GEOGRAPHIC LOCATION AND DECENTRALIZATION OF INNOVATION ACTIVITY

Aija Leiponen
Department of Applied Economics and Management
251 Warren Hall
Cornell University
Ithaca, NY 14853-7801

Constance E. Helfat
Tuck School of Business at Dartmouth
100 Tuck Hall
Hanover, NH 03755

September 2006

PRELIMINARY

This paper benefited from the comments of colleagues at a workshop at Cornell University and at the Academy of Management meetings in Atlanta, particularly Harbir Singh. The usual caveat applies.

GEOGRAPHIC LOCATION AND DECENTRALIZATION OF INNOVATION ACTIVITY

ABSTRACT

Companies face an expanding set of choices about where to locate their innovation activity, both abroad and within their home countries. This location choice also requires firms to make a simultaneous choice about the organizational structure of innovation activity: almost by definition, multiple geographic locations per firm imply some degree of decentralization. In research to date, the questions of whether to conduct innovation activity in multiple geographic locations and whether to decentralize technological innovation have remained largely separate. Using uniquely detailed firm-level data on innovation output and the location of research and development (R&D) activity, we shed new light on the relationship between firms' multiple geographic locations for R&D and their innovation success. Our results indicate that geographic decentralization of R&D at the firm level is generally beneficial in terms of the extent and breadth of innovation success—product or process—and the effect is strongly related to the knowledge sourcing strategies that firms employ. These results are consistent with the interpretation that R&D location decisions are driven by the desire of firms to access external sources of knowledge for innovation activities. We also find that the benefits of multiple R&D locations do not apply to novel, or new-to-the-market, innovations. This has important implications for the many studies that use patent data, which reflect only novel technologies. Our results suggest that when analyzing technological innovation, it is important to distinguish between novel and imitative innovations, since the results may differ for these two types of innovation.

INTRODUCTION

Technological innovation has become ever more important to the success of business enterprises. No industry is immune. Technological innovation in information technology has revolutionized the retailing industry. Technological innovation in techniques for locating and extracting oil has vastly increased available fossil fuel reserves. Technological advance affects all industries and firms, not just those considered hi-tech. At the same time as technological innovation has become pervasive, globalization of business has increased rapidly. Companies face an expanding set of choices about where to locate their innovation activity, both abroad and within their home countries. This location choice also requires firms to make a simultaneous choice about the organizational structure of innovation activity: almost by definition, multiple geographic locations per firm imply some degree of decentralization. We use the term "geographic decentralization" to denote the dual choice of location and organizational structure.

In research to date, the questions of whether to conduct innovation activity in multiple geographic locations and whether to decentralize technological innovation have remained largely separate. Typically, the issue of geographic location of innovation is viewed through the lenses of related literatures on foreign direct investment and the knowledge-based view of the firm. Literature on organizational structure that deals with the question of centralization versus decentralization, however, also can be brought to bear on the issue of the geographic location of innovation activity. The predictions of these two literatures regarding the most effective way to organize and locate innovation activity are sometimes conflicting and other times congruent. One of the issues on which these two theories differ involves the effects on geographic decentralization of R&D on innovation outcomes. Using uniquely detailed firm-level data on innovation output and the location of research and development (R&D) activity, we can shed new light on whether firms that have multiple geographic locations also have greater innovation success. We do not investigate other aspects of decentralization, such as which units have budgetary control over R&D decisions. Although our data contain unusually detailed information about R&D locations within firms, we do not have additional information regarding the structure of command and control inside organizations.

Our analysis makes several contributions. First, we bring together two literatures that rarely meet, in order to examine the impact of geographic decentralization on innovation output. Second, we examine geographic location choices of innovation activity largely within a single country, and find that even though much of the theory was developed in the international setting, the theoretical predictions apply within countries as well. Third, we compare the effect of geographic location choice on novel versus imitative (new-to-the-market versus new-to-the-firm) innovation output. Although research on innovation often focuses on patents, which reflect novel technologies, a great deal of business research and development (R&D) is directed toward imitative innovation. Fourth, we can overcome some well-known problems of using patent data. Patents are meaningful measures only of intermediate innovation output and in a relatively small number of industries (Griliches, 1990). In contrast, our dataset contains information about sales of commercialized innovations, and is close to representative of the manufacturing sector as a whole. Finally, we test whether any positive impact on innovation output of geographic decentralization of R&D is accounted for by an accompanying breadth of search for new knowledge, as one prevailing theory would predict.

In what follows, we first compare and contrast the relevant literatures and develop testable hypotheses. Then we describe our data, variables, and empirical methodology. Finally, we report and discuss the empirical results, and conclude with implications for future research.

GEOGRAPHIC DECENTRALIZATION OF INNOVATION

The question of geographic location of technological innovation activity is often framed as one of foreign direct investment (Kuemmerle, 1999; Penner-Hahn and Shaver, 2005). This framing derives from the early literature on how firms should best access foreign markets, whether by foreign direct investment or by other means such as exporting. Since the literature on foreign direct investment is very large and impossible to summarize in a few short paragraphs, we highlight a few key points relevant to our analysis here.

Early literature argued that firms should use foreign direct investment (FDI) when they had idiosyncratic assets that they sought to exploit in foreign markets. Following Hymer (1970), Buckley and Casson (1976), and Dunning (1981), scholars of the "internalization school" emphasized the advantages of internalizing economic activity when firms sought to employ special advantages abroad. Technology transfer was a chief example of such activity (Teece, 1977). Foreign direct investment in order to transfer technology abroad led firms to have multiple geographic locations for productive activity. More recently, scholars have argued that in addition to exploiting special advantages such as technological knowledge, firms often seek to acquire technological knowledge abroad (Kogut, 1991). Kuemmerle (1999) has termed these two motives "home base exploiting FDI" and "home base augmenting FDI," respectively.

The "internalization" view has a close relationship to the predictions of transaction cost theory, which states that firms should use internal organization rather than markets when transactions run a large risk of ex post opportunism. Since technology transfer, and transfer of proprietary knowledge more generally, has a high risk of ex post opportunism, transaction cost theory implies that firms should use internal organization to transfer proprietary technology abroad (Teece, 1980). Thus, internalization theory and transactions cost theory are closely aligned.

Kogut and Zander (1993) subsequently argued that other factors could predict technology transfer through foreign direct investment, independent of internalization and transactions cost arguments. They noted that a great deal of technological knowledge is tacit or not well codified, and that even the use of codified knowledge requires an understanding of the context in which the knowledge is employed. Under these circumstances, firms are superior to markets in transferring technology and information because organizations consist of communities of practice among their members and therefore have shared understanding. This in turn greatly reduces the costs of transferring information internally relative to using the market. This general argument regarding the superiority of firms as a means of knowledge transfer is one of the central tenets of what has come to be known as the "knowledge-based view of the firm" (Kogut and Zander, 1992; Grant, 1996).

The knowledge-based view has clear implications for geographic decentralization of innovation activity. Particularly when firms seek to access knowledge specific to other locales, such as other countries, they may need to establish R&D facilities in these locales if the relevant technological knowledge diffuses slowly (Kogut, 1991). Since much technological knowledge is tacit, its transfer requires frequent interactions between the sender and receiver—which proximity facilitates (Kogut and Zander, 1993). For this reason, firms are likely to have multiple locations for their innovation activity. This theory further implies that geographic decentralization in R&D will lead to greater innovation success because firms can acquire additional technological knowledge that they could not have obtained in a single R&D location. By conducting R&D in multiple locations, firms will have access to a larger number of sources of information useful in innovation. Specialized local knowledge useful in innovation can come from many different sources, including universities, research institutes, suppliers, customers, and competitors. A number of studies have documented that firms that conduct R&D outside of their home countries or that apply for patents abroad through foreign subsidiaries succeed in accessing and building upon local technological expertise (e.g., Almeida, 1996; Florida, 1997; Kuemmerle, 1999; Frost, 2001).

