transfers than R&D and sales. Because of the dearth of relevant theoretical models and the piecemeal nature of empirical findings in this area, we do not develop a set of testable hypotheses per activity or pair of activities here. Nevertheless, we do expect that *internal agglomerations vary by activity in the value chain*.

2.2. External Drivers of Location: Driving Firm Activities Apart?

Two sets of external drivers have been commonly studied in the literature: unique location endowments and agglomeration economies. Ricardo (1817) was the first to posit that a stochastic distribution of natural resources across the geographic space drives economic exchanges, an idea that is at the core of economic geography and international economics today. Some studies have since expanded Ricardo's view of endowments to encompass a location's institutional features, such as IP regimes, labor regulation, and the unique technological knowledge present in universities. This literature primarily predicts that firms will flock to locations where they can tap abundant inputs at low costs (e.g., labor in China), rare and unique resource (e.g., access to a port), or location-specific incentive programs (e.g., tax policies).

Marshall (1920) introduced a second and related external driver to explain location choices: agglomeration economies. He argued that collocated firms would enjoy higher productivity because geographic concentrations of firms would attract larger pools of specialized labor and suppliers, and they would also facilitate the flow of knowledge from one firm to another. These agglomeration economies exert a multiplier effect on firm productivity by increasing the benefits that firms would otherwise receive only from a location's physical and institutional endowments, a concept that has been adopted and expanded by researchers in various fields, including Porter (1998) in strategy, Jacobs (1984) in urban economics, and Krugman (1991) in international economics. Over time, a growing literature has broadened the set of agglomeration drivers, including local demand conditions and specialized institutions (Porter 1998), the organizational structure of regional business (Saxenian 1994), and social networks among firms and individuals (Storper 1995, Sorenson and Audia 2000, Dahl and Sorenson 2012).

Although some papers have identified negative effects from external agglomerations in the form of incentives to avoid collocating with other firms (see, e.g., Shaver and Flyer 2000 and Alcácer and Chung 2007), most empirical and theoretical models suggest that the effect of agglomeration economies on firm performance and on location choices will be positive (see, e.g., Henderson 2003, Glaeser and Kerr 2009, and Delgado et al. 2014).

Previous research also suggests that the strength of external location drivers will vary by value-chain activity. For example, Audretsch and Feldman (1996) found that innovation in the United States is more concentrated than manufacturing, whereas Alcácer (2006) found that competition and external agglomerations exert different effects on interfirm collocation for R&D, manufacturing, and sales in the wireless handset industry.

From this literature on the external drivers of location choices, we can establish our null hypothesis: that

only the potential for external agglomeration influences location choices. Additionally, we can expect that firms will adopt a basic level of geographic dispersion for different value-chain activities because activities may be attracted to different features in the external environment.[2]

Summarizing the last two sections, we establish that intra- and interfirm linkages are important for multiunit firms' performance, argue that location choices are the result of *both* external and internal drivers, and conceptualize the location decision as a function of these two sets of forces. The first set is internal agglomerations, centripetal forces that drive within-firm collocation; the second set is external agglomerations, centrifugal forces that drive firms to disperse activities geographically in search of the best external environment. Empirically, these forces may work in the same direction, as when, for example, a firm is already located in the best external environment: internal *and* external agglomerations would induce collocation. However, it is also plausible that these forces present firms with a trade-off: stay in a suboptimal location to preserve intrafirm links or move some activities to a better external environment.

3. Empirical Design

Our empirical design examines the location decisions of new establishments of biopharmaceutical firms during 1993–2005. The biopharmaceutical sector is an appropriate setting to examine firm location choices through the value chain, for several reasons. First, previous research has shown that biopharmaceutical firms tend to cluster, often near universities (Zucker et al. 1998, Furman and MacGarvie 2007) and other firms (Feldman 2003). These studies find evidence of external agglomerations (knowledge spillovers) from universities and incumbent firms for drug discovery at the firm level and at the firm–therapeutic-class level (see, e.g., Feldman and Schreuder 1996, Furman et al. 2005, Aharonson et al. 2008). Second, prior work suggests that internal agglomerations can arise within and across activities in biopharmaceuticals. For example, Chacar and Lieberman (2003) found that geographic centralization of same-firm pharmaceutical laboratories facilitated innovation, and although not accounting for the spatial organization of firms, Henderson and Cockburn (1996) found strong evidence of scale and scope economies across R&D projects that resulted in more drug discoveries. In terms of links across activities, Pisano (1997) found high levels of interdependencies among R&D and manufacturing activities in the value chain, suggesting the presence of potential internal agglomerations. Ketokivi (2006) suggested that R&D in biopharmaceuticals often involves significant manufacturing of drugs for testing during clinical trials. These clinical trials vary in their complexity (Azoulay 2004, Berndt and Cockburn 2012), and those that are more knowledge intensive (versus data intensive) are more likely to be developed in-house (Azoulay 2004), with the potential for exploiting internal agglomerations. Finally, although little is known about the nature of the interactions of sales with other activities, commercialization has become an important part of the value chain of pharmaceutical firms.

3.1. Sample

Our primary data source is the establishment-level Longitudinal Business Database (LBD), maintained by the U.S. Census Bureau. The LBD provides annual observations of the universe of U.S. establishments with payroll, including each establishment's date of entry, physical location, industry code, and number of employees.

Our first challenge was to identify firms in biopharmaceuticals. The traditional approach has been to select firms in a particular set of Standard Industrial Classification (SIC) codes. For example, Toole (2003) and Cortright and Mayer (2002) defined biopharmaceutical firms as those with establishments in SIC code 2830 (manufacturing of medicinal chemicals and botanical products (SIC code 2833), pharmaceutical preparations (SIC code 2834), diagnostic substances (SIC code 2835), and biological products (SIC code 2836)) and SIC code 8731 (commercial R&D in physical, engineering, and life sciences research). Another approach is to identify firms using firm directories (e.g., Zucker et al. 1998). Both approaches have pros and cons. SIC-based sampling captures all firms, well known or not, but it may include firms whose scope reaches beyond biopharmaceuticals.[3] Directory-based sampling guarantees that firms are in biopharmaceuticals, but it does not cover all firms.

We followed a hybrid approach to define our sample of multiunit biopharmaceutical firms. We matched LBD data to BioScan (1992), a detailed (but not exhaustive) directory of worldwide biopharmaceutical firms also used by Zucker et al. (1998), to obtain frequencies of establishments' SIC codes for the BioScan-matched firms (see Table 1). Three features in Table 1 deserve to be highlighted. First,

[2] For example, large service-oriented cities offer the best external environment for headquarters, but plants tend to locate in smaller cities or rural areas where wages are lower (Henderson and Ono 2008). Analogously, locations with top universities and a large number of Ph.D.'s may be ideal for R&D labs but not for manufacturing (Ketokivi and Ali-Yrkkö 2009, Tecu 2013).

[3] This is particularly the case with firms in SIC code 8731, an aggregated industry that includes labs that may not be associated with biopharmaceuticals.

Table 1 Top 10 Industry Codes (SICs) of BioScan Firms in 1992

Activity	SIC label	SIC code	No. of firms	% of firms
Bio R&D	Commercial physical and biological research	8731	168	43
Bio mfg	Includes SIC codes 2833–2836	2830[a]	127	32
Bio mfg	Diagnostic substances	2835	55	14
Bio mfg	Pharmaceutical preparations	2834	51	13
Bio mfg	Biological products	2836	47	12
Bio sales	Drugs, proprietaries, and sundries—wholesale	5120	45	15
Support	Holding companies	6719	42	11
Non-bio mfg	Industrial organic chemicals, not elsewhere classified	2869	40	10
Non-bio sales	Medical and hospital equipment—wholesale	5169	37	9
Support	Medical laboratories	8071	30	8
Non-bio mfg	Surgical medical instruments	3841	29	8

Notes. Sample of 395 BioScan firms; 112 of them are multiunit firms. Sources of data are LBD data and BioScan (1992).
[a]This includes firms in codes 2833–2836.

approximately 75% of BioScan-matched firms had at least one establishment in SIC codes 2830 or 8731. Thus, we use membership in either of these two SIC codes as the criterium to identify whether a multiunit firm in the LBD data is in biopharmaceuticals. Second, the frequency and definitions of SIC codes allowed us to identify activities in the biopharmaceutical value chain (referred to hereafter as the "bio value chain"): R&D (associated with SIC code 8731), manufacturing (associated with SIC code 2830), and sales (associated with SIC code 5120).[4] Third, there were BioScan firms with establishments in SIC codes outside the bio value chain, suggesting that firms can diversify into non-bio activities (e.g., medical devices and chemicals). Thus, we defined other value chains (referred to hereafter as "non-bio value chains") that include non-bio R&D (SIC codes 8732–8734), manufacturing (SIC codes 2000–3900, except 2830), and sales (SIC codes 5000–5900, except 5120). The rest of firm activities (i.e., any other SIC code) were broadly classified as support for any value chain. The main support activities included business services (e.g., headquarters); medical labs; and financial, insurance, and real estate services (e.g., holding companies).

The resulting sample spanned 563 multiunit firms in the LBD database, with at least one establishment in either SIC code 2830 or 8731 as of 1992.[5] Among these, our analysis focuses on 157 firms that opened new biopharmaceutical establishments during our sample period, 1993–2005. Table 2(a) shows the descriptive statistics for this baseline sample. These firms varied significantly in size in 1992, with a median of 879 employees and nine establishments, and often they had employment in both bio and non-bio activities (median of 324 and 63 employees, respectively).[6]

3.2. Econometric Specification: A Location Choice Model

Because internal agglomerations elude direct observation, and their effects on firm performance may be endogenous to the choice of location, we infer the existence of internal agglomerations from location choices (an approach that is commonly used in the agglomeration literature). Thus, our empirical analysis examines the relationship between internal and external agglomeration drivers and the (continental) U.S. locations of new establishments opened by biopharmaceutical firms. The baseline econometric specification is as follows:

$$Location_{firt} = \beta \ln(X^{\text{Internal}}_{frt-1}) + \gamma_r + \gamma_{fit} + \varepsilon_{firt}, \qquad (1)$$

where $Location_{firt}$ is equal to 1 if firm f chooses economic area (EA)[7] r to open a *new establishment* in focal bio activity i (R&D, manufacturing, or sales) at time t; $X^{\text{Internal}}_{frt-1}$ is a set of variables that captures firm f's geographic footprint of activities (i.e., for all activity i in bio value chain or non-bio value chains) in EA r at time $t-1$; γ_r is a vector of EA fixed effects; γ_{fit} is a vector of firm-bio activity-year fixed effects (i.e., a dummy variable for each combination of these three dimensions); and ε_{firt} is the error term.

