NBER WORKING PAPER SERIES

MEASURING SUCCESS OF ADVANCED TECHNOLOGY PROGRAM PARTICIPATION USING ARCHIVAL DATA

Lynne G. Zucker
Michael R. Darby

Working Paper 9780
http://www.nber.org/papers/w9780

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2003

This paper reports on research partially supported by the ATP Economic Assessment Office, National Institute of Standards and Technology, U.S. Department of Commerce. We also are grateful to the UC Industry-University Cooperative Research Program for funding the task of linking ATP-awardees to our archival data. Raw data and analysis data sets used here were constructed by Zucker and Darby as part of a commercial joint venture between UCLA and a private firm and are proprietary to that joint venture. We are indebted for help in doing the work reported here to a remarkably talented team of postdoctoral fellow David M. Waguespak and research assistants David Johnson and Mark Junkunc, and also Stephanie Hwang, Andrew Jing, Qiao Liu, Henry Tang, and Xiaogang Wu. This paper is a part of the NBER's research program in Productivity. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research, the University of California, or the U.S. Department of Commerce. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

Measuring Success of Advacned Technology Program Participation Using Archival Data
Lynne G. Zucker and Michael R. Darby
NBER Working Paper No. 9780
June 2003
JEL No. O31, O38, L5, C81

ABSTRACT

This paper examines the value of collecting archival data to evaluate the Advanced Technology Program's (ATP) impact on participants' short- and long-term business success. We use two types of indicators of business success: patenting activity which can be tracked for all participants, and financial market data which is extensive for public firms but limited for start-up and other private firms to receipt of venture capital, membership in joint ventures and strategic alliances, and going public in issuing stock. We compare effects of program design differences, primarily joint venture versus single participant projects, on changes in the rate of patenting before and after participation in ATP. The discussion of patent archives serves to document data for later analyses; discussion of other data sources is intended both to guide other researchers and to inform administrative decisions about collecting similar archival data as part of routine assessment activity. We find that patenting rates generally increase after ATP participation under a number of different program and participant variations. Joint venture participants consistently show increases in patenting after beginning ATP participation, while results vary with definitions for single participants. We also demonstrate that it is possible to identify the timing and amounts of receipt of venture capital by private firms participating in ATP.

Lynne G. Zucker
Professor of Sociology &
Director, Center for International Science, Technology, and Cultural Policy
School of Public Policy & Social Research
University of California, Los Angeles
Los Angeles, CA 90095-1551
and NBER
zucker@ucla.edu

Michael R. Darby
Cordner Professor of Money & Financial Markets
Depts. of Management, Economics & Policy Studies
Anderson Graduate School of Management
University of California, Los Angeles
Los Angeles, CA 90095-1481
and NBER
darby@ucla.edu

Table of Contents

Measuring Success of Advanced Technology Program Participation Using Archival Data

by
Lynne G. Zucker and Michael R. Darby
University of California, Los Angeles and National Bureau of Economic Research

Executive Summary

This paper examines the value of collecting archival data to evaluate the Advanced Technology Program's (ATP) impact on participants' short- and long-term business success. We lay out the methodology for collecting archival data and the lessons learned for possible future work by ATP staff or other researchers. The most important lesson learned is that integrating ATP information with archival data from other sources is both feasible and useful.

Here we start with archival data that UCLA and NBER had already collected for other purposes. We combine that data under support by the UC President's Office with data on the ATP participants from the National Institute of Standards and Technology (NIST) web site and from the Business Reporting System developed by ATP's Economic Assessment Office.

This paper focuses on two types of indicators of business success as indicators of the usefulness of archival data for ATP evaluations: patenting activity which can be tracked for all participants, and financial market data which is extensive for public firms but limited for start-up and other private firms to receipt of venture capital, membership in joint ventures and strategic alliances, and going public in issuing stock.

We compare effects of program design differences, primarily joint venture versus single participant projects, on changes in the rate of patenting before and after participation in ATP. We also describe both our success and some of the problems encountered in matching ATP participants into the patent files and a number of other well-known archival data sources. The discussion of patent archives serves to document data for later analyses; discussion of other data sources is intended both to guide future researchers and to inform ATP administrative decisions about collecting similar kinds of archival data as part of routine assessment activity.

Data on ATP project-specific intellectual property, including patents and other kinds of innovations, are routinely collected by ATP's Economic Assessment Office. In this paper, we begin to extend this assessment to see if ATP projects have a more general effect on formation of new intellectual property within the firm or non-profit, an "internal knowledge spillover." Our main indicator is whether overall rates of patent application by a firm or non-profit increase after participation in ATP begins. Using a patent count measure from archival data provided to us by colleagues at the NBER, we find that patenting rates generally increase after ATP participation under a number of different program and participant variations. Joint venture participants consistently show increases in patenting after beginning ATP participation, while single participants do so in the one-year window but not in the two-year window. We conclude this section with a brief discussion of comparison group issues.

We also demonstrate that it is possible to identify the timing and amounts of receipt of venture capital by private firms participating in ATP. This success is particularly encouraging because it demonstrates that there are indicators of firm success besides patent activity which are available without burdensome collection requirements on firms that have not yet gone public and hence begun regular public financial reporting.

The remainder of the paper further documents the arduous process we followed in order to get positive firm and non-profit identification and to screen for changes in name or ownership. We use extensive archival data proprietary to a commercial joint venture in order to make positive identification of ATP participants in each archive; similar data is available through licensing from commercial vendors. We conclude with several suggestions about improving NIST internal data collection and integration.

I. Introduction

The Advanced Technology Program (ATP) of the National Institute of Standards and Technology (NIST) has recently celebrated its tenth year of operation and has funded research conducted by over 1000 participant organizations and their subcontractors. There are now sufficient observations to attempt a quantitative assessment of both the overall effects of the program and the effects of at least one of the program's design elements, the encouragement of research joint ventures.

Our focus is on the use of archival data located in proprietary data bases that can be licensed, purchased from certified re-sellers of vendor data, or developed internally from the original sources (i.e. the U.S. Patent and Trademark Office patent files). Our research is designed both to prove by demonstration the feasibility of retrieving archival data on ATP awardees and to demonstrate its utility in program evaluation in two major analyses. After exploring a number of different success measures utilizing archival data sets assembled as part of a proprietary joint venture, we selected patent data as providing the most complete assessment of ATP effects. Financial disclosures of public firms, as collected by such value-added providers as COMPUSTAT, are readily available to ATP and other researchers willing to confine their analyses to public firms. While it is impossible to find similar sources for the many startups and other non-publicly-traded (private) firms participating in ATP, we do demonstrate that receipt of venture capital investment can be tracked well enough to provide an alternative measure of business success.

We focus first on the number of patents applied for before the ATP award is received and the number of patents applied for during and after the ATP award period. We first describe the basic data on patents used here, already collected and in SAS data sets under other support prior to the NIST work. We then provide some initial analyses of patent data for ATP awardees, examining total numbers of patents granted before, during, and after ATP award.

We next report our efforts to link ATP participant firms to archival data collected by others for use by investors and analysts. Over 44 percent of ATP-participant firms are publicly traded, and were linked successfully to the Compustat database which includes identifiers to readily link these firms further to many specialize databases. These public firms account for the bulk of activity by ATP-participant firms in terms of sales, employees, and patenting since they include six sevenths of the large firms and two thirds of the medium sized firms. Nonetheless, it would be misleading to ignore the private third of the medium firms and especially the private three quarters of small-firm ATP participants. We show that we can link substantial numbers of these firms (over 60% of the small firms beginning ATP participation during 1990-1992) into existing databases with extensive information on firms that have received venture capital, issued new securities (including IPOs), or participated in joint ventures or strategic alliances.

We follow these initial analyses with a description of our work necessary to combine data on ATP awardees and subcontractors with these success indicators (supported by the UC President's Office). There are potentially three modalities for analysis: the ATP project, the ATP participants, or the entire firm or other organization participating in ATP. ATP already conducts ongoing assessments at the project level and has the means and leverage to conduct appropriate surveys there. ATP participants as recorded by ATP are a mixture of single-location firms and specific local sub-units of larger corporations, as well as universities and other non-profit organizations which partner with firms in joint ventures.[1] Generally, patents as well as financial data are available only for the organization as a whole and not for individual locations

[1] Sub-units of organizations sometimes report their own patents to the USPTO, but it is more common that the whole organization files patents from corporate headquarters or the main research location.

of multi-location firms. As a result, for analysis of whether participation in ATP had a positive effect on firms, we must move to the firm/organization as our basic unit of analysis.[2]

For single unit firms, it makes no difference whether we focus on ATP participants at the local "establishment" level or at the whole firm/organization level: this is a distinction without a difference and the only difficulty is in locating these smaller, often non-public firms. For other ATP participants, it takes considerable care to identify the parent firm associated with each establishment level participant. The same firm frequently has multiple sub-units participating in one or more ATP projects.[3]

We conclude with a discussion of the most important issues we faced in our research work, and a discussion of issues our work raises for internal NIST administrative data collection on ATP awardees.

[2] The only source of data on businesses at the local-unit or "establishment" level known to us is the U.S. Census Bureau. It might be possible for ATP to have individual staff or outside researchers be sworn as census workers so as to attempt to exploit the establishment level data. We note, however, that the impact on firm success might well be missed by an establishment-level analysis since research and development conducted in one location might well be applied by the firm in other locations.

[3] In research conducted after this report was substantively complete, we discovered that in following participants over time it is necessary to develop methods for dealing with the non-negligible (5 to 10% depending on period of analysis) number of participant firms which are acquired or merged during the period of analysis. We were able to create consolidated patenting rates for the combined firms, but other methods might be more appropriate in other contexts.

