Federal Reserve Bank of Philadelphia

Working Papers
Research Department

WP 19-42
October 2019
https://doi.org/10.21799/frbp.wp.2019.42

Localized Knowledge Spillovers: Evidence from the Spatial Clustering of R&D Labs and Patent Citations

Kristy Buzard
Syracuse University and
Visiting Scholar, Federal Reserve Bank of Philadelphia Research Department

Gerald A. Carlino
Federal Reserve Bank of Philadelphia Research Department

Robert M. Hunt
Consumer Finance Institute, Federal Reserve Bank of Philadelphia

Jake K. Carr
The Ohio State University

Tony E. Smith
University of Pennsylvania and
Visiting Scholar, Federal Reserve Bank of Philadelphia Research Department

ISSN: 1962-5361

Localized Knowledge Spillovers: Evidence from the Spatial Clustering of R&D Labs and Patent Citations*

Kristy Buzard
Syracuse University

Gerald A. Carlino and Robert M. Hunt
Federal Reserve Bank of Philadelphia

Jake K. Carr
The Ohio State University

Tony E. Smith
University of Pennsylvania

October 2019

Abstract

Buzard et al. (2017) show that American R&D labs are highly spatially concentrated even within a given metropolitan area. We argue that the geography of their clusters is better suited for studying knowledge spillovers than are states, metropolitan areas, or other political or administrative boundaries that have predominantly been used in previous studies. In this paper, we assign patents and citations to these newly defined clusters of R&D labs. Our tests show that the localization of knowledge spillovers, as measured via patent citations, is strongest at small spatial scales and diminishes with distance. On average, patents within a cluster are about two to four times more likely to cite an inventor in the same cluster than one in a control group. Of import, we find that the degree of localization of knowledge spillovers will be understated in samples based on metropolitan area definitions compared to samples based on the R&D clusters. At the same time, the strength of knowledge spillovers varies widely between clusters. The results are robust to the specification of patent technological categories, the method of citation matching, and alternate cluster definitions.

Keywords: spatial clustering, geographic concentration, R&D labs, localized knowledge spillovers, patent citations

JEL codes: O31, R12

---

* Correspondence: jerry.carlino@phil.frb.org. This paper has benefited from outstanding research assistance by Adam Scavette. The views expressed here are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. This paper is available free of charge at www.philadelphiafed.org/research-and-data/publications/working-papers/.

1. INTRODUCTION

Since the seminal work of Jaffe, Trajtenberg, and Henderson (1993), hereafter JTH, patent citations are a commonly used indicator of knowledge spillovers among inventors. The primary activity at R&D establishments is knowledge based, making concentrations of R&D labs indicative of places in which localized knowledge spillovers would occur. A recent study by Buzard et al. (2017), hereafter BCHCS, shows that R&D labs are, indeed, highly spatially concentrated even within a given metropolitan area. BCHCS introduce the multiscale core cluster procedure in which the boundaries of the core clusters are determined by interrelationships among the sample of R&D labs in two major U.S. R&D regions: the northeastern corridor and California. These clusters should therefore reflect the boundaries in which knowledge spillovers are most likely to be at work more accurately than administrative boundaries. In that sense, the geography of their clusters is better suited for studying knowledge spillovers than are states, metropolitan areas, or other political or administrative boundaries that have been predominantly used in previous studies. In this paper, we extend BCHCS by assigning patents and citations to the R&D clusters they identify and test for evidence of localized knowledge spillovers in patent citations.[1]

We provide evidence that the clustering of R&D labs is related to knowledge spillovers by studying the relative geographic concentration of citations to patents originating in the BCHCS clusters.[2] To do this, we construct treatment versus control tests for the localization of patent citations in the spirit of those found in JTH. For labs in the northeastern corridor, our baseline results indicate that citations are on average about two to four times more likely to come from the same cluster as earlier patents than one would predict using a (control) sample of otherwise similar patents. For California, the baseline results suggest that citations are on average twice as

[1] Rather than using fixed geographic units, such as counties or metropolitan areas, BCHCS use continuous measures to identify the spatial structure of the concentrations of R&D labs. Specifically, they use point pattern methods to analyze locational patterns over a range of selected spatial scales (within 5 miles, 10 miles, 20 miles, etc.). This approach allows them to consider the spatial extent of the agglomeration of R&D labs and to measure any attenuation of clustering with distance more accurately.

[2] Earlier research — e.g., Jaffe, Trajtenberg, and Henderson (1993), Thompson and Fox-Kean (2005), Kerr and Kominers (2015), and Murata et al. (2014) — documents patterns of spatial concentration (often described as localization) in patent citations. Murata et al. (2014) also question whether administrative units are appropriate geographical boundaries for testing for knowledge spillovers.

likely to come from the same cluster as earlier patents than one would predict using the control sample.

It is important to compare the matching rate for BCHCS core clusters with a matching rate closer to that used by JTH, given our view that the BCHCS clusters more accurately reflect the boundaries in which knowledge spillovers are most likely at work than administrative boundaries. To facilitate such a comparison, we apply the data on patents and citations in our sample using the broader metropolitan area definitions instead of the core cluster definitions. In every case, the locational differentials are smaller when the metro area definitions are used compared to the corresponding core clusters. These findings suggest that the degree of localization of knowledge will be understated in samples based on metropolitan area definitions compared to samples based on the R&D clusters.

We can also speak to the question of whether the transmission of knowledge attenuates with distance. We add to the mounting evidence from studies using alternative data that knowledge spillovers begin to attenuate at distances ranging from just a few blocks to a few miles:[3] The localization of knowledge spillovers in our data appears strongest at small spatial scales (5 miles or less) and diminishes with distance. This attenuation reinforces our view that the magnitude of the localized knowledge spillovers documented by studies that use state and metropolitan area data may be understated, and the exact geography that is driving the spillovers may not be well identified.

We also perform a number of robustness tests. In obtaining the baseline results, we use all citations to a given patent. One concern is that this approach fails to distinguish between inventor-added citations and citations added by examiners. Inventor citations represent an acknowledgement of knowledge spillovers, while examiner-added citations likely reflect a patent's value. Using the total citations a patent receives may understate the degree of localization of knowledge spillovers. In our sample, examiner-added citations account for on average around 9 percent of all citations, suggesting that any downward bias is likely to be small.

[3] See, for example, Kerr and Kominers (2015), Elvery and Sveikauskas (2010), Arzaghi and Henderson (2008), Agrawal et al. (2008), Keller (2002), Rosenthal and Strange (2001), Adams and Jaffe (1996), and Audretsch and Feldman (1996).

In fact, we find essentially no downward bias when we redo our matching rate tests using only inventor-added citations.

We also perform a robustness check using alternative boundaries. Our test for knowledge spillovers is whether the citation matching frequency is significantly greater than the control matching frequency. Put differently, our test is whether citations are more localized relative to what would be expected given the existing distribution of technologically related activity. We perform a robustness check using the alternative cluster definitions developed by BCHCS in which the backcloth is based on STEM workers, in case R&D labs follow knowledge workers instead of manufacturing workers, and find that the baseline results are qualitatively unchanged.

As an additional robustness check, we follow Thompson and Fox-Kean (2005) — hereafter TFK — and add the requirement that the controls must share at least one technology subclass with both the patent of interest and its citing patent in addition to matching on the three-digit patent class we use to identify controls in our main analysis. The results are found to be highly robust with respect to such controls, suggesting that they are not solely a consequence of technical aggregation. Finally, we show that our results persist when we use coarsened exact matching as an alternative method to select the controls.

We conclude that there is robust evidence of localization of patent citations, and thus knowledge spillovers, within the BCHCS clusters based on R&D labs. Our methodology allows us to bring to light interesting variation across clusters with a high degree of statistical precision, and we show that the BCHCS clusters compare favorably with the analogous CMSAs in JTH.

2. THEORY

Much of the theoretical literature on urban agglomeration economies has focused on externalities in the production of goods and services rather than on inventive output itself. Nevertheless, the three formal mechanisms primarily explored in the literature — sharing, matching, and, *especially*, knowledge spillovers — are also relevant for innovative activity.[4] A recent paper by Davis and Dingel (2019) provides a useful way for us to think about how the spatial concentration of economic activity facilitates the exchange of ideas among workers and firms. In

[4] See Duranton and Puga (2003) for a more thorough discussion of the micro-foundations of urban agglomeration economies.

their model, individuals with heterogeneous skills divide their time between production and exchanging ideas with each other in order to raise their productivity. Workers in a city allocate their time in accordance with the expected gains from exchanging knowledge with others in their city. Cities with more numerous high-ability workers are better environments for fostering the exchange of ideas. Thus, the gains from the exchange of ideas are greater in larger clusters offering better opportunities for knowledge transfers because their workers are more skilled, are more abundant, and devote more time to exchanging knowledge.

One issue empirically with using administrative boundaries, such as cities, is that there is mounting evidence from studies using alternative data that the transmission of knowledge begins to attenuate at distances ranging from a few blocks to just a few miles.[5] For example, Arzaghi and Henderson (2008) show that for an ad agency in Manhattan, knowledge spillovers and the benefits of networking with other nearby agencies are large but the benefits dissipate very quickly with distance from other ad agencies and are gone after roughly one-half mile.

R&D, more than most industries, depends on new knowledge. Often, the latest knowledge about technological developments is valuable to firms but only for a short time, and the reciprocal exchange of information among co-located firms engaged in innovation can reduce uncertainty (Feldman, 1993). Thus, it behooves innovative firms to locate near sources of information and each other. As already noted, BCHCS shows that R&D labs are highly spatially concentrated even within a given metropolitan area. Given the rapid distance decay in knowledge spillovers identified in these studies, researchers using labor market boundaries (such as MSAs) or administrative boundaries (such as cities) run the risk of underestimating the importance of knowledge spillovers in the location of innovative activity. Murata et al. (2014) and BCHCS use distance-based approaches and find substantial evidence supporting the localization of patents and patent citations. As indicated, we use the clusters identified by BCHCS, as they reflect appropriate boundaries in which knowledge spillovers are most likely to be at work more accurately than administrative boundaries.

[5] See for example, Arzaghi and Henderson, 2008; Agrawal et al., 2008; Conley et al., 2003; Moretti, 2004a and 2004b; Audretsch and Feldman, 1996; and Adams and Jaffe, 1996.

3. SPATIAL CLUSTERS

We use R&D cluster definitions from BCHCS, which cover California and a 10-state area in the northeastern corridor.[6] BCHCS use continuous methods (based on Ripley's (1976) *K*-function) to assess the concentration of R&D labs relative to a baseline of manufacturing employment.[7] In their analysis, BCHCS start with the following null hypothesis:

Hypothesis H_0: *R&D labs locations are no more concentrated than manufacturing at the zip code level and then no more concentrated than total employment within each zip code.*

BCHCS use **local *K*-function analysis** to identify clustering in the neighborhood of specific R&D labs. This local *K*-function for a given point i is just the count of all additional labs within distance *d* of *i*, denoted, $C_i(d)$.

More precisely, the *local K-function*, $\hat{K}_i$, at point i for each distance, *d*, is:

$$\hat{K}_i(d) = C_i(d)$$

The corresponding simulated values, $\hat{K}^S(d), s = 1\ldots, N$, under the null hypothesis are derived by generating point patterns $X^S = (x_i = (x_j^S) \text{ for } j = 1,\ldots n-1), S = 1,\ldots, N$, representing all $n - 1$ points other than point *i*. The *p*-values for a one-sided test of H_0 with respect to point *i* is given by:

$$P_i(d) = \frac{N_i^0(d)}{N+1}, i = 1,\ldots n.,$$

where $N_i^0(d)$ denotes the number of the $N + 1$ draws that generate values at least as large as $\hat{K}_i^0(d)$.

[6] The 10 states are Connecticut, Delaware, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Virginia, plus the District of Columbia; they contain 1,035 R&D labs. There are 645 R&D labs in California.

[7] BCHCS develop an alternative benchmark or backcloth for analyzing R&D clustering with respect to STEM workers to address the concern that R&D labs may follow knowledge workers. We will provide patent citation results using this alternative backcloth as well as for the manufacturing employment backcloth.

A lab is considered to be *locally agglomerated* at a scale of *d* miles if it has more neighboring labs within distance *d* than would be expected (statistically) based only on the distribution of manufacturing employment.

Of special interest are what BCHCS refer to as *core points*. The *core point* at scale *d* is defined as those labs exhibiting maximally significant local agglomeration at this scale and having at least four neighboring labs within that distance.[8] An important property of these local tests is that the *p*-values for each point i can be *mapped* as shown in Figures 1a and 1b. Notice that Figures 1a and 1b show very low *p*-values indicating regions of significant clustering.