We estimated Equation (1) using conditional (fixed-effects) logit models of the location decisions for each activity in the bio value chain.[8] The fixed effects (or groups) in the model are firm-activity-year (γ_{fit}), and within a group the firm chooses among multiple regions (134 EAs) to locate the focal activity i.[9] The fixed effects will capture unobserved firm-activity attributes that affect the location decision (e.g., the size of a firm's bio R&D effort). Because our sample

[4] SIC code 5120 is drugs, drug proprietaries, and druggists' sundries, including wholesale distribution of prescription drugs, proprietary (patent) drugs, druggists' sundries, and toiletries.

[5] We also impose that the firms have at least 1% of total employment in the bio value chain, to avoid capturing firms that are marginally in biopharmaceuticals. Of the 563 firms, 112 also appeared in the BioScan sample.

[6] Alternatively, we considered the samples of firms that experienced organic expansions in biopharmaceuticals through new establishments or growing existing establishments (335 firms), firms that acquired biopharmaceutical establishments (98 firms), and all the bio firms.

[7] The Bureau of Economic Analysis (BEA) defined 179 economic areas (EAs) spanning the United States that reflect meaningful economic regions and ensure comprehensive regional coverage (Johnson and Kort 2004). We excluded EAs in Alaska and Hawaii to minimize concerns about transportation costs.

[8] There were 1,226 location choices for new establishments in bio activities: 726 entries in R&D, 92 in manufacturing, and 408 in sales.

[9] The choice set corresponds to the group of 134 EAs selected for expansion in our sample in 1993–2005. The selected EAs vary by

Table 2(a) Attributes of the 157 Firms Opening New Bio Establishments as of 1992

	Employment			No. of establishments		
Variable	Mean	Std. dev.	Median[a]	Mean	Std. dev.	Median[a]
FirmTotal	4,711.185	8,133.530	879.385	37.841	94.554	9.400
FirmTotalBioValueChain	1,482.446	2,833.193	324.154	8.783	23.942	3.000
FirmTotalBioR&D	839.172	2,169.278	5.118	1.866	3.606	0.282
FirmTotalBioMfg	337.076	1,178.988	1.282	3.847	22.071	0.277
FirmTotalBioSales	306.198	1,007.901	0.000	3.070	9.125	0.000
FirmTotalNon-BioValueChains	2,103.567	4,752.606	62.923	19.191	78.554	1.357
FirmTotalSupport	364.529	1,660.097	0.000	7.510	39.483	0.000

[a]The pseudo-median is computed using the range of percentiles 46–54 to avoid disclosure problems.

may include multiple location choices from the same firm (i.e., firms may open multiple establishments over time), and because a firm's subsequent location decisions are likely to be correlated, we clustered the standard errors by firm.

3.2.1. Internal Agglomerations. Internal agglomerations are intrafirm linkages that are geographically bounded (i.e., facilitated with proximity among units of the firm). To capture the role of internal agglomerations in a firm's location strategies, we developed various firm-region-year variables that account for the spatial organization of all the firm's activities. Because biopharmaceutical firms can be horizontally diversified into other products (e.g., medical devices), we grouped same-firm employment by type of value chain (bio value chain and non-bio value chains). Then, to examine internal agglomerations across biopharmaceutical activities, we grouped same-firm employment by activity (bio R&D, bio manufacturing, bio sales, and support) across regions. The main internal agglomeration variables $X^{\text{Internal}}_{frt-1}$ are defined as follows:

$$\textit{FirmBioValue-ChainEmployment}_{frt-1} = \sum_{k\in\text{BioActivities}} emp_{fkrt-1}, \quad (2)$$

$$\textit{FirmNon-BioValue-ChainsEmployment}_{frt-1} = \sum_{k\in\text{Non-BioActivities}} emp_{fkrt-1}, \quad (3)$$

$$\textit{FirmSupportEmployment}_{frt-1} = \sum_{k\in\text{Support}} emp_{fkrt-1}, \quad (4)$$

where k indexes all firm activities (four-digit SIC codes) that map into the bio value chain (Equation (2)), into non-bio value chains (Equation (3)), and into support (Equation (4)); and emp_{fkrt-1} is the employment of firm f in activity k in region (EA) r the year prior to expansion. Recall from §3.1 that the range of industry SIC codes used for bio value chain, non-bio value chains, and support were mutually exclusive. To examine within- and across-activity internal agglomerations, the *FirmBioValue-ChainEmployment* variable is also broken down into its component activities (*FirmBioR&DEmployment*, *FirmBioMfgEmployment*, and *FirmBioSalesEmployment*). Table 2(b) shows the descriptive statistics for each of these variables.

3.2.2. External Agglomerations. Our baseline specification controls for external location drivers using EA fixed effects. These region dummies capture external economies of agglomerations (e.g., the specialization of the region in biopharmaceuticals) as well as regional endowments and policies that may influence the extent of external benefits (e.g., physical endowments, policies that favor manufacturing activities, wages).

Alternatively, in supplementary analysis we define variables that capture potential external agglomerations in each region the year before an establishment expanded. We built on Delgado et al. (2014) to compute our main variable of external economies of agglomeration. External agglomeration benefits may arise from the specialization of a region in a cluster of related biopharmaceutical activities. To capture this, we defined the specialization of region r in biopharmaceuticals using the employment location quotient, i.e., the share of regional employment in biopharmaceuticals compared with the share of U.S. employment in biopharmaceuticals. We excluded the focal firm's own biopharmaceutical employment at the region level to avoid cofounding the external and internal agglomerations. To be consistent with our definition of the bio value chain at the firm level, we defined the biopharmaceutical cluster as including the same industry codes. More formally,

$$\textit{RegionBioValue-ChainSpecialization}_{frt-1} = \frac{\sum_{k\in\text{BioActivities}}(emp_{krt-1} - emp_{fkrt-1})}{emp_{rt-1}} \times \left[\frac{\sum_{k\in\text{BioActivities}} emp_{kUSt-1}}{emp_{USt-1}}\right]^{-1}, \quad (5)$$

biopharmaceutical activity (121 EAs for R&D, 43 EAs for manufacturing, and 79 for sales). Note that firms may open establishments in multiple EAs in a given bio activity-year (i.e., there could be multiple positive outcomes within a group).

Table 2(b) Variables and Descriptive Statistics: 157 Firms Opening New Bio Establishments

Variable	Definition	Mean	Std. dev.
$Location_{firt}$[a]	Dummy equal to 1 if the firm chooses EA r to locate a new establishment in bio activity i	0.02	0.15
Internal Agglomerations$_{t-1}$			
FirmBioValue-ChainEmployment	Firm bio value-chain (SIC codes 8731, 2830, 5120) employment in EA	21.84	291.44
FirmBioR&DEmployment	Firm bio R&D (SIC code 8731) employment in EA	7.11	161.17
FirmBioMfgEmployment	Firm bio manufacturing (SIC code 2830) employment in EA	9.79	199.05
FirmBioSalesEmployment	Firm bio sales (SIC code 5120) employment in EA	4.94	99.91
FirmNon-BioValue-ChainsEmployment	Firm non-bio value-chains employment in EA	14.97	137.41
FirmSupportEmployment	Firm support-activity employment in EA	4.13	86.41
External Agglomerations$_{t-1}$			
RegionBioValue-ChainSpecialization	EA specialization in bio value chain (*excluding a firm's own bio employment*)	0.69	0.90
RegionEmployment (outside bio value chain)	EA employment (excluding bio value-chain employment)	729,977.5	1,134,466

Notes. The number of observations is 50,920: 134 EAs in the choice set by 280 firm-bio activity-year groups. Source of data is LBD data. The employment data include any employee in payroll (full time or part time).
[a]The number of location decisions is 1,226: 726 in R&D, 92 in manufacturing, and 408 in sales.

where k indexes all activities that map into the bio value chain; emp_{krt-1} and emp_{kUSt-1} are employment in bio activity k at time $t-1$ for EA r and the United States, respectively; emp_{fkrt-1} is the focal firm's own employment in bio activity k at r; and emp_{rt-1} and emp_{USt-1} are total employment for r and for the United States.

A value of *RegionBioValue-ChainSpecialization* greater than 1 indicates that biopharmaceuticals was overrepresented (in terms of employment) in EA r. For example, Table A.1 in the appendix shows EAs with high bio value-chain specialization (and a high share of U.S. biopharmaceuticals employment) in the base year (1992), including New York–Newark–Bridgeport (New York–New Jersey–Connecticut–Pennsylvania) and Raleigh–Durham–Cary (North Carolina) (with a *RegionBioValue-ChainSpecialization* of 2.12 and 1.95, respectively).

Finally, we added a control for the role of region size in firm location using the EA employment outside the bio value chain at time $t-1$ (*RegionEmployment*). This variable captures the potential benefits of locating in larger regions, including access to customers and urbanization economies.

4. Empirical Analysis

Our analysis starts with a descriptive examination of the location-choice patterns of new biopharmaceutical establishments. We then examine the existence of internal agglomerations and document how the effect of external agglomerations changes when internal agglomerations are factored in (in §4.1) and the differential effect of internal agglomerations by activity (in §4.2). In §4.3 we explore alternative explanations (other than internal agglomerations) for same-firm collocation, an exercise we complement with an analysis of the role of internal agglomerations in the survival of establishments opened during our time period (in §4.4). Finally, although our main analysis focuses on the location of new establishments, we also consider the implicit location choice associated with growing an existing establishment and examine location patterns for all types of establishments, new and existing (in §4.5).

Table 3 provides a simple tabulation of the location choices of new biopharmaceutical establishments based on the intensity of external and internal agglomerations in the chosen EA. The 1,226 location choices during 1993–2005 are divided into six categories based on whether firms chose EAs with low or high external agglomerations and low, medium, or high internal agglomerations the year prior to the expansion. In Table 3, external agglomerations in a location are defined as *high* if the location choice is among the top 10 EAs by share of U.S. biopharmaceuticals that also have high specialization. These top 10 biopharmaceutical clusters account for about 50% of total U.S. biopharmaceuticals employment in 1992 (e.g., see Table A.1 for the list of top biopharmaceutical EAs). The rest of the EAs are broadly defined as locations with *low* external agglomerations. Internal agglomerations in a location are defined as high if the location is the firm's base EA for biopharmaceuticals (i.e., EA with the largest firm bio employment), medium if the location has some firm employment (but is not the base EA), and low if the location is new to the firm (no preexisting firm presence).

Table 3 shows that a majority of location choices for new bio activities (63%) occur in EAs with same-firm presence (consistent with internal agglomerations), with the rest (37%) in new locations (external

Table 3 Location Choices for New Bio Establishments, 1993–2005: Distribution by Intensity of Internal and External Agglomerations

			External Agglomerations$_{t-1}$		
			Low	High: Top 10 EAs[a]	
Internal Agglomerations$_{t-1}$	Low:	EAs with zero firm employment	372 (30%)	84 (7%)	37%
	Medium:	EAs with any firm employment	419 (34%)	213 (17%)	52%
	High:	EAs with the largest firm bio employment	54 (4%)	84 (7%)	11%
			69%	31%	

Notes. In parentheses are shown the percentage of all firm-bio activity-year location choices through new establishments (of 1,226). See Table A.1 for a list of top 10 EAs.
[a]EAs is the regional unit.

drivers). Perhaps surprisingly, many location choices (69%) occur outside the top 10 biopharmaceutical EAs. For these cases of low-external locations, most expansions (38%) occur in locations with same-firm presence. There are also expansions in locations where both potential internal and external agglomerations are high (7%), but these are a small number, reducing the concern about firms choosing a top biopharmaceutical regional cluster for their first location and subsequently expanding into that same location. Importantly, very similar patterns occur if we drop the location choices for bio sales, which could be more geographically dispersed than manufacturing and R&D.[10] Overall, these findings suggest that *both* internal and external agglomerations matter for the location choices of biopharmaceutical firms.