II. Patenting by ATP-Participant Firms and Organizations

ATP-participant firms are at the forefront of technological progress and are major users of the patents system. The 1,011 main R&D participants in ATP from 1990 through January 1999 represent 649 unique organizations. These organizations were assigned 34 percent of the U.S. patents with U.S. assignees at issue during 1993-1996.[4] Filing costs alone run in the range of $0.5 to $1 billion per year, or an average of over $1 million per firm.[5] These facts simultaneously indicate both the importance which participant organizations put on acquiring intellectual property rights to their R&D discoveries and the astonishing connection of the relatively small ATP program with the core firms driving America's national innovation system.[6]

Whether or not ATP participation increases patenting activity by these firms is an excellent indicator of the impact of the program on the participants' research productivity and long-term impact on business success. Patenting also has the signal advantage of being publicly disclosed and recorded in machine readable form for all types of ATP participants, whether publicly traded firms or privately held, whether universities, federal labs, or other non-profits.

Patenting might seem an unlikely indicator of ATP impact since only 40 patents from 1993-1996 were reported to ATP as resulting from ATP-funded projects. This is less than 0.1 percent of total patenting by ATP participant organizations. A possible consideration in

---

[4] The reported patent data are based on matching into a beta-test version of the 1981-1996 Derwent patent files as licensed and cleaned at the NBER by Bronwyn H. Hall, Rebecca Henderson, Adam B. Jaffe, Manuel Trajtenberg, and their colleagues. See Hall, Jaffe, and Tratjenberg (2001) for the final version of the data which is now available on-line and on a CD-ROM. The name cleaning process which got us from 1,011 project participants as counted by ATP to 649 unique organization is detailed in sub-section III.B and the Appendix to this paper. If the ATP were to take on following patents for participant organizations based on the results of this pilot study, either the USPTO or a value-added data provider would be a more appropriate source to use.

[5] If preparation and filing costs amount to about $50,000 per patent application, the cost for issued patents would average nearly $730 million per year. Allowing for patents applied for but not granted would raise the total filing cost to $1 billion or more.

[6] Detailed analysis of patenting by these organizations follows below. While universities and other non-profit organizations are numbered among the participant organizations, firms account for 88% of the organizations and an even higher percentage of the total patents assigned to participant organizations.

assessing the rate of reported patenting is that participant firms may have incentive to conservatively report patents to ATP in order to limit the giving of royalty-free patent use rights to the federal government. Aghion and Tirole have emphasized the difficulty that firms have in writing contracts that effectively induce researchers to disclose valuable inventions resulting from their research, and the same incentives to avoid reporting may be present here.[7]

A broad explanation for observing increases in patenting as a result of ATP project participation would be "internal knowledge spillovers" which occur through transfer of knowledge from one person or one sub-unit to another within the same organization. Moreover, internal competitive behavior within firms or non-profits may also increase patenting: other sub-units will imitate if rewards appear to flow to the unit with the ATP project, or if positive advantages appear to accrue to units more successful in developing new intellectual property.

The proof is in the results, so we present simple but powerful evidence of increased patenting as a result of ATP participation in sub-section II.B.[8] We first consider in sub-section II.A issues involved in making valid comparisons over time given the upward trend in patenting observe in the 1990s. Other measurement issues are deferred to subsection III.B and section IV.

II.A. Before and After Measures of Patenting by ATP Participants

There are two main issues in making before and after comparisons of ATP-participant patenting rates: selection of the appropriate before and after periods and deflation of patent rates so as to avoid attributing overall increases in patenting to ATP.

[7] Aghion, Philippe, and Jean Tirole, "The Management of Innovation," *The Quarterly Journal of Economics*, November 1994, 109(4): 1185-1209.

[8] These results could be due to a third factor which increases patenting by participants while also leading them to apply for ATP funding. We examine these effects in a separate report which we believe substantially strengthens the case that these patenting increases are indeed effects of ATP: Darby, Michael R., Lynne G. Zucker, and Andrew Wang, "Universities, Joint Ventures, and Success in the Advanced Technology Program", National Bureau of Economic Research Working Paper No. 9463, January 2003.

Selection of the Before and After Periods

We experimented with different alternative windows for defining the before and after patenting rates of ATP. We report here the results of two alternative pairs of before and after windows that serve to illustrate the robustness of the results to different criteria. We were somewhat constrained in our choices by the availability to us of archival patent data only through 1996. Subsequent to the results reported in this paper, we have extended the patent database up to mid-1999 and obtained similar results.

"One-year window" comparisons of patenting rates compare counts of patents granted 0 to 365 days before the start of ATP funding (the one-year "before" period) with those granted in the 365 days beginning two years after the start of ATP funding (the "after" period).[9] This allows for a minimum two-year lag in carrying out ATP-funded research and resulting patents.

"Two-year window" comparisons of patenting rates compare counts of patents granted between 365 days before and 365 days after the start of ATP funding (the two-year "before" period) with those granted in the two years (730 days) beginning two years after the start of ATP funding. Patents granted during the first year of funding cannot be attributed to ATP participation, but going back two years before the start raised concerns that recently founded companies might be too young to have any patents granted at least in the second year before the start of funding.

Deflating Patents for Comparison across Time

On average as shown below, patents increase (with allowance for lags between application and grant) after beginning ATP participation in comparison to patents before participation. However, the value of patents and the ease of obtaining them affect the overall

[9] For organizations that participated in multiple projects, the organization is in the sample only once based on the first participation in ATP.

rate of patenting.[10] In recent years Congress and the courts have strengthened patent rights and the U.S. Patent and Trademark Office has hired more patent examiners. As a result, both the rate of patent application and the speed with which patents are granted have increased. Thus, a simple before and after comparison is subject to criticism as reflecting trend increases rather than any real effect.

Accordingly, we developed two "deflated" patent-count measures. Our alternative deflators counts are the total-patents deflator and the patents-per-assignee deflator. The total-patents deflator is the ratio of the total number of U.S. patents with a U.S. assignee at issue in a given year to the number of those patents in 1990. The patents-per-assignee deflator is the ratio of total number of patents with a U.S. assignee at issue in a given year divided by the number of U.S. assignees in that year to the same calculated patents per assignee in 1990. Since the patents-per-assignee deflator is a measure of the rate of patenting by individual firms, it is our preferred deflator. The total-patents deflator confounds increases in the rate of patenting per firm with increases in the number of firms in the economy. Thus, it over deflates patent counts for individual firms.

Table 1 reports data on the total number of U.S. patents with a U.S. assignee at issue, the number of U.S. assignees, and the calculated values of patents-per-assignee deflator and the total-patents deflator. As can be seen from the table, patents per U.S. assignee have increased by nearly 10 percent while total patents with U.S. assignees at issue have increased nearly 35 percent. Thus, it is important to deflate patent counts to eliminate the upward trend in patenting per firm, but the total patent deflator appears to overdo this correction.

[10] Griliches, Zvi, "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, December 1990, 28:1661-1707.

Table 1: Patents per Assignee and Total Patents Deflators

Year	Total US Patents	Total US Assignees	Patents-Per-Assignee Deflator	Total-Patents Deflator
1989	36708	9314	1.0224	1.0665
1990	34419	8929	1.0000	1.0000
1991	37513	9339	1.0420	1.0899
1992	38892	9634	1.0473	1.1300
1993	40297	9855	1.0608	1.1708
1994	42585	10405	1.0617	1.2373
1995	42110	10499	1.0405	1.2235
1996	46421	10991	1.0957	1.3487

II.B. Increase in Patenting after Beginning ATP Participation

The following tables compare deflated patent counts in the before and after periods for ATP parent organizations whose first project started before 1993, subdivided by various categories. These tables provide examples of analytical uses of information derived from combining ATP information with that available in other archival sources. The basic goal is to see whether or not patenting behavior is different in the before and after periods. The sub-categorizations, such as whether the project is a joint venture or single applicant, are used as independent variables to further explain variation in before and after patenting. These tables are only intended to demonstrate the value to ATP of establishing regular processes to collect this information.

Table 2 shows substantial increases in patenting rates are associated with beginning participation in the ATP and that this result does not depend on whether one-year or two-year before and after windows are used nor on which, if any, patent deflator is used. For this research, we only had access to patents for matching ATP firms through 1996. As a result there are generally fewer observations available for which the two-year window is within the data set. Extending the data range would add considerably to the number of observations and the ability to measure statistically significant changes comparing before and after ATP.

Table 2: Before and After Patenting*

	DEFLATOR USED	BEFORE ATP	AFTER ATP
One-year Window	None (raw means)	58.28	70.09
	Patents per Assignee	55.51	66.39
	Total U.S. Patents	51.61	57.89
	Number of Cases	129	129
Two-year Window	None (raw means)	136.86	169.19
	Patents per Assignee	131.53	159.66
	Total U.S. Patents	124.27	136.95
	Number of Cases	104	104

* The one-year patent windows are defined as:
Before = patents issued 0 to 365 days from start of ATP research
After = patents issued 731 to 1096 days from start of ATP research
The two-year patent windows are defined as:
Before = patents issued 365 days before to 365 days after the start of ATP research
After = patents issued 731 to 1460 days from the start of ATP research

Tables 3 through 8 illustrate how patenting increases after ATP vary according to the nature of the participants' organizations and the conditions of their participation in ATP. The tables suggest hypotheses and variables for further analysis.

Table 3: Before and After Patenting Rates by Single and Joint-Venture Participants

WINDOW	PARTICIPANT GROUPS	DEFLATOR USED	BEFORE ATP	AFTER ATP
One-year Window	Single Participant	None (raw means)	1.50	1.67
		Patents per Assignee	1.42	1.59
		Total U.S. Patents	1.30	1.35
		Number of Cases	36	36
One-year Window	Joint Venture Participant	None (raw means)	80.26	96.57
		Patents per Assignee	76.45	91.47
		Total U.S. Patents	71.08	79.78
		Number of Cases	93	93
Two-year Window	Single Participant	None (raw means)	4.45	3.45
		Patents per Assignee	4.25	3.27
		Total U.S. Patents	3.95	2.76
		Number of Cases	22	22
Two-year Window	Joint Venture Participant	None (raw means)	172.38	213.66
		Patents per Assignee	165.68	201.62
		Total U.S. Patents	156.55	172.95
		Number of Cases	82	82

Table 3 examines whether there seems to be a greater effect on patenting rates for organizations participating in joint ventures than for single participants. This is a relevant design question since ATP actively encourages joint ventures. The first thing that is obvious in the table is that joint-venture participants are typically much larger and patent more than single participants. Patenting increases for every group after beginning ATP participation except for single participants using the two-year window. This exception likely reflects the very small and early sample of firms that can be included in this cell without extending the patent data.