More formally, BCHCS use local cluster analyses to group these points into local clusters of labs. That is, this approach allows them to show on a map the locations where the clustering of labs is occurring. This is accomplished by performing simulations using $N = 999$ test patterns of size $n-1$ for each of the *n* (=1,035 in the northeastern corridor and 645 in California) R&D locations in the observed pattern, X^0. BCHCS find substantial variation in significance at different spatial scales. They argue that the clearest patterns of distinct clustering are captured by the three representative distances, $D = \{1,5,10\}$. Of these three, they argue that the single best distance for revealing the overall clustering pattern in the entire data set appears to be 5 miles, as illustrated for the northeastern corridor and California in Figures 1a and 1b, respectively. As seen in the legend, those lab locations, i, exhibiting maximally significant clustering [$P_i(5) = 0.001$] are shown in black, and those with *p*-values not exceeding 0.005 are shown as dark gray. BCHCS argue that essentially all of the most significant locations occur in four distinct clusters in the northeastern corridor, which can be roughly described (from north to south) as the "Boston," "New York City," "Philadelphia," and "Washington, D.C.," agglomerations.[9] In California, their approach identifies three distinct groups, roughly described (from north to south) as "San Francisco Bay Area," "Los Angeles area (mainly Irvine)," and "San Diego."

[8]Using *K*-function permutation tests based on 1,000 permutations, BCHCS define maximal significance to be the smallest *p*-value obtainable under that test, namely $p = 0.001$. Focusing on labs with very high statistical significance mitigates the incidence of false positives due to the multiple testing problem. The condition requiring core points to have at least four other labs at a given distance excludes isolated labs that happen to be in areas with little or no manufacturing employment.

[9] Two exceptions are the small but significant agglomerations identified in the analysis — one in Pittsburgh and one in Buffalo, NY.

While the local cluster analysis provides information about where clustering is most significant at each spatial scale, it cannot identify specific "clusters" of labs. To identify distinct clusters at scale *d*, BCHCS create buffers of radius *d* around each core point in ArcMap and designate the set of labs in each connected component of these buffer zones as a *core cluster* of points. Each distinct cluster thus contains a given set of "connected" core points along with all other points that contributed to their maximal statistical significance. BCHCS refer to this procedure as the *multiscale core-cluster approach*.

An overall depiction of core clusters for the northeastern corridor and for California is shown in Figures 2a and 2b, respectively. Figure 2a shows the four major clusters identified for the northeastern corridor (one each in Boston, New York City/northern New Jersey, Philadelphia/Wilmington, and Washington, D.C.), while Figure 2b shows the three major clusters in California (one each in the Bay Area, Los Angeles, and San Diego). To see how the BCHCS multiscale core clustering approach works, consider the Bay Area in California shown in Figure 3a. The approach identifies one 10-mile cluster covering most of the Bay Area. There is a dominant 5-mile core cluster that is completely nested in the 10-mile cluster, commonly referred to as Silicon Valley. Finally, as the figure shows, there are numerous 1-mile clusters running from the Stanford Research Park area to San Jose at the center of Silicon Valley.[10]

BCHCS similarly identify other clusters of R&D labs such as the ones centered on Cambridge, MA, and the Route 128 corridor that correspond to the most well-known concentration of R&D labs (Figure 3b). Notice that the entire Boston area is itself a single 10-mile cluster. Within this area, there is again a dominant 5-mile core cluster containing the five major 1-mile clusters in the Boston area. The largest of these is concentrated around Cambridge, while the others are centered at points along Route 128 surrounding Boston.

These examples illustrate the attractive features of the multiscale core-cluster approach. First and foremost, this approach adds a scale dimension not present in other clustering methods. In essence, it extends the multiscale feature of local *K*-functions from individual points to clusters

[10] Note that while the clusters in Figure 3a tend to be nested by scale, this is not always the case. For example, the 5-mile "Livermore Lab" cluster in Figure 3a is seen to be mostly outside the major 10-mile cluster. Here, there is a concentration of six R&D labs within 2 miles of each other, although Livermore is relatively far from the Bay Area. So, while this concentration is picked up at the 5-mile scale, it is too small by itself to be picked up at the 10-mile scale.

of points. The clusters identified by BCHCS are well suited for studying knowledge spillovers. In the next section, we extend BCHCS by assigning patents and citations to the R&D clusters they identify and test for evidence of localized knowledge spillovers in patent citations. The ultimate value of the clusters identified by BCHCS can be determined only by testing their economic significance — to which we now turn.

4. CLUSTERING OF R&D LABS AND CLUSTERING OF PATENT CITATIONS

In this section, we test for evidence of localized knowledge spillovers by assigning patents and citations to the core clusters identified by BCHCS. More specifically, we study the relative geographic concentration of citations to patents originating in the clusters. These citations are a concrete indication of the transmission of information from one inventor to another.

We follow the general approach developed in JTH but modify it to reflect the geographic clustering of R&D labs we identify in this paper. JTH test for the "localization" of knowledge spillovers by constructing measures of geographic concentration of citations contained in two groups of patents — a treatment group and a control group. The treatment group represents a set of patents that cite a specific earlier patent obtained by an inventor living in a particular geographic area (in the JTH study either a state or a metropolitan area). For each treatment patent, JTH use a process to select a potential control patent that is similar to the treatment patent but does not cite the earlier patent. For patents in the treatment and control groups, JTH calculate the proportion of those patents obtained by an inventor living in the same geographic area as the inventor of the earlier patent. The difference of these two proportions is a test statistic for the localization of knowledge spillovers. In their study, JTH found that, relative to the pattern reflected in the sample of control patents, patent citations were two times more likely to come from the same state and about two to six times more likely to come from the same metropolitan area.

We construct a comparable test statistic, with several refinements, and we substitute the R&D clusters identified in BCHCS for the state and metropolitan area geography used by JTH. This provides us with an alternative way to test for possible localized knowledge spillovers at much smaller spatial scales than are found in much of the preceding literature. Recall that the boundaries of the core clusters are determined by interrelationships among the R&D labs in our sample and, therefore, should more accurately reflect the appropriate boundaries in which knowledge spillovers are most likely to be at work. In that sense, the geography of our clusters should be better suited for studying localized knowledge spillovers than states, metropolitan areas, or other political or administrative boundaries.

4.1 Construction of the Citations Data Set

For this analysis, we use data obtained from the NBER Patent Data Project.[11] The data span the years 1996–2006. We identify the inventors on a patent using data on inventor codes found in the Patent Network Dataverse (Lai, D'Amour, and Fleming, 2009). Patents are assigned to locations based on the zip code associated with the *residential* address of the first inventor on the patent.[12] We do not use the address of the assignee (typically the company that first owned the patent) because this may not reflect the location where the research was conducted (e.g., it may be the address of the corporate headquarters and not the R&D facility). While it's possible that an inventor's home lies outside a cluster while his professional work takes place inside a cluster, this type of measurement error would bias our results against finding significant location differentials. As a robustness check, we repeated our main analysis using the zip code of the second inventor on the patent. While the sample size is smaller because not all patents list two or more inventors, the results were virtually the same as we report below.[13]

For our tests, we rely primarily on the boundaries identified by the 5-mile and 10-mile core clusters located in the northeastern corridor and in California. For each core cluster at a given scale, we assemble four sets of patents. The first set, which we call *originating patents*, represents those patents granted in the years 1996–1997 by an inventor living in the cluster.[14] We call the second set of patents *citing patents*. These consist of all subsequent patents — including patents for which the residential address of the first inventor is located outside the U.S. — that cite one or more of the originating patents, after excluding patents with the same inventor or that

---

[11] See https://sites.google.com/site/patentdataproject/. We use the files pat76_06_assg.dta and cite_7606.dta.

[12] We used the location information contained in the file inventors5s_9608.tab downloaded from http://dvn.iq.harvard.edu/dvn/dv/patent. Note that this approach implies that our inventors are located at the centroid of the zip code where they live. We have zip code information for almost 99 percent of the patents with a first inventor residing in the United States.

[13] Results are available from the authors upon request.

[14] The range of 1996–1997 is chosen because the lab data on which the clusters are based are from 1998.

were initially assigned to the same company as the originating patent. We exclude these self-citations because these are unlikely to represent the knowledge spillovers we seek to identify.[15]

For every citing patent, we attempt to match it to an appropriate control patent. When we are successful, we include the citing patent in a set we call *treatment patents* and the matched patent in a set we call *control patents*. We select control patents using the following approach. For a given citing patent, the set of potential control patents must have an application date after the grant date of the originating patent that is cited. Potential control patents also cannot cite the originating patent. The application date of potential control patents must be within one year (six months on either side) of the application date of the treatment patent. Finally, as was done by JTH, potential control patents must have the same three-digit primary patent class as the treatment patent.[16] In this way, potential controls are drawn from patents in the same technological field.

The set of potential control patents for a given treatment patent may overlap with the set of potential controls for other treatment patents. To rule out any possibility that this overlap may affect our tests, we randomized the selection of a specific control patent when there was more than one potential control patent from which to choose.[17] The results reported below allow for the selection of control patents with replacement. In other words, a given control patent may be matched to more than one citing patent.

4.2 The Test Statistics

For any given cluster scale, $d\ (=5,\ 10)$, let η_o denote the number of *originating patents* indexed $\{o_i : i = 1, \cdots, \eta_0\}$ that were granted to inventors living in one of the core clusters at scale d in the years 1996–1997.[18] Let η_i denote the number of subsequent citations $\{c_{ij} : j = 1, \cdots, \eta_i\}$ to o_i

---

[15] We do this using the pdpass variable in the data set pat76_06_assg and the Invnum in the Consolidated Inventor Dataset. For details, see Lai, D'Amour, and Fleming (2009).

[16] We match on the variable class in the data set pat76_06_assg. This is the original primary classification of the patent. We feel it is important to use a "real time" classification because these are what other researchers might rely upon around the time a patent was issued.

[17] We find all the patents that are potential controls for a given patent and then randomly choose among them. In JTH, when multiple potential control patents exist, they select the one with a grant date that is nearest to the grant date of the treatment patent as the control the patent.

[18] The following formulation of the proportions used for testing purposes is based largely on Murata et al. (2014).

(after removing self-citations) over the years 1996–2006. For each of these citing patents, c_{ij}, designated as *treatment patents*, we attempted to identify a unique *control patent*, $\tilde{c}_{ij}$, with the same three-digit patent class and with an application date within one year of the treatment patent (see previous description). We are not always successful in doing so. Let $\tilde{\eta}_i (\leq \eta_i)$ denote the number of treatment patents, c_{ij}, for which a control, $\tilde{c}_{ij}$, was found.

Among these $\tilde{\eta}_i$ treatment patents, we count the number of patents, m_i, for which the residential address of the first inventor on the citing patent is located in the *same* core cluster as the originating patent it cites. The fraction of all such patents at scale *d*, i.e., the *treatment proportion*, is given by[19]

$$p = \frac{\sum_{i=1}^{\eta_o} m_i}{\sum_{i=1}^{\eta_o} \tilde{\eta}_i} = \frac{1}{\tilde{\eta}} \sum_{i=1}^{\eta_o} m_i \,. \tag{1}$$

Similarly, let $\tilde{m}_i$ denote the number of matched control patents, $\tilde{c}_{ij}$, in which the residential address of the first inventor is located in the same cluster as the originating patent cited by the treatment patent. The *control proportion* is then given by

$$\tilde{p} = \frac{\sum_{i=1}^{\eta_o} \tilde{m}_i}{\sum_{i=1}^{\eta_o} \tilde{\eta}_i} = \frac{1}{\tilde{\eta}} \sum_{i=1}^{n_o} \tilde{m}_i \,. \tag{2}$$

The resulting test statistic is simply the difference between these proportions, i.e., $p - \tilde{p}$. Under the null hypothesis of "no localization of knowledge spillovers," this difference of independent proportions is well known to be asymptotically normal with mean zero and thus provides a well-defined test statistic.

We run the test 999 times and present a *p*-value as a measure of significance. The *p*-value is the number of samples in which the control proportion is at least as large as the treatment proportion, where the treatment proportion is taken as an additional sample under the null hypothesis, divided by 1,000. Thus a *p*-value of 0.001 indicates that none of the samples had a control

[19] The dependency of fraction, *p* (and all other quantities in (1)) is taken to be implicit.

proportion as large as the treatment proportion. Note that earlier studies in the literature typically draw one set of controls — in contrast to the 999 sets of controls drawn here — and thus present a t-statistic as a measure of significance.