4.1. Baseline Results: Same-Firm Collocation

To confirm the patterns identified in Table 3, we conducted a more systematic analysis of the location choices of new establishments. Table 4 shows the results of estimating Equation (1) using a conditional logit specification. Recall that the coefficients are identified using across-location variation in both external and internal agglomeration variables, controlling for firm-activity-year fixed effects. Coefficients are transformed into odds ratios to facilitate comparisons.

Model (1) introduces our main measure of external drivers: EAs fixed effects. These region dummies capture external agglomerations and endowments that drive the location choices of firms. Not surprisingly, these external drivers are important to explain location choices of new establishments, as suggested by the fit of the model (log-likelihood of −3,119 and pseudo R^2 of 0.22). Model (2) adds our main measures of internal agglomerations: same-firm employment in the bio value chain, non-bio value chains, and support activities. Including these variables improves the model fit; the log likelihood goes up significantly

Table 4 Location of New Establishments: Collocation Within and Outside the Bio Value Chain (157 Firms; 50,920 Observations; Conditional Logit; Odds Ratios)

	Y_{firt} = Location choices for new establishment in bio activity i			
	(1)	(2)	(3)	(4)
ln(*FirmBio Value-ChainEmployment*)		1.365** (0.052)		1.394** (0.043)
ln(*FirmNon-Bio Value-ChainsEmployment*)		1.126** (0.038)		1.107* (0.044)
ln(*Firm SupportEmployment*)		1.118** (0.042)		1.142** (0.040)
Region (EA) fixed effects	Yes	Yes	No	No
ln(*RegionBioValue-ChainSpecialization*)			1.377** (0.111)	1.214* (0.111)
ln(*RegionEmployment*) (outside bio value chain)			2.418** (0.123)	1.878** (0.096)
Groups: Firm-bio activity-year	380	380	380	380
Pseudo R^2	0.22	0.26	0.16	0.22
Log-likelihood	−3,119	−2,935	−3,330	−3,092

Note. Standard errors are clustered by firm.
**Significant at the 1% level; *significant at the 5% level.

(at the 1% level), suggesting the importance of considering internal agglomerations in location choices. The coefficients for internal agglomerations variables are all positive and significant: a 1% increase in *FirmBioValue-ChainEmployment* in a location leads to a 0.36% increase in the probability of choosing that location. The magnitude is statistically larger for firm bio value-chain activities than for other value chains and for support.[11] The relatively larger collocation with biopharmaceutical activities is consistent with the hypothesis that more meaningful interdependencies will occur among activities in the same value chain than among activities in different value chains.

Although our baseline econometric specification controls for external drivers using EA fixed effects, we

[10] The main difference is that the percentages of low-internal and low-external locations decline slightly, to 25%, with an increase in the percentage of location choices in EAs with same-firm presence.

[11] The difference in the estimated coefficients of *FirmBioValue-ChainEmployment* and *FirmNon-BioValue-ChainsEmployment* is positive and significant at 1% ($\chi^2 = 13.97$).

alternatively use region-year external agglomeration variables in Table 3 (models (3) and (4)) to compare external and internal agglomerations. The external agglomeration variables are the specialization of the region in biopharmaceuticals (ln(*RegionBioValue-ChainSpecialization*)) and the size of the region (ln(*RegionEmployment*)). Model (3) only includes the external agglomeration variables. Both the specialization of the region in biopharmaceuticals and, particularly, the size of the region have a positive influence on the location choice of new biopharmaceutical establishments. Model (4) adds the internal agglomeration variables of same-firm employment in the bio value chain, non-bio value chains, and support activities. The inclusion of these variables increases the fit of the model significantly (at the 1% level). Importantly, the estimated coefficients of the external agglomeration variables decline, especially for region size (the odds ratio declines from 1.38 to 1.21 for regional specialization and from 2.42 to 1.88 for the region size). This decrease in the effect of external agglomerations may be due to firms that are located in strong biopharmaceutical clusters and decide to stay in them. These firms can exploit both internal and external agglomerations in the same location, suggesting that the effect of external agglomerations in firm locations could be overestimated if the role of internal agglomerations is not taken into account.

Regarding internal agglomeration effects, our findings are similar to those reported in model (2). Furthermore, we find that the coefficients for *FirmBioValue-ChainEmployment* and *RegionBioValue-ChainSpecialization* are not statistically different, suggesting that both internal and external agglomerations are important drivers of location choices.

4.2. Same-Firm Collocation by Activity

Table 5 shows the results for location choices by bio activity. Note that although the internal agglomeration variables still have a positive and significant effect on location, the coefficients' magnitude varies by activity. Specifically, the importance of collocating a firm's own bio value chain activities is especially larger for manufacturing expansions (model (2)), followed by R&D (model (1)) and sales expansions (model (3)).[12] Indeed, for manufacturing expansions this is the only type of internal agglomeration that seems to matter. This suggests strong linkages across new and existing plants within the same value chain that may not be easy to transfer across different value chains.

By contrast, internal agglomerations in the non-bio value chains seem to matter only for the location of R&D and sales. These positive estimates for *FirmNon-BioValue-ChainsEmployment* suggest that R&D and sales activities may tap into complementary knowledge and tasks that reside in other value chains, or that there may be economies of scope associated with both activities. For example, the same sales force can be used to market-related products in different industries. Finally, support activities only seem to matter for the location of new R&D establishments. The lack of significance for *FirmSupport Employment*, which includes employment in headquarters (HQs), in the location of manufacturing (model (2)) seems inconsistent with those studies that highlight the benefits for plant performance of proximity to the headquarters. One potential explanation for this discrepancy is that these studies examine the effect of the proximity between HQs and manufacturing abstracting from other activities of the value chain,[13] which reinforces the importance of exploring agglomeration economies considering all activities instead of isolated pairs.

Our analysis has shown strong evidence of same-firm collocation within the bio value chain. Now we turn to explore the extent of within- and across-activity collocation in biopharmaceuticals. To do so, in Table 5 (models (4)–(6)) we break down the variable *FirmBioValue-ChainEmployment* into its component activities (*FirmBioR&DEmployment*, *FirmBio-MfgEmployment*, and *FirmBioSalesEmployment*). Results suggest strong within-activity interdependencies for R&D and especially for manufacturing. Manufacturing establishments are more likely to be located in EAs where there is a higher presence of same-firm manufacturing in biopharmaceuticals (positive and significant coefficient for *FirmBioMfgEmployment* in model (5)). Similarly, R&D establishments seem to collocate with same-firm bio R&D activity (model (4)). This within-activity collocation is consistent with the idea that relevant information exchanges take place within the same activity of a firm (Kleinbaum et al. 2008, Van den Bulte and Moenaert 1998). For example, if R&D relies on the continuous exchange of tacit knowledge, we may expect bio R&D sites to collocate (Chacar and Lieberman 2003). Even if geographical dispersion may be desirable, firms need communication technology and sophisticated management practices to be able to break R&D or manufacturing activities across locations (see Fort 2011). Thus, same-activity collocation may in part reflect the lack of

[12] To compare the estimated coefficients across activities in models (1)–(3), we simultaneously estimate the three models using seemingly unrelated regression (SUR). The estimated coefficient of *FirmBioValue-ChainEmployment* is significantly higher at the 10% level for manufacturing activities.

[13] For example, if we reestimate model (5) in Table 5 including only internal agglomerations in *FirmSupportEmployment*, the estimated coefficient of this variable becomes positive and significant (odds ratio of 1.33), suggesting positive collocation between new manufacturing facilities and support activity.

Table 5 Location of New Establishments by Type of Bio Activity (Conditional Logit; Odds Ratios)

	Y_{first} = Location choices for new establishment in bio activity i					
	Bio R&D (1)	Bio mfg (2)	Bio sales (3)	Bio R&D (4)	Bio mfg (5)	Bio sales (6)
ln(*FirmBioValue-ChainEmployment*)	1.286**	1.578**	1.253**			
	(0.093)	(0.130)	(0.065)			
ln(*FirmBioR&DEmployment*)				1.282**	1.534**	1.105
				(0.101)	(0.224)	(0.112)
ln(*FirmBioMfgEmployment*)				1.394**	1.529**	1.350**
				(0.177)	(0.099)	(0.094)
ln(*FirmBioSalesEmployment*)				0.917	1.039	1.007
				(0.250)	(0.153)	(0.070)
ln(*FirmNon-BioValue-ChainsEmployment*)	1.306**	1.046	1.093*	1.303**	1.040	1.106**
	(0.100)	(0.047)	(0.040)	(0.105)	(0.075)	(0.039)
ln(*FirmSupportEmployment*)	1.141*	1.070	1.035	1.131*	0.996	0.999
	(0.071)	(0.109)	(0.076)	(0.070)	(0.089)	(0.073)
Region (EA) fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Groups: Firm-bio activity-year	192	72	116	192	72	116
Pseudo R^2	0.26	0.21	0.22	0.27	0.21	0.22
Log-likelihood	−1,583	−257	−841	−1,581	−257	−841
Firms	92	48	60	92	48	60
Observations	23,232	3,096	9,164	23,232	3,096	9,164

Note. Standard errors are clustered by firm.
**Significant at the 1% level; *significant at the 5% level.

proper technology to overcome distance. By contrast, for sales establishments we find weak within-activity interdependencies. Sales activities that aim to cover geographically dispersed customers may account for low within-activity collocation.

We also find evidence of across-activity collocation for each type of bio activity. R&D and manufacturing benefit from across-activity collocation as well as within-activity collocation. New R&D establishments are more likely to be located in EAs where there is a presence of same-firm manufacturing in biopharmaceuticals (positive and significant coefficient for *FirmBioMfgEmployment* in model (4)); manufacturing establishments are also more likely to be located in EAs where there is presence of same-firm R&D in biopharmaceuticals (positive and significant coefficient for *FirmBioR&DEmployment* in model (5)).[14]

Note that new sales activity collocates with existing, same-firm manufacturing facilities, but new manufacturing activity does not collocate with sales. The asymmetry of internal agglomerations between sales and manufacturing is puzzling. The fact that sales offices are opened in existing manufacturing locations could be in part explained by firms beginning to commercialize their products (i.e., the firm has no prior bio sales activity). The extent of collocation of sales with manufacturing also may depend on the type of customer (other businesses or final consumers), but we cannot assess this with our data.

