Regional Industrial Identity and Spatial Arrangements in the U.S. Biotherapeutics Industry, 1976-2004*

Olga M. Khessina
The Johnson School of Management
Cornell University
Ithaca, NY 14853
omk4@cornell.edu
607-255-7284

Elaine Romanelli
McDonough School of Business
Georgetown University
Washington, D.C. 20057
romanele@georgetown.edu
202-687-4188

March 5, 2007

* We are grateful for suggestions from Glenn Carroll, Jerker Denrell, Gerry George, Bill Greene, Mike Hannan, Jeff Macher, Keith Ord, Martin Ruef, and seminar participants at Cornell University, Georgetown University, Harvard Business School, London Business School, the University of Maryland, the University of Southern California, and the University of Washington. We thank Martin Doyle and Alex Gilev for help with data collection.

Regional Industrial Identity and Spatial Arrangements in the U.S. Biotherapeutics Industry, 1976-2004

Abstract

To extend understanding of the spatial arrangements of industries, we examined the influence of regional industrial identity, defined as the shared understandings of audiences about legitimate places for particular types of organizations, on the entry rates of firms and investment rates of venture capitalists into U.S. regional biotherapeutics industries, 1976 through 2004. Controlling for important resources, regional industrial identity increases the entry rates of non-local *de novo* firms into regions and the rates of investment by venture capitalists with limited experience in either biotherapeutics overall or in a regional biotherapeutics industry. Our study shows that regional industrial identity affects the cross-regional movements of entrepreneurs, firms, and investors and, thus, the location and persistence of industry clusters.

Interest in the spatial arrangements of industries has grown in recent years based on evidence of collocation benefits to firms (Krugman, 1991; Jaffe, Trajtenberg, and Henderson, 1993) and to the regions in which they reside (Saxenian, 1994; Porter, 1998). Studies have examined the effects of regional resources (Zucker, Darby, and Brewer, 1998), competition among firms producing similar products and services (Baum and Haveman, 1997; Shaver and Flyer, 2000; Chung and Kalnins, 2001), knowledge spillovers (Jaffe, et al. 1993; Almeida and Kogut, 1999; Sorenson and Audia, 2000), and characteristics of regional organizational communities (Barnett and Carroll, 1987; Audia, Freeman, and Reynolds, 2006; Freeman and Audia, 2006) on organizational locations. Little research, however, has considered how economic actors make sense of regional environments, comparatively, or how their understandings, which may be only partially grounded in knowledge about objective regional characteristics, affect their location decisions.

Arthur (1990) pointed out that substantial uncertainty surrounds understandings about the nature of best resources, and thus best regions, for organizing, especially during the early period of an industry's development. Many regions may possess, or appear to possess, important resources. Under such conditions, economic actors may rely on heuristics, such as, for example, the number of firms in an industry that have already located in a region, for assessing the relative suitability of regions for particular types of organizations. Heuristics may play a role even later in the life of an industry since important information about resources is often private and thus difficult to collect (Hayek, 1945, Uzzi, 1999) and because of bounded rationality, which limits individuals' capacities to acquire and process complete information (Simon, 1947). If actors are assessing the relative suitability of regions for particular types of organizations, then better understanding of the ways in which they make sense of regional environments, on a comparative basis, is needed to explain the location and persistence of industry spatial arrangements.

Romanelli and Khessina (2005) introduced the idea of regional industrial identity, which they defined as shared understandings about the suitability of regions for particular types of organizations, as a useful construct for characterizing regional industrial characteristics in a parsimonious and comparative way. For example, London, England has a strong identity as a world center for financial services and Houston, Texas as a U.S. center of the oil industry. Audiences associate these regions with these industries even without being able to count the number of firms that populate these regional industries, assess the amount of relevant resources, or evaluate the importance of related industries that inhabit a region. When regional industrial identities are strong, i.e., when large and diverse audiences perceive tight associations between the region and a particular industry, they attract attention and resources to a greater extent than other regions. Thus, Romanelli and Khessina argued that regional industrial identity substantively affects how entrepreneurs, managers, and investors perceive and make sense of economic potentials of regions.

In this paper, we develop and test the argument that regional industrial identities significantly influence the location of firms and investments beyond the influence of regional resources and community characteristics, which may not be fully known by external observers. Specifically, we examine the effects of regional industrial identity on the location decisions of entrepreneurs, the managers of established organizations, and venture capital investors in the U.S. biotherapeutics industry from its beginnings in 1976 through 2004. Controlling for important regional resources, our findings reveal that the entry rates of non-local *de novo* firms and investment targets of regionally and industrially inexperienced venture capital investors are strongly influenced by regional industrial identities. These findings support the argument that regional industrial identity is an important factor in the spatial arrangements of industries and set the stage for future research on the development of regional industrial identities and the patterns and processes of cross-regional movements of entrepreneurs and resources that influence the rise of regionally concentrated industries.

BACKGROUND

Cities, and the larger metropolitan regions of which they are often a part, are complex social entities that develop distinctive and enduring "characters" (Molotch, Freudenburg, and Paulsen, 2000) based on the investments of local business and political elites (Molotch, 1976). Such investments produce distinctive patterns in the cultural norms and physical arrangements of cities that, in turn, systematically affect the kinds of investments that will be made in the future. Detailed case histories of particular regions, e.g., Silicon Valley and Route 128 in Boston (Saxenian, 1994) and Las Vegas (Gottdiener, Collins, and Dickens, 1999), support the case for an enduring distinctiveness of regional characters that affect the types of industry that will likely arise and prosper.

Studies of regional characters, however, have focused mainly on explaining the uniqueness of particular cities and regions, emphasizing the enduring, distinguishing influences of business and political investments. For example, Molotch, et al (2000) explored the development of substantial differences in the economic, cultural, and infrastructural characters of geographically adjacent Ventura and Santa Barbara counties in California. Taken as a whole, regions are likely to be unique, but they also share with other regions certain important features, such as industry clusters, that help observers understand and compare regional similarities and differences.

The concept of identity, developed by organizational ecologists to characterize organizational forms (Pólos, Hannan, and Carroll, 2002; Hsu and Hannan, 2005; Hannan, Pólos, and Carroll, 2007), captures the notion of complex characters of social entities while also permitting their comparison. Identities are defined as social codes that summarize and communicate the key features of social entity and thus influence the decisions of important economic actors.

Organizational scholars have explored the concept of identity as an important characteristic of organizations (Albert and Whetten, 1985; Dutton, Dukerich, and Harquail, 1994) and organizational forms (Pòlos, Hannan, and Carroll, 2002; Hannan, Pòlos, and Carroll, 2007) that significantly influences both the commitments of employees to an organization and the ability of members of an organizational form to attract resources and maintain the boundaries of the form. According to these arguments, identity emerges not only from the personal identifications of individuals, which affect their perceptions of membership in groups, organizations, or other social entities (Cerulo, 1997), but also from the common understandings of audiences, especially external audiences, about key features of the social entities (Pólos, Hannan, and Carroll, 2002). When the salient features of an identity are well understood, audiences can compare the features of a particular entity to those of the identity.

As common understandings about an identity become widely held and institutionalized, both the members of an industry category and external audiences command conformance with the features as a condition of continued interaction and investment (Hannan and Freeman, 1984;

Zuckerman, 1999). Organizations that violate identity categories in significant ways are penalized by the withdrawal of resources. For example, Zuckerman (1999) showed that firms exhibiting features that were outside of categories understood by security analysts were not covered in the analysts' reports and, as a result, were traded at a discount rate on the stock market. Other studies have showed that identity boundaries, when they are strong, affect inter-population competition (Carroll and Swaminathan, 2000) and mutualism (Dobrev, Ozdemir and Teo, 2006); when they are weak, diminish the ability of populations to establish a distinctive organizational form (McKendrick, et al 2003).

Thus, Pólos, Hannan, and Carroll (2002) theorized identity as a social code that both parsimoniously describes and effectively proscribes the key features of an organization or organizational form. Social codes simultaneously provide a theoretical basis for characterizing both the rich and diverse features of organizations and organizational forms and for classifying them as members (or not) of a socially constructed cognitive category. External audiences, including buyers, suppliers, investors, analysts, and others with interests in an organization, are especially important in this construction because they control access to important symbolic and material resources. Through this control, they influence the characteristics of organizations and organizational forms that are collectively determined to be both salient and effective. As external audiences interact with the members of a group or an organization, they develop common understandings about the key features of the social entity through direct sharing of information or common exposure to organizational exemplars of the social entity.

THE CONCEPT OF REGIONAL INDUSTRIAL IDENTITY

Romanelli and Khessina (2005) drew on these arguments to propose the concept of regional industrial identity, which they defined as a social code reflecting audiences' shared understandings of regions as legitimate places for particular types of organizations. Regions, like organizations, are complex social entities that present multiple faces to audiences. For example, San Diego, California may be a beach town to surfers and sun lovers, a major U.S. naval base to members of the military and its suppliers, and a biotechnology center to microbiologists and others interested in that industry. Many regions, we suspect, have little industrial identity beyond what their residents know and experience on a day-to-day basis. Other regions, however, such as Los Angeles in the motion picture industry, have strong industrial identities, indicating that most observers, including those who are not directly involved in the industry or reside in a region, associate an industry with a region.

Regions have stronger or weaker identities based on the number and sizes of audiences, including both local and distant audiences, who share perceptions about regional suitability for a particular industry. Strong industrial identities arise from a variety of sources, including: formal and informal interactions between the managers, scientists, and residents of a region with individuals outside of the region, media reports about managers, organizations, and general information about the location of resources and the number of firms that are located in a region. For example, Chiles, Meyer, and Hench (2004) described the rise of Branson, Missouri as a U.S. center of the country music theatre industry based on the directed activities of residents who invested in the art form by establishing theatres and on public promotion of the industry to outsiders. As the theatres began to attract well-known performers, they also began to attract audiences from outside the region. Over time, Branson, Missouri was routinely mentioned in the public media as a country music theatre center in the U.S.

Once a region becomes taken for granted as a suitable place for particular organizations, its industrial identity is likely to persist. Self-reinforcing processes, including ongoing local entrepreneurial activity driven by thick local industry knowledge and contacts (Sorenson and Audia, 2000), the attraction of new organizational entrants and other resources (Arthur, 1990; Chiles et al 2004), and more general public acceptance of the region as an important home for an industry (Saxenian, 1994), sustain the region's identity. Regional identity is not always or necessarily an accurate reflection of regional suitability for an industry. For example, Jaffee (2003) showed that despite substantially higher failure rates of law firms in the cities of Palo Alto, San Jose, and Santa Clara than in other, less crowded areas of Silicon Valley, they remain known as "good places" for corporate law practices and continue to attract disproportionately greater numbers of new firms. In that study, the identity of the region was sufficient to maintain the inflow of new entrants to the region.