In the tables that follow (except for Tables 14 and 15, which rely on a different methodology for finding control patents), Column E (Matched Citing Patents), Column F (Matched Citing Patents for the same cluster), Column H (Control Patents), and Column I (Control Patents from the same cluster) all represent averages of the 999 samples.

4.3 Main Results

Table 1a presents the results of our localization or matching rate tests for the nine 5-mile clusters identified by BCHCS for the northeastern corridor, while Table 1b shows the results for the four 10-mile clusters they identify. As the last row of Table 1a shows, inventors living in the 5-mile clusters obtained 8,526 patents in 1996–1997 (Column A). Those patents subsequently received 76,669 citations from other patents during the sample period (Column B). Our matching algorithm, with replacement, was able to match essentially all of the citing patents with an appropriate control patent (Column H). Among the treatment patents, 3.68 percent (Column G) had a first inventor living in the same cluster as the patent cited; this is the treatment proportion. Among the control patents, only 0.93 percent (Column J) had a first inventor living in the same cluster as the patent cited by the treatment patent; this is the control proportion. As shown in the next to the last column of the table, on average, a given patent citing an earlier patent in a 5-mile cluster is 3.9 times as likely to have a first inventor living in that cluster than would be expected by chance alone. This value is on the higher side of the range reported by JTH for their test of localization at the metropolitan-area level.[20] As the last row of the Table 1a shows, the difference between the treatment and control proportions is highly statistically significant (Column L). In addition, the location differential — defined as the ratio of treatment and control proportions — is at least 2.8 for every 5-mile cluster.

Table 1b presents the results of our localization tests among 10-mile clusters in the northeastern corridor. At a somewhat larger spatial scale, we find there are more originating patents, more

[20] JTH find a significant "home bias" in patent citations. Excluding self-citations, citations are two to three times (for the corporate samples) to six times (for the university sample) more likely than control patents to come from the same metropolitan area.

citing patents, and, thus, more treatment and control patents. Both the treatment and control proportions (Columns G and J) are higher than was found among the 5-mile clusters. At the same time, the location differential is somewhat smaller. On average, a given patent citing an earlier patent in a 10-mile cluster is 2.6 times as likely to have a first inventor living in that cluster than would be expected by chance alone. This value is on the lower side of the range reported by JTH for their test of localization at the metropolitan-area level. There are a number of specific clusters where this differential is substantially higher. For example, the location differential is almost twice the four-cluster average in the Washington, D.C., and Philadelphia clusters, and 15 percent higher in the Boston cluster.

Tables 2a and 2b present the results of our localization tests among 5- and 10-mile clusters, respectively, in California identified by BCHCS. Compared with the northeastern corridor, we find many more originating patents, citing patents, and, therefore, treatment and control patents. The treatment proportions (Column G) among the California clusters are much higher than those found in the northeastern corridor. However, this is driven almost entirely by the Palo Alto–San Jose cluster associated with Silicon Valley. The control proportions (Column J) are also larger than those found in the northeastern corridor. The *p*-values for the difference in treatment and control proportions (Column L) are highly significant for all the 5-mile and 10-mile clusters. On average, a given patent citing an earlier patent in a 5- or 10-mile cluster in California is twice as likely to have a first inventor living in that cluster than would be expected by chance alone.

It is worth noting that there is significant cross-cluster variation. For 5-mile clusters in the northeastern corridor, the location differentials for the Conshohocken–King of Prussia–West Chester, PA cluster is more than twice the average, and that for the Silver Spring–Bethesda, MD–McLean, VA cluster is three-quarters greater than the average. Excluding the Dublin–Pleasonton, CA cluster, the largest location differential among our baseline results is 8.7 for the 5-mile Los Angeles cluster; this is more than four times the average for 5-mile clusters in California.

To summarize, the clusters of R&D labs identified by the multicore approach appear to coincide with the geographic clustering of patent citations, an often-cited indicator of knowledge spillovers. The following section develops these results further and discusses a number of robustness checks.

5 Additional Results and Robustness Checks

5.1 The Relationship Between Citation Location Differentials and Spatial Scale

The statistics in the preceding tables suggest that there may be a systematic relationship between the size of the clusters we study and the magnitude of the location differentials we find. To explore this further, we extended our analysis to consider clusters at spatial scales of 20 miles. We summarize the results in Table 3a and Table 3b.

A number of patterns are evident from the table. First, the rate of increase in the number of originating patents associated with larger core clusters falls off because a number of clusters that are significant at smaller spatial scales are not significant at the larger spatial scales. The treatment and control proportions tend to increase as we consider larger core clusters. At the same time, the location differential falls as the geographic size of the clusters increases, especially in the northeastern corridor. These results suggest that the core clusters are picking up knowledge spillovers over a variety of spatial scales. Nevertheless, the localization effects appear to be largest at spatial scales of 5 miles and perhaps less. The attenuation in the localization differential as cluster size increases is a typical finding in studies examining localized knowledge spillovers.[21]

5.2 Alternative Approaches to Measuring Citations

5.2.1 Do Examiner-Added Citations Bias the Results?

In obtaining the baseline results, we use all citations to a given patent. One concern with the use of all citations is that the approach fails to distinguish between inventor-added citations and citations added by examiners. Inventor-added citations represent an acknowledgement of knowledge spillovers, while examiner-added citations likely reflect a patent's value. We can address the identification issue by distinguishing inventor-added citations from citations added by patent examiner and use each as a different measure. Specifically, because Hegde and Sampat (2009) demonstrate that a patent cited by examiners has a much stronger relationship with the payment of renewal fees than a patent cited by an inventor, we use examiner-added citations as a

[21] See Carlino and Kerr (2015) for a review of studies documenting attenuation in knowledge spillovers as cluster size increases.

proxy for patent value. We take advantage of the data available from the U.S. Patent and Trademark Office from 2001 onward that distinguishes between those citations that are added by patent examiners versus inventors.

Examiner-added citations account for only 8.6 percent of total citation in the northeastern corridor and for just 9 percent in California. The inclusion of examiner-added citations is unlikely to bias our findings, given the small share of examiner-added citations in our sample. Still, in the interest of completeness, we conduct the matching rate tests when the sample includes only examiner-added citations and when the sample consists only of inventor-added citations. The matching rate tests for examiner-added citations for the 5- and 10-mile clusters in the northeastern corridor are reported in Tables 4a and 4b, and those for California are shown in Tables 5a and 5b. Compared with the baseline findings, the magnitudes of the location differentials for examiner-added citations are somewhat reduced but are still greater than 1.0 for all clusters except Dublin–Pleasonton. As the last row of Column K in Table 4a shows, examiner-added citations are 2.4 times as likely to have a first named inventor living in a typical 5-mile cluster than would be expected by chance alone. This is much lower than the baseline matching rate test of 3.9 for the northeastern corridor. For the average 5-mile cluster in California, the matching rate test for examiner-added citations falls to 1.4 from 2.1 in the baseline case. Similar declines in the matching rate tests for examiner-added citations are found for the 10-mile clusters in both the northeastern corridor and California. Although the magnitudes are smaller for the examiner-added citations, all except that for Dublin–Pleasonton are significantly different from 1.0, an indication that the patents in these clusters are of higher value than suitably defined control patents.

The matching rate tests for inventor-added citations for the 5- and 10-mile clusters in the northeastern corridor are reported in Tables 6a and 6b, and those for California are shown in Tables 7a and 7b. As the last row of Column K in Table 6a shows, inventor-added citations are 4.0 times as likely to have a first named inventor living in a typical 5-mile cluster in the northeastern corridor than would be expected by chance alone. This is similar to the results of the matching rate tests for the baseline rate of 3.9. For the average 5-mile cluster in California, the matching rate test for inventor-added citations at 2.1 is identical to the baseline case. Similarly, the matching rate tests for inventor-added citations for the 10-mile clusters both in the northeastern corridor and in California are virtually identical to those found for the baseline cases. These findings suggest that there is essentially no downward bias associated with using all citations (as in the baseline results) compared with using only inventor-added citations. Given the lack of bias, we use total citations for the analysis undertaken in the remainder of the paper.

5.2.2 Are Patents Obtained in the Core Clusters More Influential?

We now investigate whether patents obtained by inventors living within a core cluster are somehow more important, or at least better known, than patents obtained outside of these clusters. We rely on a common metric of patent quality: the number of citations received. We develop a "counterfactual" region for each of the 10-mile core clusters identified in Section 3. For example, the New York City cluster is compared with the region outside of that cluster contained in New York State, Connecticut, and northern New Jersey. The Boston cluster is compared with the region outside the cluster in the states of Massachusetts, New Hampshire, and Rhode Island.

In Table 8, we report a simple difference in means test for the number of citations per patents received by patents located inside or outside our clusters. For all our clusters, the average number of citations received by patents is greater inside the cluster compared with the average citations received outside the respective cluster; this difference in citations is statistically significant in all clusters except one (Philadelphia).[22]

5.3 Alternative Approaches to Identifying Cluster Boundaries

5.3.1 STEM Workers

It is possible that R&D labs follow knowledge workers instead of manufacturing workers, making knowledge workers in an area a better backcloth than manufacturing employment in that location. As we have shown, one important concentration of R&D labs is found around Cambridge, MA, and another important clustering is found in Silicon Valley. These labs are close to large pools of STEM graduates and workers, the very workers that R&D activity

---

[22] An anonymous referee pointed out that if more "valuable" patents originate inside core clusters than outside them, then patents inside the core clusters would tend to receive more examiner-added citations than those outside, thereby upwardly biasing the difference in means tests. As the previous subsection shows, if anything, examiner-added citations would tend to have a slight downward bias on these tests.

requires. Manufacturing activity tends to employ a more general workforce than does innovative activity and may therefore be more geographically dispersed compared with innovative activity.

To address the concern that knowledge workers may be a better reference group than manufacturing employment, we use an alternative set of clusters developed by BCHCS based on a measure of STEM workers by location.[23] For the backcloth of these clusters, they replace the number of manufacturing employees in each zip code area with an estimate of the number of STEM workers. This is constructed using the proportion of STEM jobs in each four-digit NAICs industry multiplied by the number of jobs in each industry reported in the Zip Code Business Patterns. We report the results of this alternative test for 5- and 10-mile clusters in the northeastern corridor (Tables 9a and 9b) and in California (Tables 10a and 10b). Note that the cluster definitions change when the backcloth changes, so the list of clusters in these tables differs from those in Tables 1a, 1b, 2a and 2b. With the exception of the 5-mile clusters in the northeastern corridor, the average location differentials using the STEM worker backcloth are virtually the same as for the baseline findings. The location differential falls to 3.0 for the 5-mile clusters in the northeastern corridor when considering the clusters based on STEM workers from 3.9 for the baseline results. For the most part, the findings reported for the location differentials in the baseline (and subsequent analysis) suggest little, if any, concern with using manufacturing employment as the backcloth.

5.3.2 MSAs

Our study differs from that of JTH along a number of important dimensions. The geographical scope, sample, and time periods of the analysis differ between the current study and that of JTH. JTH use patents originating in 1975 and 1980 together with citations made to these patents in 1989. Recall that the current study uses more recent samples: patents originating in 1996 and 1997 with citations to these patents made by 2006. As far as the metropolitan areas are concerned, JTH use the 1981 definitions of metropolitan statistical areas (MSAs) and 17 consolidated metropolitan statistical areas (CMSA) plus one constructed "CMSA" consisting of the location of inventors not already included in one of the MSAs or CMSAs, while the current

[23] They use the taxonomy of STEM occupations found at http://www.bls.gov/oes/stem_list.xlsx. This taxonomy is mapped to the 2010 vintage of the Standard Occupational Classifications (SOCs). We map back to the 2000 vintage of the SOCs so we can use the 2002 job counts from the Occupational Employment Statistics to calculate STEM employment "intensity" by industry.

study's focus is on innovative clusters within two regions. A comparison of the findings based on the core clusters with those of the JTH approach might be more informative if we compared the matching rate for our core clusters with a matching rate closer to that of JTH by applying our current data on patents and citations but using the MSA definitions.[24] After all, the underlying socioeconomic environment should greatly differ for the different sample periods used in these studies.