Table 4: Before and After Patenting by Organization Type

WINDOW	ORGANIZATION TYPE	DEFLATOR USED	BEFORE ATP	AFTER ATP
One-year Window	Large Business	None (raw means)	263.54	311.15
		Patents per Assignee	251.10	294.78
		Total U.S. Patents	233.74	257.37
		Number of Cases	26	26
One-year Window	Medium Business	None (raw means)	11.67	19.05
		Patents per Assignee	11.14	18.01
		Total U.S. Patents	10.43	15.85
		Number of Cases	21	21
One-year Window	Small Business	None (raw means)	0.88	1.19
		Patents per Assignee	0.84	1.13
		Total U.S. Patents	0.79	0.98
		Number of Cases	58	58
One-year Window	University	None (raw means)	18.50	29.08
		Patents per Assignee	17.53	27.70
		Total U.S. Patents	16.01	23.64
		Number of Cases	12	12
Two-year Window	Large Business	None (raw means)	526.68	637.84
		Patents per Assignee	506.27	601.90
		Total U.S. Patents	478.48	516.51
		Number of Cases	25	25
Two-year Window	Medium Business	None (raw means)	26.12	42.65
		Patents per Assignee	25.21	40.26
		Total U.S. Patents	24.02	34.87
		Number of Cases	17	17
Two-year Window	Small Business	None (raw means)	2.57	3.24
		Patents per Assignee	2.48	3.06
		Total U.S. Patents	2.35	2.63
		Number of Cases	42	42
Two-year Window	University	None (raw means)	38.82	66.27
		Patents per Assignee	36.98	62.47
		Total U.S. Patents	34.31	52.55
		Number of Cases	11	11

In Table 4 we examine size and type of organization directly by looking at small, medium, and large firms and universities. Here, whether the one-year or two-year window is used, patenting increases after beginning ATP participation for every one of these groups. The increases are sizable in percentage terms for all organization types and in absolute amounts for all organization types other than small firms. Again whether and how patents are deflated for trend does not alter these results qualitatively.

Table 5: Before and After Patenting by ATP Award Size Category*

WINDOW	AWARD CATEGORY	DEFLATOR USED	BEFORE ATP	AFTER ATP
One-year Window	0-$500K	None (raw means)	22.26	27.53
		Patents per Assignee	21.15	26.17
		Total U.S. Patents	19.47	22.47
		Number of Cases	38	38
One-year Window	$500K - $1M	None (raw means)	105.38	140.92
		Patents per Assignee	100.81	132.94
		Total U.S. Patents	94.92	117.41
		Number of Cases	39	39
One-year Window	$1M – $1.5M	None (raw means)	25.60	19.27
		Patents per Assignee	24.21	18.42
		Total U.S. Patents	22.06	15.70
		Number of Cases	15	15
One-year Window	$1.5M+	None (raw means)	67.16	68.16
		Patents per Assignee	63.59	65.04
		Total U.S. Patents	58.08	55.49
		Number of Cases	32	32
Two-year Window	0-$500K	None (raw means)	45.29	57.00
		Patents per Assignee	43.26	53.80
		Total U.S. Patents	40.38	45.50
		Number of Cases	38	38
Two-year Window	$500K - $1M	None (raw means)	290.23	406.85
		Patents per Assignee	281.14	384.48
		Total U.S. Patents	270.01	333.99
		Number of Cases	26	26
Two-year Window	$1M – $1.5M	None (raw means)	66.75	45.75
		Patents per Assignee	63.48	43.11
		Total U.S. Patents	58.54	36.19
		Number of Cases	12	12
Two-year Window	$1.5M+	None (raw means)	178.83	184.61
		Patents per Assignee	170.15	173.58
		Total U.S. Patents	157.30	145.57
		Number of Cases	23	23

* For single participants, award is the amount granted to the organization for its first appearance in ATP. For joint-venture participants, the award is the average amount granted to participants in its first ATP joint venture.

Table 5 examines whether the increase in patenting is related to the size of the grant. Surprisingly, there is a substantial increase only for award sizes less than $1 million. For awards between $1 and $1.5 million patenting actually decreases after beginning ATP while for awards over $1.5 million patenting increases only for the undeflated and per-assignee deflated measures.

(In separate analysis, we find that this inverse relationship between award size and impact on patenting does not hold up when controls are added for other firm characteristics.)

ATP records include a participant-reported item on whether or not the participant is a publicly traded firm (PUBLIC = 1). Table 6 compares the patenting behavior of these public firms with all other ATP-participating organizations. Not surprisingly, public firms on average are about 25 times larger than all other ATP-participant organizations as measured by patenting activity. More interestingly, patenting in public firms increases after beginning ATP participation more than in other participants in absolute terms but considerably less in percentage terms.

Table 6: Before and After Patenting by Public Firms and All Other ATP Participants

WINDOW	PUBLIC/PRIVATE	DEFLATOR USED	BEFORE ATP	AFTER ATP
One-year Window	Participant Organizations Other Than Public Firms	None (raw means)	5.47	7.14
		Patents per Assignee	5.18	6.76
		Total U.S. Patents	4.69	5.75
		Number of Cases	77	77
One-year Window	Public Firms	None (raw means)	136.48	163.29
		Patents per Assignee	130.05	154.68
		Total U.S. Patents	121.08	135.11
		Number of Cases	52	52
Two-year Window	Participant Organizations Other Than Public Firms	None (raw means)	10.85	16.22
		Patents per Assignee	10.36	15.30
		Total U.S. Patents	9.66	12.92
		Number of Cases	55	55
Two-year Window	Public Firms	None (raw means)	278.29	340.90
		Patents per Assignee	267.54	321.69
		Total U.S. Patents	252.91	276.16
		Number of Cases	49	49

To indicate the effect of different joint venture attributes on patenting behavior, the next two tables include data on only organizations that have been ATP joint venture participants. Previous work indicates that collaborations with university scientists are very important to firm success in biotechnology.[11] Table 7 provides only mixed support for the value

[11] Lynne G. Zucker, Michael R. Darby, and Jeff Armstrong, "Geographically Localized Knowledge: Spillovers or Markets?", *Economic Inquiry*, January 1998. 36(1): 65-86.

Table 7: Before and After Patenting within JVs by Organizations with University JV Partners

WINDOW	JOINT VENTURE TYPE	DEFLATOR USED	BEFORE ATP	AFTER ATP
One-year Window	Without a University Partner	None (raw means)	137.05	160.58
		Patents per Assignee	130.51	152.21
		Total U.S. Patents	121.29	132.71
		Number of Cases	40	40
One-year Window	With a University Partner	None (raw means)	37.40	48.26
		Patents per Assignee	35.66	45.64
		Total U.S. Patents	33.19	39.83
		Number of Cases	53	53
Two-year Window	Without a University Partner	None (raw means)	268.31	328.85
		Patents per Assignee	257.70	310.22
		Total U.S. Patents	243.07	265.84
		Number of Cases	39	39
Two-year Window	With a University Partner	None (raw means)	85.37	109.19
		Patents per Assignee	82.22	103.11
		Total U.S. Patents	78.07	88.69
		Number of Cases	43	43

of university-firm collaborations: Patenting increases after beginning ATP participation are higher in percentage terms for joint ventures with university partners than those without them, but just the opposite is true in terms of the absolute increase in patenting. The value of university partners is examined further in separate research.

Table 8: Before and After Patenting within JVs by Fixed/Changed JV Membership

WINDOW	JV FIXED/CHANGED MEMBERSHIP	DEFLATOR USED	BEFORE ATP	AFTER ATP
One-year Window	Membership remains fixed	None (raw means)	175.38	186.62
		Patents per Assignee	166.18	177.88
		Total U.S. Patents	151.93	151.83
		Number of Cases	13	13
One-year Window	Membership changes	None (raw means)	64.80	81.94
		Patents per Assignee	61.87	77.43
		Total U.S. Patents	57.95	68.07
		Number of Cases	80	80
Two-year Window	Membership remains fixed	None (raw means)	375.42	392.67
		Patents per Assignee	357.40	369.82
		Total U.S. Patents	330.88	310.83
		Number of Cases	12	12
Two-year Window	Membership changes	None (raw means)	137.57	182.97
		Patents per Assignee	132.81	172.78
		Total U.S. Patents	126.66	149.31
		Number of Cases	70	70

It seems plausible that changes in the membership of a joint venture would at least be an indicator if not a cause of lower success. Surprisingly, Table 8 indicates that just the opposite is true: Participants in joint ventures which experience turnover in their membership on average experience larger percentage and absolute increases in patenting after beginning ATP participation. *Ex post* discussions suggest we can rationalize this finding as illustrating the value of strong leadership which permits weeding out of non-performing partners, but the results remain puzzling to us.

The principal lesson of this sub-section is that patenting appears to be a good indicator of the effect of ATP participation upon the success of the firm's research productivity and hence overall success. Furthermore, this indicator shows that participation in ATP leads in time to a substantial increase in firm patenting, apparently reflecting not only direct effects of increased R&D expenditures but also an element of internal "spillovers" and competition.

II.C. Technical Issues on Measuring Patenting by ATP Participants

Our work in linking the patent files to the ATP participants illustrates three major issues in combining archival data from different sources collected for different purposes: developing variant-to-preferred name lists for various organizational levels, deciding on whether or not to treat missing values as zeroes, and reconciling different values for apparently similar concepts. We have already introduced the first issue and will discuss it at length in subsection III.B and section IV below. The other two are considered here.

Missing Data and Zeroes

The ATP participants and the U.S. patent assignees lists have the excellent features for matching that each represents the complete universe of the cases to which they refer. If both lists had a unique identifier at the firm level (e.g., a taxpayer identification number or TIN) associated with each observation and organizations never changed that identifier through merger, acquisition, or spin-off, then we would know that any ATP participant which was not a patent assignee in any given time period truly had no patents granted in that period. That is, missing values in the patent data would definitely be true zeroes. Unfortunately, no such identifiers are available and some missing patent values may be due to our inability to find the name used for assignments of patents to the firm.