4.3. Exploring Alternative Drivers of Same-Firm Collocation

In the analysis thus far we have found strong evidence of same-firm collocation of activities, even after controlling for firm-activity-year and region fixed effects. These findings are consistent with the premise of internal agglomerations as important drivers of location choices. We now turn to explore three alternative explanations to same-firm collocation mentioned in the literature: sociological factors (e.g., social capital in the region), the initial location of the firm (e.g., strength of the cluster where the firm was born), and strategic behavior in location choices (i.e., leader firms that avoid locating in strong clusters). Table 6 shows the new analyses for our baseline specifications (corresponding to models (2) and (4) of Table 4).

4.3.1. Sociological Factors. These studies suggest that entrepreneurs choose to start firms where they live because familiarity with the environment and

[14] To better understand what could be driving this across-activity collocation, we consider the subsample of firms that are diversifying into new bio activities (i.e., manufacturing firms opening their first R&D establishment or R&D firms opening their first manufacturing site). Firms that begin a new activity may need more coordination with related activities to successfully integrate the new activity with the rest of the value chain and therefore can benefit more from collocation. We estimated models (4) and (5) in Table 5, excluding the observations that correspond to diversifying into new activities, and found that the collocation of R&D and manufacturing becomes insignificant. This suggests that collocation was driven by firms diversifying into new bio activities.

Table 6 Alternative Explanations to Same-Firm Collocation: Location of New Establishments (Conditional Logit Model; Odds Ratios)

	Y_{firt} = Location choices for new establishment in bio activity i											
	Sociological factors		Initial location						Strategic considerations			
	Firm region tenure		All initial locations		Firm oldest region: Top bio cluster		Firm oldest region: Non-top bio cluster		Small firms		Large firms	
	(1)	(2)	(3)	(4)	(5a)	(5b)	(5c)	(5d)	(6a)	(6b)	(6c)	(6d)
ln(*FirmBioValue-Chain Employment*)	1.196** (0.037)	1.204** (0.028)	1.292** (0.043)	1.328** (0.036)	1.289** (0.049)	1.331** (0.038)	1.485** (0.089)	1.509** (0.097)	1.769** (0.108)	1.812** (0.112)	1.259** (0.044)	1.284** (0.037)
ln(*FirmNon-BioValue-Chains Employment*)	1.026 (0.038)	1.011 (0.032)	1.109** (0.034)	1.087* (0.041)	1.091 (0.051)	1.061 (0.057)	1.165** (0.045)	1.169** (0.046)	1.758** (0.187)	1.626** (0.168)	1.121** (0.037)	1.106** (0.042)
ln(*FirmSupportEmployment*)	1.089* (0.038)	1.116** (0.036)	1.079* (0.035)	1.108** (0.036)	1.058 (0.038)	1.114 (0.041)	1.277** (0.087)	1.267** (0.084)	1.163 (0.115)	1.194 (0.112)	1.130** (0.044)	1.157** (0.039)
ln(*FirmRegionTenure*)	1.646** (0.114)	1.700** (0.112)										
FirmOldestRegion			3.393** (0.738)	3.025** (0.644)								
Region (EA) fixed effects	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
ln(*RegionBioValue-Chain Specialization*)		1.193 (0.114)		1.216* (0.109)		1.151 (0.114)		1.245 (0.170)		1.160 (0.148)		1.195 (0.128)
ln(*EmploymentRegion*) (outside bio value chain)		1.724** (0.081)		1.865** (0.094)		1.976** (0.182)		1.519** (0.123)		1.527** (0.141)		1.847** (0.119)
Groups: Firm-bio activity-year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.28	0.24	0.27	0.23	0.27	0.23	0.21	0.18	0.29	0.26	0.25	0.21
Log-likelihood	−2,876	−3,016	−2,908	−3,068	−1,654	−1,759	−1,198	−1,249	−709	−739	−2,105	−2,219
Firms	157	157	157	157	79	79	78	78	77	77	80	80
Observations	50,920	50,920	50,920	50,920	25,929	25,929	17,005	17,005	12,168	12,168	29,120	29,120

Notes. Standard errors are clustered by firm. Models (5a)–(5d): The choice set contains 129 and 95 EAs for the subsamples of top initial cluster ((5a) and (5b)) and non-top initial cluster ((5c) and (5d)). Models (6a)–(6d): The 157 firms are divided into subsamples of small and large firms in 1992 (i.e., firm employment below/above the (pseudo) median value of 879). The choice set includes 78 EAs for the small firms and 130 EAs for large firms.

**Significant at the 1% level; *significant at the 5% level.

their built social networks facilitate transactions with suppliers, customers, and other stakeholders. These social networks were developed in the companies in which entrepreneurs previously worked as well as through links with the community in which managers and workers live (Storper 1995, Sorenson and Audia 2000, Buenstorf and Klepper 2009). Note that the empirical evidence in these studies tends to be built around new ventures; our sample includes both young firms and established and large biopharmaceutical firms. As such, some of the mechanisms behind this literature may not be as relevant.

Nonetheless, to control for sociological factors that can influence same-firm collocation of activities, we draw on Dahl and Sorenson (2012) and include a measure of firm tenure in a particular region (ln(*FirmRegionTenure*)). This variable is defined as the age of the oldest establishment of the firm in the EA since 1992 and up to the year prior to the location decision—i.e., age of the oldest establishment in the EA between 1992 and $t-1$.[15] Table 6 shows the results of this new analysis (models (1) and (2)). Both the coefficients of internal and external agglomerations decline since ln(*FirmRegionTenure*) is a proxy for potential interfirm connectivity in a location that positively correlates with our external and internal agglomeration variables. However, note that even after accounting for the positive influence of region tenure, firms collocate bio activities, further suggesting the existence of internal agglomerations.

4.3.2. Initial Location. As noted previously in §2, our analysis focuses on incumbent multiunit firms. As a result, we abstract from modeling the initial location of a firm (start-up location), its first-time expansion, and how it was founded (spin-off or not). This raises some potential endogeneity concerns regarding the role of the initial location. For example, firms that are born in a strong biopharmaceutical cluster may tend to grow within that location, and then we cannot properly separate the effect of internal versus external agglomerations.

We address this concern in two ways. First, we identify a firm's "initial location" and include in our model a dummy equal to 1 for that location. Note that although we do not observe the actual location of birth, we use as an approximation the EA with the oldest establishment of the firm in the base year 1992

[15] Note that in the LBD data, the establishment age is truncated at 1976 (i.e., as of 1992 the maximum age is 17 years). If the firm had zero employment in a location from 1992 to $t-1$, then region tenure at $t-1$ is 0. To incorporate the zeros, we scale the firm region tenure variable adding 1.

(*FirmOldestRegion* variable). This variable will control for regional attributes that may have induced firms to be born and continue expanding into that location (e.g., the initial strength of the cluster or regional policies to promote biopharmaceuticals implemented when the firm was born).[16] Our core findings are robust to the inclusion of the *FirmOldestRegion* variable (models (3) and (4)): coefficients for internal and external agglomerations are positive and statically significant. It is noticeable that the initial location also has a positive and significant effect on the location choice of new bio establishments.

Second, to more directly address the concern of firms born in a strong cluster, we classify a firm's initial location as either having or not having a high potential for external agglomerations (i.e., it is one of the top 10 biopharmaceutical clusters identified in Table A.1 (e.g., the Boston Economic Area)). We then estimate our models for these two subsamples. Table 6 shows the results for firms that were born in a top bio cluster (models (5a) and (5b)) and in a non-top bio cluster (models (5c) and (5d)). Note that our variable of internal agglomerations in the bio value chain (ln(*FirmBioValueChain*)) is significant in both subsamples. Importantly, this variable is significant (and its magnitude statistically higher) in the subsample of non-top initial cluster,[17] suggesting that same-firm collocation of bio activities is not driven by the external environment of the initial location. Overall, these analyses show that the conditions of the initial location do not explain our main findings.

4.3.3. Strategic Behavior in Location Choices. Previous studies argue that location choices may be driven by strategic considerations concerning firms' internal strengths relative to the presence of competitors in a location. For example, Shaver and Flyer (2000) show that leader firms shy away from clusters to avoid generating externalities that may help competitors. Incorporating their findings into our framework is not straightforward for two reasons. First, the empirical evidence comes from firms' first-time location choices in a host country, so there is no potential for internal agglomerations (Alcácer and Chung 2014). Second, the conceptual arguments explain why a leader firm *would not* locate in a strong cluster but do not provide predictions of where the firm *would* locate instead; it could choose locations with same-firm employment or locations where the firm has never been before. Although this literature does not provide an explicit argument that links competition to persistence in firm location, one may argue that another interpretation of same-firm collocation is that firms may not move to strong clusters but rather collocate same-firm activity in existing locations to avoid competitors. To explore this interpretation, we follow Shaver and Flyer (2000) and Alcácer and Chung (2014) and estimate our baseline models in two subsamples of small and large firms with size defined using total employment (note that previous research has proxied strategic behavior by differences in findings between small and large firms). Specifically, firms are classified as small or large if their employment in 1992 is below or above the pseudo median value for the full sample. Results for this analysis are shown in Table 6 (columns (6a) and (6b) for small firms and (6c) and (6d) for large firms) and are similar to our baseline results.[18] Note that for both subsamples, internal agglomerations in the bio value chain matter. Interestingly, the effect is significantly higher for smaller firms (i.e., the coefficient of *FirmBioValue-ChainEmployment* is statistically higher at 5% based on SUR estimates). External drivers also matter for both subsamples, but the size of the region matters significantly more for larger firms. These findings suggest that in biopharmaceuticals, same-firm collocation is not driven by strategic considerations by larger firms that are trying to avoid clusters, and internal agglomerations are also important for smaller firms.

4.4. Same-Firm Collocation and Survival

The analysis so far has focused on location choices and relies on the assumption that evidence of location should capture the value firms expect to obtain from agglomeration economies. However, there could be other unobservables (outside the ones explored in §4.3) that could impact location choices beyond expected gains from internal agglomeration. Examples of such unobservables are satisficing behavior or preferences for the familiar and obvious. To condition out these potential factors, we explore the survival of the new establishments that were opened during the period of analysis (i.e., the 1,226 new bio activities examined in the baseline location models in Table 4).[19]

[16] Note that this analysis also speaks to the social mechanisms because the initial location is the region with the longest tenure of the firm, and thus, it is associated with the earliest social network of the firm.

[17] The coefficient is significantly higher in model (5c) than in (5a), and in model (5d) than (5b), at the 10% significance level (using SUR estimates to compare coefficients across subsamples).

[18] The same findings hold when we consider a more extreme definition of small and large firms (bottom and top quartiles in employment size).

[19] We choose survival as the performance variable because it is available at the establishment level and is the same measure across activities in the value chain (i.e., other performance indicators such as productivity are more appropriate for manufacturing than R&D).