The internalization, transactions cost, and knowledge-based view arguments all contain the same implication: when firms seek to transfer or access external technology, there are clear advantages to doing so internally using multiple geographic R&D locations. Firms that use geographic decentralization for purposes of "home base augmenting" R&D will be able to access a greater number of knowledge sources than they could access in a single R&D location. This in turn improves the likelihood of innovation success. Moreover, if geographic decentralization yields a wider range of knowledge on which to build and recombine, the resulting set of innovations should span a wider range of applicability as well (e.g., product and process innovations, rather than just one or the other). These arguments regarding the benefits of geographic decentralization of innovation, although developed largely in the context of foreign expansion of innovation activity, also apply to geographic locations within a single country if these locations have types of knowledge relevant to innovation that are difficult to obtain from afar (Kuemmerle, 1999). Thus, we frame the hypotheses arising out of the foregoing literatures to include all geographic location choices, whether foreign or domestic.

H1: Geographic decentralization of innovation activity is associated with greater innovation success.

H2: Any positive association between geographic decentralization of innovation activity and innovation success reflects a positive correlation between geographic decentralization and the number of knowledge sources accessed outside of the boundaries of the organization.

H3a: Geographic decentralization of innovation activity is associated with a wider range of applicability of innovation output.

A literature largely separate from that of FDI and the knowledge-based view concerns the organizational structure of the firm in general, and the question of whether to use a centralized versus a decentralized organization form in particular. The use of multiple geographic locations for innovation activity within an organization necessarily brings with it some degree of decentralization. As a practical matter, a firm cannot operate organizational units from a distance without delegating at least some responsibility for day-to-day operations to the geographically distant units. Headquarters may or may not choose to delegate longer-term operational or strategic decisions to organizational units in other geographic locations. But even if headquarters delegates no long-term decisions at all, the firm will need an operational manager on-site in its geographically distant units to carry out the decisions of headquarters. In short, a certain amount of decentralized authority is unavoidable when firms have multiple geographic locations. We next examine the implications of the literature on organizational structure for the geographic decentralization of innovation activity. Like the literature on foreign direct investment, the literature on organizational form is very large. Here we abstract a few key points for our analysis.

As documented by Alfred Chandler (1962), many large, diversified firms in the U.S. adopted a decentralized organizational structure, in contrast to earlier structures that organized tasks centrally according to function. Within this decentralized structure, operational decision making and control occurred at the level of divisions organized according to product markets, and strategic decision making took place at the top of the organization. In this type of an organization, which Oliver Williamson (1975) later referred to as the M-form, delegation of operational control reduced the burden on the bounded rationality of top managers, enabling them to focus on strategic issues for the company as a whole. Moreover, by using divisions as profit centers, top management could more easily obtain information about business unit performance and use these data to reward managers. This created higher-powered incentives and reduced agency problems and opportunism within the corporation (Williamson, 1991).

This logic, when applied to the organization of R&D, suggests that decentralization has the advantage that it minimizes unnecessary coordination of R&D decisions across divisions and allocates operative control of R&D projects on site (Argryes and Silverman, 2004). Then, each

R&D unit can pursue the best innovation opportunities available in its environment. This transaction cost argument in favor of decentralization of R&D meshes well with the application of transaction cost logic to international R&D described earlier, which implied that firms should geographically decentralize. Both with regard to organizational form and location choices, transaction cost analysis leads to hypothesis H1 above.

By delegating operational decisions to division management, including decisions regarding R&D, the M-form firm minimizes on what Williamson (1975) termed "information impactedness." That is, the M-form insures that the division managers who are closer to customers and markets, and therefore have superior information about those markets (relative to top management), also have responsibility for product R&D directed towards those customers and markets (Argyres and Silverman, 2004). The literature on "user-based" innovation (Von Hippel, 1998) also implies that decentralization of innovation activity will benefit firms by bringing them closer to their customers. Von Hippel (1988) observed that companies often received valuable ideas for innovation from their customers who had additional or different needs than their current products or services provided. Additionally, a recent strand of the literature on organization form suggests that by decentralizing, firms can better adapt to changes in markets through "patching" (involving a modular decentralized organization that facilitates rapid movements into and out of evolving markets) (Brown and Eisenhardt, 2000). These benefits that a decentralized organizational form provides through closer contact with markets suggest that decentralization is particularly appropriate for more applied, market- and customer-oriented R&D. Therefore, we expect that firms that conduct non-novel (imitative) R&D in particular would benefit from decentralization of innovation activity.

Although a decentralized organizational structure has many benefits for R&D, it does not in itself contain a mechanism for achieving economies of scope across decentralized R&D

locations, except insofar as top management or cross-divisional teams transfer information from one location to another. R&D, however, may be subject to economies of scale and scope, as has been documented in a number of settings (e.g., Henderson and Cockburn, 1994, 1996). Thus, decentralized R&D is more appropriate for research that is specific to individual product markets than for more fundamental "non-specific" research that potentially can be applied in multiple businesses (Argyres and Silverman, 2004; Kay, 1988). Evidence from patents, which by definition involve novel technologies, shows that international dispersion of patent inventors within firms does not result in firms having more highly cited patents, unless mechanisms are put in place to bring disparate sources of knowledge together (Singh, 2005). Thus, novel innovations, and fundamental R&D directed toward these innovations, may not benefit from geographic decentralization. This logic leads to the same conclusion as the arguments related to user-based innovation and patching. Essentially, decentralization is well suited for more applied (less fundamental) customer- and market-oriented R&D. The following hypothesis reflects this train of logic and modifies H1:

H4: Geographic decentralization of innovation activity is associated with greater innovation success for non-novel (imitative) innovations than for novel (new-to-the-market) innovations.

Because "non-specific" and presumably centralized R&D is directed toward more fundamental innovations with potentially broader applicability, Argyres and Silverman (2004) proposed that centralized R&D activity will support broader search processes for new innovations. Therefore, firms with more centralized R&D activity of the broader "non-specific" type will search more broadly outside of their organizational boundaries for information relevant to innovation (Argyres and Silverman, 2004). Then, firms with more decentralized R&D (including via geographic decentralization) will search less broadly outside of their organizational boundaries. This conflicts directly with hypothesis H2.

As argued earlier with regard to hypothesis H3, wider search for knowledge useful in innovation is likely to result in a wider range of applicability of any resulting set of innovations. Argyres and Silverman (2004) make this point as well, but argue in contrast to H3a that since more centralized (broader) R&D is directed toward—and results in—broader, more fundamental innovations, the resulting set of innovations will have wider applicability as well. By implication, decentralization of R&D will result in innovations that have a narrower range of applicability. We test the following hypothesis against H3a:

H3b: Geographic decentralization of innovation activity is associated with narrower applicability of innovation output.

THE SURVEY DATASET, VARIABLES, AND STATISTICAL INFERENCE

The data used to test the hypotheses come from the Finnish R&D Surveys conducted in 1999 and 2003, and the Finnish Community Innovation Survey (CIS) conducted in 2001. Statistics Finland, the national statistical agency, administered all of the surveys. The R&D surveys, carried out every other year, target all Finnish firms that engage in research and development (R&D). The innovation survey was coordinated with the statistical agency of the European Union, Eurostat, which sponsors CIS surveys in several member countries. Eurostat coordinated the development of the survey instrument and the data collection techniques. Data collection takes place every four years.

Within the manufacturing sector, all firms with more than 100 employees received the R&D survey, as did a stratified random sample of firms with 10-99 employees. The sample for the smallest firms with fewer than 10 employees is not random; it includes only firms that are known to perform R&D, from information such as earlier surveys or firms' public R&D funding applications. This survey collected detailed information about investments in R&D, both in-house and external, including the geographic locations of R&D activities. We use the information on R&D locations to construct our main explanatory variables of interest. Control variables gleaned from the R&D survey data include gross R&D expenditures and number of employees per firm. Finally, our innovation output measures for the panel dataset originate from the two R&D surveys.

For the Finnish CIS, Statistics Finland surveyed all Finnish manufacturing firms with more than 100 employees, as well as a random sample stratified by size and industry of the remainder of the population of Finnish manufacturing companies. The sample included separate observations for subsidiaries of larger companies; the latter are termed “business groups.” Certain Finnish industrial companies, for example the basic metals company Outokumpu Inc., are organized as business groups. Then, the divisions, in this case Outokumpu Stainless Steel and Outokumpu Technology, are wholly owned subsidiaries of the parent company. Survey respondents were either independent (not part of a business group), a subsidiary of a business group, or in a few instances, the parent company itself (the business group). As a result, the firms in the sample were not widely diversified. Because the data are confidential, the firms in the dataset are not identified by name. The CEO or the R&D manager of each firm filled out the survey, and the response rate was 50 percent.