Since the patent file covers some years before 1985, we are able to identify in earlier years patents by some firms which later participated in ATP but had no patent assignments in 1985-1996. We are thus confident that their absence from the patent file represents a true zero for the 1985-1996 period. Unfortunately, as illustrated in Table 9, of the 41.3 percent of ATP participant organizations with no patents in 1985-1996, only 1.8 percent can be so classified as definite zeroes. For the other 39.4 percent of ATP participant organizations, there is an element of doubt whether their being missing from the patent list is due to their lack of patents or their patenting in an undiscovered name.

Table 9: Comparison of Patent Identification Rates in 1985-1996 and in Unrestricted Time Frame for ATP Participant Organizations

	Frequency	Percent	Cumulative Frequency
Zero or missing	256	39.4	256
Definite Zero	12	1.8	268
Positive Value	381	58.7	649

Table 10 gives us reason to believe that treating the zero or missing cases as true zeroes is acceptable in this case. For large firms which we would expect most likely to have patents, 94 percent have either positive or definite zero values. Some of the 6 percent of large firms with no discovered patents may truly have had none, while others may have been missed. Nonetheless, if we know what is going on with 94 percent of the large firms, little error is introduced by treating the other 6 percent as zeroes. University participants in ATP have similar percentages of definite zeroes, positive values, and zeroes or missing.

Table 10: Comparison of Patent Identification Rates in 1985-1996 and in Unrestricted Time Frame for ATP Participant Organizations by Organization Type

		Frequency	Percent	Cumulative Frequency
Large Business	Zero or missing	6	6.4	6
	Definite Zero	3	3.2	9
	Positive Value	85	90.4	94
Medium Business	Zero or missing	31	22.3	31
	Definite Zero	2	1.4	33
	Positive Value	106	76.3	139
Small Business	Zero or missing	188	55.8	188
	Definite Zero	5	1.5	193
	Positive Value	144	42.7	337
University	Zero or missing	2	5.4	2
	Definite Zero	1	2.7	3
	Positive Value	34	91.9	37

For medium and small sized firms, 22 and 56 percent, respectively, cannot be identified at all among the patent assignees up through 1996. While these percentages are much higher than observed for large firms and universities, it is certainly reasonable that these firms are in fact less likely to have patent assignments. Furthermore, referring back to Table 4, we see that increasing the patenting rates of medium firms by 28.7 percent (22.3/0.777 = 28.7) and small firms by 126.2 percent would have only a small effect on the overall rate of patenting by ATP participant organizations. While the matching seems to have caught essentially all of the major patenting firms, there is some reason to view comparisons of patenting by medium and small firms as possibly affected by misclassification of missing values as zeroes.

Differences in Related Measurements across Data Sources

There is reason to expect that total firm patenting would be little affected by participation in ATP since relatively few patents are reported by firms as resulting from ATP funding – only 40 such patents were reported by 1996 as detailed in Table 11. Further disaggregate analysis of total patenting by firms could possibly allow better assessment of reported project-level patenting.

Table 11: Comparison of Yearly Patent Counts for ATP Participant Organizations with the Number of Patents Reported to ATP as Resulting from an ATP Project

Year	UCLA Total Patent Count	Patents Reported *
1993	13781	2
1994	14561	6
1995	14255	14
1996	15636	18

* The number of patents reported to ATP by participants

III. Financial Markets Database Information for ATP-Participant Firms

ATP participant firms fall into two distinct classes: well established, largely public firms of large or medium size and young, rapidly growing firms that are initially private and go public if they succeed in their business plan. The latter start out as small businesses, but the best of them become medium and then large size in relatively short order.

Publicly traded firms are required to make extensive disclosures of accounting and other material data. This data is available in a number of forms, but most researchers find the Compustat database to be the state of the art. The next sub-section indicates the feasibility of linking the ATP and Compustat databases. Many industrial-organization researchers find that restricting their firm sample to public firms and using Compustat or similar sources is adequate to answer most important questions, and we have already seen that the bulk of patenting is concentrated in the large (mostly public) firms.

ATP plays an important role in fostering research at young or start-up firms that are often set up by outstanding scientists unable to interest more established firms in commercializing their ideas – some of which truly amount to scientific and technological breakthroughs. Since small firms play a major role in bringing innovations to the economy, it would be a serious error to not try to evaluate the effect of ATP participation on these (initially) small private firms.[12] Fortunately, these small private firms are the subjects of intense interest on the part of both venture capitalists looking for investment clients and investment bankers looking for firms to take public. As a result, there is significantly more archival information on small, private,

[12] On the contribution to innovation of small firms, see Zoltan J. Acs, and David B. Audretsch, "Innovation in Large and Small Firms: An Empirical Analysis," *American Economic Review*, September 1988, 78: 678-690.

high-tech firms than industrial-organization researchers are used to. We illustrate the availability of this data and the possibility of matching into it in sub-section III.B concentrating on certain Securities Data Corporation databases that we had licensed for other purposes.

III.A. Matches to the Compustat Database

ATP sets a PUBLIC flag to 'Yes' for firms that self-report public status. This flag is set to 'Yes' for at least one establishment in 209 firms. It is not surprising that different establishments of the same firm might differ in their interpretation of whether they are publicly traded. For example, is a wholly owned subsidiary of a publicly traded firm private or public? That depends on whether one thinks of the particular corporation or the entire organization filing a consolidated income tax return.

Table 12: Firms Matched to Compustat Public Firms by Whether ATP's Public Flag Is Set

Firms by Compustat Match	Value of PUBLIC Variable	
	No*	Yes
Not matched to Compustat	303	15
Matched to Compustat	58	194
Total number of firms	361	209

* The PUBLIC variable is a flag for public firms so no is inferred from absence of the flag.

Fortunately, we can directly match firm names to those in Compustat and then clean for name variants as is our general methodology (discussed in detail in Section IV below). If we succeed in matching to Compustat, it is straightforward to supplement the data there with specialized data sources using CUSIP numbers which identify the firm's securities. We were able to match 93 percent of the companies listed as public by ATP to the Compustat data base (see Table 12); the rest may be due to foreign public parents of American subsidiaries or to parents with very different names not identified in the final manual cleaning process. Surprisingly, 16 percent of the participant firms in ATP files as nonpublic are also matched to the Compustat database. This apparent error rate likely reflects both firms that go public subsequent to (and perhaps because of the success of) their ATP project and also such ambiguous cases as wholly owned subsidiaries of public firms.

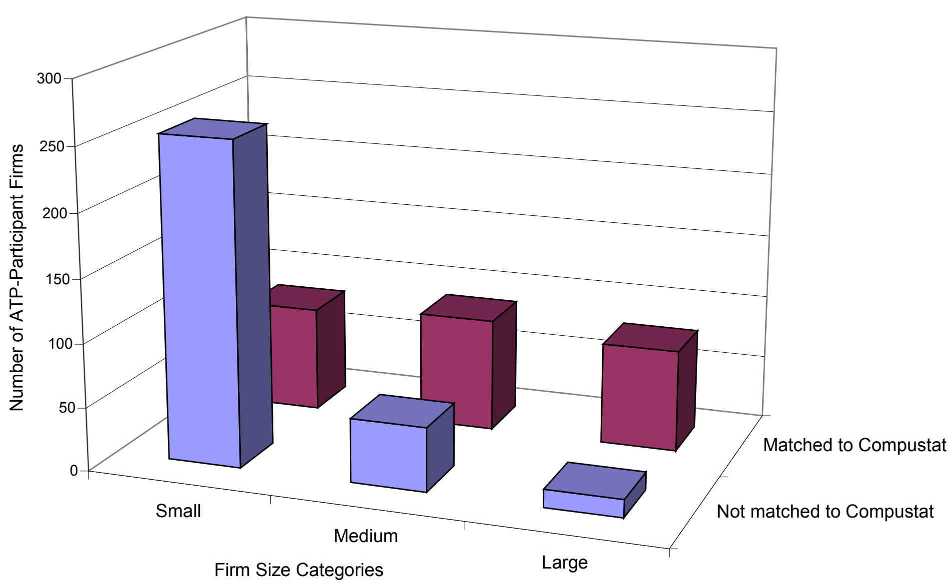
Table 13 reports the distribution of ATP-participant firms by size and whether or not they are matched to Compustat database. Figure 1 illustrates that medium and especially large firms

Table 13: Firms by Size and Whether Matched to Compustat Public Firms

Firms by Compustat Match	Small Firms	Medium Firms	Large Firms
Not matched to Compustat	254	50	14
Matched to Compustat	83	89	80
Total number of firms	337	139	94

are mostly matched to the extensive annual financial data in Compustat, while small firms mostly are not, even though small firms account for 33 percent of the participant firms matched to Compustat. We conclude that full accounting data is available for the main mass of ATP participant firms, when firms are considered weighted by patenting, sales, or employees.

Figure 1. Number of ATP-Participant Firms by Size Matched or Not Matched to Compustat



We believe that it is still important to investigate the small innovative firms as well as the larger public firms. For example, it would be particularly interesting to see if participation in ATP was a significant factor in the success of the firm and a subsequent initial public offering (IPO). We next turn to databases particularly relevant to smaller and nonpublic firms.

III.B. Matches to Securities Data Corporation Databases

Under other funding, we had licensed academic access to three Securities Data Corporation (SDC) databases: Venture Capital Financing (begins 1960), Global Corporate New Issues (begins 1970), and Global Joint Ventures/Strategic Alliances (begins 1988). Each of these databases has many data fields describing particular deals and the characteristics of the firms involved. Unfortunately, the databases had their origins in different predecessor firms and consistency across databases is incomplete. However, matching to any one of the databases provides identifiers which permit matching to the other databases that contain observations on the given firm.

Methodology of Matching within and across Databases

We will describe the process of matching to the above three SDC databases (and the patent assignees database) in some detail as it also explains the history underlying some of the methodological lessons reported in Section IV below. Further details on creation of unique parent organization identifiers, filtering and matching algorithms, and cleaning organization names are contained in the Appendix to this paper.