Over time, strong regional industrial identities crowd out attention to regions with weaker identities and, thus, influence the spatial concentration of industries. To the extent that audiences ascribe stronger or weaker industrial identities to regions, their decisions about how to allocate important resources should be directed toward regions with stronger industrial identities. Thus, regional industrial identity establishes a basis for classifying and comparing regions without necessary resort to full objective information. We explore these arguments in the hypotheses below.

HYPOTHESES

Regional Identity and Organizational Entry

Starting a new firm or opening a new subsidiary requires that prospective entrepreneurs and managers of established firms (1) acquire information about opportunities for a new business and (2) mobilize resources necessary to build the business. Although many theorists have argued that economic actors learn about opportunities for new organizations and obtain resources primarily in their local environments (Sorenson and Audia, 2000; Stuart and Sorenson, 2003a), decisions about where to locate a new firm or subsidiary may also involve comparisons of regional suitability for particular kinds of organizations.

Theorists have argued that information about the location and amount of industry-relevant resources in regions and the density of industry specific firms serve as potent indicators of regions' relative suitability for particular industries. However, most prospective entrepreneurs and managers do not routinely or systematically scout the location of resources or the densities of regional populations. While researchers may collect data about regional resources, organizational densities, knowledge spillovers, product heterogeneity, and many other characteristics, to specify and compare their effects on various phenomena of interest, most observers, even interested observers, will not have access to such information. Due to bounded rationality (Simon, 1947) and the private nature of important information (Hayek, 1945), most observers will have only highly generalized impressions of regional industries, developed through myriad informal readings, conversations, and visits to a region. Even the managers of large organizations, who may commission regional studies, devoting significant resources, will tend to focus attention on a few regions that fit their impressions as possible good places for a particular kind of business. Thus, regions are routinely perceived in terms of socially constructed, shared understandings, i.e., in terms of their industrial identities, and they should be compared on the basis of such identities.

Although common understandings about regional resources may be only partially accurate, regional industrial identity supplies a cheap, readily available, and easily understood basis for comparing regional industrial characteristics. To the extent that relevant audiences tend to focus their attention on regions with strong identities and ignore those with weak identities, relying on regional identity effectively helps to narrow down the number of locations considered in the pre-entry analysis and, therefore, makes the comparison of business opportunities across different regions feasible. The need for such a heuristic is especially strong when there is a significant uncertainty about the location of best business opportunities and the acquisition of objective information is too costly or impossible.

Regional industrial identity may also influence the amount of resources that individuals and firms can mobilize for starting a new business in the region for two main reasons. First, regions with stronger identities as legitimate places for a particular type of business are likely to have a larger pool of relevant resources. Second, shared understandings about legitimate locations for a particular industry should affect the availability of resources because audiences may be less likely to invest in businesses founded outside of the strong identity regions. Thus, we argue that there will be a positive relationship between the strength of a region's identity for a particular type of business and the rate of entry of new firms into the regional industry.

> **Hypothesis 1.** The stronger a region's identity for a particular industry, the greater the entry rate of firms into the regional industry.

So far we have assumed that all potential entrants are subject to imperfect information to the same extent and that all potential entrants depend equally on audiences for resources. This stylized assumption is convenient for initial arguments, but it is not very realistic. We relax the assumption now and take into consideration two facts. First, potential entrants possess different amount of information about the suitability of regions for particular types of businesses. Second, not all potential entrants depend on their audiences for resources. Below, we consider two characteristics of prospective entrants that may affect their access to information and resources: (1) the pre-entry location of a firm's founders or parent organizations, and (2) the mode of a firm's entry into an industry.

Regional Identity and Pre-entry Location of Entrants

Prospective entrepreneurs and diversifying firms can enter an industry either locally, i.e., in the region in which they already reside, or from another region. Local entrants are likely to have both more accurate information about business opportunities and greater access to resources in their home region than non-local entrants for several reasons. First, the residents of a region attend more acutely to local events and trends because they are more directly affected by those events than non-residents (Hilgartner and Bosk, 1988; Hoffman and Ocasio, 2001). Second, local entrants enjoy information benefits derived from participation in both informal and formal social networks (Sorenson and Audia, 2001; Sorensen, 2003a). The emergence and maintenance of social networks, based either on work, friendship, family or community ties, is facilitated by the spatial proximity of actors (Festinger, Schachter, and Back, 1950; Kono et al., 1998; Sorenson and Stuart, 2001). Since social networks often serve as conduits for scarce and private information (Granovetter, 1974; Uzzi, 1996; Stuart and Sorenson, 2003a), local entrants acquire information that is both more accurate and more complete than information available to prospective non-resident entrants.

Non-local entrants, by contrast, have greater difficulty in collecting information about distant regions both because they do not attend as strongly to local information and because they are not privy to information that flows primarily over social networks. Thus, they will tend to rely on cognitive categories that designate some regions rather than others as good places for particular kinds of industries.

Local entrants also have an advantage over non-local entrants in their ability to mobilize resources because they can call on their formal and informal network memberships based on strong ties. For example, relationships with local financial institutions and other holders of capital reduce the asymmetry of information and, therefore, increase entrants' chances of receiving financial resources (Fried and Hisrich, 1994; Uzzi, 1999; Shane and Cable, 2002). Friendship and family ties help entrants recruit employees by engendering trust (Ruef, Aldrich, and Carter 2003; Sorenson, 2003b). Social connections to local universities and strong ties with local suppliers, buyers and incumbent firms facilitate the acquisition of knowledge and intellectual capital (Saxenian, 1994; Rosenkopf and Almeida, 2003; Powell, et al. 2005). In contrast, non-resident entrants, who cannot rely on local networks for resource mobilization, may have to conform to the perceptions and expectations of outside audiences to persuade them to invest (Zuckerman 1999). If important investors and suppliers believe that firms founded in strong identity regions are more likely to be successful, they will differentially direct their investments to firms that are located in such regions.

Thus, because of needs for acquiring information and resources from outside audiences, non-local entrants should be more strongly influenced by regional industrial identity than local entrants. We predict the following:

> **Hypothesis 2***:* The effect of regional industrial identity will be stronger for the entry rates of *non-local* firms than for the entry rates of *local* firms.

Regional Identity and Entry Mode of Non-local Entrants

Non-local entrants may also vary in the degree of their need for information and resources. Non-local entrants can be classified into three categories: (1) *de novo* entrants, defined as new firms founded by entrepreneurs, (2) *de alio* entrants, defined as diversifying firms that enter an industry which is different from their established industry, e.g., traditional pharmaceutical firms that enter biotechnology, and (3) relocators, defined as established organizations within a focal industry that move to a different region. In considering the cross-regional movement of firms, relocators can be considered as a special type of *de alio* entrants. *De alio* firms have been shown to enjoy significant advantages in performance and survival because at the time of entry they possess more resources and greater experience than *de novo* firms (e.g., see Carroll and Khessina, 2005 for review).

We expect that the entry into a region by non-local *de novo* firms will be more affected by the region's industrial identity than the entry by non-local *de alio* firms for two main reasons. First, *de novo* entrants have fewer resources than *de alio* firms for gathering objective information or analyzing the relative benefits of multiple locations (Schoonhoven, Eisenhardt, and Lyman, 1990; Henisz and Macher, 2004). Second, *de novo* firms are more dependent on outsiders, including venture capitalists, for important resources (Stinchcombe, 1965; Carroll, et al 1996).

Resource-holding agents, however, experience significant difficulty in evaluating new ventures that do not possess track records. They are more likely to provide resources to non-local founders who conform to their expectations by entering a region with a strong identity for the industry.

By contrast, non-local *de alio* entrants can usually obtain resources from their parent companies (Penrose, 1958; Mitchell, 1994). Even if there is a need for outside resources, established reputations and relationships with suppliers and buyers in the industries of prior activities make the procurement of outside resources easier (Carroll, et al. 1996). Resource-holding agents that have a history of relationships with *de alio* firms also have less need to rely on heuristics such as regional identity to make investment decisions and are more forgiving when *de alio* firms violate their expectations about the region of entry (Zuckerman et al 2003).

Based on the above reasoning, we argue that the geographic destinations of non-local *de novo* entrants are more affected by regional identity than those of non-local *de alio* entrants.

> **Hypothesis 3***:* The effect of regional industrial identity will be stronger for the entry rates of *non-local de novo* firms than for the entry rates of *non-local de alio* firms.

Regional Identity and Venture Capital Investments

Studies of venture capital investments have shown that location plays an important role. For example, Florida and Smith (1993) and Gompers, *et al*. (1998) showed that, while venture capital firms are widely distributed in geographic space, investments are spatially concentrated. According to Gompers, *et al*, demand for venture capital is the most important influence on location of investment, and such demand in the U.S. has been historically concentrated in four states: California, Massachusetts, New York, and Texas. They showed that state-level industrial and academic R&D spending, as well as gross state product per capita, significantly and positively influenced the amount of venture capital investments in states. Thus, regional resources are related to the location of venture investments.

Like entrepreneurs and managers, however, venture capitalists are limited in their abilities to assess the economic conditions of regions, and no study that we could find suggested that the location of venture capital investments was systematically driven by such assessments. Rather, venture capitalists tend to specialize in industries or technology sectors to improve their abilities to discover, evaluate, and monitor their investments (Gompers and Lerner, 2001). To specialize, venture capitalists develop broad networks of industry contacts, including scientists, engineers, and the current and former senior managers of firms in an industry (Bygrave and Timmons, 1992). These networks improve the quality of investment opportunities as trusted contacts recommend certain entrepreneurial businesses, help to evaluate the technology and management of other entrepreneurial businesses, and serve directly as a source of managerial talent businesses that receive venture capital investments.

Venture capital networks take time to develop and maintain, however, and are probably dependent on a high amount of face-to-face communication, both to develop trust and to maintain intellectual currency in developing new technologies. Physical distance from the source of information exacerbates the challenge (Lerner, 1995; Sorenson and Stuart, 2001). As industry specialists, however, venture capitalists have interest in discovering entrepreneurial opportunities outside their home regions. To facilitate such discovery, and also to diversify their portfolios, venture capitalists syndicate deals with investors in both their home and other regions. Sorenson and Stuart (2001) showed that venture capitalists were more likely to invest in industries that were outside the scope of their normal investments and in regions that were distant from their home regions when they had participated previously in deals with other venture capitalists that were more experienced in the industry or the region.

These patterns suggest that venture capitalists are at least somewhat limited in both the industrial and geographic scope of their investments. Although, in theory, sophisticated investors would seek to be broadly knowledgeable about investment opportunities across all geographic regions, time and geographic constraints on their abilities to establish good information networks likely restrict their focus to one or a few regions. As a result, venture capitalists, as boundedly rational decision makers, are likely to be influenced by common understandings about the differential suitability of regions for the new industry. Thus, we argue that capital investors will be significantly influenced by a region's industrial identity for a particular kind of business.