The innovation survey included questions about innovation output, R&D activity, and knowledge sources related to innovation. The questions about innovation output ask whether the firm introduced technological innovations (product or process), what percent of firm sales derived from the introduction of technologically new products (either new to the firm, including imitative as well as novel innovations, or new to the market, referring to only novel innovations), and what type and extent of innovation impact firms achieved (e.g., reduction of material costs, expansion of current markets, fulfillment of regulatory requirements). As the CIS data have become available, scholars have begun to use them to measure innovation output as a complement to more traditional measures such as patents (e.g., Leiponen, 2000; Leiponen, 2005; Mairesse and Mohnen, 2002; Cassiman and Veugelers, 2002).[1]

The innovation survey also included questions about the utilization of nine different sources of knowledge. These included the firm's internal sources, sources within other units of the business group (if applicable), customers, suppliers, competitors, universities, non-profit research institutes, professional meetings and publications, and trade fairs and exhibitions.

We combined the 1999 R&D and 2001 CIS surveys to obtain a cross-sectional sample of firms that had data in both surveys. The R&D survey reflects data as of 1998 and the CIS survey covers the period 1998-2000. We also constructed a smaller panel data set by combining the cross-sectional sample with data from the 2003 R&D survey. The latter survey reflects data as of 2002. We first describe the cross-sectional sample and then describe the panel.

The cross-sectional sample includes 469 manufacturing firms with activity directed toward innovation, and encompasses all of the manufacturing industries in Finland. Table 1 compares the industry distribution of firms in our sample with the industry distribution of the Finnish manufacturing sector as a whole. Relative to the manufacturing sector as a whole, the forest sector (wood, pulp, paper, printing and publishing) is underrepresented in our sample, while the chemicals and electronics sectors are overrepresented. This may occur because our sample includes only firms with some activity directed toward technological innovation, and innovation activities are relatively less frequent in the forest-related sector and relatively more frequent in

[1] Although patents reflect success in creating something new, they do not necessarily result in commercially viable innovations (Griliches, 1990). Moreover, in most industries, firms do not rely heavily on patents (Levin et. al., 1987). The CIS data provide a direct measure of success in commercializing innovations for a broad range of industries. Kleinknecht, Montfort and Brouwer (2002) found that CIS innovation output (measured as the share of sales revenue per employee derived from innovative products) was not correlated with the number of patent applications per employee. This finding suggests that the CIS data provide useful complementary measures of innovation success that more traditional measures may not capture.

chemicals and electronics. Otherwise, the sample used here is very similar to the actual distribution of Finnish manufacturing industries.

Table 1 Industry Representation

NACE class	Industry	Percentage in sample	Percentage in the manufacturing sector as a whole
15-16	Food products, beverages, and tobacco	7.5	8.7
17-19	Textiles, textile products, leather, and leather products	3.4	5.1
20-22	Wood, wood products, pulp, paper, paper products, printing, publishing	10.0	18.7
23-25	Coke, refined petroleum products, nuclear fuel, chemicals, chemical products, manmade fiber, rubber, plastics	15.8	10.1
26	Nonmetallic mineral products	6.0	4.5
27-28	Basic metals, fabricated metal products	13.2	15.8
29	Machinery and equipment	17.5	16.0
30-33	Electrical and optical equipment	17.1	11.1
34-35	Transport equipment	5.5	4.2
36	Furniture, other manufacturing, recycling	4.1	5.9
	Total manufacturing	100	100.0

When we combine data from the 1999 R&D survey with the data from the CIS innovation survey, we experience some sample attrition. Compared to R&D active firms in the manufacturing sector as a whole, firms in our cross-sectional sample are larger (a mean of 383 employees in 2000 versus 182 employees in the manufacturing sector as a whole) and slightly more R&D intensive (R&D expenditures equal to 3.3% versus 2.9% of sales). But surprisingly, the firms in our sample are less innovative than R&D active firms in the manufacturing sector as a whole (67% are successful product innovators in our sample versus the sectoral average of 75%). Although the firms in our sample are substantially larger than those in the broader Finnish economy, the sample does not appear to be consistently or strongly biased towards firms with more innovation success.

We use the cross-sectional data as our primary sample. In order to control for potential unobserved heterogeneity, we also created a short panel that included firms for which we have two consecutive observations of R&D locations and innovation success by combining the cross-sectional sample with observations from the 2003 R&D survey (reflecting 2002 data). The restriction that firms appear in both the cross section and the later R&D survey reduced the sample size to 262 firms for a total of 524 observations. Although this dataset has the disadvantage that it includes many fewer firms than the cross-section, the panel enables us to conduct robustness checks on the cross-sectional results.

Variables

Dependent Variables

To measure the amount of innovation success per firm, we utilize binary variables as well as sales revenues. These variables are available in each of the three surveys. The binary (0,1) variables indicate whether the firm introduced any technological innovations during the preceding three-year period. The survey provided a detailed explanation to respondents of what constituted a technological innovation, as explained in the footnote below.[2]

We constructed three different binary variables relevant to this study. The first indicates whether the firm introduced technological innovations of any type, which could have been product or process innovations or both. These innovations were new to the firm, and may or may not have been new to the market (novel). The innovation survey also asked respondents whether they

[2] The survey defines a product innovation as including both a technologically new product and a technologically significant product improvement. A technologically new product is one whose purpose or technological characteristics are clearly distinct from those of the existing products of the firm. The new product can be based on a new technology, a new application of existing technologies, or application of new knowledge. A technologically significant product improvement significantly improves on the characteristics or performance of an existing product of the firm, and may include improvements in components, materials, or subsystems. The survey defines a process innovation as one that is technologically new or that contains a fundamentally improved method of production or product distribution. A process innovation may include (but is not limited to) improvements based on changes in equipment, instruments, organization of production, or new knowledge. We also require that the firm indicate that it developed the innovation primarily by itself, and that any associated research and development was performed internally rather than outsourced or performed in collaboration with other firms.

introduced any novel product innovations (new to the market). In order to test Hypothesis 4, the second binary variable indicates whether or not the firm had any novel product innovations. For purposes of comparison with the latter variable, we also utilized a third binary variable that indicated whether the firm had product innovations of any sort, regardless of their novelty. In the cross sectional sample, 67% of the firms introduced product innovations. Approximately three quarters of these firms (and 52% of all firms in the sample) introduced novel (new-to-the-world) product innovations, and the remaining firms introduced only imitative product innovations (16% of all firms). Fewer firms introduced process innovations than product innovations (46% of the sample).

The second type of dependent variable measures sales from innovative products. The surveys contain information about the percentage of total firm sales revenues in the year of the survey that derived from the introduction of technologically new products in the preceding three-year period (up to and including the final year covered by the survey). We multiplied the percent of sales from new products by total firm sales to obtain the value of sales from innovative products in that year. This variable reflects sales from all new products, novel or not. We also constructed a second variable for the value of sales from novel products only. Firms in the cross-sectional sample derived 15% of their sales from innovative products of any type, and approximately 10% of sales from novel product innovations. Both sales variables are expressed in logarithmic terms.

The product sales variables have the advantage that they provide measures of the extent of commercial success, in contrast to the binary innovation variables which provide only a minimum measure of innovation success (commercialization of at least one innovation). Among the firms that innovated, approximately 90 percent introduced product innovations, indicating that use of a product sales variable is appropriate. Since over half of the product-innovating firms also had process innovations, however, product sales do not fully reflect innovation success. By using both the binary innovation and the product sales variables, we obtain a fuller picture of innovation success. If the results for both types of dependent variables are similar, we can place greater confidence in the results.