The goal of the matching process is to identify specific observations across different databases in which the organizations are either identical or related in a known or learnable way, such as parent and subsidiary. In some cases where there is nothing like a variant-to-preferred name list, it is equally important to identify matches within a single data set, as we have done in reducing the original 1,011 main R&D participants to 649 unique organizations with establishments included in the 1,011. When identification of one of these 649 organizations is reported in this sub-section, it means only that we have observed the organization in an outside dataset. This does not necessarily imply that the identified matches occur within the period of interest for analysis of the effects of ATP. However, it does mean that we can be more confident that the absence of matches during the relevant period can be viewed as a definite zero rather than a possible false negatives (see sub-section II.C. above).

Matching revolves around construction of a variant-to-preferred list of alternative names used for a given firm and for its associated subsidiaries and local establishments. Besides names, other identifying information such as addresses, phone numbers, CUSIP numbers, and other codes are built into this identification database. The process is iterative in that we repeatedly compare the items in the identification database with targets for matching and with the source file of 1,011 main ATP R&D participants. This permits us to use linkages discovered in one database to find new linkages in the other databases, gradually building up the identification database. If ATP were to adopt creation and maintenance of linked archival-derived database for ATP participants, the accuracy, speed, and economy of this process would be much improved by establishing a central database recording preferred and variant names and other identifying information at the organization level and at the level of these organizations' establishments.

We can describe the process in more detail by reference to Table 14. The table refers to moving from the original computer processing of the source file of 1,011 main ATP R&D participants through three successive processes. The original computer processing looks for exact name matches and that alone reduces the number of main R&D ATP participants from

1,011 to 804 unique organizations. The column labeled 'original' reports the identification frequency and rate for exact name matches to these 804 names. Overall, the number of matches is low with only 22.3 percent of the names matching exactly to names in one or more of the tree SDC databases and the patent databases. Recall, however, that this data did not standardize names for ATP repeaters, nor separate subdivision names from parent organization names.

Table 14: Comparison of Identification (ID) Rates for Original, Original/Filtered, Cleaned/Filtered, and Original/Filtered/Hand Identified (Main R&D Participants Only)

	Original		Original/Filtered		Cleaned/Filtered		Cleaned/Filtered/Hand Identified	
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
Venture Capital ID	86	10.7	119	14.8	134	20.7	164	25.3
New Issues ID	2	0.2	128	15.9	165	25.5	179	27.6
Joint Venture ID	14	1.7	204	25.4	281	43.4	302	46.5
Patent Assignee ID	108	13.4	257	32.0	337	52.0	395	60.9
Any ID	179	22.3	341	42.4	430	66.4	476	73.3
N	804		804		648		649	

The next step in the matching process is to filter out of the 804 names to be matched relatively irrelevant terms like Company, Corporation, Corp., Incorporated, and Inc., which are frequent causes of erroneous failed matches. Using this computer filtered version of the original names increased the rate of matching to one or more of the target files to a much more respectable 42.4 percent, as indicated in the column of Table 14 labeled "Original/Filtered."

The next step involved identifying names which refer to the same organization and selecting a preferred organization name for the 648 unique organizations identified. Thus a unique organization is identified by a variety of names, such as IBM and International Business Machines. Additional names for the same organization were added from lists in the matched databases. Computer matching with filtered versions of these preferred and variant names resulted in two thirds of the unique organizations being identified in one or more of the target databases and the identified organizations on average are identified in 2.1 different targets.

The final step, which takes most of the time, is the investigation by research assistants of all the organizations that were not identified in previous rounds as well as uncertain "fuzzy" matches of similarly named participants tentatively treated as a single organization in the preceding step. This process involved searching out whether organizations with matching name parts were part of the same or different organizations and finding instances in which establishments with dissimilar names were in fact subsidiaries or divisions of the same parent organization. The net effect was to increase to 649 the number of unique organizations with one or more establishments which were main R&D participants in ATP during 1990-1998 as indicated in the column of Table 14 labeled "Cleaned/Filtered/Hand Identified." Nearly three quarters of these unique organizations were identified in one or more of the target databases and the identified organizations on average are identified in 2.2 different targets.

These 649 unique organizations represent a clean and internally consistent variant-to-preferred organization list based on linkages previously identified in the source and target databases as supplemented by external information. Construction of this list took four months longer than we had estimated from previous matching experience. The reason for this unusually arduous process is that ATP itself maintains no consistent variant-to-preferred organization list which would impose consistent spelling and abbreviation. Lack of full street addresses and zip codes – used to confirm uncertain matches – also posed major problems for us.

<u>Patterns in Identification Rates</u>

Table 15 reports the identification rates in the cleaned data for the 4 largest categories of ATP-participant organizations. We achieve very high matching percentages (around 90 percent overall) for large and medium firms and universities but somewhat lower rates for small firms. On the other hand, given the quality of the matching procedures in the other size categories, we can be quite confident that we have identified essentially all the small-firm ATP participants which received venture capital investments, went IPO or subsequently issued securities, and/or engaged in joint ventures or strategic alliances with other firms. The ability to measure those variable alone for (initially) private firms is a substantial achievement.

Table 15: Identification Rates for Cleaned Data by Organization Type (Main R&D Participants Only)

	Large Business		Medium Business		Small Business		University	
	Freq	Pct	Freq	Pct	Freq	Pct	Freq	Pct
Venture Capital ID	18	19.1	43	30.9	102	30.3	0	0.0
New Issues ID	48	51.1	66	47.5	64	19.0	0	0.0
Joint Venture ID	76	80.9	82	59.0	109	32.3	25	67.6
Patent Assignee ID	88	93.6	108	77.7	149	44.2	35	94.6
Any ID	93	98.9	124	89.2	207	61.4	35	94.6
N	94		139		337		37	

Table 16: Variation in Identification by 3-Year Cohort (1990-1998 Main R&D Participants Only)

	96-98 Cohort	93-95 Cohort	90-92 Cohort
Venture Capital ID	35.4%	18.9%	33.7%
New Issues ID	23.8%	24.8%	41.3%
Joint Venture ID	32.3%	45.6%	71.2%
Patent Assignee ID	41.5%	63.1%	76.9%
Any ID	60.8%	73.5%	88.5%
N	130	355	104

Table 16 shows that, in general, identification improves for firms that began participation in ATP earlier.[13] This pattern of identification is consistent with the scenario in which small firms rapidly mature and begin to receive venture capital and enter into alliances with other firms ultimately go public. Table 17 confirms this interpretation by breaking out firms by size category from the three 1990-1998.[14] For small firms, with the exception of the surprisingly large number receiving venture capital in the most recent cohort, the rate of identification

[13] Note that these three three-year cohorts exclude January 1999 when 60 of the 649 unique ATP-participant organizations started their first projects. This leaves 589 unique participant organizations active 1990-1998.
[14] Note that there are 52 firms among the 60 unique organizations beginning ATP participation in January 1999. This leaves 518 unique firms active during 1990-1998 as recorded in Table 17.

declines (and the rate of implied zeroes increases) for each database as the cohorts become more recent.

Table 17: Variation in Identification by 3-Year Cohort and Firm Type (1990-1998 Main R&D Participants Only)

		96-98 Cohort	93-95 Cohort	90-92 Cohort
Large Business	Venture Capital ID	50.0%	8.5%	28.0%
	New Issues ID	50.0%	46.8%	64.0%
	Joint Venture ID	66.7%	85.1%	92.0%
	Patent Assignee ID	83.3%	95.7%	96.0%
	Any ID	100.0%	100.0%	100.0%
	N	12	47	25
Medium Business	Venture Capital ID	43.5%	22.1%	41.2%
	New Issues ID	60.9%	39.5%	58.8%
	Joint Venture ID	47.8%	57.0%	82.4%
	Patent Assignee ID	60.9%	82.6%	70.6%
	Any ID	78.3%	90.7%	88.2%
	N	23	86	17
Small Business	Venture Capital ID	33.0%	24.6%	50.0%
	New Issues ID	11.0%	18.3%	40.5%
	Joint Venture ID	23.1%	32.6%	57.1%
	Patent Assignee ID	31.9%	47.4%	66.7%
	Any ID	51.6%	62.3%	85.7%
	N	91	175	42

In Section II, we saw that participants in ATP joint-venture projects tended to be larger than single participants and have a larger increase in patenting rate after beginning to participate. Table 18 shows how identification rates vary by whether or not the organization was ever a member of an ATP joint venture. Generally, JV participants have a higher rate of identification than single participants consistent with their larger average size. Intriguingly, however, single participants have a higher rate of receiving venture capital, suggesting that small firms going it alone actually do have a hot technology so that it makes sense that they are unwilling to share with potential JV partners even though the size of awards to single participants is capped at $2 million while there is no such cap on per firm awards to joint ventures.

Table 18: Variation in Identification by JV Membership (1990-1998 Main R&D Participants Only)

	ATP Organization was never a JV member		ATP Organization was a JV member at least once	
	Frequency	Percent	Frequency	Percent
Venture Capital ID	56	32.7	92	22.0
New Issues ID	41	24.0	121	28.9
Joint Venture ID	64	37.4	214	51.2
Patent Assignee ID	92	53.8	266	63.6
Any ID	116	67.8	316	75.6
N	171		418	

Table 19 reports the variation in identification by the amount of money granted per participant for those beginning during 1990-1998.[15] The larger the amount of the ATP award the more likely the recipients will be known to the financial markets and patent office.

Table 19: Variation in Identification by Total ATP Award Money* (Main R&D Participants Only)

	0-$500K total ATP award money		500K – 2M total ATP award money		2M - 5M total ATP award money		5M+ total ATP award money	
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
Venture Capital ID	11	13.6	90	27.4	40	28.2	7	22.6
New Issues ID	20	24.7	76	23.2	49	34.5	17	54.8
Joint Venture ID	29	35.8	141	43.0	79	55.6	27	87.1
Patent Assignee ID	49	60.5	184	56.1	94	66.2	27	87.1
Any ID	55	67.9	233	71.0	109	76.8	31	100.0
N	81		328		142		31	

* Award totals are available for projects only. For Joint Ventures, the award for each participant was calculated as the average amount per JV participant.