Table 7 Survival of New Bio Activity (157 Firms; 6,027 Observations; Cox Model; Hazard Ratio)

	Y_{firt} = Exit of new bio activity i (157 firms and 1,226 subjects)	
Variables specified at t	(1)	(2)
ln(*FirmBioValue-ChainEmployment*)	0.925*	0.916**
	(0.030)	(0.027)
ln(*FirmNon-BioValue-ChainsEmployment*)	1.048	1.037
	(0.036)	(0.035)
ln(*FirmSupportEmployment*)	0.966	0.967
	(0.046)	(0.041)
Region (EA) fixed effects	Yes	No
ln(*RegionBioValue-ChainSpecialization*)		0.957
		(0.064)
ln(*RegionEmployment*)		1.115
(outside bio value chain)		(0.070)
Firm fixed effects	Yes	Yes
Bio activity fixed effects	Yes	Yes
Log-likelihood	−3,238	−3,257

Notes. Standard errors are clustered by firm. Failures: 528 (out of 1,226). The internal agglomeration variables exclude the new bio establishments' own employment.

**Significant at the 1% level; *significant at the 5% level.

Specifically, we estimate the following equation:

$$Exit_{firt} = \beta \ln(X_{firt}^{\text{Internal}}) + \gamma_r + \gamma_f + \gamma_i + \varepsilon_{firt}, \qquad (6)$$

where *Exit* is equal to 1 if firm f's new bio activity i that was introduced in location r during our time period exits (because of death or acquisition), and $X_{firt}^{\text{Internal}}$ is a set of variables that captures firm f's geographic footprint of activities in EA r at time t (excluding establishments' own employment). The model also includes EA (γ_r), firm (γ_f), and bio activity (γ_i) fixed effects. We estimated Equation (6) using a Cox proportional hazard model of the survival of each new biopharmaceutical activity. Table 7 shows the results of this analysis. Firms were less likely to close a bio establishment in a region where the firm had other bio activities (i.e., the hazard ratio is less than 1 for ln(*FirmBioValue-ChainEmployment*)). The strength of biopharmaceutical firms in the region has a positive but insignificant effect. This could capture the fact that acquisitions (i.e., exits) will happen in strong clusters if acquiring firms are trying to access good locations. (See the drivers of acquisitions in §5.2.)

4.5. Expanding in an Existing Establishment

So far, our analysis has focused on location choices for *new* establishments. However, internal agglomerations may also stem from expansion in *existing* establishments. For example, when firms grow within the same establishment, coordination and control costs will be the lowest and interdependencies the highest. Thus, failing to recognize that an expansion in an existing establishment is equivalent to choosing again the same location could underestimate the relevance of internal agglomerations. Furthermore, if the locations where a firm is already present are rich with external agglomerations, abstracting from growth in existing establishments may also underestimate the importance of external agglomeration. Therefore, to further understand the role of internal agglomerations on location choices, we estimate our baseline models using the location decisions for both new and existing establishments.

We define the location for an existing establishment as equal to 1 if the firm chooses EA r to increase employment in an existing establishment in bio activity i at time t at a level that is larger than the median size of new establishments opened during 1993–2005 for the same bio activity.[20] There were 2,337 expansions in existing establishments in the bio value chain in our sample. Perhaps not surprisingly, expansions via new establishments are a much rarer phenomenon than expansions via existing establishments. Finally, we generate the variable $LocationAll_{firt}$, which is equal to 1 if firm f either opens a new establishment or increases employment in an existing establishment above the threshold in EA r, in focal bio activity i at time t.

Table 8 introduces results of estimating Equation (1) for expansions in both existing and new establishments (*LocationAll*), which includes 335 firms (60% of the biopharmaceutical firms) and more than 3,300 location choices. Model (1) only includes the EA fixed effects; model (2) adds our measures of internal agglomerations. Including these variables improves the model fit (the log-likelihood goes up significantly relative to model (1)). As expected, the main internal agglomerations occur within the firm bio value chain, and the estimated coefficient is significantly larger than when we only model the locations of new establishments. Consistent with our prior findings, there are also internal agglomerations in other value chains and in support activities.[21]

[20] The pseudo-median size of new establishments was computed using the range of percentiles 40–60 to avoid disclosure problems. The thresholds corresponded to values of 7, 45, and 21 employees for bio R&D, manufacturing, and sales, respectively. We considered alternative thresholds to define internal expansions that take into account the size of the expanding firms. The new thresholds are the median size of new biopharmaceutical establishments by bio activity and firm size (firms are coded as small (large) if their size is below (above) the median size of expanding firms in 1992). The number of expansions only increases slightly with this new criterion.

[21] We considered two alternative samples for expansions in existing establishments. First, we dropped internal expansions of establishments that were created after 1992 (25% of the expansions) because new establishments are expected to grow in a given location across time. The results in Table 8 are very similar after dropping these observations. Second, as described above, we considered alternative thresholds to define internal expansions that take into account the size of the expanding firms; the same findings hold.

Table 8 Location of New and Existing Establishments (Conditional Logit Model; Odds Ratios)

	Y_{first} = Location choices for new and existing establishments, bio activity i		
	(1)	(2)	(3)
ln(*FirmBioValue-Chain Employment*)		2.520** (0.100)	2.575** (0.108)
ln(*FirmNon-BioValue-Chains Employment*)		1.098** (0.033)	1.060 (0.035)
ln(*FirmSupportEmployment*)		1.082* (0.035)	1.076* (0.032)
Region (EA) fixed effects	Yes	Yes	No
ln(*RegionBioValue-Chain Specialization*)			1.177* (0.065)
ln(*RegionEmployment*) (outside bio value chain)			1.506** (0.071)
Groups: Firm-bio activity-year	1,490	1,490	1,490
Pseudo R^2	0.26	0.57	0.55
Log-likelihood	−9,511	−5,556	−5,790
Firms	335	335	335
Observations	216,050	216,050	216,050

Notes. Standard errors are clustered by firm. Results are robust to using the weighted internal agglomeration variables. The choice set includes 145 EAs selected for expansion.
**Significant at 1% level; *significant at 5% level.

In summary, our findings offer relevant new insights. First, same-firm collocation plays a role in the location of biopharmaceutical activity and they arise in the bio value chain, in other value chains, and in support activities. Second, the effect of same-firm collocation varies by activity (R&D, manufacturing, and sales). Third, evidence of same-firm collocation emerges both within and across activities. Fourth, although external agglomerations are also important, their effect may be overestimated if we abstract from the spatial organization of the firm.

The results described in §§4.3 and 4.4 show that same-firm collocation in the bio value chain persists even after controlling for sociological drivers, initial location, and strategic location behavior, and it is positively associated with survival. These findings suggest that same-firm collocation captures internal agglomerations. Additionally, variations in the findings by value chain type and by activity may not be explained by some of the alternative explanations. For example, there are no clear arguments linking firm initial location or sociological drivers to the finding that same-firm collocation is smaller for sales than R&D and manufacturing, or to the finding that for any type of activity collocation with support activity is smaller than collocation with the bio value chain.

5. Exploring Firm Heterogeneity and Robustness Analysis

Before settling on these results, we conducted several robustness tests using different ways to capture firm heterogeneity, samples, and explanatory variables.

5.1. Exploring Firm Heterogeneity

Whereas our model accounts for firm-activity-year fixed effects, it is possible that additional unobserved firm heterogeneity influences firms' location decisions. We addressed this concern in several ways. First, we used "weighted" internal agglomeration with firm-specific weights to account for the extent of relatedness between activities of the value chain. Second, we explored the geographical diversification and activity diversification of firms because location choices in these cases could be more idiosyncratic, as explained below.

5.1.1. Weighted Internal Agglomerations: Exploring Firm Heterogeneity in the Value Chain. Our main internal agglomeration variables used so far assume that all activities in a firm value chain are equally related. Here, we develop a new set of internal agglomeration variables that account for firm-level heterogeneity in the extent of relatedness between value chain activities. For a firm, two activities (say, R&D and manufacturing) may have meaningful and geographically bounded linkages (knowledge flows, input–output links, etc.) and could then benefit from collocation, whereas other pairs of activities of the firm may be less related (say, R&D and sales). To account for this, we developed internal economies variables that weight more heavily the employment that seems most relevant for the focal bio activity i. Similar to Equations (2)–(4) (in §3.2.1), the set of weighted internal agglomeration variables is defined as follows:

$$FirmBioValue\text{-}ChainEmployment(weighted)_{first-1} = \sum_{k\in \text{BioActivities}} w_{fikt-1} emp_{fkrt-1}, \quad (7)$$

$$FirmNon\text{-}BioValue\text{-}ChainsEmployment(weighted)_{first-1} = \sum_{k\in \text{Non-BioActivities}} w_{fikt-1} emp_{fkrt-1}, \quad (8)$$

$$FirmSupportEmployment(weighted)_{first-1} = \sum_{k\in \text{Support}} w_{fikt-1} emp_{fkrt-1}, \quad (9)$$

where i is the focal bio activity of a new establishment subject to location choice; k indexes all activities that map into the bio value chain (Equation (7)), into non-bio value chains (Equation (8)), and into support (Equation (9)); emp_{fkrt-1} is the employment in firm-EA-activity k the year prior to expansion; and w_{fik} is the firm-specific relatedness between pairs of activities i and k (e.g., bio R&D and bio man-

Table 9 Location of New Establishments: Alternative Internal Agglomeration Variables (157 Firms; 50,920 Observations; Conditional Logit Model; Odds Ratios)

	Y_{firt} = Location choices for new establishment in bio activity i			
	Internal Agglomeration$_{frt-1}$: *Weighted employment*		*Internal Agglomeration*$_{frt-1}$: *Specialization*	
	(1)	(2)	(3)	(4)
ln(*FirmBioValueChain*)	1.286** (0.042)	1.312** (0.035)	1.195** (0.044)	1.221** (0.038)
ln(*FirmNon-BioValueChains*)	1.118** (0.033)	1.103** (0.038)	1.097** (0.017)	1.107* (0.018)
ln(*FirmSupport*)	1.109** (0.036)	1.128** (0.032)	1.100** (0.021)	1.114** (0.017)
Region (EA) fixed effects	Yes	No	Yes	No
ln(*RegionBioValue-Chain Specialization*)		1.214* (0.109)		1.256** (0.093)
ln(*RegionEmployment*) (outside bio value chain)		1.857** (0.094)		1.875** (0.088)
Groups: Firm-bio activity-year	380	380	380	380
Pseudo R^2	0.26	0.22	0.25	0.22
Log-likelihood	−2,935	−3,088	−2,970	−3,124

Note. Standard errors are clustered by firm.
**Significant at the 1% level; *significant at the 5% level.

ufacturing in Equation (7)).[22] Although we do not observe the relatedness between a pair of activities of a firm, we infer it from their collocation pattern across locations. Building on Porter (2003), we calculated w_{fikt-1} using the "locational correlation" of employment. This measure is the correlation coefficient between firm employment in activity i and firm employment in activity k across regions (EAs) where the firm had any positive employment: $LC_{fikt-1} = Correlation(emp_{firt-1}, emp_{fkrt-1})$. The greater the collocation of a pair of firm activities across regions in the past (LC_{fik}), the greater the extent of geographically bounded linkages between those activities. For a given pair of activities, some firms may have developed management practices or communication technologies that allow them to break down the value chain across locations (Fort 2011). In such a case, even if the pair of activities has meaningful linkages, the optimal level of collocation might be reduced (resulting in a lower LC_{fik} for these firms). The locational correlation weights were then transformed to values between 0 and 1 to compute the internal agglomeration variables ($w = (1 + LC)/2$).[23]

[22] For example, if the activity i of a new establishment was bio R&D, the *FirmBioValue-ChainEmployment* variable would be computed as follows: $emp_{firt-1} + w_{f,i,\text{SIC-2830},t-1} emp_{f,\text{SIC-2830},r,t-1} + w_{f,i,\text{SIC-5120},t-1} emp_{f,\text{SIC-5120},r,t-1}$, where $w_{f,i,\text{SIC-2830},t-1}$ and $w_{f,i,\text{SIC-5120},t-1}$ capture the firm-specific relatedness of R&D with manufacturing and with sales, respectively.