There are seven metropolitan areas in our study that roughly correspond to the 10-mile clusters identified by BCHCS. Using the official 1990 definitions, we find that all but the San Diego metropolitan area had a CMSA definition. For San Diego we use the official 1990 MSA definition. For ease of exposition, we will refer to all seven metro areas as CMSAs. We reproduced our matching rate tests for these CMSAs using patents originating in 1996 and 1997 with citations to these patents made by 2006. Table 11 presents the location differentials for the seven CMSAs in our sample, based on 1990 definitions. Column K of the table gives the matching rate tests for the seven CMSAs using our data. For ease of comparison, Column M of Table 11 shows the location differentials for the 10-mile clusters from our baseline results. A comparison of Column K to Column M reveals that the location differentials for each of the seven CMSAs are smaller in magnitude than those found for the corresponding 10-mile buffer clusters reported in Table 1b and Table 2b and reproduced in Column M. In several cases (Los Angeles, Philadelphia, and Washington, D.C.), the locational differential found using the CMSA definitions are substantially smaller than the differentials found for the corresponding 10-mile buffer clusters. Furthermore, in no instance are the locational differentials larger when the CMSA definitions are used compared to the corresponding 10-mile buffer clusters. These findings suggest that the degree of localization of knowledge will be understated in samples based on MSA/CMSA definitions compared to samples based on the R&D clusters.

[24] We thank an anonymous referee for suggesting that we replicate the JTH experiment using our data.

5.4 Alternative Approaches to Identifying Control Patents

5.4.1 Disaggregated Subclasses

As previously discussed, there has been some debate in the literature as to the best way of implementing a technological similarity requirement based on patent classifications. As already noted, JTH find evidence supporting localized knowledge spillovers for states and for CMSAs. JTH identify potential control patents, as we have in this study, within the same three-digit primary patent class as the treatment patent. TFK are critical of the JTH approach, suggesting that their method for selecting the control group using three-digit technology classifications may not adequately control for the existing geographic distribution of industrial activity and may introduce spurious evidence of localized knowledge spillovers. Instead, TFK use a finer six-digit technology subclass and stipulate that the originating, citing, and control patents should all share at least one technology subclass. They find that tests using this alternative approach reduce the size and significance of the localization ratios, especially at smaller geographies. Murata, et al. (2014) add to the debate, as we do, by questioning whether CMSAs are a relevant geography for testing for evidence of localized knowledge spillovers. Rather than using aggregated administrative boundaries, such as states and CMSAs, they develop a distance-based test of localized knowledge spillovers that embeds the concept of control patents and find evidence supporting localized knowledge spillovers even when using six-digit controls.

The results presented thus far use the BCHCS clusters based on continuous measures to identify the spatial structure of the concentration of R&D labs. We have followed the JTH approach of limiting potential control patents to ones that share the same three-digit primary class as the citing patent. As a robustness check, we implement one version of the matching requirements tested in TFK. We restrict potential control patents to ones that share the same primary class, and for which there is at least one subclass in common between the originating-citing-control triads.[25] Our methodology is otherwise the same as we describe in Section 4.2.

[25] This is similar to the test reported in Table 3, Column (7) in TFK. The following are differences between our methodology and theirs: TFK use only patents with “corporate assignees” for computational ease, while we do not make this restriction. We randomize between all controls within a +/- six-month window, whereas they first check within a +/- one-month window, then +/- three-month window, and finally a +/- six-month window. We use first inventor location only, while they randomly choose an inventor to assign location. Finally, our clusters only partially overlap the 17 CMSAs they use, many of which are low-innovation areas.

We report the results of this alternative test for 5- and 10-mile clusters in the northeastern corridor (Tables 12a and 12b) and in California (Tables 13a and 13b). With the exception of the Stanford–Milford, CT 5-mile cluster, where there are insufficient data to construct a matching rate test, the difference between the treatment and control proportions is highly statistically significant (Column L). In addition, the location differential — defined as the ratio of treatment and control proportions — is at least around 1.7 for every 5-mile cluster and 10-mile cluster. Comparing these results with our baseline results (Tables 1a and 1b) and (2a and 2b), the location differentials are somewhat larger for the northeastern corridor and slightly smaller for California. We conclude that not only are our results robust to the choice of technology controls, we also find evidence, similar to Murata, et al. (2014), supporting localized knowledge spillovers even when using six-digit controls.

5.4.2 Coarsened Exact Matching

More recently, methods for constructing a matched sample of treatment and control groups have evolved. Specifically, coarsened exact matching (CEM) (Iacus, King, and Porro, 2011) can be used to improve the balance between the treated group (citing patents) and the control group. In addition to matching on the application year of the patent and the patent's three-digit technology classification, we also matched discrete bins on two additional variables: 1) the year the patent was granted, and 2) the number of citations a patent received (all cites). We relied upon the CEM algorithm in STATA to coarsen the matched bins based on the optimization of an objective function rather than arbitrarily assigning cut points to the bins.

We use the CEM matched controls in several ways. First, we follow the JTH location differential approach used in producing Tables 1a, 1b, 2a and 2b, our baseline findings, but use the CEM controls. For this approach, we exclude patents with the same inventor or that were initially assigned to the same company as the originating patent.[26] The results are reported in Table 14 (for the northeastern corridor) and Table 15 (for California). On average, the location differentials are somewhat larger than we previously reported for the broad cluster in the

[26] For this approach, the set of potential control patents for a given treatment patent may overlap with the set of potential controls for other treatment patents. To rule out any possibility that this overlap may affect our tests, we randomized the order in which treatment patents were matched to control patents, and we randomized the selection of a specific control patent when there was more than one potential control patent from which to choose. The results reported below allow for the selection of control patents with replacement.

northeastern corridor and in California. On average, a given patent citing an earlier patent in a 5-mile cluster in the northeastern corridor is 4.5 times as likely to have a first inventor living in that cluster than would be expected by chance alone, compared with a differential of 3.9 reported in our baseline results. The location differential in California's 5-mile cluster increases to 2.4 when using the CEM matched controls from 2.1 reported for the baseline. The location differential in the northeastern corridor 10-mile cluster increases slightly to 2.8 when using the CEM-matched controls from 2.6 reported for the baseline. In the California 10-mile cluster, the location differential rises to 2.5 from 2.0 reported for the baseline.

In our second approach, we estimate a logistic model of the likelihood that a patent in cluster h cites an originating patent in that cluster. More formally, if for any given patent, we let T_h denote the indicator variable that this patent cites at least one originating patent in cluster *h*, and similarly, let D_h indicate whether this patent itself originates in cluster *h*, then the conditional likelihood, $\Pr(T_h = 1 | D_h)$, of citing patents in cluster *h* given D_h is postulated to be of the logit form:

$$\Pr(T_h = 1 | D_h) = \frac{\exp(\alpha_h + \beta_h D_h)}{1 + \exp(\alpha_h + \beta_h D_h)} .$$

In this setting, it should be clear that citations of patents in cluster *h* are more likely for (treatment) patents in cluster *h* than for (control) patents not in cluster *h*, i.e., $\Pr(T_h = 1 | D_h = 1) > \Pr(T_h = 1 | D_h = 0)$, if and only if $\beta_h > 0$. The estimated coefficients, $(\hat{\beta}_h)$, are reported in Table 16 (along with robust standard errors for these estimates).[27] As seen from the table, the estimated coefficients for *all* clusters are significantly positive (at the 1% level), and thus provide strong support for the findings in Tables 14 and 15.

Finally, to facilitate comparison, the main results found for location differentials are summarized in Table 17. The table shows the results when R&D clustering is analyzed with respect to (i)

[27] For this approach, we do not exclude patents with the same inventor or that were initially assigned to the same company as the originating patent. The observations are weighted based on the number of CEM-matched controls found for each treated observation.

manufacturing employment (baseline), (ii) citations added by examiners only, (iii) citations added by inventors only, (iv) STEM workers, (v) using CMSA definitions, (vi) when the controls are alternatively selected to have at least one subclass in common with both the citing patent and the originating patent, or (vii) when the controls are selected using coarsened exact matching.

Regardless of the specification chosen to construct the location differentials, we find that citations are at least about twice as likely to come from the same cluster as earlier patents than one would predict using a control sample of otherwise similar patents.

6. CONCLUDING REMARKS

In this paper, we use the local clusters identified in BCHCS to measure the degree to which patent citations are localized in these clusters — tangible evidence that knowledge spillovers are geographically mediated. For labs both in the northeastern corridor and in California, we find, on average that citations are about two to four times more likely to come from the same cluster as earlier patents than one would predict using a (control) sample of otherwise similar patents.

We believe that the BCHCS clusters more accurately reflect the boundaries in which knowledge spillovers are most likely at work than administrative boundaries, such as MSAs/CMSAs and states. It is, therefore, important to compare the matching rate for BCHCS clusters with a matching rate closer to that of the CMSAs used by JTH. To facilitate such a comparison we apply the data on patents and citations in our sample using the broader metropolitan area definitions instead of the core cluster definitions. In every case, the locational differentials are smaller when the metro area definitions are used compared to the corresponding core clusters. These findings suggest that the degree of localization of knowledge spillovers will be understated in samples based on metropolitan area definitions compared to samples based on the R&D clusters.

We can also speak to the question of whether the transmission of knowledge attenuates with distance. The localization of knowledge spillovers in our data appears strongest at small spatial scales (5 miles or less) and diminishes with distance. This attenuation reinforces our view that the magnitude of the localized knowledge spillovers documented by studies that use state and metropolitan area data may be understated, and the exact geography that is driving the spillovers may not be well identified.

Our results are robust to whether we use total citations or only inventor-added citations, the specification of patent technological categories, the method of citation matching, and alternate cluster definitions.

REFERENCES

Adams, J., and Jaffe, A. "Bounding the Effects of R&D: An Investigation Using Matched Establishment-Firm Data." *RAND Journal of Economics*, 27 (1996), pp. 700-21.

Agrawal, Ajay, Devesh Kapur, and John McHale. "How Do Spatial and Social Proximity Influence Knowledge Flows? Evidence from Patent Data," *Journal of Urban Economics*, 64 (2008), pp. 258-69.

Arzaghi, Mohammad, and J. Vernon Henderson. "Networking off Madison Avenue," *Review of Economic Studies*, 75 (2008), pp. 1011-38.

Audretsch, David B., and Maryann P. Feldman. "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 86 (1996), pp. 630–40.

Buzard, Kristy, Gerald A. Carlino, Robert M. Hunt, Jake K. Carr, and Tony E. Smith, "The Agglomeration of American R&D Labs," *Journal of Urban Economics*, 101 (2017), pp. 14-26.

Carlino, Gerald A., and William R. Kerr. "Agglomeration and Innovation," in Gilles Duranton, J. Vernon Henderson, and William Strange (Eds.), *Handbook of Regional and Urban Economics*, Vol. 5A (2015), North Holland, Amsterdam.

Conley, T., Flyer, F., Tsiang, G. "Spillovers from local market human capital and the spatial distribution of productivity in Malaysia." *Advances in Economic Analysis & Policy*, 3 (2003), pp. 1-45.

Davis, Donald, and Jonathan Dingel. "A Spatial Knowledge Economy," *American Economic Review*, 109 (2019), pp. 153-70.

Elvery, Joel A., and Leo Sveikauskas. "How Far Do Agglomeration Effects Reach?" Unpublished Paper, Cleveland State University (2010).

Feldman, Maryann P. "An Examination of the Geography of Innovation," *Industrial and Corporate Change*, 2 (1993), pp. 451–470.

Hegde, Deepak, and Bhaven Sampat. "Examiner Citations, Applicant Citations, and the Private Value of Patents," *Economics Letters*, 105 (2009), pp. 287–289.

Iacus, Stefano, Gary King, and Giuseppe Porro. "Causal Inference without Balance Checking: Coarsened Exact Matching," *Political Analysis* (2011).

Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108 (1993), pp. 577–98.

Keller, W. "Geographic Localization of International Technology Diffusion," *American Economic Review*, 92 (2002), pp. 120-42.

Kerr, William R., and Scott Duke Kominers. "Agglomerative Forces and Cluster Shapes," *Review of Economics and Statistics*, 97 (2015), pp. 877–99.

Lai, Ronald, Alexander D'Amour, and Lee Fleming. "The Careers and Co-authorship Networks of U.S. Patent Holders Since 1975," mimeo, Harvard Business School (2009).

Moretti, E. “Human Capital Externalities in Cities,” in Henderson, J.V., Thisse, J-F. (Eds.), *Handbook of Urban and Regional Economics*, Vol. 4. (2004a), North-Holland, Amsterdam.

Moretti, E. “Workers’ Education, Spillovers and Productivity: Evidence from Plant-Level Production Functions,” *American Economic Review*, 94 (2004b), 656–90.

Murata, Yasusada, Ryo Nakajima, Ryosuke Okamoto, and Ryuichi Tamura. “Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach,” *Review of Economics and Statistics*, 96 (2014), pp. 967–985.