III.C. Conclusions on Matching to Financial Market Databases

This research has clearly demonstrated the ability to obtain valuable information for assessing the effects of the Advanced Technology Program by matching participants into financial market databases. For many purposes, analysts may wish to concentrate on only the publicly traded firms which can be matched to Compustat or similar databases with rich accounting and other data. We were able to match into Compustat 44 percent of all the unique

[15] ATP's participant file included 7 firms for which there is no record of receiving funds or of cancellation of their project. Since we cannot identify their funding class, they are excluded from Table 19, leaving 582 organizations.

firms participating in ATP and 73 percent of those with 500 or more employees (medium and large firms). This may serve as a useful standard for other exercises concentrating on public ATP participants. We also note that the public variable in ATP's files is accurate in identifying public firms at least 93 percent of the time when it indicate the firm is public. However, false negatives are more of a problem since we identified as public firms 16 percent of those firms with no flag indicating that they were public.

We also experimented with matching ATP participants to three Securities Data Corporation (SDC) databases: Venture Capital Financing (begins 1960), Global Corporate New Issues (begins 1970), and Global Joint Ventures/Strategic Alliances (begins 1988). The match rates were generally too low – except for large public firms for which Compustat is a better source – to recommend relying on exploiting the detailed firm information as a way to explore the effects of ATP participation on small, non-public firms. On the other hand, identifying whether any of these firms have received venture capital and if so how much of what type, whether any have gone public and on what terms, or formed joint ventures or strategic alliances with other firms – all provides important measures of the success of private firms not otherwise available. Furthermore, matching is sufficiently reliable that absence of matches can be treated as a true zero with acceptable confidence. Thus, we would recommend that ATP license access to these SDC databases – as well as Compustat – for the purpose of following program participants and assessing their success.

IV. Lessons Learned in the Matching Process

Any empirical assessment of the Advanced Technology Program – overall or focused on particular elements -- is necessarily constrained by the database upon which it rests. The Business Reporting System (BRS) has been the major data resource for internal evaluations of ATP. It has the advantage of including confidential information such as patent applications and the disadvantages of limited and non-continuous reporting – with the potential difficulty of strategic behavior influencing responses (e.g., to minimize government rights to resultant patents). We have shown that there is considerable information available to ATP from archival sources which can complement the existing internal data resources.

Our most important methodological lessons have been presented above in the context of explaining how ATP participants were linked to a patent history and various sources of financial data. In this section we discuss several remaining issues while presenting additional information on both participants and the data sets we have examined in trying to learn about them. We begin with a discussion of the possible units of analysis for ATP assessments and its implications for database construction. We next turn to a survey of ATP's existing data resources which get high marks for accuracy on what they do cover, although we have some suggestions for organization and reduction in reporting burden. Lastly, we compare and contrast the characteristics of participants as viewed at the project level, the establishment level, and the firm or organization level.

IV.A. Units of Observation in a Flexible Database

ATP data are for the most part organized around the specific projects which it has funded. Projects have different attributes, such as whether there is a single participant or a joint venture with multiple participants, the amount of funding, the project's beginning date and duration, whether the project was selected through the general or a focused competition. Since projects are funded to develop specific technologies, the project unit of analysis might be a good way of organizing data aimed at seeing whether the specific technology was achieved and then following where it is applied regardless of the particular participants which were involved in the project. On the other hand, ATP is unlikely to be a success unless it contributes to the success of those who participate in the projects. So we find it natural to look at the firms and other organizations that participate in ATP.

The Census Bureau generally analyzes firm data at two levels: the establishment level, and the firm or organization level. We follow their lead here. Establishments refer to a single geographic location, such as a building, where economic activity occurs. The identity of the establishment is maintained as long as the economic activity continues. For example, sale of a particular plant or store would not change the identity of the establishment so long as the same basic people and capital continued to do pretty much the same thing. Purchase of a hardware store which is liquidated and replaced with a video rental store would change the identity of the establishment, of course. Thus, the Census Bureau can analyze how changes in ownership affect the productivity of particular establishments. Unfortunately, there are few sources of establishment-level data other than in the highly confidential files of the Census Bureau. Nonetheless, we frequently see two or three different establishments belonging to the same firm listed separately as participants in the same or different ATP projects.

The firm or organization level of analysis focuses on all the activity of an economic organization regardless of where they are located. As a practical matter, one can define the firm as the reporting unit for tax or financial disclosures. For small firms all economic activity is usually located in a single establishment so that what is known about the firm can be directly related to that establishment. For larger firms, there is usually no way to disentangle the data referring to particular establishments from the aggregates reported for the entire organization. In some cases, sub-organizations such as subsidiaries or divisions may report some data separately, but this is not the rule. Of particular relevance to ATP, patents are generally but not always assigned to the firm (or parent firm) and not to individual subsidiaries or divisions.[16]

A flexible database should be constructed with identifiers that permit analysts to choose the level of analysis appropriate to the nature of their problem and the availability of data. Consider, for example, ATP participants. A participant is the organization awarded an ATP grant. Each project has at least one participant, some have several. For instance, when a joint venture project has seven members we count each member as a separate and unique ATP participant, regardless of whether any of those organizations have ever participated in the ATP before. Generally participants correspond to particular establishments and a single firm may have multiple participants in a particular project. Establishments accordingly are associated with particular projects during particular period and also with particular firms. Due to sales, mergers, and acquisitions the association of an establishment with a firm may change at particular dates. For project analysis, we can say that the seven participants in our example correspond to four firms or organizations. For analysis of firm success, we can instead look at every project in

[16] It is important for the database to permit an intermediate or sub-organizational level of reporting because there are instances where they are important such as firms with tracking stocks or unconsolidated subsidiaries. University systems are also empirically inconsistent. For example, most university patents are assigned to the particular

which any establishment of the firm participated. Further discussion of creation of unique parent organization identifiers and associated variant-to-preferred name lists are in Appendix A.1.

IV.B. ATP's Internal Data Resources

Because ATP data has been organized strictly at the project level, we found that records from different projects pertaining to a particular establishment or firm were frequently inconsistent. Reorganizing the data into a flexible database could simultaneously reduce errors and reporting burden by maintaining a single correct file for each establishment and firm.[17]

Information on participants in the ATP files are identified by project number and an identifying letter which distinguishes among the joint venture participants on a single project. We would supplement this be adding a field with the establishment identifier from our proposed master file of establishments and organizations. The only identifier useful for matching to external archive data is the Dun and Bradstreet code numbers. These Dun and Bradstreet codes are available for about 62 percent of the participants but only 54 percent of the participating organizations (see Table 20). We do not believe that Dun and Bradstreet data would be of sufficient consistency and quality to support rigorous empirical research and did not attempt to match the remaining participant establishments or organizations to that database.

Table 20: Participants and Organizations with an ATP Provided Dun Bradstreet Number

Unit of Analysis	Dun Bradstreet Number?	Frequency	Percent	Frequency
Participant	Yes	629	62.2	629
	No	382	37.8	1011
Parent Organization	Yes	349	53.8	349
	No	300	46.2	649

university campus, but all of the University of California patents are assigned to the Regents of the University of California [system], not to the individual campuses or laboratories.

[17] It is important to note that such a database should distinguish between corrections of data and changes that occur over time. Prior records should be retained for changes that occur over time so that a history is accumulated.

Table 21 is constructed -- using selected data at the participant level for establishments that are all part of the same firm, the 3M Company – to illustrate variations in coding of key variables. The 'subdivision', 'employee code', and 'SIC code' variables come from ATP's Business Reporting System (BRS). Subdivision is "Y" if the respondent reports the participating unit is a subdivision. Employee code is a self-reported total employment code, with 7 meaning greater than 1000 employees, 1 less then 20, and other values ranging between. SIC code is a single self-reported Standard Industry Code for the participating establishment. The variation in the values in this table suggests that BRS respondents are sometimes responding with information that is relevant to the corporate parent, and at other times with information that is relevant to the subdivision the respondent works with. This variation – possibly due to imprecision about the appropriate unit of analysis in the survey instrument – makes determining the proper attributes of ATP organizations problematic using only the internal information.

Table 21: Example of Variability of Coding for a Single Organization on Key Variables over Repeated Participations

Participant ID	Participant Organization Name	Location	Research Technology	Subdivision?	Employee code	SIC code*
94010305A	3M Company	3M Center. Building 224-2S-25. St. Paul, MN	Polymers	N	7	3081
94040027	3M Company Health Information Systems	3M Center. St. Paul, MN	Computer Software	N	5	7372
94040028A	3M Company Health Information Systems	12501 Prosperity Drive, Suite 150. Silver Spring, MD	Information/Computers/ Communication/Entertainment Systems	Y	5	8731
95030018A	3M Company	3M Center. Building 220-14E-11. St. Paul, MN	Storage--Magnetic	N	6	3572
95100025	3M Company	575 W. Murray Boulevard. Murray, UT	Computer Software	N	5	7372
95080006G	3M Company	3M Center. St. Paul, MN	Materials	N	7	265

* Documentation for the BRS describes the SIC code as the "4-digit DoC Census industry code of the participant's establishment."

Table 22 demonstrates that 3M is not unique in the variability in reports of public status, employment, and SIC code. For each variable, "Varies" is equal to "Yes" if one reported value for this parent organization on this variable is different from any other values reported for this parent organization on this variable. "Missing values" is equal to "Yes" if any of the repeated values of the variable in question are missing for this parent organization.[18] Such variability in employment and SIC code would be appropriate if the respondents were consistently reporting establishment values, but it appears from the actual values reported that in fact the respondents report a mixture of establishment and firm level data.