> **Hypothesis 4.** The stronger a region's identity for a particular industry, the greater the investment rate of venture capital firms in the regional industry.

Regional Identity and Investment Experience of Venture Capital Firms

Capital providers have different degrees of experience with investments in particular industries and in particular regions. The level of experience may influence the extent to which venture capitalists rely on regional industrial identities for making decisions about the location of their investments. Prior investments in an industry directly affect the attention that venture capitalists pay to the industry as well as the number of contacts they have for acquiring additional information about new investment opportunities and for assessing the quality of investment opportunities (Gompers and Lerner, 2001; Sorenson and Stuart, 2001). Investors who have not previously invested in an industry firm are more likely to attend to regional industrial identity for similar reasons as we have described for non-local and non-local *de novo* firms. At greater distances, and in the absence of direct experience and contacts in an industry, venture capitalists will tend to rely on common understandings about the best locations for a particular industry. Thus, we predict that venture capitalists who are inexperienced with investing in a particular industry will be more strongly influenced by regional industrial identity than venture capitalists who have previously invested in an industry.

> **Hypothesis 5.** The effect of regional industrial identity will be stronger for the investment rates of venture capital firms who have never invested in an industry than for the investment rates of venture capital firms who have previously invested in the industry.

METHOD

Research Setting

We tested these hypotheses on the population of biotherapeutics firms that existed in the United States from 1976 through 2004. The biotechnology industry is generally dated to the early 1970s when commercial organizations began to investigate the use of bio-technology for the development of new or improved chemicals, foods, and biological drug components. Biotherapeutics is a subset of the more general biotechnology industry that includes the development of drugs for the prevention and treatment of human disease using biotechnological techniques. The founding of Genentech, in 1976, which was established to exploit the gene-splicing technology developed by Herbert Cohen and Stanley Boyer in 1973, marks the beginning of the biotherapeutics population and also the first year of our study period.

Biotherapeutics is an appropriate setting for testing our theory for several reasons. First, the industry is relatively new, which makes it possible to track the entry dates and locations of most of the firms that have entered the industry from the outset. Second, the industry has experienced high rates of entry by different types of firms, including *de novo* and diversifying entrants as well as firms with local and non-local backgrounds. Although three regions—San Francisco, Boston, and San Diego—have become home to very large biotherapeutics populations, the industry has nonetheless been characterized by substantial geographic dispersion. Thus, the industry is well suited for investigating location differences in entry of new firms and diversified companies.

Like all single-industry studies, our focus on just one population, biotherapeutics, poses limits to generalization. Potentially, our findings may be limited to knowledge-intensive industries where transportation and large-scale manufacturing are not significant issues.

Data Sources for Biotherapeutics Firms

We identified biotherapeutics organizations located in the U.S., including the subsidiaries of non-U.S. organizations, primarily using listings provided in *BioScan* (1987 through 2006). *BioScan* is an industry trade publication that includes a comprehensive listing of biotechnology firms in the U.S. and elsewhere. We also consulted several other listings of biotherapeutics firms, including Dibner's *Biotechnology Guide* (1999) and the *Venture Expert* database, to ensure the most complete listing possible. We added approximately 30 biotherapeutics firms to our data set that were not captured in *BioScan*. Finally, we conducted *Lexis-Nexis* and Internet searches for every firm marked as a candidate for inclusion based on the *BioScan* listings to verify the firms' actual work in the development of therapeutics using biotechnology techniques.[1] The result, we believe, is the most accurate database on U.S.-located biotherapeutics organizations possible.

We restricted our analyses to the period 1976 through 2004 to ensure that firms founded very recently, which may not yet have been captured by any data source, did not bias results. We also ran analyses using different end points to define our observation period, 1976-2002, 1976-2003 and 1976-2004. Since the results were very similar, we report only analyses based on the observation period of 1976-2004.

Operationalization of Variables

Regional Boundaries. Regions exist within several levels of geographic location, for example, cities, counties, states, nations, though they are often not perfectly defined by political boundaries. The specification of a region, which may be accurate for analytical purposes at any of these geographic levels, depends on the interests of observers. In this study, we focus on *metropolitan* regions, e.g., greater Chicago, Illinois, as physical locations in geographic space that are embedded in larger social, business, and political environments, but that are also spatially distinct from other metropolitan regions.

To define metropolitan regions, we relied on designations of regions as Metropolitan or Micropolitan Statistical areas (MSAs) developed by the U.S. Census Bureau. Metropolitan and Micropolitan Statistical Areas are designated on the basis of commuting patterns between the counties in which people live and the counties in which they work. The objective is to capture economically integrated regions in physical space

[1] For these identifications, we relied on the scientific knowledge of Martin Doyle, an MBA student with a Masters degree in microbiology and substantial experience in the biotechnology industry.

without regard for state boundaries. For example, the Chicago-Naperville-Joliet, IL-IN-WI Metropolitan Statistical Area includes contiguous counties in three states. We used U.S. Census Bureau classifications published in 2003 (*OMB Bulletin 03-04*). These classifications cover about 83 percent of the U.S. population.

MSAs Included in This Study. We treated a region as a host for the biotherapeutics industry if the region had at least one biotherapeutics firm at any time during the study period. Over the study period, 86 regions were home to at least one biotherapeutics firm, including 81 Metropolitan Statistical Areas, four Micropolitan Statistical Areas, and one U.S. county that has not been designated as part of either a Metropolitan or Micropolitan Statistical Area. Since our theory does not depend on the physical sizes of regions or the sizes of their human population, we treated Metropolitan and Micropolitan Statistical Areas and the single county as equivalent for designating the locations of biotherapeutics firms.

We included in our analyses only regions that hosted at least one biotherapeutics firm during the observation period, i.e., were active in biotherapeutics, based on the following statistical considerations. As we explain below, we rely on event-count methodology, the Poisson framework, to test the hypotheses. Provided that the only MSAs we exclude from the analyses have zero entries for all years of the observation period, i.e., are non-active MSAs, the model predicting zero entries into an MSA takes the form:

$$Pr(0)=Pr(\text{non-active MSA})+Pr(0 \mid \text{active MSA}) = (1-\alpha) + \alpha \exp(-\lambda).$$

Here, $Pr(\text{non-active MSA}) = (1-\alpha)$ and λ is the Poisson mean, so the second term is the probability of a zero entry in an active MSA . Thus, removing non-active MSAs will increase α

but should not affect the estimate of λ (Johnson, Kotz and Kemp, 1992). Therefore, we can exclude non-active MSAs from analyses without biasing estimates.[2]

Starting and ending events for the regional biotherapeutics industry. We defined the beginning of the biotherapeutics industry in a region as occurring in the year when the first biotherapeutics firm entered the region. Most dates of firm entry into specific region are given only with precision to the year of entry. To make the analysis tractable, all the information about timing was annualized. We defined the ending of the biotherapeutics industry in a region as occurring in the year when the last biotherapeutics firm exited the region. We discovered four regions that had significant gaps as hosts of biotherapeutics firms; typically these regions were home to just one biotherapeutics firm that either failed or relocated to another region and then, several years later, hosted another biotherapeutics firm. We treated these regions as having exited the biotherapeutics industry when they first became empty. We coded the secondary entry of these regions into biotherapeutics as a new event. Empirical results from the analyses do not depend on this treatment and are practically identical to those received if the four regions were treated as if they never exited the biotherapeutics.

From 1976 to 2004, 938 biotherapeutics firms entered into 86 unique U.S. regions located in 39 different states. Approximately 30 percent of the entries took place as cross-regional movements of one kind or another, i.e., the move of an entrepreneur to a new region to found a new biotherapeutics firm, the relocation of an established biotherapeutics firm from one MSA to

[2] We are grateful to William H. Greene and Keith Ord for help in making this decision.

another MSA, or the acquisition or creation of a new biotherapeutics subsidiary by an established firm outside of its headquarters region. The data include 1,325 MSA-year observations.

Dependent Variables. There are two primary dependent variables in this study. The first variable measures the entry rates of firms into the regional biotherapeutics industries. The second variable measures the rates of investment by venture capital firms into the regional biotherapeutics industries. All variables were updated annually, unless noted otherwise.

<u>Firm entry rates</u>. Firms were identified as having entered a regional biotherapeutics industry whenever (a) a new firm was founded in a region to develop biotherapeutics products, (b) an established firm entered the biotherapeutics in an MSA by creating a new division, subsidiary, or joint venture, or through the acquisition of or merger with an existing biotherapeutics firm, or (c) an established biotherapeutics firm relocated from one MSA to another. The *rate of all firm entry* was constructed by counting the number of all biotherapeutics firms that entered a particular MSA in a given year.

To test hypotheses about the effects of regional identity on the entry rates of firms with local and non-local origins, we coded the pre-entry MSA locations of entrepreneurs and established organizations. The pre-entry locations of *de alio* firms were given directly by the addresses of the firms' headquarters. For *de novo* firms, an extensive search of the career histories of their founders was undertaken. Information was obtained from many sources, including biographies of managers and boards of directors that are reported on websites and in annual reports, 10-Ks and prospectuses, and newspaper articles. We identified the names of the organizations that employed entrepreneurs immediately prior to the founding of the biotherapeutics firm, then looked up the organizations' addresses and assigned the pre-entry location to an MSA.

The *entry rate of all local firms* was constructed by counting the number of entrants into a region in a given year for which at least one founder or parenting organization was located in the region where the entry into biotherapeutics occurred. The *entry rate of all non-local firms* was measured by counting the number of entrants into a region in a given year for which all founders or parent organizations were located in the region that was different from the one in which the biotherapeutics entry occurred. This represents a conservative measure of the amount of movement in the industry as many *de novo* and *de alio* entries were formed through the collaboration of entrepreneurs or firms from several MSA locations. Out of 938 entrants into the biotherapeutics industry in different MSAs, 642 firms had local origins, 263 firms had non-local origins, and 33 had unidentifiable origins.

Three additional variables measure the entry rates of different types of non-local firms. The e*ntry rate of non-local de novo firms* was constructed by counting the number of start-ups, defined as firms with no prior business existence of any kind, launched in a region in a given year by founders who previously resided outside of the focal region. The *entry rate of non-local de alio firm*s was measured as the number of firms with prior existence of any kind that entered a region in a given year whose parent organizations were located outside of the region of entry. Finally, the e*ntry rate of relocating firms* was created by counting the number of biotherapeutics firms that moved their headquarters from one metropolitan region to another. During the observation period, out of 263 non-local entrants, 128 were *de novo* firms, 49 were *de alio* entrants, and 86 were firms that relocated from one MSA to another.