National Science Foundation. *Research and Development in Industry: 1998*, Arlington, VA: National Science Foundation, Division of Science Resources Studies (2000).

Ripley, B.D. “The Second-Order Analysis of Stationary Point,” *Journal of Applied Probability*, 13 (1976), pp. 255–66.

Rosenthal, Stuart, and William C. Strange. “The Determinants of Agglomeration,” *Journal of Urban Economics*, 50 (2001), pp. 191–229.

Thompson, Peter, and Melanie Fox-Kean. “Patent Citations and the Geography of Knowledge Spillovers: A Reassessment,” *American Economic Review*, 95 (2005), pp. 450–460.

U.S. Patent and Trademark Office. Overview of the U.S. Patent Classification System (USPC).

Washington, D.C. (2012), http://www.uspto.gov/patents/resources/classification/overview.pdf.

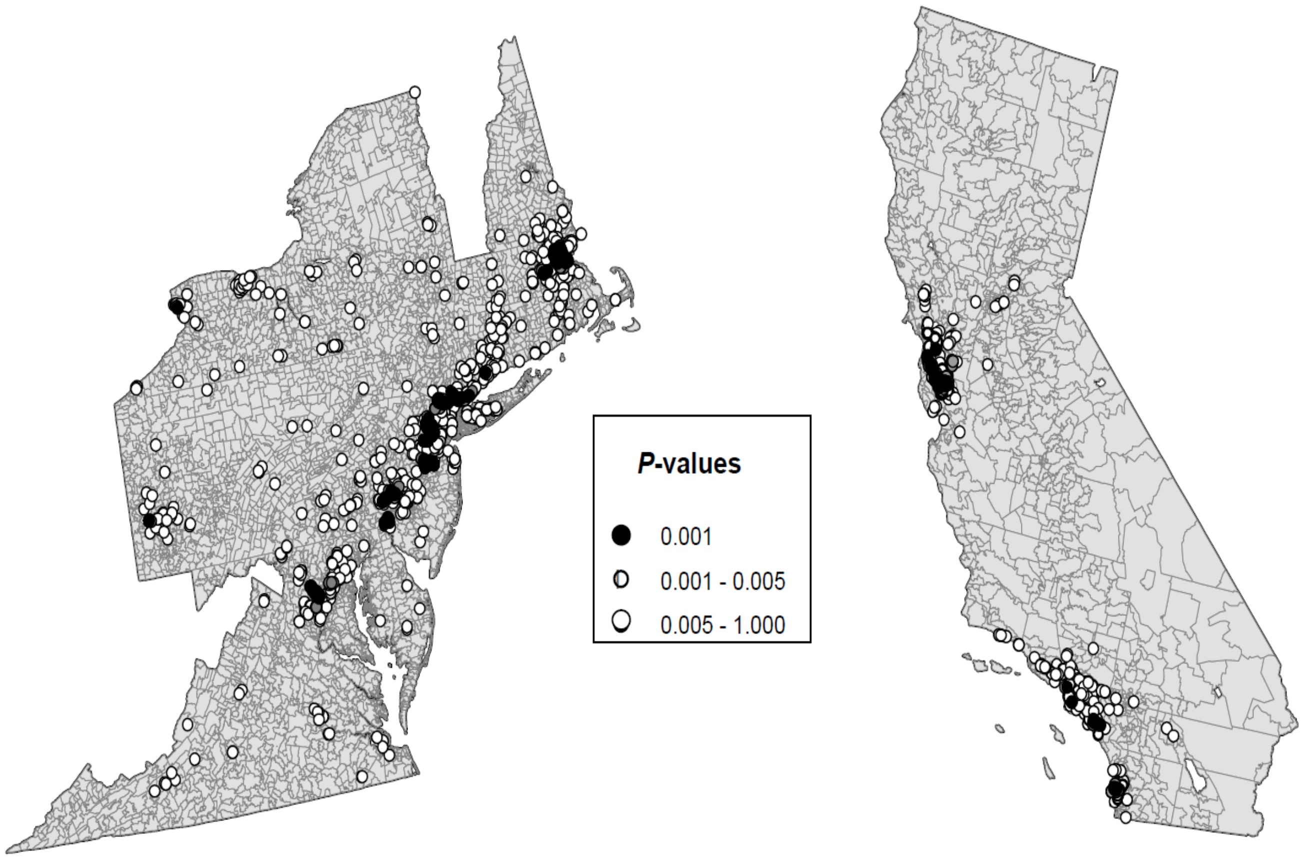
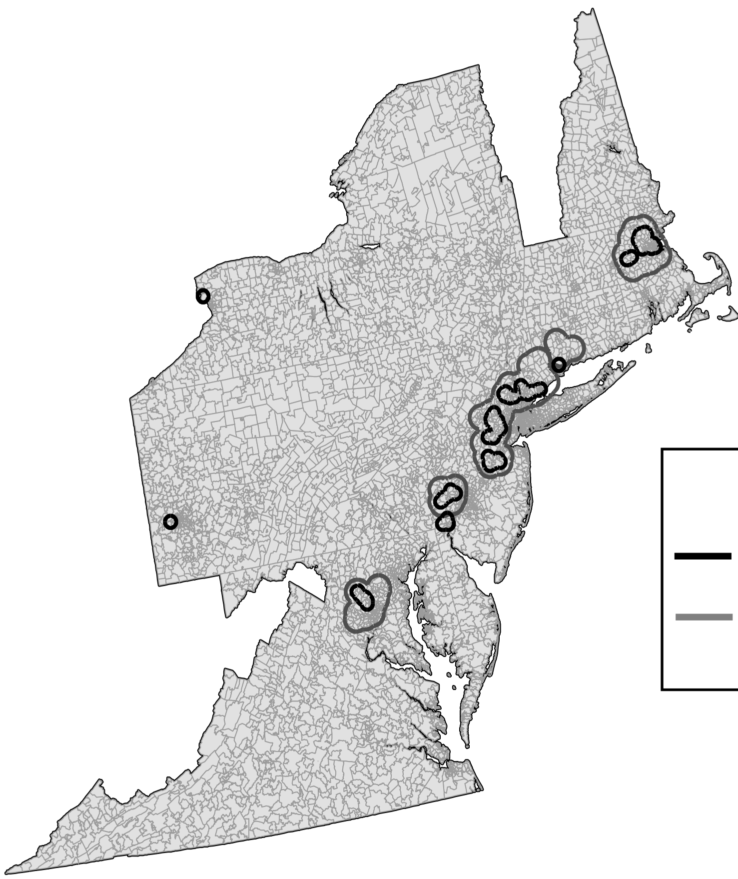


Figure 1a: Northeast Corridor *P*-values at *d* = 5 miles

Figure 1b: California *P*-values at *d* = 5 miles

Source: BCHCS Figure 4a and Figure 4b



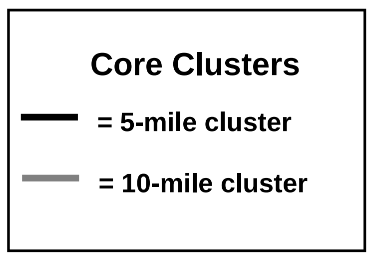


Figure 2a: Northeastern Corridor Core Clusters,
***d* = 5, 10**

Figure 2b: California Core Clusters
***d* = 5, 10**

Source: BCHCS Figure 7a and Figure 7b

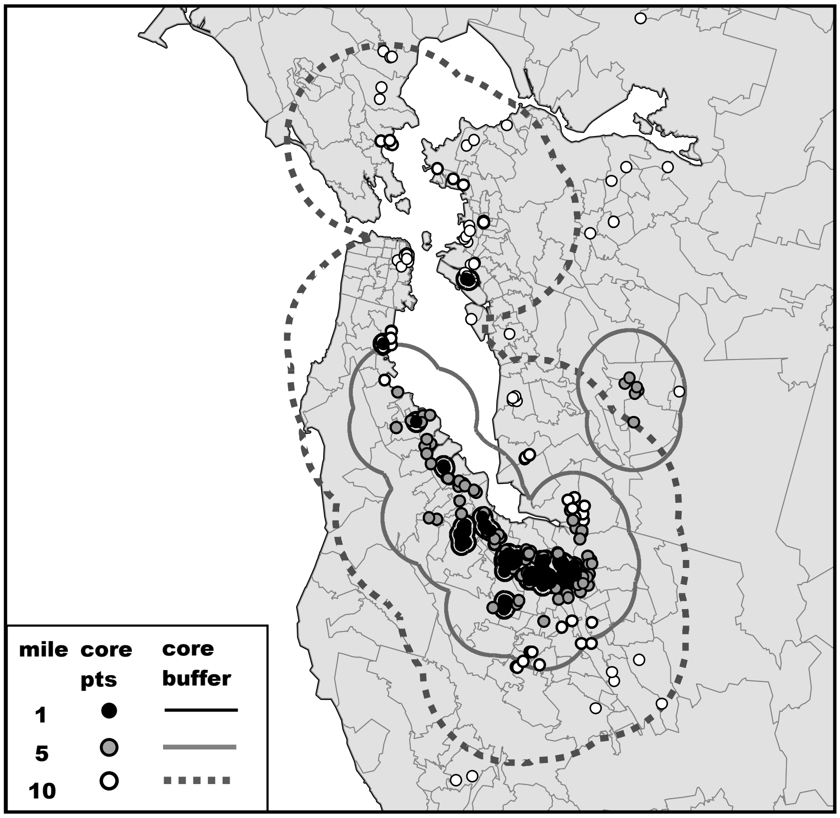


Figure 3a: Multiscale Core Clusters in the San Francisco Bay Area

Source: BCHCS Figure 5

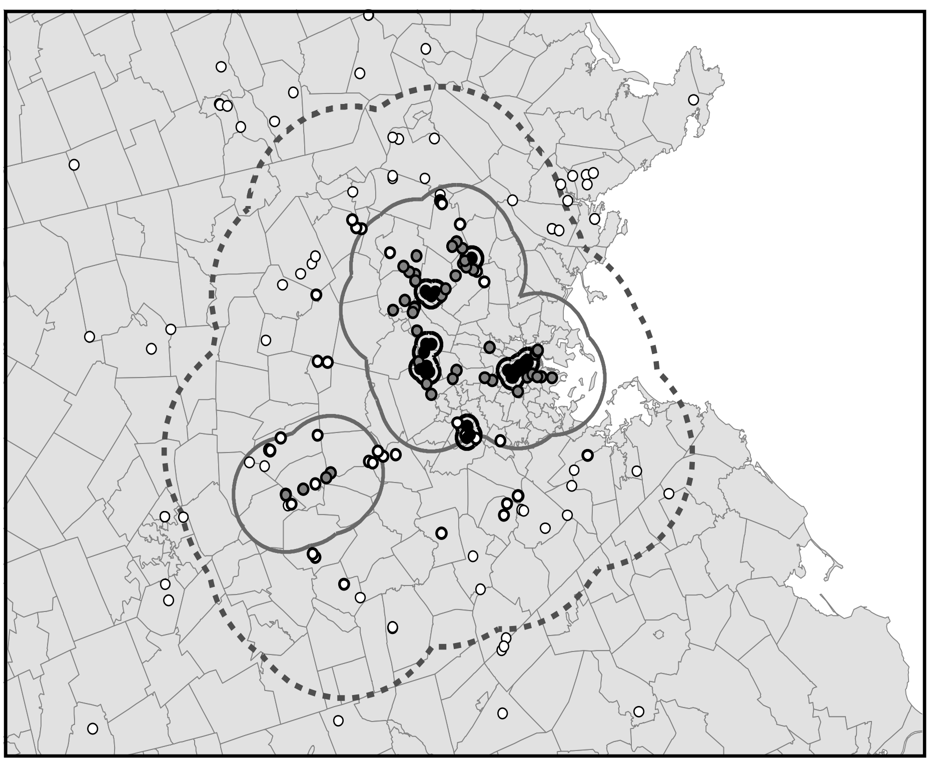


Figure 3b: Multiscale Core Clusters in Boston

Source: BCHCS Figure 6a

Table 1a: 5-Mile Clusters in the Northeastern Corridor, Baseline Results

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
Framingham–Marlborough–Westborough, MA	323	3493	104	2.98 %	3490	104	2.98 %	3490	10	0.29 %	10.4	0.001
Boston–Cambridge–Waltham–Woburn, MA	2634	27642	1715	6.20 %	27634	1715	6.21 %	27634	469	1.70 %	3.7	0.001
Silver Spring–Bethesda, MD–McLean, VA	367	3423	89	2.60 %	3414	89	2.61 %	3414	13	0.38 %	6.8	0.001
Trenton–Princeton, NJ	889	9018	260	2.88 %	9018	260	2.88 %	9018	41	0.45 %	6.3	0.001
Parsippany–Morristown–Union, NJ	1710	14555	358	2.46 %	14551	358	2.46 %	14551	106	0.73 %	3.4	0.001
Greenwich–Stamford, CT–Scarsdale, NY	1205	11209	141	1.26 %	11206	141	1.26 %	11206	51	0.46 %	2.8	0.001
Stratford–Milford, CT	235	1482	12	0.81 %	1481	12	0.81 %	1481	1	0.07 %	12.0	0.001
Conshohocken–King of Prussia–West Chester, PA	539	2348	67	2.85 %	2348	67	2.85 %	2348	8	0.34 %	8.4	0.001
Wilmington–New Castle, DE	624	3499	72	2.06 %	3497	72	2.06 %	3497	15	0.43 %	4.8	0.001
Total	8526	76669	2818	3.68 %	76639	2818	3.68 %	76639	714	0.93 %	3.9	0.001