[23] We assumed that if a firm had only one EA, $w = 1$ for all pairs of activities in the location. We also assumed that the firm's new bio activities (i.e., the firm diversified into new activities) are related to its existing activities ($w = 1$).

Table 9 shows the results of estimating Equation (1) using the weighted internal agglomeration variables. Findings using the weighted internal agglomeration variables are robust in statistical significance to those using the unweighted variables, but weighted internal economies tend to result in slightly lower coefficients. For example, the estimated odds ratio for *FirmBioValue-ChainEmployment* is 1.286 when weighted (model (1), Table 9) and 1.365 when unweighted (model (2), Table 4). This suggests that accounting for firm-specific interdependencies across activities (and the associated unobserved management practices that mitigate internal agglomerations) may offer a more accurate estimate of the role of internal agglomerations.

5.1.2. Firm Geographical Diversification. There may be some unobserved attributes that drive firms to be specialized in one location (EA). Firms that have a single location could be problematic since the initial external drivers that pulled them to that location may not be disentangled from internal forces, inducing them to expand in the same location. Very few firms in our sample had a single EA the year prior to the expansion, and our findings are robust to their exclusion.

We explored further the role of the geographic diversity of firms by dividing them into two groups: those with few EAs versus those with many EAs in 1992 (i.e., firm number of EAs below/above the pseudo-median value of five EAs in 1992). For both subsamples, internal agglomerations in the bio value chain matter, and the effect is significantly higher

for firms with few locations. This finding would be consistent with geographically diversified firms having better managerial capabilities to break their value chain across distant locations, and so their potential for internal agglomerations is reduced.

Finally, we also controlled for the diversification of firms into new locations by including a dummy equal to 1 if the firm had no employment in a location at $t-1$ (*NewRegion*). With the inclusion of the *NewRegion* dummy, we address the concern that firms may locate where they have some prior presence because of decision-making biases rather than because of the expected internal agglomerations associated with the level of same-firm employment present in the region (i.e., if the firm has a relevant presence in the region, the potential internal agglomerations should be greater than if the firm has a marginal presence in the region). As expected, we find that firms are less likely to choose locations with no firm presence (negative and significant coefficient of the *NewRegion* variable). Importantly, the coefficients of our main internal agglomeration variables continue to be positive and statistically significant (i.e., all the findings through Tables 4 and 5 are robust). This analysis reinforces the importance of internal agglomerations and reduces the concerns about firm persistence in location choices resulting from other factors.

5.1.3. Firm Activity Diversification. We examined the vertical and horizontal diversification of biopharmaceutical firms. Smaller biotech firms may be more specialized (often in R&D), whereas big pharmaceutical firms tend to be more diversified, with multiple activities in the bio value chain as well as diversification into other value chains (e.g., medical devices, downstream chemical products). The extent of firm diversification may capture an unobserved strategy choice of the firm when it was founded (Gans et al. 2013). For example, some firms' strategy is to specialize in some core activities and rely on other firms for complementary assets. This may be the case for many small biotech firms. By contrast, other firms may choose a more vertically (and horizontally) integrated value chain.

We explored how a firm's degree of *horizontal diversification* influences the extent of internal agglomerations by considering the location choices of two types of firms that open new bio establishments: *HorizontallyDiversified* versus *BioSpecialized*. *HorizontallyDiversified* firms had activities in both bio and non-bio value chains at $t-1$; the rest of firms were classified as *BioSpecialized*. We estimated Equation (1) for both subsamples and found similar patterns to those in our baseline results: internal agglomerations matter and arise primarily within the bio value chain. However, the collocation with same-firm bio value chain is significantly smaller for horizontally diversified firms than for bio-specialized firms.[24] This would be consistent with diversified firms having managerial capabilities to coordinate activities across locations, perhaps reducing the need to collocate same-value chain activities and instead exploiting internal agglomerations across value chains (i.e., positive and significant effect of ln(*FirmNon-BioValueChainsEmployment*) on the location of new bio activity).

Similarly, we explored the role of a firm's degree of *vertical diversification* in the extent of internal agglomerations by considering two types of firms: *BioDiversified* and *Bio-ActivitySpecialized*. *BioDiversified* firms have two or more bio activities (e.g., SIC codes 8731 and 2830) at $t-1$, and *Bio-ActivitySpecialized* firms are the subset of firms that are most specialized with employment in a single bio activity at $t-1$ (e.g., the firm only has employment in SIC code 8731). We estimated Equation (1) and found that for both subsamples internal agglomerations arise within the bio value chain, but the effect is smaller for biodiversified firms. Overall, the analysis suggests that internal agglomerations within the bio value chain matter for specialized and (vertically and horizontally) diversified firms, but the effect seems smaller for diversified firms.

5.2. Alternative Samples and Variable Definitions

To address some additional econometric concerns, we estimated additional models using different samples and variable definitions. In terms of sampling, we considered three additional samples of bio firms and their location choices. First, we examined the subsample of firms that expand through acquisitions of bio establishments. This analysis allowed us to contrast the location choices of new establishments with those of acquisitions and to better assess the role of internal and external agglomerations in location choices. The goal of many acquisitions is to buy technology, patents, or other assets of a firm (Furman et al. 2005) rather than to make a location choice motivated by internal agglomerations. In other cases, acquisitions may be motivated by access to a strong biopharmaceutical cluster. We ran our baseline location models and found that internal and external agglomerations in the bio value chain matter for the location of acquisitions (see Table A.2 in the appendix, models (1) and (2)).[25] However, the magnitude of the coefficient for internal agglomerations is statistically smaller for

[24] This finding is robust to controlling for *FirmRegionTenure* and *FirmOldestRegion* (see §4.3).

[25] The finding is also robust for the larger sample of expansions through new bio establishments or acquisitions (198 firms; not reported).

the location of acquisitions than for new establishments.[26] The effects of the external agglomeration variables are larger for the sample of acquisitions than for the sample of new establishments (but the difference is not significant). Thus, our findings are consistent with acquisitions being less motivated by internal agglomerations than are greenfield investments.

Second, we examined the expansion and location decisions of all 563 bio firms to address potential selection bias from excluding firms that did not expand during our time period. To do so, we specified the following linear probability model:

$$LocationAll_{firt} = \beta \ln(X^{\text{Internal}}_{firt-1}) + \gamma_r + \gamma_f + \gamma_i + \gamma_t + \varepsilon_{firt}. \quad (10)$$

The dependent variable considers the location of both new and existing bio establishments. For firms with no organic expansion in a given bio activity-year, the *LocationAll* variable is always 0 across all EAs in the choice set (145 EAs).[27] The model includes the same explanatory variables as in Equation (1) and firm, bio activity, and year fixed effects. We estimated Equation (10) by the ordinary least squares (OLS) method and found collocation of same-firm bio value-chain activities and a positive effect for the external agglomeration variables (Table A.2, models (3) and (4)).[28] This suggests that our core findings are not driven by selection issues.

Third, we applied specific criteria to identify biopharmaceutical firms from the LBD data. Because it may be possible that some of these firms are not actually in biopharmaceuticals, we reestimated our models using only firms that were in the BioScan directory. Results using this subsample of firms are very similar to those obtained from the extended sample across specifications.

In terms of the definition of the location choice, the analysis so far has focused on a categorical variable (whether or not to enter a location). Alternatively, we modeled the intensity of the entry decision by using as a dependent variable the level of employment created by a firm in a region-activity-year through new bio establishments (the *EntryEmployment* variable). We estimated a negative binomial model of *EntryEmployment* using the same independent variables as in our baseline specifications and including firm-bio activity-year fixed effects (i.e., the same groups as in our conditional logit model). Our core findings remained robust: internal agglomerations arise in the bio value chain, and the strength of biopharmaceuticals in the region also positively influences the level of entry in a region (Table A.2, models (4) and (6)).

In terms of alternative internal agglomeration variables, we also developed internal agglomeration variables that measure the extent of employment specialization. The goal was to assess whether a firm locates in regions where the firm's activities are overrepresented. These variables will further confirm that firm location choices are not random but instead driven by internal agglomerations. Our measure of firm-level regional specialization in the bio value chain (*FirmBioValue-ChainSpecialization*) captures whether a firm's employment in these activities is overrepresented in a particular location (given the size of the firm and its portfolio of activities and locations), and it is defined as follows:

$$FirmBioValue\text{-}ChainSpecialization_{frt-1} = \frac{\sum_{k\in \text{BioActivities}} emp_{fkrt-1}}{emp_{frt-1}} \times \left[\frac{\sum_{k\in \text{BioActivities}} emp_{fkt-1}}{emp_{ft-1}}\right]^{-1}, \quad (11)$$

where k indexes all firm activities (four-digit SIC codes) that map into the bio value chain, and the employment variables are as follows: emp_{fkrt-1} and emp_{frt-1} are, respectively, employment in firm-activity-EA and in firm-EA at $t-1$; emp_{fkt-1} aggregates employment by firm-activity; and emp_{ft-1} is firm total employment. For example, if a firm locates 100% of its bio value-chain employment in location r and has 50% of its total employment in the bio value chain, then this variable will equal 2 in that location (and 0 in all other locations). Similarly, we compute *FirmNon-BioValue-ChainsSpecialization* and *FirmSupportSpecialization*.

Table 9, models (3) and (4) show the estimates using these alternative variables. The findings are robust in statistical significance to those in Table 4, but internal agglomerations based on specialization tend to result in lower coefficients. More broadly, all our findings across Tables 4–8 are robust to using these alternative variables.

Finally, in terms of defining value chains, we tested whether our method of mapping bio value-chain activities to SIC codes drives our findings on within- and across-activity collocation. Specifically, we defined new internal agglomeration variables considering a single value chain and examined collocation across broadly defined activities (i.e., without separating bio and non-bio value chains). This redefinition has the benefit of reducing sparsity for the explanatory variables as well as relaxing the definition of the core value chain of a biopharmaceutical firm. Findings with this new

[26] The comparison of the acquisitions and new establishments models is based on SUR estimates.