Venture capital investment rates. The *investment rate of all VC firms* into a regional biotherapeutics industry was constructed by counting the number of venture capital firms that invested in at least one biotherapeutics firm in a region in a given year. The *investment rate of industry-experienced VC firms* was coded as a count of venture capital firms that had previously invested in a biotherapeutics firm anywhere in the U.S.; conversely, the *investment rate of industry-inexperienced VC firms* was measured as the count of firms that had not invested in a biotherapeutics firm to date. Finally, we created two additional variables measuring investments rates of experienced and inexperienced venture capital firms to reflect their prior investments in the biotherapeutics industry in a specific region. The *investment rate of regional industry-experienced VC firms* was coded as the number of venture capital firms investing in the regional biotherapeutics who had invested in a biotherapeutics firm in the region previously; the *investment rate of regional industry-inexperienced VC firms* was coded as the number of venture capital firms that had never previously invested in a biotherapeutics firm in the region, but may or may not invested in a biotherapeutics firm in another region. The data to construct the venture capital variables was obtained from *Venture Expert* database.

Independent Variable. The independent variable in this study is the strength of a region's biotherapeutics identity defined as the strength of shared understandings of audiences about the legitimacy of a region for biotherapeutics businesses. Operationalizing shared understandings presents a significant empirical challenge because it is not feasible to measure them directly. We took an indirect approach. Shared understandings are a part of public discourse, and it is quite common in the literature to measure public discourse through analyses of mass media reports (Ruef, 2000; Pollock and Rindova, 2003; Fiss and Hirsh, 2005; Sine, Haveman, and Tolbert, 2005). According to the communication literature, media reports reflect public knowledge and opinion at the same time as they focus public attention on certain issues (Cohen, 1963; Kosicki, 1993; McCombs, et al. 1998). Drawing on this literature, we believe that media attention to regional industries both focuses and reflects shared understandings of audiences that certain regions are suitable places for certain industries.

Ruef (2000) used media counts to measure the identity of emerging organizational forms. In similar fashion, in order to construct a measure of regions' biotherapeutics identities, we counted the number of *New York Times*' articles that either (1) mentioned or discussed a biotherapeutics firm in an MSA, (2) quoted a biotherapeutics firm manager or a scientist located in an MSA, or (3) discussed the growth, decline, or general suitability of a region for biotherapeutics activity. We excluded from the counts articles with recurring lists such as scientists accepted into the National Academy of Scientists, firms that had recently completed an initial public offering, wedding notices, and paid obituaries, as they did not reflect any discriminating attention of *New York Times'* reporters to the region. We chose *The New York Times* to construct a regional identity measure because it is a national coverage newspaper that is delivered nationally and that appeals to and reflects understandings of broad and diverse types of audiences.

We measured the strength of a region's biotherapeutics identity as the *cumulative number of New York Times' articles about biotherapeutics in an MSA* from the beginning of the industry in 1976 until a given year. Two important considerations support this decision. First, regional industrial identities take time to develop. Regions may have similar coverage in any given year, but may have very different cumulative coverage. Regions with greater cumulative coverage have stronger identities. Second, regional industrial identity may not be significantly influenced by a single year's attention. Regional industries may attract attention in a given year for many reasons that represent idiosyncratic, one-time events or developments that do not reflect understandings about the region as a legitimate place for the industry. For example, the death of 18-year-old Jesse Gelsinger, in 1999, from a gene-therapy clinical trial based on research from Philadelphia's Wistar Institute, generated a large spike (for about one year) in the number of *New York Times'* articles that mentioned biotherapeutics in connection with that region. While such spikes contribute to regional industrial identity, we believe that only sustained attention establishes the region as a legitimate location for a particular industry in the minds of external audiences.

The regional identity variable was logged to reduce extreme skewness. Taking a log of the variable also helped to achieve a better empirical measure of the theoretical construct. Although we theorize that the strength of regional identity increases with the number and size of audiences that share understandings that a region is a legitimate place for a certain industry, we believe that, similar to the process of legitimation of new organizational forms (Carroll and Hannan, 2000), this dependence is not strictly linear. Initially, when few audiences perceive a region as a suitable place for a particular kind of activity, the returns to identity of adding a new audience or expanding an existing audience are great. When the number and size of audiences who take for granted that a region is a suitable place for a given industry reaches a certain critical level, however, further increases in audience size and numbers do not significantly add to the strength of the region's identity for the industry. In other words, an increase in the number of understandings shared by each additional audience significantly strengthens regional identity when it is weak, but adds only marginally to its strength when it is very strong. The logged version of the identity variable takes this dynamic into account by allowing each additional *New*

York Times article to have a more profound effect at lower levels of regional identity strength than at higher levels. Finally, the variable is lagged one year to insure exogeneity.

Control Variables. Other factors besides regional identity may affect the entry rates of firms into a given region. We constructed a set of variables to control for these influences. Unless otherwise noted, all controls are updated annually. All time variant controls are lagged one year to ensure exogeneity.

<u>Density controls</u>. We controlled for effects of competition on biotherapeutics entry rates in two ways. First, we operationalized the *MSA density of biotherapeutics firms (and its square)* as the number of biotherapeutics firms operating in an MSA in a given year to control for effects of regional legitimation and competition processes on the entry rates of firms. A large body of literature has demonstrated a non-monotonic effect of organizational density on firm entry rates in organizational populations at several geographic levels (see Carroll and Hannan 2000 for a review). Second, when multiple regions host biotherapeutics firms, potential founders and diversifying organizations have a choice about where to locate their organizations, which affects the likelihood of entry into any one region. To control for this influence we created variable the *number of MSAs hosting biotherapeutics firms*, which is the number of MSAs that have at least one biotherapeutics firm in a given year.

<u>Controls for MSA resources</u>. We constructed several variables to control for regional differences in resources relevant to biotherapeutics firms. We measured *venture capital investment into biotherapeutics in an MSA* as the U.S. dollar amount invested by venture capitalists into the biotherapeutics firms in an MSA in a given year. The variable was logged to adjust for

skewness. The *amount of NIH grants received by an MSA* was measured as the U.S. dollar amount of grants by National Institutes of Health to firms and research organizations in an MSA in a given year. The variable was also logged to adjust for skewness. To control for an MSA's scientific resources, we used a time-invariant measure from Zucker et al (1998), the *number of "top quality" universities in biosciences in an MSA*, which are universities with biotechnology relevant programs (biochemistry, cellular/molecular biology, and microbiology) with scholarly quality reputation rankings of 4.0 or higher in 1982 National Research Council Survey. Since previous research has suggested that size of population is associated with land costs and, thus, may affect firm entry rate (Figueiredo et al, 2002), we measured *human population in an MSA* as the number of people residing in the MSA in a given year. The variable was logged to adjust for skewness. The data were obtained from the U.S. Census bureau.

<u>Other controls.</u> The above measures, while they capture many potentially important effects on the location of biotherapeutics firms, may not capture all important influences. For example, in the U.S., state regulations, e.g., the prevalence and enforcement of non-compete agreements (Stuart and Sorenson, 2003b), may influence the rate of entry into a regional organizational population. States may also differ in the extent of institutionally supported entrepreneurship that may influence firm entry rates as well (Sine and Lee, 2005). To control for these and other possible state-level differences, we created a set of *state-specific dummy variables*. Temporal changes in the macroeconomic conditions of a region may also affect the entry rate of regions into an MSA. We created *dummy variables for each calendar year* to control for unobserved time-varying changes.

Finally, one possible concern with using *New York Times* articles to measure the strength of a region's biotherapeutics identity is that it may provide greater coverage for the New York region. To address this concern we constructed a dummy variable for the *New York-Newark-Edison MSA* which takes a value of one if the MSA is New York-Newark-Edison, and zero otherwise.

Model Specification

Firm entry rate models. To test the hypotheses that entry rates of biotherapeutics firms into a region are affected by the region's biotherapeutics identity, it is necessary to take into account that many regions experience entry by more than one biotherapeutics firm in a given year and that the entries can happen at any time during the year. Since our data is precise only to the year, it is not possible to determine the exact waiting time between all events of firm entry. It is common to analyze event information with this form as event-count data (Carroll and Hannan, 2000: 129-31; 146-49). Event-count data is analyzed by assuming that firm entry rates are events governed by stochastic processes. The instantaneous rate of event occurrence (i.e., the rate at which an MSA experiences its next firm entry) is defined as

$$\lambda_n(t) = \lim_{\Delta t \to 0} \frac{Pr[Y(t+\Delta t)-Y(t) \geq 1 \mid Y(t) = n]}{\Delta t} ,$$

where $Y(t)$ is a random variable denoting the cumulative number of firms that entered into an MSA by time t. The stochastic process of interest, the firm entry rate, is $[Y(t) \mid t \leq 0]$ with state space equal to $[0,1,2,\ldots]$. The fundamental parameter of such a process is firm entry rate, the rate of arriving at state $n+1$ at time t. The analysis focuses on year-to-year variations in counts.

The two common estimation models used in event-count analysis are the Poisson regression model and negative binomial regression model. The latter is used if the data is overdispersed, meaning that the variance in counts is not equal to the mean. Overdispersion in this data arises from the years in which no firms entered a given MSA, i.e., from years with zero entry events. We used a zero-inflated Poisson model designed to deal with excess zeros in the data (Greene 2003: 749-752, 779-780). The model for Y_{it}, the number of firms entered into an MSA *i* at year *t*, is given by two equations: $Y_i=0$ with probability q_i and $Y_i \sim Poisson\ (\lambda_i)$ with probability $1\text{-}q_i$. This implies that $Pr(Y_i = 0) = q_i + [(1\text{-}q_i)\ R_i(0)]$, and $Pr(Y_i = j > 0) = (1\text{-}q_i)\ R_i\ (j)$, where $R_i(y) = exp(-\lambda_i)\lambda_i / y_i!$, the Poisson probability, and the rate of events $\lambda_\iota = exp(\beta' X_i)$. The state probability q_i is specified as $q_i \sim$ logistic (v_i), where $v_i = \gamma' Z_i$ (Carroll and Swaminathan, 2000).

In the zero-inflated Poisson model, X_i is a vector of variables that affects the occurrence of non-zero counts, years when an MSA experienced one or more firm entries, and Z_i is a vector of variables that affects zero counts, years in which no new firms entered the MSA. In this study, we expect that the likelihood of zero firm entry increases with industry age as a result of market saturation. So, we included the variable of industry age into the part of the model that predicts the occurrence of zero events, i.e., the so-called inflated model.