In addition to the binary and sales variables that measure the amount of innovation success, we created two dependent variables that reflect the range of applicability of each firm's innovations, in order to test Hypotheses 3a and 3b. The first variable is binary, indicating whether the firm had both product and process innovations, as opposed to only one or the other, or neither. Three quarters of the entire sample introduced either product or process innovations. Of this general group of innovators, about half introduced *both* product and process innovations. The latter variable is a measure of the range of innovation impact. The 2001 innovation survey also asked the firms to use a four-point Likert scale to assess the impact of their innovations in nine different categories: expanding product range, improving product quality, extending market share or opening up new markets, improving production flexibility, expanding production capacity, reducing labor costs, reducing materials costs, reducing environmental effects, and fulfilling government regulations and standards. The survey asked firms to evaluate each of the foregoing effects on a four-point scale ranging from zero (not applicable or no impact) to three (a very substantial impact). To measure the range of impact of a firm's innovations, we constructed a variable that reflects the number of different effects that firms viewed as important or very important, using the same procedure as that described below for knowledge sources. In the cross-sectional dataset, the binary innovation, product innovation sales, and innovation impact variables derive from the innovation survey. In the panel dataset, the binary and product innovation sales variables derive from the two waves of the R&D survey. Innovation impact questions were not included in the R&D survey questionnaires.

Explanatory Variables

Our key explanatory variables relate to: 1) the number of locations where firms carried out R&D activities and, 2) the number of different external sources of knowledge that firms reported were important in their innovation activities. Here we focus on our main cross-sectional data set, and discuss the variables in the panel data set when we report those results. The location variables derive from the R&D survey and reflect 1998 data. In the cross-sectional sample, the number of locations per firm varies between zero and thirteen, but the average is just 1.1 locations. Most of the firms centralized their R&D activities: 7.5% of the firms had two R&D locations and 5.3% had three or more locations in 1998. Only 1% of firms had six or more locations. Given this distribution of the location data, we formed two binary indicators, one indicating whether a firm had two locations and another indicating whether a firm had three or more locations.

R&D locations are identified in the data as separate counties where a firm's R&D activity took place within Finland. Although Finland is a small country in terms of population, it is geographically dispersed and diverse. While the Helsinki metropolitan area is the most important commercial, industrial, and intellectual center, regional centers of Turku and Tampere also provide significant markets and high quality universities. Moreover, some technological concentrations can also be found, for example, electronics around Oulu University in the North, forest sector around Lappeenranta University of Technology in the Southeast, and medical research around University of Turku and Åbo Akademi in the city of Turku. A firm founded in the periphery thus may consider expanding R&D activities into a major market, to the vicinity of a major university, or to locate close to its peers in a technology "hot spot."

The innovation survey asked respondents to identify the importance of each of seven different external sources of information used in innovation activities: customers, suppliers, competitors, universities, non-profit research institutes, professional meetings and publications, and trade fairs and exhibitions. We use this information to test Hypothesis 2 concerning the relationship between geographic decentralization of R&D and sourcing of external knowledge. For each knowledge source listed in the survey questionnaire, each firm was asked to “evaluate the importance of the following sources of information for the innovation activities of your firm” on a four-point Likert scale from zero (not important at all/not used) to three (very important). To account for the varying importance of different knowledge sources, we adopted the approach introduced by Cohen and Malerba (2001) in their analysis of industry level innovation activity and subsequently adopted by Leiponen and Helfat (2005) for firm-level analysis. For each of the seven knowledge sources external to the firm, we created a binary indicator for whether the survey response indicated that the item was important to the firm. A survey response of either two (important) or three (very important) for a knowledge source received a binary value of one; survey responses of zero (not important at all/not used) or one (some importance) received a binary value of zero. This variable has a maximum value of seven. The most important external sources in the sample are customers and suppliers, followed by professional knowledge and trade meetings. Interestingly, universities and competitors are considered less important than professional and trade sources.

As noted above, we also used this same procedure to create a variable for the number of important effects of innovation outputs. For this variable, we assigned a binary value of one if an individual innovation effect received a value of two (substantial impact) or three (a very substantial impact), and a zero otherwise. We then summed the binary values for the nine possible innovation effects.

The use of binary measures of importance for the individual knowledge sources and innovation effects helps to alleviate potential measurement error that might arise from use of a Likert scale in the survey questions (Cohen and Malerba, 2001). When filling out a survey, it is easier to discriminate between important and not very important sources or effects than it is to make fine-grained distinctions between very important and important sources or effects, or between unimportant and not very important sources or effects. For this reason, we use these measures rather than the original Likert scale survey answers.

Control Variables

In the cross-sectional analyses and in the random effects panel analyses it is very important to control for other firm-specific characteristics. We include the following control variables in all models: firm size measured as the natural logarithm of number of employees, natural logarithm of R&D expenditures, natural logarithm of firm revenues derived from exports, a binary variable indicating whether or not the surveyed firm is part of a business group of companies, a binary variable indicating whether the surveyed firm is responding as the parent business group company (the units of these firms did not respond as business groups), a binary variable indicating whether sales grew by 10% or more due to mergers and acquisitions, a binary variable indicating whether sales shrank by 10% or more due to divestments, and industry dummy variables. The control variables are described in more detail below.

Logarithm of Number of Employees. Firm size is likely to be a particularly important predictor for the binary (0,1) innovation variables. Because larger firms have access to greater financial and human resources, these firms may have a greater ability to achieve at least a single innovation. Larger firms also may derive greater sales from a single innovation, since the firms have a larger base of customers on which to build. In addition, since larger firms may be able to support more R&D locations, it is important to control for firm size. Finally, including firm size in the analysis controls for potential non-response bias associated with firm size, since the firms that responded to the innovation survey were larger than those in the target population.

Logarithm of R&D Expenditures. Because R&D spending is explicitly directed toward the development of new products and processes, greater R&D expenditures may increase the probability of successful innovation. In addition, since the number of R&D locations may be correlated with the amount of R&D spending, it is important to control for scale of R&D. Although R&D spending does not include all funds devoted to innovation activity, which can take other forms (e.g., process engineering), it should capture most direct spending on innovation. R&D expenditures are measured in natural logarithmic form, since they tend to increase with firm size.

Business group subsidiary and *business group parent.* These binary (0,1) variables come from the innovation survey and measure whether a firm was either a subsidiary of a larger firm or a business group parent company at the time of the survey. These variables are mutually exclusive. The excluded dummy variable is that indicating independent firms. The business group variables control for whether the firm was part of a diversified company, which can affect R&D locations or innovation success. For example, firms in business groups might have access to internal knowledge sources that influence their decisions to decentralize R&D or introduce innovations. In addition, the business group parent variable controls for the possibility that these firms might report a larger number of R&D locations due to the inclusion of their subsidiaries in their survey responses.

Logarithm of Export Revenues. The potential for greater sales outside of Finland may increase the incentive to innovate. Additionally, firms with a high volume of exports might have stronger motivation to introduce new innovations or expand their R&D activities than domestically oriented firms, because of perhaps more intense international competition. Therefore, it is important to control for this aspect of firm structure.

Sales Growth Due to Mergers and Acquisitions and *Sales Reductions due to Divestitures.* These binary (0,1) variables come from the innovation survey and measure whether or not the firm had mergers or acquisitions or whether or not the firm had divestitures that increased or decreased sales revenues by 10% or more during the three-year period covered by the survey. These variables control for the possibility that the number of R&D locations could reflect recent large acquisitions or divestitures. The excluded dummy variable is that for no (or less than 10%) change in sales from mergers, acquisitions, and divestitures.

Industry of Operation. Industry level factors such as technological opportunity, appropriability of the returns to innovation, and customer demand for new products also may affect the innovation success of individual firms. We include a (0,1) dummy variable for each of the two-digit level NACE industries in the sample (excluding the non-metallic minerals industry dummy).[3]

Table 2 presents descriptive statistics for the variables used in the analyses, as well as for additional variables describing the characteristics of the firms in the cross sectional sample. In 1998, the average firm had about 350 employees and sales of 64 million euros. 4.3% of sales were invested in R&D activities. Moreover, Finnish manufacturing firms were highly internationalized. Average sales revenues from exports were 35 million euros, or 42% of average total revenues per firm. About half of the sampled firms were part of a business group and 5% were business group parents. Finally, 10 percent of the firms experienced sales growth from mergers and acquisitions, and 4 percent experienced sales declines from divestitures.

[3] The data make it difficult to utilize a more fine-grained classification of industry membership, which would result in a significant number of industries with just one or a few firms. NACE (Nomenclature Actuariel dans la Communauté Européenne) is the European standardized industry classification system that is similar to the North American Industrial Classification System.