[27] This specification allows us to add all the zeros (i.e., for existing firms that never expand or that do not expand in a specific bio-activity-year).

[28] Findings are the same if we define location choices based only on the opening of new establishments and using our alternative internal agglomeration variables.

definition are reported in Table A.3 in the appendix and confirm prior results in Table 5 regarding within- and across-activity internal agglomerations.

6. Conclusions

This paper examined the extent to which the geographical location of distinct activities in the value chain (R&D, manufacturing, or sales) is explained by internal agglomerations (i.e., geographically bounded intrafirm linkages) and external agglomerations. We proposed a conceptual framework where both internal and external agglomerations drive location choices. Both forces can work in the same direction (when a firm is already located in a cluster) or in opposite directions. In the latter scenario, external agglomerations act as centrifugal forces that drive firms to disperse their activities geographically; internal agglomerations act as centripetal forces that drive within-firm collocation.

Based on a review of relevant papers across disciplines, we identified five mechanisms through which geographic proximity may lead to internal agglomerations: coordination, control, knowledge and information flows, economies of scope and scale in internal labor markets, and access to intermediate inputs. The last three parallel the mechanisms identified as external agglomerations by Marshall (1920). Although the focus of our paper is empirical, and we cannot unbundle these alternative mechanisms, we believe that our conceptual framework is a first step toward building theory that advances our understanding of internal agglomerations and their impact on firm performance.

We tested our conceptual framework using data on the location of new establishments opened by biopharmaceutical firms in the continental U.S. during 1993–2005. Our findings offer several relevant insights. First, internal agglomerations have a relevant, positive impact on location. They are especially important within the focal value chain (bio value chain), and they also arise in related value chains of the firm. Second, the effects of internal agglomerations vary by activity, and they arise both within an activity (e.g., among plants) and across activities (e.g., between sales and manufacturing). Third, external agglomerations also have a relevant positive impact on location, but its effect declines if internal agglomerations are considered. Note that this decline in the estimated effect of external agglomerations suggests that in our analysis the external and internal agglomeration variables are positively related, and firms may be exploiting both in the same location (e.g., firms that are located in a strong cluster). Our findings remain after considering alternative explanations of same-firm collocation, such as sociological drivers, initial location, strategic location behavior, and firm heterogeneity.

These insights offer important contributions to the literature. Because both internal and external agglomerations play a role in the location of firms, focusing on only one type (external or internal) may produce biased estimates as a result of omitted variables. Regarding internal agglomerations, failing to consider *all* activities in the value chain may lead to omitted relationships that can also bias results. For example, the location of firm R&D may be influenced not only by the presence of same-firm R&D and manufacturing but also by the presence of same-firm sales and support activities. Furthermore, to better understand the links between activities (e.g., manufacturing and R&D), we need to examine the location decision of each activity. Finally, considering just one category of expansions (new establishments) may also underestimate the importance of internal agglomerations. For example, we show that if we also consider firms' expansions by increasing activity levels in existing establishments, the estimated effect of internal agglomerations is greater. After all, the default alternative to a new location is to stay in an existing location—an insight that is absent in most of the location literature.

All of these issues—the relevance of internal and external agglomerations, the distinct relationships between pairs of activities in the value chain, and the fact that effects may vary by the type of expansion—emphasize the need for a comprehensive framework, both at the theoretical level and at the empirical level, to understand the spatial organization of firms.

Several avenues for further research remain. First, although our results suggest the existence of internal agglomerations, we do not measure them directly. Future research that isolates each of the five mechanisms through their detailed measurement is a desirable direction for future papers. Based on more recent similar efforts in papers focused on external agglomerations (Alcácer and Chung 2007, 2014; Glaeser and Kerr 2009), this process will not be straightforward. Besides isolating mechanisms, future work must also further examine the effects of internal and external agglomerations on the performance of new facilities and the firm as a whole. Although our survival analysis is a first step in this direction, more refined performance measures could be adopted. These measures may need to vary by activity to capture the fact that, for example, R&D labs may maximize patents while manufacturing facilities focus on productivity. Second, the current empirical analysis focuses on location decisions of biopharmaceutical firms. One could extend the analysis to other industries with different degree of modularity (Baldwin and von Hippel 2011). Biopharmaceuticals has lower modularity than other industries, such as semiconductors and automotives, and

this could affect the extent and types of agglomerations. Third, since industry maturity influences firms' location choices (e.g., Dumais et al. 2002, Duranton and Puga 2001), the effect of internal and external agglomerations may vary across time, and our results may reflect a specific stage in the life cycle of biopharmaceuticals. Relatedly, the life cycle of firms may also influence location choices. For example, external agglomerations are important for start-ups (Glaeser and Kerr 2009, Delgado et al. 2010) and smaller firms (Henderson 2003, Rosenthal and Strange 2010), but as firms get larger, the trade-off between internal and external agglomerations may vary. Future research that explores these industry and firm dynamics would greatly enrich our understanding of the spatial organization and performance of firms. Fourth, strategic choices that firms make regarding activity diversification versus specialization can also influence the ability to exploit internal and external agglomerations. We found that internal agglomerations matter for horizontally and vertically diversified firms and, especially, for specialized firms. More research will be required to understand the relative strength of internal versus external agglomeration and the particular agglomeration mechanisms used by specialized versus diversified firms. Fifth, although we have accounted for the spatial organization of firms with a novel approach, we did not examine the business practices that shape firms' spatial organization. The location choice and performance of firms may depend on firm management practices that facilitate intra- and interfirm interactions even with geographic separation (e.g., outsourcing practices, labor mobility practices, monitoring and control practices, information and communications technology investments; see, e.g., Baldwin and von Hippel 2011, Choudhury 2010, Fort 2011). The examination of these management practices could help identify particular internal and external agglomeration mechanisms. Finally, we did not look at overseas expansions because of data limitations, and the exclusion of foreign expansions may bias our internal agglomeration estimates. For example, firms that are expanding globally are likely to have the managerial capabilities and communication technology to break R&D or manufacturing across locations, and so they may benefit less from internal agglomerations. Consistent with this argument, we found that for firms that are geographically diversified within the United States (which are more likely to open facilities abroad), internal agglomerations do matter for location choices, but to a lesser extent than they matter to geographically specialized firms. Future research will be needed to evaluate the role of internal and external agglomerations when firms expand globally (e.g., Beugelsdijk et al. 2010).

Acknowledgments

For helpful comments, the authors are grateful to the editor and anonymous reviewers, to Ashish Arora, Pierre Azoulay, Ian Cockburn, Wes Cohen, Jeff Furman, Alfonson Gambardella, Mat Marx, Will Mitchell, Ram Mudambi, Joanne Oxley, David Ridley, Scott Stern, and participants in seminars at Bocconi University, the Carlson School of Management, the Fuqua School of Business, Duke University, the University of Minnesota, the Rotman School of Management, MIT Sloan School of Management, the University of Toronto, the National Bureau of Economic Research Productivity Seminar, the U.S. Census Bureau Center for Economic Studies, the DRUID conference, the Fox School of Business/Kauffman Foundation Workshop on Empirical Entrepreneurship, and the Academy of Management and Strategic Management Society annual meetings. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. M. Delgado gratefully acknowledges support from the Center for International Business Education and Research (CIBER) at Temple University.

Appendix

Table A.1 Top 10 Economic Areas by Share of U.S. Biopharmaceuticals Employment in 1992 (with High Specialization)

Economic area name	Share U.S. bio employment (%)	Bio specialization	Bio employment
New York–Newark–Bridgeport, NY–NJ–CT–PA	18.5	2.12	101,972
Chicago–Naperville–Michigan City, IL–IN–WI	6.7	1.58	36,661
Washington–Baltimore–Northern Virginia, DC–MD–VA–WV	4.7	1.52	26,010
Boston–Worcester–Manchester, MA–NH	4.7	1.37	25,678
Philadelphia–Camden–Vineland, PA–NJ–DE–MD	4.4	1.61	24,149
San Jose–San Francisco–Oakland, CA	4.3	1.29	23,858
Raleigh–Durham–Cary, NC	1.7	1.95	9,362
Indianapolis–Anderson–Columbus, IN	1.7	1.36	9,107
Knoxville–Sevierville–La Follette, TN	1.5	4.20	8,274
San Diego–Carlsbad–San Marcos, CA	1.5	1.69	8,262

Notes. Authors' calculations based on County Business Patterns data. "Bio" refers to the biopharmaceuticals value chain. The criterion for high bio specialization is a value greater than the 80th percentile of the Location Quotient.

Table A.2 Location Choices: Alternative Samples and Dependent Variable

	Location of acquisitions (conditional logit)		Location all (OLS)		Entry employment (negative binomial)	
	(1)	(2)	(3)	(4)	(5)	(6)
ln(*FirmBioValue-ChainEmployment*)	1.157**	1.170**	0.0165**	0.0166**	1.448**	1.446**
	(0.040)	(0.041)	(0.0007)	(0.0007)	(0.020)	(0.019)
ln(*FirmNon-BioValue-ChainsEmployment*)	1.026	1.028	0.0001	0.0001	0.970	0.946**
	(0.039)	(0.031)	(0.0001)	(0.0001)	(0.017)	(0.016)
ln(*FirmSupportEmployment*)	1.149*	1.148*	0.0010**	0.0011**	1.180**	1.175**
	(0.068)	(0.070)	(0.0003)	(0.0003)	(0.022)	(0.021)
Region (EA) fixed effects	Yes	No	Yes	No	Yes	No
ln(*BioValue-ChainSpecRegion*)		1.387**		0.0001**		1.134**
		(0.153)		(0.0000)		(0.045)
ln(*EmploymentRegion*) (outside bio value chain)		2.179**		0.0001**		1.520**
		(0.224)		(0.0000)		(0.046)
Groups: Firm-bio activity-year	208	208	No	No	380	380
Firm fixed effects			Yes	Yes		
Bio activity fixed effects			Yes	Yes		
Year fixed effects			Yes	Yes		
R^2	0.20	0.17	0.08	0.08		
Log-likelihood	−1,128	−1,163			−8,007	−8,152
Firms	98	98	563	563	157	157
Observations	18,304	18,304	2,431,215		50,920	50,920

Notes. Models (1) and (2) report odds ratios; standard errors are clustered by firm. In models (3) and (4), standard errors are clustered by firm-bio activity-year (16,767 groups). Models (5) and (6) report incidence rate ratios.