Table 1b: 10-Mile Clusters in the Northeastern Corridor, Baseline Results

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	**-value**
Boston, MA	4719	48275	4257	8.82 %	48259	4257	8.82 %	48259	1436	2.98 %	3.0	0.001
Washington, D.C.	926	9727	327	3.36 %	9715	327	3.37 %	9715	72	0.74 %	4.5	0.001
New York, NY	7768	67923	4735	6.97 %	67908	4734	6.97 %	67908	2200	3.24 %	2.2	0.001
Philadelphia, PA	1594	9022	409	4.53 %	9020	409	4.53 %	9020	91	1.01 %	4.5	0.001
Total	15007	134947	9728	7.21 %	134902	9727	7.21 %	134902	3799	2.82 %	2.6	0.001

Sources: NBER Patent Data Project and authors' calculations

*The subset of citing patents for which we obtained a similar control patent. See text for details.

**Control Patents are chosen to have the same three-digit technology classification as the citing patent, and their application date must be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned.

Table 2a: 5-Mile Clusters in California, Baseline Results

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
San Diego	444	3430	77	2.24 %	3429	77	2.25 %	3429	13	0.38 %	5.9	0.001
Los Angeles	454	3640	104	2.86 %	3638	104	2.86 %	3638	12	0.33 %	8.7	0.001
Palo Alto–San Jose	11318	145359	26667	18.35 %	145281	26651	18.34 %	145281	13037	8.97 %	2.0	0.001
Dublin–Pleasonton	283	3894	127	3.26 %	3880	127	3.27 %	3880	9	0.23 %	14.1	0.001
Total	12499	156323	26975	17.26 %	156228	26959	17.26 %	156228	13071	8.37 %	2.1	0.001

Table 2b: 10-Mile Clusters in California, Baseline Results

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
San Diego	2099	20062	970	4.84 %	20059	970	4.84 %	20059	325	1.62 %	3.0	0.001
Los Angeles	1266	10668	609	5.71 %	10662	609	5.71 %	10662	113	1.06 %	5.4	0.001
San Francisco	14963	188784	44169	23.40 %	188673	44144	23.40 %	188673	22329	11.83 %	2.0	0.001
Total	18328	219514	45748	20.84 %	219394	45723	20.84 %	219394	22767	10.38 %	2.0	0.001

Sources: NBER Patent Data Project and authors' calculations

*The subset of citing patents for which we obtained a similar control patent. See text for details.

**Control Patents are chosen to have the same three-digit technology classification as the citing patent, and their application date must be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned.

Table 3a: Citation Location Differentials and Spatial Scale (Northeastern Corridor)

Cluster Size	# of Clusters	Originating Patents	Citing Patents	Treatment Proportion	Control Proportion	Location Differential
5-Mile	9	8526	76669	3.68 %	0.93 %	3.9
10-Mile	4	15007	134947	7.21 %	2.82 %	2.6
20-Mile	3	191502	18795	9.82 %	4.86 %	2.0

Sources: NBER Patent Data Project and authors' calculations

Table 3b: Citation Location Differentials and Spatial Scale (California)

Cluster Size	# of Clusters	Originating Patents	Citing Patents	Treatment Proportion	Control Proportion	Location Differential
5-Mile	4	12499	156323	17.26 %	8.37 %	2.1
10-Mile	3	18328	219514	20.84 %	10.38 %	2.0
20-Mile	2	223089	50241	22.52 %	11.34 %	2.0

Sources: NBER Patent Data Project and authors' calculations

Control Patents are chosen to have the same three-digit technology classification as the citing patent, and their application date must be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned.

Table 4a: 5-Mile Clusters in the Northeastern Corridor, Examiner-Added Citations

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
Framingham–Marlborough–Westborough, MA	323	328	4	1.22 %	327	4	1.22 %	327	1	0.31 %	4.0	0.013
Boston–Cambridge–Waltham–Woburn, MA	2634	2353	84	3.57 %	2353	84	3.57 %	2353	37	1.57 %	2.3	0.001
Silver Spring–Bethesda, MD–McLean, VA	367	336	7	2.08 %	335	7	2.09 %	335	1	0.30 %	7.0	0.001
Trenton–Princeton, NJ	889	793	7	0.88 %	793	7	0.88 %	793	3	0.38 %	2.3	0.061
Parsippany–Morristown–Union, NJ	1710	1288	16	1.24 %	1288	16	1.24 %	1288	8	0.62 %	2.0	0.009
Greenwich–Stamford, CT–Scarsdale, NY	1205	922	8	0.87 %	921	8	0.87 %	921	4	0.43 %	2.0	0.025
Stratford–Milford, CT	235	100	0	0.00 %	100	0	0.00 %	100	0	0.00 %	N/A	0.056
Conshohocken–King of Prussia–West Chester, PA	539	123	2	1.63 %	123	2	1.63 %	123	0	0.00 %	N/A	0.037
Wilmington–New Castle, DE	624	189	2	1.06 %	189	2	1.06 %	189	1	0.53 %	2.0	0.096
Total	8526	6432	130	2.02 %	6429	130	2.02 %	6429	55	0.86 %	2.4	0.001

Table 4b: 10-Mile Clusters in the Northeastern Corridor, Examiner-Added Citations

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
Boston, MA	4719	4145	216	5.21 %	4144	216	5.21 %	4144	110	2.65 %	2.0	0.001
Washington, D.C.	926	811	23	2.84 %	810	23	2.84 %	810	6	0.74 %	3.8	0.001
New York, NY	7768	5511	284	5.15 %	5508	284	5.16 %	5508	152	2.76 %	1.9	0.001
Philadelphia, PA	1594	504	14	2.78 %	504	14	2.78 %	504	3	0.60 %	4.7	0.001
Total	15007	10971	537	4.89 %	10966	537	4.90 %	10966	271	2.47 %	2.0	0.001

Sources: NBER Patent Data Project and authors' calculations.

*The subset of citing patents for which we obtained a similar control patent among citations made by patent examiners (2001 onwards). See text for details.

**Control Patents are chosen to have the same three-digit technology classification as the citing patent, and their application date must be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned.

Table 5a: 5-Mile Clusters in California, Examiner-Added Citations

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
San Diego	444	273	3	1.10 %	273	3	1.10 %	273	1	0.37 %	3.0	0.024
Los Angeles	454	258	3	1.16 %	258	3	1.16 %	258	1	0.39 %	3.0	0.032
Palo Alto–San Jose	11318	12306	1603	13.03 %	12304	1603	13.03 %	12304	1160	9.43 %	1.4	0.001
Dublin–Pleasonton	283	277	1	0.36 %	277	1	0.36 %	277	1	0.36 %	1.0	0.440
Total	12499	13114	1610	12.28 %	13112	1610	12.28 %	13112	1163	8.87 %	1.4	0.000

Table 5b: 10-Mile Clusters in California, Examiner-Added Citations

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
San Diego	2099	1791	72	4.02 %	1791	72	4.02 %	1791	30	1.68 %	2.4	0.001
Los Angeles	1266	732	15	2.05 %	732	15	2.05 %	732	7	0.96 %	2.1	0.005
San Francisco	14963	15738	2631	16.72 %	15736	2631	16.72 %	15736	1943	12.35 %	1.4	0.001
Total	18328	18261	2718	14.88 %	18259	2718	14.89 %	18259	1980	10.84 %	1.4	0.001

Sources: NBER Patent Data Project and authors' calculations

*The subset of citing patents for which we obtained a similar control patent among citations made by patent examiners (2001 onwards). See text for details.

**Control Patents are chosen to have the same three-digit technology classification as the citing patent, and their application date must be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned.

Table 6a: 5-Mile Clusters in the Northeastern Corridor, Inventor-Added Citations

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
Framingham–Marlborough–Westborough, MA	323	2910	93	3.20 %	2908	93	3.20 %	2908	9	0.31 %	10.3	0.001
Boston–Cambridge–Waltham–Woburn, MA	2634	23522	1528	6.50 %	23519	1528	6.50 %	23519	406	1.73 %	3.8	0.001
Silver Spring–Bethesda, MD–McLean, VA	367	2843	69	2.43 %	2839	69	2.43 %	2839	11	0.39 %	6.3	0.001
Trenton–Princeton, NJ	889	7676	225	2.93 %	7676	225	2.93 %	7676	33	0.43 %	6.8	0.001
Parsippany–Morristown–Union, NJ	1710	12151	312	2.57 %	12150	312	2.57 %	12150	89	0.73 %	3.5	0.001
Greenwich–Stamford, CT–Scarsdale, NY	1205	9602	118	1.23 %	9601	118	1.23 %	9601	45	0.47 %	2.6	0.001
Stratford–Milford, CT	235	1243	12	0.97 %	1243	12	0.97 %	1243	1	0.08 %	12.0	0.001
Conshohocken–King of Prussia–West Chester, PA	539	2006	61	3.04 %	2006	61	3.04 %	2006	6	0.30 %	10.2	0.001
Wilmington–New Castle, DE	624	3106	65	2.09 %	3105	65	2.09 %	3105	14	0.45 %	4.6	0.001
Total	8526	65059	2483	3.82 %	65047	2483	3.82 %	65047	614	0.94 %	4.0	0.001

Table 6b: 10-Mile Clusters in the Northeastern Corridor, Inventor-Added Citations

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
Boston, MA	4719	40714	3824	9.39 %	40705	3824	9.39 %	40705	1229	3.02 %	3.1	0.001
Washington, D.C.	926	8164	271	3.32 %	8157	271	3.32 %	8157	59	0.72 %	4.6	0.001
New York, NY	7768	57604	4019	6.98 %	57601	4019	6.98 %	57601	1856	3.22 %	2.2	0.001
Philadelphia, PA	1594	7936	373	4.70 %	7935	373	4.70 %	7935	81	1.02 %	4.6	0.001
Total	15007	114418	8487	7.42 %	114398	8487	7.42 %	114398	3225	2.82 %	2.6	0.001

Sources: NBER Patent Data Project and authors' calculations

*The subset of citing patents for which we obtained a similar control patent among citations made by inventors (2001 onwards). See text for details.

**Control Patents are chosen to have the same three-digit technology classification as the citing patent, and their application date must be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned.

Table 7a: 5-Mile Clusters in California, Inventor-Added Citations

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
San Diego	444	2909	71	2.44 %	2909	71	2.44 %	2909	12	0.41 %	5.9	0.001
Los Angeles	454	3054	84	2.75 %	3054	84	2.75 %	3054	10	0.33 %	8.4	0.001
Palo Alto–San Jose	11318	121058	22804	18.84 %	121016	22798	18.84 %	121016	10820	8.94 %	2.1	0.001
Dublin–Pleasonton	283	3294	100	3.04 %	3293	100	3.04 %	3293	7	0.21 %	14.3	0.001
Total	12499	130315	23059	17.69 %	130272	23053	17.70 %	130272	10849	8.33 %	2.1	0.001

Table 7b: 10-Mile Clusters in California, Inventor-Added Citations

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
San Diego	2099	16741	838	5.01 %	16741	838	5.01 %	16741	276	1.65 %	3.0	0.001
Los Angeles	1266	8989	488	5.43 %	8988	488	5.43 %	8988	99	1.10 %	4.9	0.001
San Francisco	14963	157711	38009	24.10 %	157656	37997	24.10 %	157656	18534	11.76 %	2.1	0.001
Total	18328	183441	39335	21.44 %	183385	39323	21.44 %	183385	18909	10.31 %	2.1	0.001

Sources: NBER Patent Data Project and authors' calculations

*The subset of citing patents for which we obtained a similar control patent among citations made by inventors (2001 onwards). See text for details.