Table 2 Descriptive statistics for the cross-sectional sample (N = 469)

Variable	Mean	Std. deviation	Min.	Max.
Employees	354.122	1553.888	5	22000
Log(employees)	4.443	1.394	1.609	9.461
Exports	35797.550	190147.300	0	3410100
Log(exports)	7.949	2.309	0	15.042
Sales growth due to mergers or acquisitions (M&A)	0.102	0.303	0	1
Sales shrinking due to divestitures	0.038	0.192	0	1
Business group subsidiary	0.510	0.500	0	1
Business group	0.051	0.221	0	1
R&D investments	5800.192	20733.340	0	311149
R&D intensity	0.043	0.095	0	0.933
Log(R&D)	6.361	2.831	0	12.648
Product innovations	0.674	0.469	0	1
Process innovations	0.458	0.499	0	1
Novel innovations	0.546	0.498	0	1
Any innovations	0.742	0.438	0	1
Product and process innovations	0.390	0.488	0	1
Sales share from newly introduced products	0.151	0.212	0	1
Log(product innovation sales)	5.192	3.996	0	13.038
Sales share from newly introduced novel products	0.103	0.185	0	1
Log(novel innovation sales)	3.950	4.043	0	13.038
R&D locations	1.149	1.107	0	13
Two locations	0.075	0.263	0	1
Three or more locations	0.053	0.225	0	1
Sum of important innovation effects	3.188	2.730	0	9
Sum of important external knowledge sources	2.525	2.015	0	7

Statistical Inference

Surveys can pose issues related to non-response bias and common method variance. Since our main sample consists of a cross-section, it also is important to assess the extent of potential endogeneity of the explanatory and control variables. We next discuss these issues before reporting our empirical results.

The innovation survey forms the primary basis for our samples. Because the firms that responded to both the innovation survey and the R&D surveys were twice the size of the set of firms that received the survey, it is important to control for firm size in the regressions. We also noted earlier that despite the difference in average firm size of respondents and non-respondents, respondents had less innovation success than the targeted population as a whole. This suggests that any non-response bias does not inflate the survey measures of innovation success.

Both the innovation and the R&D survey had a single respondent per firm, suggesting the need to check for common method variance, which could inflate any observed correlations between the dependent and independent variables. Since the variables in our regressions came from three different surveys, and the same person may not have filled out all forms, this reduces the potential for common method variance. Nevertheless, as insurance we conducted a standard check for common method variance. We used Harmon's one-factor test to assess common method bias. If common method variance is a serious problem, a factor analysis would produce a single factor that accounts for most of the correlation between the dependent and independent variables (see Podsakoff and Organ, 1986).

We performed a principal components analysis for the cross sectional sample that included all explanatory and control variables and the dependent product innovation sales variable. We used the sales dependent variable rather than the binary dependent variable, because principal component analyses tend not to work as well for binary variables. The analysis retained nine factors with eigenvalues greater than 1.00, and no factor explained more than 20 percent of the variance. Moreover, the dependent variable did not load most strongly on the same component as did the explanatory variables of location. Excluding the industry dummy variables, the principal component analysis still retained five components with eigenvalues above 1.00, and the first component explained 25% of the variance. R&D location variables continued to load most strongly on different components than the dependent variable. The factor analyses thus indicate that our regression results are not subject to an inherent common method bias in the responses to the survey.

In the cross-sectional regressions, most of our explanatory and control variables reflect data for 1998 and the dependent variables reflect data for 2000 (sales variables) and 1998-2000 (binary variables), thus alleviating simultaneity issues. In particular, our main explanatory variables for number of locations date from 1998, as do R&D expenditures, number of employees, and export sales. The knowledge source variable reflects the years 1998-2000, which suggests greater potential for simultaneity. A common solution to this form of potential endogeneity utilizes instrumental variables. Unfortunately, we do not have data with which to construct appropriate instruments. The use of product sales as a dependent variable mitigates this concern to some extent, however, since the variable includes only sales in 2000.

ESTIMATION METHODOLOGY, RESULTS, AND DISCUSSION

We first report the cross-sectional regression results and then the robustness tests using the panel data that enable us to mitigate the issue of unobserved firm heterogeneity. Table 3 displays the results of the probit estimation for the binary dependent variables. We find that having two R&D locations is positively and statistically significantly associated with the probability that firms introduced innovations of any type or introduced at least some product innovations; three or more R&D locations does not differ statistically significantly from zero. As expected, R&D expenditures are positively and statistically significantly associated with the probability of innovation success. The only other significant variables are the industry dummy variables (particularly in the product innovation regression) and the divestment variable (although this reflects a relatively small number of firms).

In table 4, continuous dependent variables measuring sales from innovative products focus our attention on the economic impact of innovation on the firm. Largely the same variables are significant in the main product sales regression as for the binary innovation indicators, particularly the R&D location variables. The logarithm of the number of employees is positive and significant as well, and the M&A variable rather than the divestment variable is significant. These results lend support for H1 that geographic decentralization of R&D, up to a point, is associated with greater innovation success. At least for the relatively small sized firms in this sample, having more than two R&D locations does not appear to have an additional benefit.

In contrast to the above results, tables 3 and 4 also show that introduction of novel (new to the market) product innovations and sales from novel product innovations do not appear to benefit from geographic decentralization of R&D. The R&D location variables are not significant in either of the novel product innovation regressions, but they are significant in the regressions that include imitative innovations as well. This suggests that imitative imitations in particular benefit from geographic decentralization of R&D. These results support H4, in that geographic decentralization of innovation activity is associated with greater innovation success for non-novel innovations than for novel innovations.

Table 3 Estimation results for binary innovation indicators

Dependent variable:	Product innovation			Any innovation			Novel innovation		
Explanatory variables	Coeff.	S.E.	ME	Coeff.	S.E.	ME	Coeff.	S.E.	ME
Constant	-1.055***	0.356		-0.567	0.375		-1.247***	0.344	
Log(employees)	0.054	0.076	0.019	0.095	0.079	0.029	-0.008	0.073	-0.003
Log(R&D)	0.108***	0.027	0.038***	0.115***	0.027	0.035***	0.117***	0.027	0.046***
Log(exports)	0.023	0.041	0.008	-0.014	0.043	-0.004	0.031	0.041	0.012
Business group	-0.309	0.332	-0.114	-0.093	0.356	-0.029	0.076	0.316	0.030
Business group subsidiary	0.103	0.149	0.036	0.050	0.154	0.015	0.109	0.143	0.043
M&A	0.217	0.231	0.072	0.286	0.251	0.079	-0.003	0.216	-0.001
Divest	-0.578*	0.327	-0.220*	-0.581*	0.339	-0.205*	-0.243	0.321	-0.097
Two locations	0.726**	0.321	0.204**	0.917**	0.394	0.195**	0.277	0.249	0.107
Three or more locations	-0.007	0.328	-0.003	-0.163	0.348	-0.052	0.137	0.307	0.054
Food	0.483	0.347	0.147	0.271	0.370	0.075	0.385	0.330	0.146
Textiles	0.214	0.405	0.070	-0.027	0.422	-0.008	-0.046	0.411	-0.018
Forest	-0.278	0.315	-0.102	-0.136	0.331	-0.043	-0.222	0.315	-0.088
Chemical	0.235	0.290	0.078	0.104	0.311	0.031	0.288	0.285	0.111
Metals	0.506	0.302	0.156*	0.248	0.320	0.070	0.544*	0.298	0.203*
Machine	0.732***	0.297	0.217**	0.389	0.315	0.106	0.719**	0.287	0.264**
Electronics	0.567*	0.295	0.175*	0.252	0.313	0.072	0.367	0.285	0.141
Vehicles	-0.188	0.362	-0.068	-0.260	0.379	-0.085	-0.040	0.358	-0.016
Other, furniture	0.921**	0.420	0.234**	0.733	0.456	0.164	0.723*	0.391	0.255*
Log likelihood	-258.821			-239.559			-290.445		

S.E. = Standard error. ME = Marginal effect. Probit ML estimation method.
*** implies significance at the 99% level, ** at the 95% level, and * at the 90% level.