**Significant at the 1% level; *significant at the 5% level.

Table A.3 Location of New Establishments: Collocation Within and Across Activities with a Single Value Chain (Conditional Logit; Odds Ratios)

	Y_{firt} = Location choices for new establishment in bio activity *i*		
Single value chain	Bio R&D (1)	Bio mfg (2)	Bio sales (3)
ln(*FirmR&DEmployment*)	1.304**	1.374**	1.162
	(0.104)	(0.166)	(0.113)
ln(*FirmMfgEmployment*)	1.215*	1.471**	1.243**
	(0.111)	(0.104)	(0.069)
ln(*FirmSalesEmployment*)	1.214	1.077	1.062
	(0.136)	(0.124)	(0.061)
ln(*FirmSupportEmployment*)	1.150*	1.034	1.023
	(0.075)	(0.085)	(0.069)
Region (EA) fixed effects	Yes	Yes	Yes
Groups: Firm-bio activity-year	192	72	116
Pseudo R^2	0.26	0.21	0.22
Log-likelihood	−1,585	−258	−840
Firms	92	48	60
Observations	23,232	3,096	9,164

Notes. Standard errors are clustered by firm. Findings are robust to using the alternative internal agglomeration variables (weighted and based on specialization).

**Significant at the 1% level; *significant at the 5% level.

References

Adams J, Jaffe A (1996) Bounding the effects of R&D: An investigation using matched establishment-firm data. *RAND J. Econom.* 27(4):700–721.

Aharonson B, Baum J, Plunket A (2008) Inventive and uninventive clusters: The case of Canadian biotechnology. *Res. Policy* 37(6–7):1108–1131.

Alcácer J (2006) Location choices across the value chain: How activity and capability influence collocation. *Management Sci.* 52(10):1457–1471.

Alcácer J, Chung W (2007) Location strategy and knowledge spillovers. *Management Sci.* 53(5):760–776.

Alcácer J, Chung W (2014) Location strategies for agglomeration economies. *Strategic Management J.* 35(12):1749–1761.

Atalay E, Hortaçsu A, Syverson C (2014) Vertical integration and input flows. *Amer. Econom. Rev.* 104(4):1120–48.

Audretsch DB, Feldman MP (1996) R&D spillovers and the geography of innovation and production. *Amer. Econom. Rev.* 86(4):253–273.

Azoulay P (2004) Capturing knowledge within and across firm boundaries, evidence from clinical development. *Amer. Econom. Rev.* 94(5):1591–1612.

Baldwin C, von Hippel E (2011) Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. *Organ. Sci.* 22(6):1399–1417.

Berndt ER, Cockburn I (2012) Price index for clinical trial: A feasibility study. NBER Working Paper 18918, National Bureau of Economic Research, Cambridge, MA.

Beugelsdijk S, McCann P, Mudambi R (2010) Place, space, and organization: Economic geography and the multinational enterprise. *J. Econom. Geography* 10(4):485–493.

BioScan (1992) *The Worldwide Biotech Industry Reporting Service*, Vol. 6 (Oryx Press, Phoenix, AZ).

Buenstorf G, Klepper S (2009) Heritage and agglomeration: The Akron tyre cluster revisited. *Econom. J.* 119(537):705–733.

Chacar AS, Lieberman M (2003) Organizing for technological innovation in the U.S. pharmaceutical industry. Sorenson O, Baum J, eds. *Geography and Strategy*, Advances in Strategic Management, Vol. 20 (JAI Press, Greenwich, CT), 299–322.

Chandler A (1977) *The Visible Hand: The Managerial Revolution in American Business* (Belknap Press, Cambridge, MA).

Choudhury P (2010) Seeking resources or seeking knowledge? A study of mobility and knowledge creation using micro data. Working paper.

Clark KB, Chew WB, Fujimoto T (1987) Product development in the world auto industry. *Brookings Papers Econom. Activity* 3: 729–771.

Cohen W, Levinthal D (1990) Absorptive capacity: A new perspective on learning and innovation. *Admin. Sci. Quart.* 35(1): 128–152.

Cortright J, Mayer H (2002) Signs of life: The growth of biotechnology centers in the U.S. Report, Brookings Institution, Washington, DC.

Dahl MS, Sorenson O (2012) Home sweet home: Entrepreneurs' location choices and the performance of their ventures. *Management Sci.* 58(6):1059–1071.

Delgado M, Porter ME, Stern S (2010) Clusters and entrepreneurship. *J. Econom. Geography* 10(4):495–518.

Delgado M, Porter ME, Stern S (2014) Clusters, convergence, and economic performance. *Res. Policy* 43(10):1785–1799.

Di Minin A, Bianchi M (2011) Safe nests in global nets: Internalization and appropriability of R&D in wireless telecom. *J. Internat. Bus. Stud.* 42(7):910–934.

Dumais G, Ellison G, Glaeser EL (2002) Geographic concentration as a dynamic process. *Rev. Econom. Statist.* 84(2):193–204.

Duranton G, Puga D (2001) Nursery cities: Urban diversity, process innovation, and the life cycle of products. *Amer. Econom. Rev.* 91(5):1454–1477.

Feldman MP (2003) The locational dynamics of the US biotech industry: Knowledge externalities and the anchor hypothesis. *Indust. Innovation* 10(3):311–328.

Feldman MP, Audretsch D (1999) Innovation in cities: Science-based diversity, specialization, and localized competition. *Eur. Econom. Rev.* 43(2):409–429.

Feldman MP, Schreuder Y (1996) Initial advantage: The origins of geographic concentration of the pharmaceutical industry in the mid-Atlantic region. *Indust. Corporate Change* 5(3):839–862.

Fort T (2011) Breaking up is hard to do: Why firms fragment production across locations. Working Paper CES-WP-13-35, Center for Economic Studies, U.S. Census Bureau, Washington, DC.

Furman J, MacGarvie M (2007) Academic science and the birth of industrial research laboratories in the U.S. pharmaceutical industry. *J. Econom. Behav. Organ.* 63(4):756–776.

Furman J, Kyle M, Cockburn I, Henderson R (2005) Public and private spillovers, location, and the productivity of pharmaceutical research. *Ann. d'Econom. Statist.* 79/80:167–190.

Gans J, Murray F, Stern S (2013) Choosing an entrepreneurial strategy. Working paper, University of Toronto, Toronto.

Giroud X (2013) Proximity and investment: Evidence from plant-level data. *Quart. J. Econom.* 128(2):861–915.

Glaeser EL, Kerr WR (2009) Local industrial conditions and entrepreneurship: How much of the spatial distribution can we explain? *J. Econom. Management Strategy* 18(3):623–663.

Grossman SJ, Hart OD (1986) The costs and benefits of ownership: A theory of vertical and lateral integration. *J. Political Econom.* 94(4):691–719.

Hamilton B, Nickerson J, Owan H (2003) Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *J. Political Econom.* 111(3):465–497.

Hanson G (2001) Scale economies and the geographic concentration of industry. *J. Econom. Geography* 1(3):255–276.

Helfat C (1997) Know-how complementarities and knowledge transfer within firms: The case of R&D. *Strategic Management J.* 18(5):339–360.

Henderson JV (2003) Marshall's scale economies. *J. Urban Econom.* 53(1):1–28.

Henderson R, Cockburn I (1996) Scale, scope and spillovers: The determinants of research productivity in drug discovery. *RAND J. Econom.* 27(1):32–59.

Henderson JV, Ono Y (2008) Where do manufacturing firms locate their headquarters? *J. Urban Econom.* 63(2):431–450.

Jacobs J (1984) *Cities and the Wealth of Nations: Principles of Economic Life* (Random House, New York).

Johnson KP, Kort JR (2004) 2004 redefinition of the BEA economic areas. *Surv. Curr. Bus.* 84(11):68–75.

Kalnins A, Lafontaine F (2013) Too far away? The effect of distance to headquarters on business establishment performance. *Amer. Econom. J. Microeconom.* 5(3):157–179.

Ketokivi M (2006) When does co-location of manufacturing and R&D matter? Discussion Paper 1051, Research Institute of the Finnish Economy, Helsinki, Finland.

Ketokivi M, Ali-Yrkkö J (2009) Unbundling R&D and manufacturing: Postindustrial myth or economic reality? *Rev. Policy Res.* 26(1–2):35–54.

Kleinbaum AM, Stuart TE, Tushman M (2008) Communication (and coordination?) in a modern, complex organization. Working Paper 09-004, Harvard Business School, Boston.

Krugman P (1991) Increasing returns and economic geography. *J. Political Econom.* 99(3):483–99.

Mansfield E (1988) The speed and cost of industrial innovation in Japan and the United States: External vs. internal technology. *Management Sci.* 34(10):1157–1168.

Marshall A (1920) *Principles of Economics; An Introductory Volume* (Macmillan and Co., London).

Pisano GP (1997) *The Development Factory: Unlocking the Potential of Process Innovation* (Harvard Business School Press, Boston).

Porter ME (1996) What is strategy? *Harvard Business Rev.* (November):61–78.

Porter ME (1998) Clusters and competition: New agendas for companies, governments, and institutions. Porter ME, ed. *On Competition* (Harvard Business School Press, Boston), 197–299.

Porter ME (2003) The economic performance of regions. *Regional Stud.* 37(6–7):549–578.

Ricardo D (1817) *On the Principles of Political Economy and Taxation* (John Murray, London).

Rosenthal SS, Strange WC (2004) Evidence on the nature and sources of agglomeration economies. Henderson JV, Thisse JF, eds. *Handbook of Regional and Urban Economics*, Vol. 4 (Elsevier, North-Holland, Amsterdam), 2119–2171.

Rosenthal SS, Strange WC (2010) Small establishments/big effects: Agglomeration, industrial organization, and entrepreneurship. Glaeser E, ed. *Agglomeration Economics* (University of Chicago Press, Chicago), 277–302.

Saxenian A (1994) *Regional Advantage: Culture and Competition in Silicon Valley and Route 128* (Harvard University, Cambridge, MA).

Shaver JM, Flyer F (2000) Agglomeration economies, firm heterogeneity, and foreign direct investment in the United States. *Strategic Management J.* 21(12):1175–1193.

Sorenson O, Audia PG (2000) The social structure of entrepreneurial activity: Geographic concentration of footwear production in the United States, 1940–1989. *Amer. J. Sociol.* 106(2):424–462.

Storper M (1995) The resurgence of regional economies, ten years later: The region as a nexus of untraded interdependencies. *Eur. Urban Regional Stud.* 2(3):191–221.

Tate G, Yang L (2015) The bright side of corporate diversification: Evidence from internal labor markets. *Rev. Financial Stud.* 28(8):2203–2249.

Tecu I (2013) The location of industrial innovation: Does manufacturing matter? Working Paper CES-WP-13-09, Center for Economic Studies, Washington, DC.

Toole AA (2003) Understanding entrepreneurship in the US biotechnology industry: Characteristics, facilitating factors, and policy challenges. Hart DM, ed. *The Emergence of Entrepreneurship Policy: Governance, Start-Ups, and Growth in the US Knowledge Economy* (Cambridge University Press, Cambridge, UK), 175–195.

Van den Bulte C, Moenaert RK (1998) The effects of R&D team co-location on communication patterns among R&D, marketing, and manufacturing. *Management Sci.* 44(11, Part 2):S1–S18.

Zucker L, Brewer M, Darby M (1998) Intellectual capital and the birth of U.S. biotechnology enterprises. *Amer. Econom. Rev.* 88(1):290–306.