**Control Patents are chosen to have the same three-digit technology classification as the citing patent, and their application date must be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned.

Table 8: Citation Differentia Between Labs Inside Clusters vs. Labs Outside Clusters (Difference in Means† Test)							
	Inside Cluster[1]			Outside Cluster[2]			
Area	Mean	Std. Dev.	n	Mean	Std. Dev.	n	*t*-statistic
Boston	12.888	18.148	4,704	9.949	14.895	2,644	7.491
New York City	11.065	16.338	8,279	9.491	14.410	10,600	6.912
Philadelphia	8.030	9.657	1,598	7.654	10.515	3,655	1.262
Washington, D.C.	11.707	17.457	1,273	7.825	10.371	1,741	7.073
Southern California	11.464	15.734	3,668	9.087	12.074	6,716	7.956
Northern California	15.532	19.845	15,106	10.811	15.110	2,680	14.155

Sources: NBER Patent Data Project and authors' calculations

†: Citations per patent granted, 1996–1997

1: Inside Cluster refers to all patents in one or more 10-mile clusters in the region.

2: Outside Cluster refers to all patents outside of the 10-mile clusters in the regions defined as follows:
Boston (Massachusetts/New Hampshire/Rhode Island), New York City (New York/Connecticut/northern New Jersey), Philadelphia (Delaware/eastern Pennsylvania/southern New Jersey), Washington, D.C. (Maryland/D.C./Virginia), southern California (10 southern counties), and northern California (remaining counties).

Table 9a: 5-Mile Clusters in the Northeastern Corridor, STEM Worker Clusters

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
Bethesda–Rockville, MD–Vienna, VA	414	4289	100	2.33 %	4280	100	2.34 %	4280	18	0.42 %	5.6	0.001
Columbia–Laurel, MD	53	494	3	0.61 %	494	3	0.61 %	494	0	0.00 %	N/A	0.001
Phoenix–Cockeysville, MD	72	419	0	0.00 %	419	0	0.00 %	419	0	0.00 %	N/A	0.118
Wilmington, DE	539	2348	67	2.85 %	2348	67	2.85 %	2348	8	0.34 %	8.4	0.001
King of Prussia, PA	974	5532	242	4.37 %	5531	242	4.38 %	5531	36	0.65 %	6.7	0.001
Philadephia, PA	81	617	6	0.97 %	617	6	0.97 %	617	0	0.00 %	N/A	0.001
Princeton, NJ–New York, NY	5124	45968	2322	5.05 %	45957	2322	5.05 %	45957	1049	2.28 %	2.2	0.001
Long Island, NY	270	1913	18	0.94 %	1913	18	0.94 %	1913	4	0.21 %	4.5	0.001
Danbury, CT	347	4405	162	3.68 %	4405	162	3.68 %	4405	11	0.25 %	14.7	0.001
Stratford, CT	240	1499	12	0.80 %	1498	12	0.80 %	1498	2	0.13 %	6.0	0.001
North Haven, CT	105	456	13	2.85 %	456	13	2.85 %	456	1	0.22 %	13.0	0.001
Hartford, CT	87	503	8	1.59 %	503	8	1.59 %	503	0	0.00 %	N/A	0.001
Hudson–Westborough, MA	255	2839	84	2.96 %	2837	84	2.96 %	2837	7	0.25 %	12.0	0.001
Boston–Cambridge, MA	2958	30895	2057	6.66 %	30885	2057	6.66 %	30885	590	1.91 %	3.5	0.001
Nashua, NH	295	2964	53	1.79 %	2963	53	1.79 %	2963	5	0.17 %	10.6	0.001
Binghamton, NY	23	332	0	0.00 %	332	0	0.00 %	332	0	0.00 %	N/A	0.048
Syracuse, NY	40	238	15	6.30 %	238	15	6.30 %	238	0	0.00 %	N/A	0.001
Buffalo, NY	91	410	1	0.24 %	410	1	0.24 %	410	0	0.00 %	N/A	0.193
Pittsburgh, PA	42	165	2	1.21 %	165	2	1.21 %	165	0	0.00 %	N/A	0.004
Pittsburgh–Verona, PA	70	426	4	0.94 %	426	4	0.94 %	426	0	0.00 %	N/A	0.001
Total	12080	106712	5169	4.84 %	106677	5169	4.85 %	106677	1731	1.62 %	3.0	0.001

Sources: NBER Patent Data Project and authors' calculations
*The subset of citing patents for which we obtained a similar control patent. See text for details.
[†]Control Patents are chosen to have the same three-digit technology classification as the citing patent, and their application date must be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned.
The clusters identified in the above table are based on STEM workers as the backcloth. Note that the cluster definitions change because the backcloth changed to STEM workers instead of manufacturing workers as used in Table 1.

Table 9b: 10-Mile Clusters in the Northeastern Corridor, STEM Worker Clusters

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
Richmond, VA	154	666	71	10.66 %	666	71	10.66 %	666	5	0.75 %	14.2	0.001
Washington, D.C.–Baltimore, MD	1376	12709	537	4.23 %	12696	537	4.23 %	12696	127	1.00 %	4.2	0.001
Hagerstown, MD	17	40	1	2.50 %	40	1	2.50 %	40	0	0.00 %	N/A	0.091
Lancaster, PA	104	565	8	1.42 %	565	8	1.42 %	565	1	0.18 %	8.0	0.001
Philadelphia, PA–Wilmington, DE–Cherry Hill, NJ	2601	14152	990	7.00 %	14150	990	7.00 %	14150	214	1.51 %	4.6	0.001
Pittsburgh, PA	921	5803	400	6.89 %	5799	400	6.90 %	5799	25	0.43 %	16.0	0.001
Binghamton, NY	329	3124	31	0.99 %	3122	31	0.99 %	3122	2	0.06 %	15.5	0.001
Syracuse, NY	130	678	44	6.49 %	678	44	6.49 %	678	1	0.15 %	44.0	0.001
Rochester, NY	1571	7979	391	4.90 %	7975	391	4.90 %	7975	75	0.94 %	5.2	0.001
Buffalo, NY	122	632	3	0.47 %	632	3	0.47 %	632	0	0.00 %	N/A	0.015
Boston, MA	4682	47925	3898	8.13 %	47908	3898	8.14 %	47908	1386	2.89 %	2.8	0.001
New York, NY–Northern NJ–CT	9514	80890	6238	7.71 %	80871	6238	7.71 %	80871	3336	4.13 %	1.9	0.001
Total	21521	175163	12612	7.20 %	175102	12612	7.20 %	175102	5172	2.95 %	2.4	0.001

Sources: NBER Patent Data Project and authors' calculations

*The subset of citing patents for which we obtained a similar control patent. See text for details.

**Control Patents are chosen to have the same three-digit technology classification as the citing patent, and their application date must be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned.

The clusters identified in the above table are based on STEM workers as the backcloth. Note that the cluster definitions change because the backcloth changed to STEM workers instead of manufacturing workers as used in Table 1.

Table 10a: 5-Mile Clusters in California, STEM Worker Clusters

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
San Diego–La Jolla	563	4129	119	2.88 %	4128	119	2.88 %	4128	21	0.51 %	5.7	0.001
Carlsbad	261	1628	43	2.64 %	1628	43	2.64 %	1628	9	0.55 %	4.8	0.001
Irvine	946	7456	375	5.03 %	7454	375	5.03 %	7454	52	0.70 %	7.2	0.001
Camarillo	199	1942	39	2.01 %	1942	39	2.01 %	1942	3	0.15 %	13.0	0.001
Santa Barbara	82	1401	55	3.93 %	1401	55	3.93 %	1401	1	0.07 %	55.0	0.001
San Jose–Santa Clara	14220	182287	42532	23.33 %	182191	42514	23.33 %	182191	21282	11.68 %	2.0	0.001
Pleasonton	283	3895	127	3.26 %	3881	127	3.27 %	3881	9	0.23 %	14.1	0.001
Santa Rosa	127	1013	29	2.86 %	1012	29	2.87 %	1012	1	0.10 %	29.0	0.001
Total	16681	203751	43319	21.26 %	203637	43301	21.26 %	203637	21378	10.50 %	2.0	0.001

Table 10b: 10-Mile Clusters in California, STEM Worker Clusters

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
San Diego	2146	20486	1056	5.15 %	20483	1056	5.16 %	20483	347	1.69 %	3.0	0.001
Anaheim–Irving	1911	15333	1063	6.93 %	15324	1063	6.94 %	15324	203	1.32 %	5.2	0.001
Oxnard–Camarillo	76	475	15	3.16 %	475	15	3.16 %	475	0	0.00 %	N/A	0.001
Santa Barbara	288	3296	129	3.91 %	3296	129	3.91 %	3296	6	0.18 %	21.5	0.001
San Francisco–Palo Alto–San Jose	14564	185477	44083	23.77 %	185379	44064	23.77 %	185379	21975	11.85 %	2.0	0.001
Santa Rosa	144	1197	54	4.51 %	1195	53	4.44 %	1195	2	0.17 %	26.5	0.001
Total	19129	226264	46400	20.51 %	226152	46380	20.51 %	226152	22533	9.96 %	2.1	0.001

Sources: NBER Patent Data Project and authors' calculations

*The subset of citing patents for which we obtained a similar control patent. See text for details.

**Control Patents are chosen to have the same three-digit technology classification as the citing patent, and their application date must be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned.

The clusters identified in the above table are based on STEM workers as the backcloth. Note that the cluster definitions change because the backcloth changed to STEM workers instead of manufacturing workers as used in Table2.

Table 11: MSAs, Baseline Results

					Treatment Group			Control Group					
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**	**Location Differential Baseline 10-mile clusters**
Boston	5732	57407	5511	9.60 %	57389	5511	9.60 %	57389	2008	3.50 %	2.7	0.001	3.0
Los Angeles	6192	47083	5155	10.95 %	47069	5155	10.95 %	47069	1709	3.63 %	3.0	0.001	5.4
New York	10663	95177	7798	8.19 %	95150	7797	8.19 %	95150	4655	4.89 %	1.7	0.001	2.2
Philadelphia	4406	26955	2024	7.51 %	26952	2023	7.51 %	26952	592	2.20 %	3.4	0.001	4.5
San Diego	2717	24387	1494	6.13 %	24382	1494	6.13 %	24382	524	2.15 %	2.9	0.001	3.0
San Francisco	18530	230392	63356	27.50 %	230264	63329	27.50 %	230264	32819	14.25 %	1.9	0.001	2.0
Washington, D.C.	2653	23907	1513	6.33 %	23891	1510	6.32 %	23891	440	1.84 %	3.4	0.001	4.5
Total	50893	505308	86851	17.19 %	505097	86819	17.19 %	505097	42747	8.46 %	2.0	0.001	2.1

Based on a 1990 CMSA definitions, with the exception of San Diego, which is based on a 1990 MSA definition

Sources: NBER Patent Data Project and authors' calculations

*The subset of citing patents for which we obtained a similar control patent. See text for details.

**Control Patents are chosen to have the same three-digit technology classification as the citing patent, and their application date must be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned.

Table 12a: 5-Mile Clusters in the Northeastern Corridor, Originating-Citing-Control Triad Subclass Restriction

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
Framingham–Marlborough–Westborough, MA	323	3493	104	2.98 %	966	25	2.59 %	966	2	0.21 %	12.5	0.001
Boston–Cambridge–Waltham–Woburn, MA	2634	27642	1715	6.20 %	7178	564	7.86 %	7178	73	1.02 %	7.7	0.001
Silver Spring–Bethesda, MD–McLean, VA	367	3423	89	2.60 %	959	26	2.71 %	959	1	0.10 %	26.0	0.001
Trenton–Princeton, NJ	889	9018	260	2.88 %	2321	97	4.18 %	2321	7	0.30 %	13.9	0.001
Parsippany–Morristown–Union, NJ	1710	14555	358	2.46 %	3949	113	2.86 %	3949	20	0.51 %	5.6	0.001
Greenwich–Stamford, CT–Scarsdale, NY	1205	11209	141	1.26 %	3147	54	1.72 %	3147	12	0.38 %	4.5	0.001
Stratford–Milford, CT	235	1482	12	0.81 %	456	4	0.88 %	456	0	0.00 %	N/A	0.001
Conshohocken–King of Prussia–West Chester, PA	539	2348	67	2.85 %	607	17	2.80 %	607	1	0.16 %	17.0	0.001
Wilmington–New Castle, DE	624	3499	72	2.06 %	964	32	3.32 %	964	1	0.10 %	32.0	0.001
Total	8526	76669	2818	3.68 %	20547	932	4.54 %	20547	117	0.57 %	8.0	0.001

Table 12b: 10-Mile Clusters in the Northeastern Corridor, Disaggregated Subclasses

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
Boston, MA	4719	48275	4257	8.82 %	12548	1214	9.67 %	12548	216	1.72 %	5.6	0.001
Washington, D.C.	926	9727	327	3.36 %	2565	123	4.80 %	2565	11	0.43 %	11.2	0.001
New York, NY	7768	67923	4735	6.97 %	18774	1475	7.86 %	18774	472	2.51 %	3.1	0.001
Philadelphia, PA	1594	9022	409	4.53 %	2238	127	5.67 %	2238	15	0.67 %	8.5	0.001
Total	15007	134947	9728	7.21 %	36125	2939	8.14 %	36125	714	1.98 %	4.1	0.001

Sources: NBER Patent Data Project and authors' calculations

*The subset of citing patents for which we obtained a similar control patent. See text for details.