We conducted a Vuong test (Greene 2003: 751-752) which showed that a zero-inflated framework was favored over either Poisson or negative binomial models. The majority of models (all models except Model 3.1) also showed that zero-inflated Poisson was favored over zero-inflated negative binomial models, i.e., event counts were not overdispersed in the zero-inflated Poisson framework.

Since observations within specific MSAs are not necessarily independent, to calculate robust standard errors we use Huber/White/sandwich estimator of variance with cluster option that relaxes the assumption of independence.

A general form of the models we estimate specifies firm entry rate into a given MSA *r(i,t)* as a log-linear function of MSA's regional industrial identity and a vector of covariates $X_{(t-1)}$:

$$\ln r(i,t) = \beta I_{i(t-1)} + \delta \Sigma X_{ni(t-1)} ,$$

where $I_{i(t-1)}$ measures the strength of industrial identity of MSA *i* in year *(t-1),* and $X_{ni(t-1)}$ summarizes all other time-varying covariates. $\beta > 0$ indicates that regions with stronger biotherapeutics identities will experience higher entry rates of biotherapeutics firms.

<u>Venture capital investment rates models</u>. To test hypotheses predicting effects of regional industrial identity on the investment rates of venture capitalists into the regional biotherapeutics industry, we relied on the event-count methodology as well. During the period of observation, more than a half of MSAs did not receive a biotherapeutics investment from a venture capitalist in one or more, or even all, years. As a result, the count data contains excess zeros. A Vuong test showed that a zero-inflated framework is favored over either the Poisson or negative binomial models. Additionally, the test revealed that event counts were overdispersed even in the zero-inflated Poisson framework. Thus, we relied on zero-inflated negative binomial models to address this problem.

A general form of the models we estimate specifies the rate of venture capital investment into a given MSA *h(i,t)* as a log-linear function of MSA's regional industrial identity and a vector of covariates:

$$\ln h(i,t) = \beta I_{i(t-1)} + \delta \Sigma X_{ni(t-1)} ,$$

where $I_{i(t-1)}$ measures the strength of biotherapeutics identity of MSA i in year $(t-1)$, and $X_{ni(t-1)}$ summarizes all other time-varying covariates. $\beta > 0$ indicates that regions with stronger biotherapeutics identities experience greater rates of venture capital investments into regional biotherapeutics firms.

We estimated zero-inflated Poisson and zero-inflated negative binomial models using the software package STATA. To estimate rate models with time-varying covariates, we constructed split-spell data breaking observed durations in year-long periods with the values of covariates updated every year.

RESULTS

Table 1a provides descriptive statistics for the variables used in the analyses. Table 1b shows correlations between the variables.

[Tables 1a and 1b about here]

<u>Analyses of Firm Entry Rates</u>. Tables 2 and 3 report estimates of zero-inflated Poisson and negative binomial count models. Control variables show common effects across different models. Not surprisingly, the curvilinear effect of MSA biotherapeutics firm density on entry rates is strongest for the entry rates of local biotherapeutics firms. Consistent with empirical evidence of density dependent processes in many other industries (Carroll and Hannan, 2000), the entry rates of firms first increase with the number of firms already present in a region; when density increases further, however, the entry rates of firms start to decrease.

The negative effect of the variable measuring the number of MSAs hosting biotherapeutics firms, which is significant in some models, points to competition between regions. The greater the number of regions that are home to at least one biotherapeutics firm in the U.S., the lower the entry rate of biotherapeutics firms into any given region. This finding makes intuitive sense since greater dispersion of firms in an industry likely obscures perceptions of the regional advantages of particular regions.

Amounts of venture capital investments and NIH grants into a region increase firm entry into the region, but mostly to an insignificant extent. The region's population rarely impacts firm entry. The presence of universities with "top quality" bioscience programs does not significantly affect firm entry into the region.

The key explanatory variable in Model 2.1, a region's biotherapeutics identity, measured as the logged cumulative number of *New York Times* articles about biotherapeutics in an MSA, shows a positive, significant effect on firm entry. Thus, the stronger a region's biotherapeutics identity, the higher the entry rate of biotherapeutics firms into the region. Hypothesis 1 is supported.

Models 2.2 and 2.3 explore how regional identity affects the entry rates of local versus non-local firms. Model 2.2 indicates that entry rates of firms with local entrepreneurial and organizational origins are not affected by the region's biotherapeutics identity. Model 2.3, however, shows that entry rates of non-local biotherapeutics firms into a region significantly increase with the strength of a region's biotherapeutics identity. Thus, regions with stronger industrial identities experience a higher rate of entry by non-local firms than regions with weaker identities. Therefore, Hypotheses 2 is supported.

[Table 2 about here]

Models 3.1 through 3.3 consider whether regional identity differentially affects the entry rates of non-local *de novo*, non-local *de alio,* and relocating firms. They demonstrate that the entry rates of all types of non-local firms into a region increase with the strength of biotherapeutics identity of the region. However, as Model 3.1 shows, only the effect of regional identity on entry rates of non-local *de novo* firms is significant, indicating that regions with stronger identities experience a greater rate of non-local *de novo* entries than regions with weaker identities. Although we hesitate to conclude that regional identity does not affect the entry rate of non-local *de alio* and relocating firms because the statistical insignificance of the effect may be caused by very few numbers of observations, Hypothesis 3, which argues that regional industrial identity will have its strongest effect on non-local *de novo* firms, is supported.

[Table 3 about here]

Analyses of Venture Capital Investment Rates. Table 4 reports zero-inflated negative binomial models aimed to estimate effects of regional biotherapeutics identity on venture capital investments in the regional industry. Control variables in Models 4.1-4.5 exhibit largely common effects. There is a strong inverted U-shape relationship between the number of biotherapeutics firms in the region and venture capital investments. The number of MSAs hosting biotherapeutics firms decreases venture capital investment, indicating a harmful effect of competition between MSAs for resources. NIH grants, however, positively affect rates of venture capital investments into regional biotherapeutics. Neither the number of universities with strong bioscience programs nor the size of human population significantly affects the rates of VC investments.

The key explanatory variable of regional biotherapeutics identity reveals interesting effects. Model 4.1 shows that the stronger a regions' identity for biotherapeutics, the higher the rate of venture capital investments in biotherapeutics firms in the region. This finding supports Hypothesis 4. Model 4.2 demonstrates that venture capital firms with no previous investment experience in the biotherapeutics industry overall are positively and significantly affected by the strength of a region's biotherapeutics identity, whereas Model 4.3 reveals that venture capitalists with experience in biotherapeutics investments are not significantly affected by regional identity. This result lends support to Hypotheses 5. Finally, Models 4.4 and 4.5 show that venture capital firms who had never before invested in a particular regional biotherapeutics industry, though they may have invested in biotherapeutics firms elsewhere, are positively affected by regional biotherapeutics identity. Venture capital firms that had previously invested in a regional biotherapeutics industry do not respond to the strength of the region's biotherapeutics identity. Overall, Table 4 shows that venture capital investors are more likely to be affected by regional identities when they lack investment experience, whether in the industry overall or in the industry in a specific region.

Sensitivity Analyses

Sensitivity analyses were conducted to check for robustness of empirical findings. We used two alternative approaches for testing the hypotheses: residual models and choice models.

Residual Models. Since the strength of regional industrial identity is probably not completely exogenous to the resources and the number of firms that are already present in the regional industry, we ran so-called residual models to insure that empirical findings reported earlier were not spurious and driven by possible endogenous links between regional identity and levels of resources and firms in the region. The residual models were executed in two steps. First, we ran a linear model predicting the strength of a region's biotherapeutics identity as a function of the number of MSAs hosting biotherapeutics firms, the density of biotherapeutics firms in an MSA and its square, the level of MSA resources (i.e., the amount of VC investments and NIH grants, the number of universities with top bioscience programs and the size of human population), state and year dummies, and the New York-Newark-Edison MSA dummy. From this regression, we obtained standardized residuals to create an alternative measure of regional identity, which we referred to as the "identity residual." The "identity residual" reflects the part of regional identity which is driven by factors other than an MSA's resources and densities of firms.

We plugged the "identity residual" into zero-inflated count models of firm entry rates and VC investment rates reported in Tables 2-4 instead of the original measure of regional identity based on the cumulative number of NYT articles. The "residual" measure of regional identity revealed effects very similar to those in Tables 2-4, both in terms of significance and relative size of coefficients, strengthening our confidence in earlier findings and providing further support for the hypotheses.

<u>Choice Models</u>. An alternative way of testing the effects of regional industrial identity strength involves examining the choices that entrepreneurs and diversified entrants make regarding locations for their businesses. Choice models have been used in several studies to investigate the location choices of individuals and firms (Carlton, 1983; Cushing and Poot, 2004; Kalnins and Chung, 2004).

We relied on McFadden's conditional logit model (McFadden, 1973; Green, 2003: 719-724) to test whether regional industrial identity affects which region an entrant chooses. We re-organized the data so that each entering firm had a choice from among any of the 86 MSAs that ever hosted a biotherapeutics firm. The dependent variable called *choice* was coded 1 if a firm entered a given region and 0 otherwise. An entrant's choice of region was modeled as a function of an MSA's biotherapeutics identity, its firm density and density squared, its resources, a dummy variable for non-local MSAs, a dummy variable indicating non-compete enforcement in an MSA, and the New York-Newark-Edison MSA dummy. Since the choice model methodology is time-invariant, the year dummy variables could not be used in these analyses. Neither could we include into models variables that did not vary by firm, such as state dummies or the number of MSAs hosting biotherapeutics at the time of firm entry.

We estimated conditional logit models using the software package STATA. First, we ran a choice model on the whole population of entrants. The model revealed that firms are less likely to enter non-local regions. However, the stronger the regional identity of a non-local region, the more likely are entrants to select such region for their businesses. This result provides further support to Hypothesis 2 that regional identity affects non-local entrants more strongly than local entrants. Next, we ran choice models separately for firms that chose to enter local vs. non-local regions, as well as for *de novo*, *de alio* and relocating firms that chose a non-local region. The results were consistent with those reported in Tables 2-3. Entrants that chose to enter their home regions were not significantly affected by regional identity, whereas firms that selected non-local regions were strongly influenced by regional identity strength. Furthermore, although *de alio* firms that chose non-local regions were not influenced by regional identities, both *de novo* and relocating firms were strongly affected. Thus, Hypotheses 2 and 3 received additional support.

Finally, we used the "identity residual" in choice models to check for robustness of identity effects and received results that were highly consistent with choice models based on the original identity measure.[3]

[Table 4 about here]

DISCUSSION AND CONCLUSION

Most studies of industrial agglomerations and the factors that explain their existence have emphasized the characteristics of regions that are known, *post hoc*, to have become homes to significant industry clusters. While an impressive body of case historical research has provided good insight into the kinds of resources and the characteristics of social networks that tend to give rise to and sustain particular industry clusters, such studies have failed to confirm that these characteristics, which may also have been present in other regions, systematically account for differences in the growth of regional industries.