Table 22: Variability and Missing Data Rates for Organizations with Multiple ATP Participations*

			Frequency	Percent	Cumulative frequency
Public	Varies	No	164	97.6	164
		Yes	4	2.4	168**
	Missing values	No	NA	NA	NA
		Yes	NA	NA	NA
Employee	Varies	No	124	73.8	124
		Yes	44	26.2	168
	Missing values	No	53	31.5	53
		Yes	115	68.5	168
SIC code	Varies	No	121	72.0	121
		Yes	47	28.0	168
	Missing values	No	51	30.4	51
		Yes	117	69.6	168

* This table does not account for the geographic location of the reporting unit.
** The number of parent organizations with repeated participations reported here, 168, is different from the number reported in following tables, 164. The difference is that below a parent organization is counted as a multiple participant only if it is in more than one project. In this table a parent organization is counted as a multiple participant if there is more than one record for the parent. The additional four parent organizations for this table only participated in one project, but had two or more establishments participating in that single project simultaneously.

Perhaps these concerns are overdrawn. To get at the question of whether variations in reported values are correct responses for different establishments (or the same establishment at different times), we divided multiple-participant organizations by whether the organization has a single establishment appearing multiple times or instead has different establishments appearing in ATP. An organization with two or more ATP appearances that occur for the same location falls into the "same" category. An organization with two or more ATP appearances by two or more different locations falls into the "different" category. If the respondents are reporting establishment data consistently, then the variability of responses for the "same" group should be sharply less than for the "different" category. In Table 23 we see that the variability rates for the "same" category are indeed smaller than in the "different" category. This evidence is encouraging, but we hope that ATP analysts can look further at the issue of consistency on these questions.

Table 23: Variability and Missing Data Rates by Division or Establishment Variation for Organizations with Multiple ATP Participations

			same division or establishment		Different divisions or establishments	
			Frequency	Percent	Frequency	Percent
Employee	Varies	No	89	79.5	35	62.5
		Yes	23	20.5	21	37.5
	Missing values	No	47	42.0	6	10.7
		Yes	65	58.0	50	89.3
SIC code	Varies	No	86	76.8	35	62.5
		Yes	26	23.2	21	37.5
	Missing values	No	45	40.2	6	10.7
		Yes	67	59.8	50	89.3
N			112		56	

Taken as a whole, ATP's internal data resources appear to be of rather high quality and consistency, although perhaps under-documented from the point of view of outside users. The main issues are to develop variant-to-preferred establishment and organization lists so that particular firms and other organizations can be followed over time and to link those firms and organizations to the archival data available from external sources.

[18] The assessment of the missing data rate does not apply for this variable, since "Public" is either missing or equal to "Yes" and we assume that missing implies the firm is private,.

IV.C. Characteristics of Establishments and Organizations Participating in ATP

This section first reports on characteristics of ATP participants, conventionally defined by ATP as each instance of participation by an establishment in a project. We then present characteristics of the 649 unique firms and organizations that have participated in ATP. Our analysis here covers all ATP projects initiated during the period 1990 through January 1999.

Table 24 makes a side-by-side comparison for participants and organizations. Whether we use the conventional project-participant unit of analysis or the organization level of analysis, firms account for about 88 percent and universities and other non-businesses for about 12 percent of the units participating. However, moving from project-participant to organization level, we see the percentage of large firms fall from 28.1 to 14.5 while the percentage of small firms rises from 40.3 to 51.9 percent of all units participating. From an accounting view, large firms are more dispersed in location and may have different units participating in ATP, while that is precluded for small businesses with a single location. Furthermore, larger firms pursue many more different lines of research and are more likely than small firms to have repeated or multiple participations in ATP.

Table 24: Main R&D Project Participants and Participant Organizations by Organization Type

Organization Type	Project-Participant Level Data			Organization-Level Data		
	Frequency	Percent	Cumulative Frequency	Frequency	Percent	Cumulative Frequency
Federal Laboratory	8	0.8	8	6	0.9	6
Independent Research Organization	1	0.1	9	1	0.2	7
Large Business	284	28.1	293	94	14.5	101
Medium Business	203	20.1	496	139	21.4	240
Non Profit Organization	54	5.3	550	35	5.4	275
Small Business	407	40.3	957	337	51.9	612
University	54	5.3	1011	37	5.7	649

Characteristics of ATP Project Participants

While ATP project participants are conventionally defined to count each establishment in each project. Table 25 presents a broad approach to defining ATP participation. In this table we include all single participants and all JV participants listed in ATP award tables, as well as subcontractors reported by project participants as receiving more than $25,000 in project budget. There are 1722 project participants and subcontractors. Resource constraints prevented us from carrying out the extensive matching/cleaning procedures for the subcontractor observations, so we limit our attention to the 1086 project participants. In addition, 72 participants were either non-R&D members of consortia (e.g., project administrators) or participants in projects that were approved but never started. Excluding these 72, we identify 1,011 main R&D project participants.

Table 25: ATP Project Participants by Participation Type, 1990-January 1999

Participation Type	Frequency	Percent	Cumulative Frequency
Joint Venture	800	46.5	800
Single Applicant	286	16.6	1086**
Subcontractor *	636	36.9	1722

* Subcontractors reported as receiving more than $25,000 in project budget.
** 72 Participants were either non-R&D consortia members, or participants in projects that never started. Excluding these 72 participants leaves 1,011 participants. The descriptive statistics reported below refer only to these 1,011 participants.

Table 26: Main R&D Project Participants by Year

Project Year	Frequency	Percent	Cumulative Frequency
1990	57	5.6	57
1991	89	8.8	146
1992	31	3.1	177
1993	49	4.8	226
1994	196	19.4	422
1995	311	30.8	733
1996	9	0.9	742
1997	101	10.0	843
1998- Jan. 1999	168	16.6	1011

Table 26 reports number of project participants by year of project start for the 1,011 main R&D project participants. The specific project start dates vary throughout the year. The low number of participants in 1996 indicates the year Congress cut funding during debate over ending the program.

Table 27 reports project type for the 1,011 main R&D project participants. While single participants are not uncommon, 73 percent of project participations are in joint-venture projects.

Table 27: Main R&D Project Participants by Participation Type

Participation Type	Frequency	Percent	Cumulative Frequency
Joint Venture	735	72.7	735
Single Participant	276	27.3	1011

Table 28 reports program type for 1,004 of the 1,011 main R&D project participants. Data on 408 ATP projects was obtained from the ATP website. (The missing cases were not available on the ATP website at the time this information was obtained.)

Table 28: Main R&D Program Participants by Program Type *

Program Type	Frequency	Percent	Cumulative Frequency
Focused Competition	605	60.3	605
General Competition	399	39.7	1004

* Frequency missing = 7

Table 29 reports the number of main R&D project participants (out of the 1,011) for the 20 states with the highest frequency of participation. California seems to have done very well in the competition, but on a per capita basis Michigan, Massachusetts, and Delaware are the standouts. The extent of participation by Massachusetts may be partially explained by the large number of scientists and engineers in the state, but Michigan and Delaware have also done very well relative to their science/engineering employment base. Florida, North Carolina, and

Washington, on the other hand, have had less success in ATP competitions whether on a per capita basis or in terms of science and engineering employment.

Table 29: Number of Main R&D Project Participants for Twenty Most Frequent States

State	Frequency	Percent	Projects per million state residents*	Projects per million state scientists and engineers*
California	193	19.1	6.15	27.23
Michigan	129	12.8	13.47	83.22
Massachusetts	85	8.4	14.10	47.28
Texas	67	6.6	3.64	21.45
New York	62	6.1	3.42	20.63
Ohio	57	5.6	5.14	35.46
New Jersey	45	4.5	5.69	25.22
Pennsylvania	43	4.3	3.57	24.15
Illinois	33	3.3	2.81	18.17
Connecticut	32	3.2	9.79	37.81
Minnesota	29	2.9	6.35	36.44
Florida	15	1.5	1.07	7.40
Colorado	14	1.4	3.83	14.63
Georgia	13	1.3	1.84	12.94
Oregon	13	1.3	4.21	29.29
Wisconsin	11	1.1	2.17	16.58
North Carolina	10	1.0	1.42	9.92
Utah	10	1.0	5.19	30.02
Delaware	9	0.9	12.74	66.53
Washington	8	0.8	1.50	6.43

* Population and scientists and engineers figures are Census estimates for 1994.

Characteristics of ATP Participant Organizations

After linking all 1,011 project participations to the firm or organization of which the participating establishment is a part, we are left with 649 unique (main R&D) participant organizations. Table 30 presents descriptive information for the participant organizations. (We were not able to obtain information for seven of the 649 participant organizations at the time this report was assembled.) Project count is the number of projects (projects where the parent organization has multiple participating establishments are counted only once). Project years is the cumulative project duration years for all projects the parent organization has participated in.

Total award and total contribution are the sum of award and organization contribution amounts, respectively, over all project participations by the parent organization.

Table 30: Selected Descriptive Statistics for ATP Main R&D Participant Organizations

Variable	N	Mean	Std. Deviation
Project Count	649	1.51	1.48
Project Years	642	5.53	6.52
Total Award	642	1984462.84	2496639.00
Total Contribution	642	1969547.19	2962839.00

Table 31 contains variables of interest from ATP databases. An interesting implication of the last row of this table in conjunction with the mean project counts in Table 30 is that although only one quarter of participant organizations participated in more than 1 ATP project, that quarter on average participated in 3.0 projects.

Table 31: Selected Frequencies for ATP Main R&D Participant Organizations

Variable	Value	Frequency	Percent
Public firm?	No	440	67.8
	Yes	209	32.2
Ever in a Joint Venture?	No	180	27.7
	Yes	469	72.3
Ever in Joint Venture with a University?	No	424	65.3
	Yes	225	34.7
Ever in a Project with University Subcontractors?	No	466	71.9
	Yes	182	28.1
Ever Participation by a Subsidiary?	No	595	91.7
	Yes	54	8.3
Ever in a Focused Program?	No	198	30.5
	Yes	451	69.5
Project Participation Count	One project only	484	74.6
	More than one project	165	25.4

* Total unique organization count is 649

Table 32 reports the number of unique ATP participant organizations beginning a project each year and the number that are beginning their first ATP project. Notice that the fraction of new participant organizations in a given year has gradually declined to around two thirds as the stock of previous participants has accumulated.