Table 4 Estimation results for sales from innovative products

Dependent variable:	Log(product innovation sales)			Log(novel innovation sales)		
	Coeff.		S.E.	Coeff.		S.E.
Constant	-4.748	***	1.365	-7.306	***	1.868
Log(employees)	0.639	**	0.287	0.338		0.388
Log(R&D)	0.417	***	0.107	0.633	***	0.149
Log(exports)	0.197		0.163	0.153		0.222
Business group	-0.510		1.183	-0.397		1.580
Business group subsidiary	0.416		0.566	0.614		0.771
M&A	1.488	*	0.798	0.732		1.087
Divest	-2.026		1.320	-1.015		1.766
Two locations	1.797	**	0.913	1.275		1.219
Three or more locations	0.674		1.157	1.020		1.535
Food	2.021		1.328	2.044		1.807
Textiles	0.486		1.636	-1.610		2.353
Forest	-1.192		1.280	-0.916		1.757
Chemical	1.061		1.154	1.429		1.575
Metals	1.722		1.197	2.410		1.634
Machine	2.817	**	1.141	3.702	**	1.551
Electronics	2.135	*	1.146	1.698		1.566
Vehicles	-0.175		1.453	-0.060		1.988
Other, furniture	3.042	**	1.524	3.854	*	2.064
Sigma	4.888		0.216	6.319		0.328
Log likelihood	-1081.801			-958.053		

S.E. = Standard error. ME = Marginal effect. Tobit ML estimation method.
*** implies significance at the 99% level, ** at the 95% level, and * at the 90% level.

Next we examine how patterns of knowledge sourcing relate to R&D location choice. Table 5 presents a subset of earlier specifications with the sum of important knowledge sources as an additional explanatory variable. We focus on the dependent variables that included non-novel innovations, since geographic decentralization of R&D was not significant in the novel innovation models. (Table 5 includes the novel product sales regression for completeness.)

The models in table 5 demonstrate the strong correlation between the extent of knowledge sourcing and geographic decentralization of R&D: adding the knowledge source variable to the regressions wipes out the previously significant positive effects of multiple R&D locations on innovation outcomes. We note that the knowledge source variable dates from the same survey as the innovation output variables, reflecting knowledge sources that firms utilized for 1998-2000. This implies that multiple locations in 1998 might have enabled firms to access a greater number of knowledge sources during 1998-2000, which is completely consistent with the literature that motivated hypothesis H2. Overall, these results support H2 by suggesting that R&D location decisions and knowledge sourcing decisions are strongly related. The positive association between geographic decentralization of R&D and innovation success appears to reflect a positive correlation between geographic decentralization and the number of knowledge sources that firms access outside of their boundaries.

We also assessed whether geographic decentralization of R&D is associated with wider versus narrower applicability of innovation success. Table 6 reports the results for two dependent variables that reflect innovation impact: a binary variable for whether the firm had both product and process innovations (as opposed to one or the other or neither) and the sum of important innovation effects. We first report the base regressions and then add the knowledge source variable. In the base regressions for both the binary indicator and the innovation effects variable, having two R&D locations (but not three or more) is positively and statistically significantly associated with wider applicability of innovation success. This result rejects H3b in favor of H3a: geographic decentralization of innovation activity is associated with wider, not narrower, applicability of innovation output. Consistent with H2, once we add the knowledge source variable, the variable for two R&D locations becomes statistically insignificant.

Thus far, the cross-sectional results suggest that, first, the extent of innovation success is positively associated with having two R&D locations rather than one, second, this relationship holds for non-novel but not for novel innovations, and, third, greater access to important knowledge sources accounts for the positive effect associated with geographic decentralization of R&D. Wider applicability of innovations also is positively associated with having two R&D locations rather than one. These regressions included a number of important controls for other firm level factors that could have affected innovation success, such as R&D expenditures, firm size, exports, mergers, acquisitions, and divestitures, and whether or not the firm was diversified. We next use our panel sample to mitigate the effect of unobserved firm heterogeneity that these controls may not pick up.

Table 5 Selected models with important knowledge sources as an additional explanatory variable

	Product innovation			Any innovation			Log(product innovation sales)		Log(novel innovation sales)	
	Coef.	S.E.	ME	Coeff.	S.E.	ME	Coeff.	S.E.	Coeff.	S.E.
Constant	-1.094 ***	0.378		-0.469	0.430		-4.460 ***	1.249	-7.087 ***	1.772
Log(employees)	0.003	0.080	0.001	0.023	0.089	0.006	0.407	0.264	0.107	0.369
Log(R&D)	0.059 **	0.029	0.020 **	0.066 **	0.031	0.016 **	0.206 **	0.100	0.404 ***	0.143
Log(exports)	0.010	0.044	0.003	-0.042	0.048	-0.010	0.155	0.150	0.122	0.211
Business group	-0.234	0.356	-0.083	0.078	0.422	0.018	-0.341	1.082	-0.308	1.497
Business group subsidiary	0.134	0.161	0.045	0.062	0.179	0.015	0.459	0.518	0.585	0.730
M&A	0.101	0.246	0.033	0.246	0.291	0.054	1.071	0.730	0.333	1.029
Divest	-0.437	0.336	-0.160	-0.407	0.369	-0.117	-0.985	1.207	0.152	1.668
Sum of important knowledge sources	0.326 ***	0.040	0.109 ***	0.475 ***	0.051	0.116 ***	1.126 ***	0.123	1.198 ***	0.175
Two locations	0.410	0.331	0.121	0.637	0.449	0.116	0.801	0.835	0.254	1.153
Three or more locations	-0.241	0.344	-0.085	-0.706 *	0.385	-0.222 *	0.235	1.053	0.517	1.450
Food	0.354	0.366	0.107	0.067	0.419	0.016	1.287	1.215	1.188	1.715
Textiles	-0.039	0.410	-0.013	-0.654	0.443	-0.204	-0.022	1.488	-1.928	2.199
Forest	-0.208	0.336	-0.073	-0.066	0.375	-0.017	-0.946	1.175	-0.659	1.668
Chemical	0.129	0.307	0.042	-0.126	0.354	-0.032	0.547	1.058	0.898	1.493
Metals	0.510	0.322	0.149	0.164	0.364	0.038	1.358	1.097	2.007	1.549
Machine	0.736 **	0.313	0.206 **	0.232	0.354	0.052	2.570 **	1.043	3.423 **	1.467
Electronics	0.490	0.314	0.146	-0.039	0.359	-0.010	1.509	1.050	1.038	1.485
Vehicles	-0.228	0.389	-0.080	-0.399	0.438	-0.114	-0.335	1.334	-0.319	1.893
Other, furniture	0.910 **	0.452	0.217 **	0.681	0.525	0.118	2.397	1.394	3.202	1.952
Sigma							4.445	0.196	5.939	0.306
Log likelihood	-222.100			-183.314			-1040.932		-934.115	

Estimation methods: Probit ML for product innovation and any innovation, tobit ML for log(innovation sales). S.E. = Standard error. ME = Marginal effect. *** implies significance at the 99% level, ** at the 95% level, and * at the 90% level.