Part of the difficulty, we believe, has been a failure to consider that entrepreneurs, managers, and capital investors may make location decisions based on understandings or beliefs about the relative suitability of regions for particular types of business activity. To the extent that entrepreneurs and investors either know or can learn about differences in regional resources, they make rational location decisions. Because much important information about regional industries is difficult to come by, however, individuals also rely on heuristics, including shared

[3] We could not use the choice model methodology to test whether regional identity affects investment choices of venture firms. The methodology is based on the assumption that choice alternatives are mutually exclusive (McFadden, 1973; Train, 1993). Because venture capital firms often invest in more than one region at the time, this assumption is violated.

understandings about regional suitability, which may or may not be accurate with respect to actual resource availability, firm survival chances, or population densities.

In this paper, we have examined the effects of regional industrial identity on firm entry rates and venture capital investment decisions over the life of the U.S. biotherapeutics industry from 1976 through 2004. We believe that regional industrial identity, while it may be socially informed by information about resources and regional population densities, is more cognitively available and, thus, has a greater theoretically compelling influence on the location decisions of prospective entrants and investors, among others, for several reasons. First, the construct does not rely on active and costly monitoring or the acquisition of accurate information about regional advantages. We think that it is unlikely that many entrants or investors systematically collect such information because of difficulties in acquiring accurate information (Hayek, 1945; Uzzi, 1999) and cognitive biases, e.g., bounded rationality (Simon, 1947) that further limit the amount of information that will be collected. Second, while the construct does assume that audiences will develop common understandings about the relative attractiveness of metropolitan regions for particular kinds of business, it also provides an explanation for how such common understandings come into existence. They emerge as entrepreneurs, investors, and other audiences share information about science, industry, and regional attractiveness. It is important to emphasize that such understandings can evolve independent of rational calculations or objective information. Simply, regions, like organizations and organizational forms, develop a taken-for-granted status as better or worse places for particular kinds of activities as audiences socially negotiate these conclusions.

Our findings show that non-local entrants, especially the founders of non-local *de novo* biotherapeutics firms, are significantly and positively influenced by regional industrial identity. For entrepreneurs who believe that their prospects will be improved by locating in a region with a stronger industrial identity, our findings are intriguing. These entrepreneurs, who have less information about the actual resources available in other regions, appear to rely, as predicted, on regional industrial identity for determining the locations of their firms. This is a strong and important result, we believe, in that it suggests a mechanism for industry geographic concentration that takes into account competition among regions for particular clusters. If prospective entrepreneurs respond to regional identities for a particular kind of business, and act on such understandings, then their migration across regions may account for changes in the concentrations of industries across regions. Regions with stronger biotherapeutics identities are more likely to attract prospective entrepreneurs from other regions, which implies the loss of these entrepreneurs to these other regions.

Similarly, venture capital investments were significantly influenced by regional industrial identity. Inexperienced venture capital investors were more influenced by regional industrial identity than experienced investors. This result suggests that regional industrial identity has a general effect on audiences' understandings about the suitability of regions for particular types of organizations. Especially in cases where accurate information is difficult to obtain, entrepreneurs and venture capitalists both rely on regional industrial identity as a salient cue about where to locate their firms or investments.

Overall, our findings demonstrate an important effect of regional industrial identity on the regional entry rates of firms in one industry and set the stage for several additional research questions. First, while our findings indicated that local entrants were not strongly influenced by the strength of their regional industrial identities, it is possible that this effect is an artifact of our using a national newspaper, *The New York Times*, to measure regional industrial identity. Potentially, local rather than national media coverage is a more salient indicator of regional industrial identity to local entrants.

Second, research is needed to explore how strong regional industrial identities emerge. The location and sizes of previously established, related industries, as well as public and private research institutions in the case of high technology industries, may play an important role in attracting attention to a region and in fostering beliefs that the region is suitable for a particular new industry. Romanelli and Khessina (2005) argued that both the dominance of one or a few regional populations may either promote or impede the emergence of a particular new industry depending on whether the established populations are related or unrelated to the emerging industry. Audia, Freeman, and Reynolds (2006) showed a strong effect of the location of related firms on firm entry rates among U.S. instruments manufacturing firms. McKendrick et al (2003) showed that the more consistent and coherent the organizational forms of regional industrial populations, the greater the ability of external audiences to perceive the organizational form and thus its location. Better understanding of regional community ecologies holds promise for increasing our understanding about both the location and evolution of new industries.

Third, research is needed to examine the effects of regional industrial identity on inter-regional competition and legitimacy. Hannan, et al (1998) showed strong cross-regional effects of legitimacy in the European automobile industry. While their study examined founding rates without regard for the origins of entrepreneurs, our study found that approximately 30 percent of regional entries were accomplished by entrepreneurs or parent organizations outside of their home regions. Although we cannot say for sure whether this result is anomalous to this industry, or perhaps only to science- or information-based industries, it suggests that prospective entrepreneurs and managers do frequently consider non-local regions as potential homes for their new businesses. Investigation of the origins and destinations of non-local entrepreneurs and firms may also provide important insights into the dynamic development of industry spatial arrangements.

Overall, our study reveals that regional industrial identity is an important influence on the entry rates of firms and investment rates by venture capitalists above and beyond the presence of objectively important regional resources. We think this construct opens the door to research that compares and, more important, explains the relative growth of regional industries especially during their early years before resources have concentrated. Moreover, because the construct does not rely on potentially idiosyncratic decisions about the nature of important industry resources, investigators can begin to compare processes of regional concentration (or clustering) across different types of industries. Geography is an increasingly important factor in both academic studies of industry development and change and practical decisions of entrepreneurs and investors. Regional industrial identity holds promise for characterizing regions in a way that both captures the essence of their industrial strengths and allows for comparative analysis.

REFERENCES

Albert, S., and D.A. Whetten.
1985 Organizational identity. In L.L. Cummings, B.M. Staw (eds.), *Research in Organizational Behavior*, 7: 263-295. Greenwich CT: JAI Press.

Almeida, P., and B. Kogut.
1999 Localization of knowledge and the mobility of engineers in regional networks. *Management Science*, 45: 905-917.

Arthur, W.B.
1990 'Silicon Valley' locational clusters: When do increasing returns imply monopoly? *Mathematical Social Sciences*, 19: 235-51.

Audia, P., J. Freeman, and P.D. Reynolds.
2006 Organizational foundings in community context: Instruments Manufacturers and their interrelationship with other organizations. *Administrative Science Quarterly*, 51: 381-419.

Barnett, W. and G.R. Carroll
1987 Competition and mutualism among early telephone companies. *Administrative Science Quarterly*, 32: 400-421.

Baum, J.A.C., and H.A. Haveman
1997 Love thy neighbor? Differentiation and agglomeration in the Manhattan hotel industry, 1898-1990. *Administrative Science Quarterly*, 42: 304-338.

Bioscan.
1987-2006 Atlanta: Thompson Bioworld.

Carlton, D.W.
1983 The location and employment choices of new firms: An econometric model with discrete and continuous endogenous variables. *Review of Economics and Statistics*, 65: 440-449.

Carroll, G.R., L.S. Bigelow, M-D.L. Seidel, and L.B. Tsai
1996 The fates of *de novo* and *de alio* producers in the American automobile industry 1885-1981. *Strategic Management Journal*, 17: 117-137.

Carroll, G.R. and M.T. Hannan
2000 *The Demography of Corporations and Industries*. Princeton Press.

Carroll, G.R. and O.M. Khessina
2005. The ecology of entrepreneurship. In S.A. Alvarez, R. Agarwal, and O. Sorenson (eds). *Handbook of Entrepreneurship Research: Disciplinary Perspectives*, Chapter 8 167-200, Springer.

Carroll, G.R. and A. Swaminathan
2000 Why the microbrewery movement? Organizational dynamics of resource partitioning in the U.S. brewing industry. *American Journal of Sociology*, 3: 715-762.

Cerulo, K.A.
1997 Identity construction: New issues, new directions. *Annual Review of Sociology*, 23: 385-409.

Chiles, T., A. Meyer, and T. Hench.
2004 Organizational emergence: The origin and transformation of Branson, Missouri's musical theatres. *Organization Science,* 15: 499-519.

Chung, W. and A. Kalnins, A.
2001 Agglomeration effects and performance: A test of the Texas lodging industry. *Strategic Management Journal*, 10: 969-988.

Cohen, B.
1963 *The Press and Foreign Policy*. Princeton: Princeton University Press.

Cushing, B. and J. Poot.
2004 Crossing boundaries and borders: Regional science advances in migration modeling. *Papers in Regional Science*, 83: 317-338.

Dibner, Mark D.
1999 *Biotechnology guide U.S.A.: Companies, data, and analyses, 5th edition*. New York: Grove's Dictionarys

Dobrev, S.D., S.Z. Ozdemir, and A. Teo
2006 The ecological interdependence of emergent and established organizational populations: Legitimacy transfer, violation by comparison, and unstable identities. *Organization Science*, 17: 577-597.

Dutton, J.E., J.M. Dukerich, and C.V. Harquail.
1994 Organizational images and member identification. *Administrative Science Quarterly*, 39: 239-263.

Festinger, L., S. Schachter, and K.W. Back.
1950 *Social Pressure in Informal Groups*. New York: Harper.

Figueiredo, O., P. Guimaraes, and D. Woodward.
2002 Home-field advantage: Location decisions of Portuguese entrepreneur. *Journal of Urban Economics*, 52: 341-361.

Fiss, P.C. and P.M. Hirsch.
2005 The discourse of globalization. *American Sociological Review,* 70: 29-52.

Florida, R., and D.F. Smith, Jr.
1993 Venture Capital Formation, Investment, and Regional Industrialization. *Annals of the Association of American Geographers*, 83: 434-451.

Freeman, J.H. and P.G. Audia.
2006 Community ecology and the sociology of organizations. *Annual Review of Sociology*, 32: 145-169.

Fried, V.H. and R.D., Hisrich.
1994 Toward a model of venture capital investment decision making. *Financial Management*, 23: 28-37.

Gottdiener, M., C.C. Collins, and D.R. Dickens.
1999 *Las Vegas: The Social Production of an All-American City*. Oxford: Blackwell.

Granovetter, M.
1974 *Getting a Job: A Study of Contacts and Careers*. Cambridge. Harvard University Press.

Gompers, P.A., and J. Lerner
2001 The venture capital revolution. *The Journal of Economic Perspectives*, 15: 145-168.

Gompers, P.A., J. Lerner, M.M. Blair, and T. Hellman.
1998 What drives venture capital fundraising? *Brookings Papers on Economic Activity, Microeconomics*, 1998: 149-204.