**Control Patents are chosen to have the same primary technology classification as the citing patent and such that they have at least one subclass in common with both the originating and the citing patent. Their application date must also be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned.

Table 13a: 5-Mile Clusters in California, Disaggregated Subclasses

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
San Diego	444	3430	77	2.24 %	1038	29	2.79 %	1038	5	0.48 %	5.8	0.001
Los Angeles	454	3640	104	2.86 %	860	23	2.67 %	860	4	0.47 %	5.8	0.001
Palo Alto–San Jose	11318	145359	26667	18.35 %	40729	7820	19.20 %	40729	4619	11.34 %	1.7	0.001
Dublin–Pleasonton	283	3894	127	3.26 %	1070	66	6.17 %	1070	2	0.19 %	33.0	0.001
Total	12499	156323	26975	17.26 %	43697	7938	18.17 %	43697	4630	10.60 %	1.7	0.001

Table 13b: 10-Mile Clusters in California, Disaggregated Subclasses

					Treatment Group			Control Group				
Column	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
Cluster	**Originating Patents**	**Citing Patents**	**From Same Cluster**	**Percent (C/B)**	**Matched Citing Patents***	**From Same Cluster***	**Percent (F/E)**	**Control Patents****	**From Same Cluster****	**Percent (I/H)**	**Location Differential (G/J)**	***P*-value**
San Diego	2099	20062	970	4.84 %	6262	318	5.08 %	6262	139	2.22 %	2.3	0.001
Los Angeles	1266	10668	609	5.71 %	2738	153	5.59 %	2738	53	1.94 %	2.9	0.001
San Francisco	14963	188784	44169	23.40 %	52987	13043	24.62 %	52987	7720	14.57 %	1.7	0.001
Total	18328	219514	45748	20.84 %	61987	13514	21.80 %	61987	7912	12.76 %	1.7	0.001

Sources: NBER Patent Data Project and authors' calculations

*The subset of citing patents for which we obtained a similar control patent. See text for details.

**Control Patents are chosen to have the same primary technology classification as the citing patent and such that they have at least one subclass in common with both the originating and the citing patent. Their application date must also be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned.

Table 14a: 5-Mile Clusters in the Northeastern Corridor, Coarsened Exact Matching

					Treatment Group			Control Group				
Column	A	B	C	D	E	F	G	H	I	J	K	L
Cluster	Originating Patents	Citing Patents	From Same Cluster	Percent (C/B)	Matched Citing Patents*	From Same Cluster*	Percent (F/E)	Control Patents†	From Same Cluster	Percent (I/H)	Location Differential (G/J)	*t*-Statistic
Framingham–Marlborough–Westborough, MA	323	3,498	104	2.97%	2,845	80	2.81%	2,845	9	0.32%	**8.9**	**7.6**
Boston–Cambridge–Waltham–Woburn, MA	2,634	27,664	1,717	6.21%	22,937	1,400	6.10%	22,937	284	1.24%	**4.9**	**27.9**
Silver Spring–Bethesda, MD–McLean, VA	367	3,424	89	2.60%	2,779	69	2.48%	2,779	15	0.54%	**4.6**	**6.0**
Trenton–Princeton, NJ	889	9,022	260	2.88%	7,453	207	2.78%	7,453	25	0.34%	**8.3**	**12.1**
Parsippany–Morristown–Union, NJ	1,710	14,567	358	2.46%	11,912	282	2.37%	11,912	91	0.76%	**3.1**	**10.0**
Greenwich–Stamford, CT–Scarsdale, NY	1,205	11,218	141	1.26%	9,277	109	1.17%	9,277	49	0.53%	**2.2**	**4.8**
Stratford–Milford, CT	235	1,484	12	0.81%	1,228	11	0.90%	1,228	2	0.16%	**5.5**	**2.5**
Conshohocken–King of Prussia–West Chester, PA	539	2,352	68	2.89%	1,964	53	2.70%	1,964	13	0.66%	**4.1**	**5.0**
Wilmington–New Castle, DE	624	3,501	72	2.06%	2,940	53	1.80%	2,940	11	0.37%	**4.8**	**5.3**
All 5-Mile Clusters	8,526	76,730	2,821	3.68%	63,335	2,264	3.57%	63,335	499	0.79%	**4.5**	**34.1**

Table 14b: 10-Mile Clusters in the Northeastern Corridor, Coarsened Exact Matching

					Treatment Group			Control Group				
Column	A	B	C	D	E	F	G	H	I	J	K	L
Cluster	Originating Patents	Citing Patents	From Same Cluster	Percent (C/B)	Matched Citing Patents*	From Same Cluster*	Percent (F/E)	Control Patents†	From Same Cluster	Percent (I/H)	Location Differential (G/J)	*t*-Statistic
Boston, MA	4,719	48,315	4,263	8.82%	39,760	3,493	8.79%	39,760	896	2.25%	**3.9**	**40.7**
Washington, D.C.	926	9,741	327	3.36%	7,851	250	3.18%	7,851	58	0.74%	**4.3**	**11.1**
New York, NY	7,768	67,982	4,738	6.97%	55,989	3,706	6.62%	55,989	1,710	3.05%	**2.2**	**27.9**
Philadelphia, PA	1,594	9,028	409	4.53%	7,603	327	4.30%	7,603	68	0.89%	**4.8**	**13.3**
All 10-Mile Clusters	15,007	135,066	9,737	7.21%	111,203	7,776	6.99%	111,203	2,732	2.46%	**2.8**	**50.7**

Sources: NBER Patent Data Project and authors' calculations

*The subset of citing patents for which we obtained a similar control patent. See text for details.

†Control patents are selected using the coarsened exact matching procedure. Control patents must have the same three-digit technology classification as the citing patent, and their application date must be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned. Control patents must have the same application year and three-digit technology classification as the treatment patents, in addition to having the same grant year and the number of citations that the treatment patent receives.

Table 15a: 5-Mile Clusters in California, Coarsened Exact Matching

					Treatment Group			Control Group				
Column	A	B	C	D	E	F	G	H	I	J	K	L
Cluster	Originating Patents	Citing Patents	From Same Cluster	Percent (C/B)	Matched Citing Patents*	From Same Cluster*	Percent (F/E)	Control Patents†	From Same Cluster	Percent (I/H)	Location Differential (G/J)	*t*-Statistic
San Diego	444	3,434	77	2.24%	2,811	58	2.06%	2,811	14	0.50%	**4.1**	**5.2**
Los Angeles	454	3,646	104	2.85%	3,019	79	2.62%	3,019	5	0.17%	**15.8**	**8.2**
Palo Alto–San Jose	11,318	145,471	26,684	18.34%	118,537	21,223	17.90%	118,537	8,962	7.56%	**2.4**	**76.5**
Dublin–Pleasonton	283	3,899	127	3.26%	3,199	87	2.72%	3,199	9	0.28%	**9.7**	**8.1**
All 5-Mile Clusters	12,499	156,450	26,992	17.25%	127,566	21,447	16.81%	127,566	8,990	7.05%	**2.4**	**77.0**

Table 15b: 10-Mile Clusters in California, Coarsened Exact Matching

					Treatment Group			Control Group				
Column	A	B	C	D	E	F	G	H	I	J	K	L
Cluster	Originating Patents	Citing Patents	From Same Cluster	Percent (C/B)	Matched Citing Patents*	From Same Cluster*	Percent (F/E)	Control Patents†	From Same Cluster	Percent (I/H)	Location Differential (G/J)	*t*-Statistic
San Diego	2,099	20,079	970	4.83%	16,392	801	4.89%	16,392	335	2.04%	**2.4**	**14.1**
Los Angeles	1,266	10,685	609	5.70%	8,915	457	5.13%	8,915	90	1.01%	**5.1**	**16.1**
San Francisco	14,963	188,943	44,215	23.40%	154,195	35,457	22.99%	154,195	14,455	9.37%	**2.5**	**104.5**
All 10-Mile Clusters	18,328	219,707	45,794	20.84%	179,502	36,715	20.45%	179,502	14,880	8.29%	**2.5**	**105.5**

Source: NBER Patent Data Project and authors' calculations

*The subset of citing patents for which we obtained a similar control patent. See text for details.

†Control patents are selected using the coarsened exact matching procedure. Control patents must have the same three-digit technology classification as the citing patent, and their application date must be within a one-year window of the citing patent's application date. These control patents are chosen with replacement sampling. We eliminate self-citations and do not allow controls to be drawn from patents assigned to the same firm to which the originating patent is assigned. Control patents must have the same application year and three-digit technology classification as the treatment patents, in addition to having the same grant year and the number of citations that the treatment patent receives.

Table 16†		
Northeastern Corridor		
Cluster Name	Coefficient on Originating Patent ($\hat{\beta}_h$)	Standard Errors
Framingham–Marlborough–Westborough, MA	2.82	0.1062*
Boston–Cambridge–Waltham–Woburn, MA	1.5	0.0300*
Silver Spring–Bethesda, MD–McLean, VA		
Trenton–Princeton, NJ	2.17	0.0737*
Parsippany–Morristown–Union, NJ	1.26	0.0603*
Greenwich–Stamford, CT–Scarsdale, NY	0.8	0.0967*
Stratford–Milford, CT	2.26	0.3235*
Conshohocken-King of Prussia-West Chester, PA	3.13	0.1321*
Wilmington–New Castle, DE	2.28	0.1335*
Boston, MA (10 mile)	1.37	0.0199*
Washington, D.C. (10 mile)	1.65	0.0652*
New York, NY (10 mile)	0.79	0.0192*
Philadelphia, PA (10 mile)	2.13	0.0574*
Broad Regions		
Northeastern corridor, 5-mile Total	0.77	0.0167*
Northeastern corridor, 10-mile Total	0.68	0.0113*
California		
Cluster Name	Coefficient on Originating Patent ($\hat{\beta}_h$)	Standard Errors
San Diego	2.34	0.1251*
Los Angeles	2.52	0.1137*
Palo Alto–San Jose	1.06	0.0107*
Dublin–Pleasonton	2.81	0.1098*
San Diego (10 mile)	1.56	0.0381*
Los Angeles (10 mile)	2.06	0.0493*
San Francisco (10 mile)	1.09	0.0093*
Broad Regions		
California, 5-mile Total	1.01	0.0103*
California, 10-mile Total	0.99	0.0086*

†The California regressions included 1,390,727 observations
The northeastern corridor regressions included 1,444,272 observations.
Robust standard errors are reported.
*Indicates significance at the 1 percent level.

Table 17: Summary of Location Differentials

	Northeastern Corridor 5-Mile Cluster	Northeastern Corridor 10-Mile Cluster	California 5-Mile Cluster	California 10-Mile Cluster
Baseline	3.9	2.6	2.1	2.0
Examiner-Added	2.4	2.0	1.4	1.4
Inventor-Added	4.0	2.6	2.1	2.1
STEM	3.0	2.4	2.0	2.1
CMSA	2.1			
Subclass Restriction	8.0	4.1	1.7	1.7
CEM	4.5	2.8	2.4	2.5

[†]Baseline results from Column K in Tables 1 and 2; Examiner-Added results from Column K in Tables 4 and 5; Inventor-Added results from Column K in Tables 6 and 7; STEM results from Column K in Tables 9 and 10; CMSA results from Column K in Table 11; Subclass Restriction results from Column K in Tables 12 and 13; and CEM results from Column K in Tables 14 and 15.