Table 6 **Effects of location on the breadth of innovation output**

	Product and process innovation			Sum of important innovation effects		Product and process innovation			Sum of important innovation effects	
	Coeff.	S.E.	ME	Coeff.	S.E.	Coeff.	S.E.	ME	Coeff.	S.E.
Constant	-1.178 ***	0.342		0.232	0.246	-1.202 ***	0.352		0.249	0.218
Log(employees)	0.189 **	0.074	0.072 **	0.022	0.052	0.162 **	0.077	0.061 **	-0.008	0.047
Log(R&D)	0.045 *	0.027	0.017 *	0.106 ***	0.024	0.011	0.029	0.004	0.035 *	0.020
Log(exports)	0.001	0.042	0.001	0.018	0.031	-0.009	0.043	-0.003	0.005	0.029
Business group	-0.142	0.313	-0.053	-0.153	0.197	-0.120	0.323	-0.045	-0.199	0.184
Business group subsidiary	-0.155	0.146	-0.059	-0.154	0.105	-0.157	0.149	-0.059	-0.143	0.091
M&A	0.382 *	0.206	0.150 *	0.345 ***	0.128	0.333	0.213	0.130	0.203 *	0.119
Divest	0.056	0.323	0.021	-0.366	0.287	0.212	0.323	0.083	-0.103	0.252
Two locations	0.431 *	0.238	0.169 *	0.291 **	0.142	0.254	0.241	0.099	0.044	0.128
Three or more locations	0.271	0.304	0.106	-0.154	0.195	0.178	0.309	0.069	-0.158	0.176
Sum of important knowledge sources						0.205 ***	0.035	0.078 ***	0.264 ***	0.021
Food	0.056	0.332	0.022	0.128	0.237	-0.055	0.343	-0.021	-0.161	0.212
Textiles	-0.593	0.419	-0.199	-0.177	0.317	-0.708	0.420	-0.226 *	-0.491	0.299
Forest	-0.686 **	0.324	-0.229 **	-0.155	0.236	-0.717 **	0.338	-0.235 **	-0.072	0.206
Chemical	-0.258	0.287	-0.095	-0.012	0.209	-0.369	0.296	-0.133	-0.078	0.183
Metals	-0.022	0.297	-0.008	0.093	0.218	-0.092	0.306	-0.035	0.036	0.191
Machine	-0.282	0.285	-0.104	-0.019	0.208	-0.344	0.294	-0.124	-0.059	0.183
Electronics	-0.335	0.287	-0.122	0.062	0.206	-0.469	0.297	-0.166	-0.040	0.182
Vehicles	-0.481	0.362	-0.167	-0.259	0.277	-0.548	0.374	-0.185	-0.333	0.246
Other, furniture	0.058	0.384	0.022	0.209	0.272	-0.043	0.396	-0.016	0.060	0.242
Delta				2.126	0.270				1.376	0.196
Log likelihood	-287.781			-1043.848		-270.281			-969.088	

Estimation methods: Probit ML for product and process innovation, negative binomial with constant dispersion for sum of innovation effects. S.E. = Standard error. ME = Marginal effect. *** implies significance at the 99% level, ** at the 95% level, and * at the 90% level

Our main explanatory variables of interest regarding the number of R&D locations come from the R&D survey. Therefore, in order to obtain a panel data set, we used data from two years of the R&D survey available to us; the surveys were conducted in 1999 and 2003. This yielded a two-period panel for a subset of firms in the cross section that also had data in the R&D survey. We supplemented these data with selected variables from the cross section, as explained below.

The dependent variables in the panel analysis include binary indicators of innovation success and sales from product innovations. For the binary indicators, we use random effects probit maximum likelihood estimation. For the product sales variables, we use random effects tobit maximum likelihood estimation. We use random effects estimation because conditional fixed effects estimation is not available for either probit or tobit, and unconditional fixed effects are inconsistent (Greene, 1993). In order to control for unobserved fixed effects that may be correlated with the explanatory variables (which the use of random effects may not accomplish), we include time-invariant explanatory variables in the regressions. Inclusion of time-invariant variables reduces the error variance in the regression and produces consistent coefficient estimates (Woolridge, 2002).

Our dependent and right-hand side variables are the same as those reported for the cross-section with the exceptions: of the innovation effects dependent variable and the knowledge sources explanatory variable. Since the R&D survey does not contain questions regarding innovation effects and knowledge sources, we cannot include these variables in the panel data set.

The product sales dependent variables and the right-hand side variables for R&D locations, R&D expenditures, and number of employees reflect data for 1998 for the first period and 2002 in the second period. The binary innovation indicators reflect data for 1996-1998 in the first period and 2000-2002. These data come from the two R&D surveys. Export revenues are those for 1998 in the first period and 2000 in the second period, based on information from the innovation survey**,** which does not contain data for 2002. The R&D surveys do not have information about exports. The time invariant variables used in the panel are binary variables for business group membership, business group parent company, mergers and acquisitions, and divestments. Any firms that were part of a business group or were a business group parent anytime during the time period covered by the panel (1998-2002) received a value of "1 for these (mutually exclusive) variables. The binary variables for mergers/acquisitions and divestments come from the innovation survey, indicating firms that had major business changes during the period 1998-2000. We also include industry dummy variables in the analysis.

Using these data, we estimate the relationship between geographic decentralization of R&D and innovation success while controlling for unobserved firm level heterogeneity. We use these data as a robustness check on the cross-sectional results, rather than as a definitive analysis for two reasons. First, the panel data do not allow us to control for potential simultaneity of R&D location and innovation success, given the dates of the variables. Second, the short panel may not be able to completely identify the unobserved firm effects. In addition, the panel contains a smaller sample of firms, and therefore is potentially less representative, than the cross section.

Table 7 reports the panel results for the binary dependent variables and table 8 reports the results for the product sales variables. Here we see that the variable for three or more R&D locations is positive and statistically significant but the variable for two R&D locations is not, in all but the novel innovation regressions. These results suggest that firms that increased the number of R&D locations from two to three had a positive change in the extent and breadth of innovation success, consistent with H1 and H3a. The support for these hypotheses is consistent with the results from the cross sectional analysis. In the cross section, the variable for two R&D locations was positive and significant. Evidently, firms that increased the number of R&D locations to three or more in the panel also met with greater innovation success. None of the R&D location variables are statistically significant in either the binary novel innovation regression or the novel product sales regression. These results also are consistent with the cross sectional analysis, and provide additional support for H4 that geographic decentralization has a greater benefit for non-novel than for novel innovations.

The cross sectional and panel results together provide evidence that geographic decentralization of R&D is positively associated with both the extent and breadth of innovation success, as hypothesized. We also find that these results do not hold for novel innovations. Hence, our results for largely domestic R&D locations are consistent with those of Singh (2005), who found that international dispersion of patenting within firms did not have a positive correlation with the extent of forward patent citations within firms. Our results regarding R&D location and novel innovation are also consistent with those of Argyres and Silverman (2004), who found that centralization rather than decentralization of R&D was positively associated with the extent and breadth of firm patenting activity. This has important implications for the many studies that use patent data, which reflect only novel technologies. Our results suggest that when analyzing technological innovation, it is important to distinguish between novel and non-novel innovations, since the results may differ for these two types of innovation.

Table 7 **Panel results for binary innovation indicators**

	Product innovation			Any innovation			Novel innovation			Product and process innovation		
	Coef.		S.E.	Coef.		S.E.	Coef.		S.E.	Coef.		S.E.
Constant	0.434		0.395	0.773	**	0.369	-0.555	*	0.332	-1.114	***	0.377
Log(employees)	-0.102		0.092	-0.007		0.092	-0.016		0.073	-0.010		0.081
Log(R&D)	0.026		0.049	0.041		0.049	-0.010		0.042	0.058		0.047
Log(exports)	0.047		0.043	-0.014		0.047	0.077	**	0.037	0.105	**	0.041
Business group	0.751		0.516	0.490		0.497	0.313		0.297	0.224		0.327
Business group subsidiary	0.015		0.176	-0.140		0.176	-0.036		0.144	-0.012		0.161
M&A	0.460		0.288	0.268		0.280	0.440	**	0.207	0.660	***	0.227
Divest	-0.687		0.523	-0.561		0.519	-0.369		0.447	-1.250	**	0.579
Two locations	0.192		0.250	-0.123		0.236	-0.130		0.194	0.189		0.208
Three or more locations	0.856	**	0.420	0.910	*	0.511	0.220		0.270	0.658	**	0.321
Food	0.594		0.419	0.363		0.396	0.201		0.330	0.357		0.375
Textiles	-0.460		0.452	-0.465		0.410	0.047		0.416	-0.012		0.453
Forest	0.022		0.349	-0.241		0.306	0.084		0.304	-0.467		0.340
Chemical	0.118		0.310	-0.006		0.271	0.438		0.273	-0.289		0.299
Metals	0.426		0.381	0.455		0.376	0.104		0.320	0.075		0.353
Machine	0.909	**	0.333	0.547	*	0.287	0.422		0.266	-0.658	**	0.297
Electronics	0.640	*	0.338	0.561	*	0.313	0.457		0.281	-0.278		0.308
Vehicles	0.023		0.384	-0.079		0.353	0.287		0.341	-0.179		0.375
Other, furniture	0.321		0.473	n.a.			-0.031		0.401	-0.108		0.440
Sigma_u	0.422		0.210	0.179		0.489	0.287		0.211	0.458		0.163
Rho	0.151		0.128	0.031		0.164	0.076		0.103	0.173		0.102
Log likelihood	-234.407			-189.012			-344.185			-321.220		
Observations	526			526			522			526		

Estimation method: Random effects probit ML. S.E. = Standard error. ME = Marginal effect. Probit ML estimation method.
*** implies significance at the 99% level, ** at the 95% level, and * at the 90% level.