Greene, W.H.
2003 *Econometric Analysis*. 5th ed. Upper Saddle River., N.J. Prentice-Hall.

Hannan, M.T., G.R. Carroll, S.D. Dobrev, and J. Han
1998 Organizational mortality in European and American automobile industries Part I: Revisiting the effects of age and size. *European Sociological Review*, 14: 279-302.

Hannan, M.T. and J. Freeman.
1984 Structural inertia and organizational change. *American Sociological Review*, 49: 149-164.

Hannan M.T., L. Pólos, and G.R. Carroll.
2007 *The Logics of Organizational Theory: Audiences, Codes and Ecologies.* Princeton: Princeton University Press, in press.

Hayek, F.
1945 The use of knowledge in society. *American Economic Review*, 35: 519-530.

Henisz, W.J. and J.T. Macher.
2004 Firm- and country-level trade-offs and contingencies in the evaluation of foreign investments: the semiconductor industry, 1994-2002. *Organizations Science*, 15: 537-554.

Hilgartmer, S. and C.L. Bosk.
1988 The rise and fall of social problem: A public arena discourse. *American Journal of Sociology*, 94: 53-78.

Hoffman A.J. and W. Ocasio.
2001 Not all events are attended equally: Toward a middle-range theory of industry attention to external events. *Organization Science*, 12: 414-434.

Hsu, G., and M.T. Hannan.
2005 Identities, genres, and organizational forms. *Organization Science*, 16: 474-490.

Jaffe, A.B., M. Trajtenberg, and R. Henderson.
1993 Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108: 577-598.

Jaffee, J.
2003 Law firm office location and firm survival in Silicon Valley, 1969 to 1998. In J. Baum and O. Sorenson (eds.), *Advances in Strategic Management*, 20: 341-376.

Johnson, N.L., S. Kotz, and A.W. Kemp
1992. *Univariate Discrete Distributions*. Wiley, New York.

Kalnins, A., and W. Chung
2004 Resource-seeking agglomeration: A study of market entry in the lodging industry. *Strategic Management Journal*, 25: 689-699.

Kono, C., D. Palmer, R. Friedland, and M. Zafonte.
1998 Lost in space: The geography of corporate interlocking directorates. *American Journal of Sociology*, 103: 863-911.

Kosicki, G.M.
1993 Problems and opportunities in agenda setting research. *Journal of Communication*, 43: 100-127.

Krugman, P.
1991 *Geography and Trade*. Leuven, Belgium: Leuven University Press and Cambridge, MA: MIT Press.

McCombs, M., J.P. Llaman, E. Lopz-Escobar, and F. Rey.
1998 Candidate images in Spanish elections: Second-level agenda effects. *Journalism and Mass Communication Quarterly*, 74: 703-717.

McFadden, D.
1973 Conditional logit analysis of qualitative choice behavior. " In P. Zarembka (ed.) *Frontiers in Econometrics*. New York: Academic Press.

McKendrick, D.G., J. Jaffee, G.R. Carroll, and O.M. Khessina.
2003 In the bud? Disk array producers as a (possibly) emergent organizational form. *Administrative Science Quarterly*, 48: 60-93.

Mitchell, W.
1994 The dynamics of evolving markets: The effects of business sales and age on dissolutions and divestitures. *Administrative Science Quarterly*, 39: 575-602.

Molotch, H.
1976 The city as a growth machine: Toward a political economy of place. *American Journal of Sociology*, 82: 309-331.

Molotch, H., W. Freudenburg, and K.E. Paulsen.
2000 History repeats itself, but how? City character, urban transition, and the accomplishment of place. *American Sociological Review*, 65: 791-823.

Penrose, E.
1958 *The Theory of the Growth of the Firm*. New York: Wiley.

Pollock, TG. and V.P. Rindova.
2003 Media legitimation effects in the market for initial public offerings. *Academy of Management Journal*, 46: 631-642.

Pólos, L., M.T. Hannan, and G.R. Carroll.
2002 Foundations of a theory of social forms. *Industrial and Corporate Change*, 11: 85-115.

Porter, M.
1998 Clusters and the new economics of competition. *Harvard Business Review*, Vol. 76(6) November-December: 77-90.

Powell, W.W., D.R. White, K.W. Koput, and J. Owen-Smith
2005 Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *American Journal of Sociology*, 110: 1132-1205.

Romanelli, E. and O.M. Khessina.
2005 Regional industrial identity: Cluster configurations and economic development. *Organization Science*, 16: 344-358.

Rosenkopf, L. and P. Almeida
2003 Overcoming local search through alliances and mobility. *Management Science*, 49: 751-766.

Ruef, M.
2000 The emergence of organizational forms: A community ecology approach. *American Journal of Sociology*, 106: 658-714.

Ruef, M., H.E. Aldrich, and N.M. Carter.
2003 The structure of founding teams: Homophily, strong ties, and isolation among U.S. entrepreneurs. *American Sociological Review*, 68: 195-222.

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

Schoonhoven, C., K. Eisenhardt, and K. Lyman.
1990 Speeding the product to market: Waiting time to first production introduction in new firms. *Administrative Science Quarterly* 35:177-207.

Shane, S. and D. Cable.
2002 Social relationships and the financing of new ventures. *Management Science*, 48: 364-381.

Shaver, J.M. and F. Flyer
2000 Agglomeration economies, firm heterogeneity, and foreign direct investment in the United States. *Strategic Management Journal*, 21: 1175-1193.

Simon, H.A.
1947 *Administrative Behavior: A Study of Decision-making Processes in Administrative Organizations*. Macmillan, Chicago, IL.

Sine, W.D., H.A. Haveman, and P.S. Tolbert.
2005 Risky business? Entrepreneurship in the new independent power sector. *Administrative Science Quarterly,* 50: 200-232.

Sine, W.D., and B. Lee.
2005 The winds of change: Technical, political, and cultural antecedents of entrepreneurial activity in the emerging U.S. wind power industry. Paper presented at the 2005 Meeting of American Sociological Association, Philadelphia, PA.

Sorenson, O.
2003a Social networks, informational complexity, and industrial geography. In D. Fornahl and C. Zellner (eds). *The Role of Labor Mobility and Informal Networks for Knowledge Transfer*, 1-19. Kluwer Publishers.

Sorenson, O.
2003b Social networks and the persistence of clusters: Evidence from the computer workstation industry. In S. Breschi and F. Malerba (eds.) *Clusters, Networks and Innovation*, Oxford University Press.

Sorenson, O. and P.G. Audia.
2000 The social structure of entrepreneurial activity: Footwear production in the United States, 1940-1989. *American Journal of Sociology*, 106: 424-462.

Sorenson, O. and T.E. Stuart.
2001 Syndication networks and the spatial distribution of venture capital investments. *American Journal of Sociology*, 106: 1546-1588.

Stinchcombe, A.L.
1965 Social structure and organizations. In March, J.G. (ed.) *Handbook of Organizations,* Chicago: Rand McNally.

Stuart, T. and O. Sorenson.
2003a The geography of opportunity: Spatial heterogeneity in founding rates and the performance of biotechnology firms. *Research Policy*, 32: 229-253.

Stuart, T. and O. Sorenson.
2003b Liquidity events and the geographic distribution of entrepreneurial activity *Administrative Science Quarterly*, 48: 175-201.

Train, K.E.
1993 *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*. Cambridge: MIT Press.

Uzzi, B.
1996 The sources and consequences of embeddedness for the economic performance of organizations. *American Sociological Review*, 61: 675-698.

Uzzi, B.
1999 Embeddedness in the making of financial capital: How social relations and networks benefit firms seeking financing. *American Sociological Review*, 64, 481-505.

Venture Expert.
1975-2006 Thompson Financial.

Zucker, L.G., M.R. Darby, and M.B. Brewer.
1998 Intellectual human capital and the birth of U.S. biotechnology enterprises. *American Economic Review*, 88: 290-306.

Zuckerman, E.
1999 The categorical imperative: Securities analysts and the illegitimacy discount. *American Journal of Sociology*, 104: 1398-1438.

Zuckerman, E., T-Y. Kim, K. Ukanwa, and J. Von Rittman.
2003 Robust identities or non-entities? Typecasting in the feature-film labor market. *American Journal of Sociology*, 5: 1018-1074.

Table 1a. Descriptive Statistics for the Biotherapeutics Industry, 1976-2004: Split-Spell File

	Mean	Std. Dev.	Min	Max
N of Entries by All Firms into an MSA (t)	.708	1.57	0	16
N of Entries by All Local Firms into an MSA (t)	.485	1.13	0	12
N of Entries by All Non-Local Firms into an MSA (t)	.198	.587	0	5
N of Entries by Non-local *De Novo* Firms into an MSA (t)	.097	.386	0	5
N of Entries by Non-local *De Alio* Firms into an MSA (t)	.037	.197	0	2
N of Entries by All Relocating Firms into an MSA (t)	.065	.302	0	3
N of All VC firms invested in biotherapeutics in an MSA (t)	4.01	10.32	0	74
N of VC firms inexperienced in biothera invested in an MSA (t)	.601	1.67	0	14
N of VC firms experienced in biothera invested in an MSA (t)	3.41	9.00	0	68
N of VC firms inexperienced in MSA biothera invested in an MSA (t)	2.15	5.35	0	39
N of VC firms experienced in MSA biothera invested in an MSA (t)	1.85	5.40	0	43
Cumulative N of NYT articles about biothera in an MSA (t-1)	28.2	67.7	0	569
Ln [Cumulative N of NYT articles about biothera in an MSA (t-1)]	1.90	1.64	0	6.34
Density of biothera firms in an MSA (t-1)	6.4	13.0	1	85
N of MSAs hosting biothera firms (t-1)	54.5	15.7	1	69
VC investments into biothera in an MSA in US$ (t-1)	1.06e+07	3.61e+07	0	4.00e+08
Ln [VC investments into biothera in an MSA in US$ (t-1)]	5.35	7.61	0	19.8
Amount of NIH grants received by an MSA in US$ (t-1)	1.36e+08	2.10e+08	0	2.00e+09
Ln [Amount of NIH grants received by an MSA in US$ (t-1)]	16.7	4.46	0	21.42
MSA human population (t-1)	2.18e+06	3.08e+06	2.33e+04	1.87e+07
Ln [MSA human population (t-1)]	13.9	1.32	10.1	16.7
N of universities with 'top quality' bioscience programs in an MSA	.358	.689	0	3
New York-Newark-Edison MSA dummy =1	.020	.141	0	1
Industry age/ Time trend (t)	17.8	6.82	0	28

N of MSAs = 90 (N of unique MSAs = 86, N of “reborn” MSAs = 4); N of MSA-years = 1325
N of entrants = 938 (local = 642; non-local = 263; unknown = 33; de novo = 600; de alio = 252; relocating = 86)