Table 32: Total Participant Organization Counts and New Participant Organization Counts by Year

Project year	Total unique organizations this year	Total unique organizations participating in first ATP	Percentage of new ATP organizations
1990	35	35	100.0%
1991	82	69	84.1%
1992	30	25	83.3%
1993	45	27	60.0%
1994	163	139	85.3%
1995	248	189	76.2%
1996	9	4	44.4%
1997	97	61	62.9%
1998- Jan. 1999	151	100	66.2%

Table 33 presents the percentage of new organizations that come to participate in ATP for the first time through the general competition. The data indicate that in the years where focused programs were active, many new entrants to ATP came through focused programs.

Table 33: New ATP Participant Organizations by Year and Program

Project year	General competition	Focused competition	Percentage of new organizations entering through general competition
1990	30	5	85.7%
1991	69	0	100.0%
1992	25	0	100.0%
1993	27	0	100.0%
1994	48	91	34.5%
1995	18	171	9.5%
1996	4	0	100.0%
1997	19	42	31.1%
1998- Jan. 1999	23	77	23.0%
Total	263	386	40.5%

V. Conclusions

This paper is of the nature of a series of snapshots of a work in progress. We report here on what are the central insights and methods we have been using in our efforts to prove the feasibility and usefulness of linking ATP's current internal data resources with a variety of external archival data sets created for very different purposes. By the time this paper was drafted that purpose had largely been achieved: We uncovered a sharp increase in the patenting rate of participants after they began to participate in ATP. We also showed that it was feasible to link firms accounting for the bulk of patents, employment, and sales by ATP participant firms to Compustat and, hence, a variety of databases with accounting and other data on publicly traded firms. Further, we showed that it was possible to identify when and on what terms private-firm participants received venture capital, entered into joint ventures or strategic alliances, and ultimately went public.

Although it is feasible and valuable to link ATP participants to particular entities for which archival data exists, the process can be difficult and tedious. The issues range from simple name spelling errors to organizational name changes to different levels of reporting for different purposes. The last issue has proven to be a substantial one, especially for patenting.

In work subsequent to drafting this paper, we are using the SDC Mergers and Acquisitions database to enhance our ability to identify name changes and recombinations of participant establishments. M&A activity complicates but does not fundamentally alter the task of building a panel analysis data set for statistical estimation and hypothesis testing.

Methodology Appendix

This appendix expands on discussions of several topics of interest to the research analyst.

A.1. Creation of Unique Parent Organization Identifiers

In order to link particular establishments to the larger organization of which the are a part, we associate a parent organization identifier which we call *pcode* with each participant record. Thus we can use *pcode* to identify when two participants have the same parent entity. Inserting these codes requires substantial hand-correction by research assistants to deal with instances where the organization has different, but equally valid names.

Table 34 reports ATP participation viewed at the level of unique parent organizations that were main R&D participants as defined in the text. The BEFORE calculation of unique organizations takes the names as given to us by ATP, and considers participants to be the same organization only when the names are exactly identical. This step is represented in Figure 2 as movement from the box with 1,011 main R&D project participants to the box with 804 unique organizations. The AFTER category calculates unique organizations based on cleaned names and correction of the organizational identifier to allow for different but equivalent organization names. The reduced count of 649 unique organizations is due to two factors. First is the case where a division was identified as the ATP participating organization. Our cleaning parsed out division name from organization. Second is the case where an organization has multiple equivalent names. IBM and International Business Machine is an example of this case. For these instances we corrected the organizational identifier to indicate that in fact this is the same organization.

Figure 2. Identifying Unique Organizations among ATP Participants

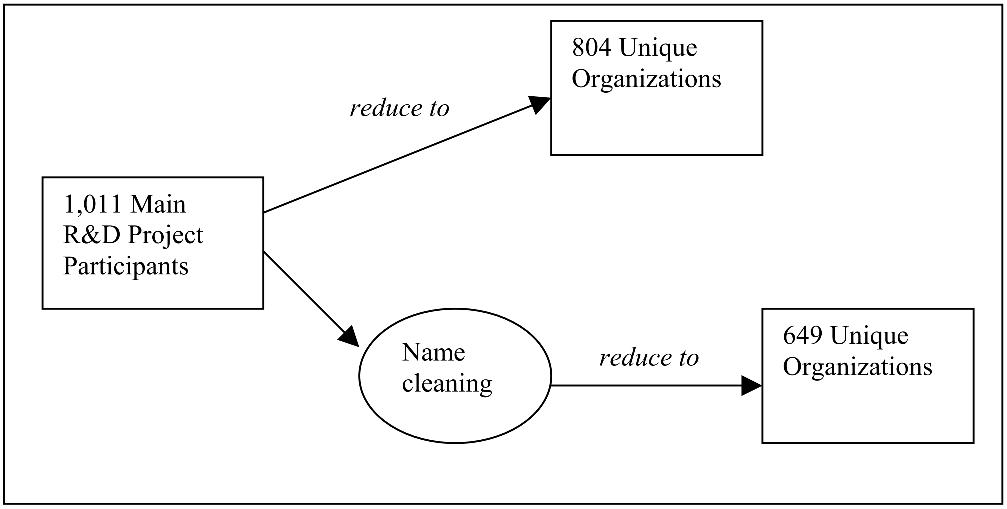


Table 34: Comparison of Counts of Unique Main R&D Parent Organizations Before and After Cleaning.

Parent Organization Type	Organization Count BEFORE Cleaning and Parent Organization Identifier Correction	Organization Count AFTER Cleaning and Parent Organization Identifier Correction
Federal Laboratory	7	6
Independent Research Organization	1	1
Large Business	191	94
Medium Business	158	139
Non Profit Organization	41	35
Small Business	353	337
University	53	37
Total Unique Organizations	804	649

A.2. Filtering and Matching Algorithms

Filtering and matching algorithms are used to enable the computer to find equivalencies which it would otherwise miss. Filtering means removing irrelevant things from names such as “the” or “inc.” which are more or less arbitrarily included or excluded depending on who is entering the data and converting the remainder to all capitals. Matching looks for exact matches to any of the filtered ATP organization names – the objective is to retrieve identifiers used in analytical data creation. Table 35 gives an example of applying filtering and matching to variant names for the Ford Motor Company. Filtering and matching is an iterative process, the goal of which is to identify systematic things that reduce the effectiveness of computer matching.

Table 35: Examples of Name Filtering

ATP Name	Filters to	Filtered Name	Filters to	Archival Name
Ford Motor Company	→	FORD MOTOR	←	Ford Motor Co
Ford Motor Company	→	FORD MOTOR	←	FORD MOTOR
Ford Motor Company	→	FORD MOTOR	←	The Ford Motor Company
Ford Motor Company	→	FORD MOTOR	←	Ford Motor

Approximate matching is used to identify cases that require a research assistant's (RA) attention. The RA researches non-matches and adds additional firm name information from our data – with this new information the process is re-run. In Table 35, although all the firm names in the ATP column and the archival column are readily identifiable as the same company by a human reader, when strictly evaluated by the computer the four pairs of names are not equal. As we are dealing with very large numbers of firms we have to rely on the computer for firm identification. The filtering mechanism standardizes names so that electronic matching is feasible.

A.3. Cleaning Organization Names

The level of organization reported as an ATP participant may or may not differ from the level of organization at which important, relevant archival data are available. Therefore, it is important to create a new field "parent organization" which may be a real parent organization or may simply repeat the participant (establishment) name for smaller unitary organizations.[19]

[19] It is sometimes appropriate similarly to identify sub-organizations for which relevant information is reported as an intermediate category between establishments and parent organizations. In the work reported here, we have relied on defining a preferred parent organization name and maintaining a list of variant to preferred names that match establishments, sub-organizations, and alternative names to the parent organization in one pass.

The first step in this process is to standardize names by removing terms indicating subdivision, acronyms, and abbreviations. We save during this process all known alternative names for the participants, their parent, and sister sub-units. This establishes a variant-to-preferred lexicon of the ways in which the firm and its components are referred to in practice.

The first round of this process can be computerized, but ultimately RAs need to examine problem organization names. Some of the participant names in ATP's database which we found most challenging are reported below in Table 36.

Table 36: Examples of Problem Organization Names

Participant ID	Participant Organization Name	Participation Type	City	State
94020039B	Advance USA	Main	Old Lyme	CT
94010382S2	APD	Subcontractor	Bristol	PA
97030061A	CHIME	Main	Wallingford	CT
98030027C	JME	Main	Shaker Heights	OH
95040026S1	M.A.D.S.	Subcontractor	Geoffstown	NH
95020009S1	Management & Eng. Tech	Subcontractor	Dayton	OH
97020028S5	MSC	Subcontractor	Oneida	TN

Certain common practices make it difficult to know the official name of the participant or its parent:

1. Abbreviations of names. Some are trivial, such as U for University, Ctr for Center, etc. Some are not so simple.
2. Acronyms used as company names. Some of these are intuitive, i.e. IBM or 3M. Some are not, and do not correspond to particular yellow page listings.
3. Incomplete records. Some records have incomplete, ambiguous or unspecified names.

References

Acs, Zoltan J., and David B. Audretsch, "Innovation in Large and Small Firms: An Empirical Analysis," *American Economic Review*, September 1988, 78: 678-690.

Aghion, Philippe, and Jean Tirole, "The Management of Innovation," *The Quarterly Journal of Economics*, November 1994, 109(4): 1185-1209.

Darby, Michael R., Lynne G. Zucker, and Andrew Wang, "Universities, Joint Ventures, and Success in the Advanced Technology Program," National Bureau of Economic Research Working Paper No. 9463, January 2003.

Griliches, Zvi, "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, December 1990, 28:1661-1707.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Tratjenberg. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," National Bureau of Economic Research Working Paper 8498, October 2001. [These data are now at http://www.nber.org/patents/.]

Zucker, Lynne G., Michael R. Darby, and Jeff Armstrong, "Geographically Localized Knowledge: Spillovers or Markets?" *Economic Inquiry*, January 1998. 36(1): 65-86.