Table 8 Panel results: continuous sales measures of innovation

	Log(product innovation sales)		Log(novel innovation sales)	
	Coef.	S.E.	Coef.	S.E.
Constant	2.093	1.429	-2.255	2.280
Log(employees)	0.166	0.308	-0.216	0.461
Log(R&D)	0.286	0.177	0.203	0.269
Log(exports)	0.324 **	0.150	1.014 ***	0.257
Business group	1.190	1.226	2.900	1.901
Business group subsidiary	0.135	0.623	0.532	0.990
M&A	1.641 *	0.850	1.072	1.251
Divest	-2.493	1.971	-11.078 **	4.446
Two locations	0.372	0.807	1.152	1.430
Three or more locations	2.423 **	1.117	1.378	1.721
Food	2.833 **	1.430	3.730	2.298
Textiles	-1.287	1.802	2.091	2.916
Forest	-0.146	1.323	-1.547	2.073
Chemical	1.276	1.171	0.937	1.832
Metals	1.909	1.383	1.219	2.216
Machine	3.335 **	1.147	2.466	1.822
Electronics	2.844 **	1.212	3.326 *	1.958
Vehicles	0.504	1.454	0.230	2.260
Other, furniture	1.543	1.705	0.762	2.638
Sigma_u	2.519	0.174	2.897	0.246
Sigma_e	4.976	0.199	6.340	0.348
Rho	0.204	0.027	0.173	0.031
Log likelihood	-1491.088		-952.955	
Observations	524		352	

Estimation method: Random effects tobit ML. S.E. = Standard error. ME = Marginal effect. Probit ML estimation method. *** implies significance at the 99% level, ** at the 95% level, and * at the 90% level.

CONCLUSION

This empirical study introduces a new dataset containing uniquely detailed data on firms' R&D locations to examine the effects of geographic decentralization of R&D on innovation outcomes. We compare and contrast hypotheses that derive from two relatively separate strands of literature, one on multinational firm location decisions and another on organizational form.

Our results indicate that geographic decentralization of R&D at the firm level is generally beneficial in terms of the extent and breadth of innovation success—product or process—but the effect is strongly related to the knowledge sourcing strategies that firms employ. These results are consistent with the interpretation that R&D location decisions are driven by the desire of firms to access external sources of knowledge for innovation activities. As a result, and consistent with depictions in the popular press, firms may locate R&D activities in major markets in order to obtain knowledge from customers, and in technology hotspots in order to tap into highly tacit research knowledge from universities, competitors, and suppliers.

Our empirical analyses also suggest that firms with multiple R&D locations tend to introduce less novel innovations. In other words, locating near relevant knowledge sources such as customers and suppliers is associated with more productive but less technologically advanced innovation output. Other data sources are necessary to assess whether geographic decentralization of R&D is welfare improving, but our results suggest that it may be profit maximizing (or at least profit-improving) for many firms, since geographic decentralization is associated with greater sales from innovative products—which in many markets is crucial for sustaining a competitive edge in the long run.

REFERENCES

Almeida P. 1996. Knowledge Sourcing by Foreign Multinationals: Patent Citation Analysis in the U.S. Semiconductor Industry. *Strategic Management Journal* 17: 155-165.

Argyres NS, Silverman BS. 2004. R&D, Organization Structure, and the Development of Corporate Technological Knowledge. *Strategic Management Journal* **25**: 929-958.

Brown S, Eisenhardt KM. 2000. *Competing on the Edge of Chaos.* Boston: Harvard Business School Press.

Buckley PJ, Casson MC. 1976. *The Economic Theory of the Multinational Enterprise.* Macmillan: London.

Cassiman B, Veugelers R. 2002. R&D cooperation and spillovers: Some empirical evidence from Belgium. *American Economic Review* 92(4): 1169-1184

Cohen WM, Malerba F. 2001. Is the Tendency to Variation the Chief Cause of Progress? *Industrial and Corporate Change* 10:587-608.

Chandler AD. 1962. *Strategy and Structure*. MIT Press: Cambridge.

Dunning JH. 1981. *International Production and the Multinational Enterprise*. Allen and Unwin: London.

Florida R. 1997. The Globalization of R&D: Results of a Survey of Foreign-Affiliated R&D Laboratories in the USA. *Research Policy* 26: 85-103.

Frost TS. 2001. The Geographical Sources of Foreign Subsidiaries' Innovations. *Strategic Management Journal* 22: 101-123.

Grant RM. 1996. Toward a Knowledge-Based Theory of the Firm. *Strategic Management Journal* **17**(Winter Special Issue): 109-122.

Greene, William H. (1993), Econometric Analysis (2nd ed. ed.). New York: Macmillan.

Griliches Z. 1990. Patent Statistics as Economic Indicators: A Survey. *Journal of Economic Literature* 28(Dec.): 1661-1707.

Henderson R, Cockburn I. 1994. Measuring Competence? Exploring Firm Effects in Pharmaceutical Research. *Strategic Management Journal* **15**(Special Issue): 63-84.

Henderson R, Cockburn I. 1996. Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery. *RAND Journal of Economics* **27**(1): 32-59.

Hymer S. 1970. The Efficiency (Contradictions) of Multinational Corporations. *American Economic Review* 60: 441-448.

Kay N. 1988. The R&D Function: Corporate Strategy and Structure. In *Technical Change and Economic Theory*. Dosi, G., Freeman, C., Nelson, R., Silverberg, G., Soete, C. (eds.) Pinter: London: 283-294.

Kleinknecht A, Van Montfort K, Brouwer E. 2002. The Non-Trivial Choice Between Innovation Indicators. *Economics of Innovation and New Technology* 11(2): 109-121.

Kogut B. 1991. Country Capabilities and the Permeability of Borders. *Strategic Management Journal*, Summer special issue 12: 33-47.

Kogut B, Zander U. 1992. Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. *Organization Science* 3: 383-397.

Kogut B, Zander U. 1993. Knowledge of the Firm and the Evolutionary Theory of the Multinational Corporation. *Journal of International Business Studies* **24**: 625-646.

Kuemmerle W. 1999. Foreign Direct Investment in Industrial Research in the Pharmaceutical and Electronics Industries--Results From a Survey of Multinational Firms, *Research Policy* 28: 179-193.

Leiponen, A. 2000. Competencies, Innovation, and Profitability of Firms. *Economics of Innovation and New Technology* 9(1):1-24.

Leiponen, A. 2005. Skills and Innovation. *International Journal of Industrial Organization* 23(5-6): 303-323.

Leiponen A, Helfat CE. 2005. *Innovation Objectives, Knowledge Sources, and the Benefits of Breadth.* Ithaca NY and Hanover, NH. Unpublished manuscript.

Levin R C, Klevorick AK, Nelson RR, Winter SG. 1987. Appropriating the Returns from Industrial Research and Development. *Brookings Papers on Economic Activity* 3: 783-820.

Penner-Hahn J, Shaver JM. 2005. Does international research and development affect patent output? An analysis of Japanese pharmaceutical firms. *Strategic Management Journal* **26**: 121–140.

Podsakoff PM, Organ DW. 1986. Self-reports in Organizational Research: Problems and Prospects. *Journal of Management* 12:531-544.

Singh J. 2005. Distributed R&D, Cross-regional Knowledge Integration and Quality of Innovative Output. Working Paper, Insead.

Teece DJ. 1977. Technology Transfer by Multinational Firms: The Resource Cost of Transferring Technological Know-How. *Economic Journal* **87**(June): 242-261.

Teece DJ. 1980. Economies of Scope and the Scope of the Enterprise. *Journal of Economic Behavior and Organization* **1**: 223-247.

Von Hippel E. 1988. *The Sources of Innovation*. Oxford University Press: Oxford.

Williamson OE. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. Free Press: New York.

Williamson OE. 1991. Comparative Economic Organization: The Analysis of Discrete Structural Alternatives. *Administrative Science Quarterly* **36**(June): 269-296.

Wooldridge, JM. 2002. *Econometric Analysis of Cross Section and Panel Data.* The MIT Press: Cambridge, MA.