Table 1b. Correlations of Key Variables

		1	2	3	4	5	6	7	8	9	10
1	N of Entries by All Firms into an MSA (t)										
2	N of Entries by All Local Firms into an MSA (t)	.94									
3	N of Entries by All Non-Local Firms into an MSA (t)	.79	.55								
4	N of Entries by Non-local *De Novo* Firms into an MSA (t)	.62	.44	.78							
5	N of Entries by Non-local *De Alio* Firms into an MSA (t)	.41	.30	.48	.12						
6	N of Entries by All Relocating Firms into an MSA (t)	.49	.33	.66	.16	.14					
7	N of All VC firms invested in biotherapeutics in an MSA (t)	.69	.63	.60	.44	.27	.43				
8	N of VC firms inexperienced in biothera invested in an MSA (t)	.57	.51	.48	.38	.23	.31	.83			
9	N of VC firms experienced in biothera invested in an MSA (t)	.69	.63	.59	.44	.27	.43	.99	.76		
10	N of VC firms inexperienced in MSA biothera invested in an MSA (t)	.68	.62	.57	.41	.28	.42	.96	.87	.94	
11	N of VC firms experienced in MSA biothera invested in an MSA (t)	.65	.59	.57	.44	.24	.40	.96	.72	.97	.85
12	Ln [Cumulative N of NYT articles about biothera in an MSA (t-1)]	.52	.47	.45	.33	.21	.34	.59	.47	.59	.57
13	Density of biothera firms in an MSA (t-1)	.70	.64	.61	.40	.28	.50	.86	.67	.86	.82
14	N of MSAs hosting biothera firms (t-1)	-.03	-.04	.04	-.02	-.03	.12	.08	-.04	.10	.08
15	Ln [VC investments into biothera in an MSA in US$ (t-1)]	.48	.44	.41	.31	.21	.27	.58	.48	.57	.57
16	Ln [Amount of NIH grants received by an MSA in US$ (t-1)]	.26	.25	.20	.15	.11	.13	.26	.23	.26	.27
17	Ln [MSA human population (t-1)]	.36	.35	.28	.19	.16	.20	.34	.30	.33	.34
18	N of universities with 'top quality' bioscience programs in an MSA	.63	.58	.51	.39	.27	.33	.65	.61	.63	.62
19	New York-Newark-Edison MSA dummy =1	.23	.23	.18	.08	.09	.19	.10	.12	.09	.12
20	Industry age/ Time trend (t)	-.04	-.06	.03	-.04	-.03	.14	.10	-.00	.11	.10

		11	12	13	14	15	16	17	18	19	20
12	Ln [Cumulative N of NYT articles about biothera in an MSA (t-1)]	.56									
13	Density of biothera firms in an MSA (t-1)	.83	.66								
14	N of MSAs hosting biothera firms (t-1)	.08	.30	.14							
15	Ln [VC investments into biothera in an MSA in US$ (t-1)]	.54	.56	.57	.06						
16	Ln [Amount of NIH grants received by an MSA in US$ (t-1)]	.24	.49	.29	.00	.34					
17	Ln [MSA human population (t-1)]	.31	.59	.42	-.08	.38	.59				
18	N of universities with 'top quality' bio programs in an MSA	.63	.56	.68	-.18	.53	.34	.50			
19	New York-Newark-Edison MSA dummy =1	.06	.19	.38	-.07	.18	.12	.31	.34		
20	Industry age/ Time trend (t)	.09	.29	.15	.94	.09	.02	-.07	-.16	-.06	

Table 2. Zero-inflated Negative Binomial and Poisson Models:
Effect of Regional Identity on **Entry Rates of All Firms and Firms with Different Pre-entry Location**
(Robust standard errors clustered on MSA shown in parentheses)

	Model (2.1)	Model (2.2)	Model (2.3)
		Dependent Variable:	
	All Entrants	***Local*** Entrants	***Non-local*** Entrants
Main Model			
Density of biotherapeutics firms in an MSA (t-1)	.036*	.044**	-.003
	(.017)	(.015)	(.018)
Density^2 of biotherapeutics firms in an MSA (t-1)	-.000	-.000*	-.000
	(.000)	(.000)	(.000)
Number of MSAs hosting biotherapeutics firms (t-1)	-.110	-.258***	-.007
	(.060)	(.059)	(.050)
Ln [VC investments into biotherapeutics in an MSA in US$ (t-1)]	.010	-.000	.039*
	(.010)	(.011)	(.018)
Ln [Amount of NIH grants received by an MSA in US$ (t-1)]	.472	1.07***	.194
	(.248)	(.273)	(.141)
Ln [MSA human population (t-1)]	-.203	-.319	-.436*
	(.140)	(.176)	(.202)
N of universities with 'top quality' bioscience programs in an MSA	-.108	-.266	.107
	(.171)	(.176)	(.185)
New York-Newark-Edison MSA dummy =1	-.270	-.712	.062
	(.459)	(.477)	(.584)
State Dummies	*Yes*	*Yes*	*Yes*
Year Dummies	*Yes*	*Yes*	*Yes*
Ln [Cumulative number of NYT articles about biotherapeutics in an MSA (t-1)]	**.296***	**.144**	**.533*****
	(.134)	**(.107)**	**(.138)**
Inflated Model			
Industry age/ Time trend	-.030	-.030	.218
	(.093)	(.093)	(.194)
Constant	-14.7***	-14.7***	-13.8
	(2.40)	(2.40)	(79.4)
Number of Observations	1229	1229	1229
Number of None-zero Observations	344	283	156
Number of Zero Observations	885	946	1073
Dispersion parameter (alpha)	.066	0	0
Log-likelihood	-836.0	-685.9	-389.9
(d.f.)	(74)	(74)	(74)

p* < .05; p** < .01; p*** < .001

Table 3. Zero-inflated Poisson Models:
Effect of Regional Identity on **Entry Rates of Non-Local Firms with Different Entry Mode**
(Robust standard errors clustered on MSA shown in parentheses)

	Model (3.1)	Model (3.2)	Model (3.3)
	Dependent Variable:		
	Non-local ***De Novo*** Entrants	Non-local ***De Alio*** Entrants	***Relocating*** Entrants
Main Model			
Density of biotherapeutics firms in an MSA (t-1)	.036	-.132**	-.019
	(.023)	(.046)	(.026)
$Density^2$ of biotherapeutics firms in an MSA (t-1)	-.001*	.001*	.000
	(.000)	(.001)	(.000)
Number of MSAs hosting biotherapeutics firms (t-1)	.027	-.542**	-.041
	(.046)	(.177)	(.062)
Ln [VC investments into biotherapeutics in an MSA in US$ (t-1)]	.036	.020	.067
	(.024)	(.029)	(.035)
Ln [Amount of NIH grants received by an MSA in US$ (t-1)]	.298	2.20**	.072
	(.176)	(.659)	(.072)
Ln [MSA human population (t-1)]	-.846***	-.579	-.242
	(.214)	(.413)	(.263)
N of universities with 'top quality' bioscience programs in an MSA	.022	-.636	.122
	(.287)	(.431)	(.252)
New York-Newark-Edison MSA dummy =1	.317	-3.51*	1.51*
	(.598)	(1.68)	(.713)
State Dummies	*Yes*	*Yes*	*Yes*
Year Dummies	*Yes*	*Yes*	*Yes*
Ln [Cumulative number of NYT articles about biotherapeutics in an MSA (t-1)]	**.759*** (.183)**	**.344 (.311)**	**.296 (.199)**
Inflated Model			
Industry age/ Time trend	.039	.018	.003
	(.505)	(.111)	(.129)
Constant	-15.8**	-16.3***	-13.5***
	(5.01)	(2.43)	(2.89)
Number of Observations	1229	1229	1229
Number of None-zero Observations	86	44	60
Number of Zero Observations	1143	1185	1169
Dispersion parameter (alpha)	0	0	0
Log-likelihood	-226.9	-117.7	-166.4
(d.f.)	(74)	(74)	(74)

p* < .05; p** < .01; p*** < .001

Table 4. Zero-Inflated Negative Binomial Models:
Effect of Regional Identity on **Investment Rates of Venture Capital Firms**
(Robust standard errors clustered on MSA shown in parentheses)

	Model (4.1)	Model (4.2)	Model (4.3)	Model (4.4)	Model (4.5)
			Dependent Variable:		
	All VC investors	***Inexperienced*** in biothera VC Investors	***Experienced*** in biothera VC Investors	***Inexperienced*** in regional biothera VC Investors	***Experienced*** in regional biothera VC Investors
Main Model:					
Density of biotherapeutics firms in an MSA (t-1)	.141*** (.023)	.096** (.029)	.141*** (.025)	.120*** (.023)	.135** (.044)
Density^2 of biotherapeutics firms in an MSA (t-1)	-.001*** (.000)	-.001* (.000)	-.001*** (.000)	-.001*** (.000)	-.001* (.001)
Number of MSAs hosting biotherapeutics firms (t-1)	-.067* (.034)	-.003 (.039)	-.087* (.034)	-.040 (.034)	-.140** (.051)
Ln [Amount of NIH grants received by an MSA in US$ (t-1)]	.114** (.038)	.074 (.060)	.130* (.052)	.125** (.040)	.138 (.092)
Ln [MSA human population (t-1)]	.084 (.171)	-.259 (.191)	.145 (.178)	-.072 (.161)	.344 (.277)
N of universities with 'top quality' bioscience programs in an MSA	-.022 (.182)	.137 (.226)	-.084 (.187)	.065 (.172)	-.157 (.313)
New York-Newark-Edison MSA dummy =1	-1.79** (.555)	.179 (.633)	-2.02** (.594)	-.890 (.525)	-2.47* (.960)
Year Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
State Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Ln [Cumulative number of NYT articles about biotherapeutics in an MSA (t-1)]	**.311* (.156)**	**.333** (.119)**	**.280 (.160)**	**.335* (.142)**	**.204 (.174)**
Inflated Model:					
Industry age/ Time trend	-.162* (.080)	-.129 (.193)	-.173* (.072)	-.232** (.077)	-.164 (.097)
Constant	-.586 (1.59)	-.169 (.933)	.443 (.963)	-.311 (1.51)	1.01 (.791)
Number of Observations	1235	1235	1235	1235	1235
Number of None-zero Observations	437	260	413	391	328
Number of Zero Observations	798	975	822	844	907
Dispersion parameter (alpha)	1.33	.254	1.00	1.14	.508
Log pseudo-likelihood (d.f.)	-1823.5 (73)	-801.9 (73)	-1675.3 (73)	-1493.1 (73)	-1204.8 (73)

p* < .05; p** < .01; p*** < .001