Using Database Linkages to Measure Innovation, Commercialization, and Survival of Small Businesses

James Onken, Andrew C. Miklos, Travis F. Dorsey, Richard Aragon, Anna Maria Calcagno

Abstract

Here, we report the results of an outcomes evaluation of the Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) Programs at the National Institute of General Medical Sciences (NIGMS). Since the programs’ inception, assessments of the SBIR/STTR programs at several federal agencies have utilized surveys of former grantees as the primary source of data. Response rates have typically been low, making non-response bias a potential threat to the validity of some of these studies’ results. Meanwhile, the availability of large publicly-available datasets continues to grow and methods of text mining and linking databases continue to improve. By linking NIGMS grant funding records, U.S. Patent and Trademark Office data, and business intelligence databases, we explored innovation, commercialization and survival for recipients of NIGMS SBIR/STTR funding. In doing so, we were able to more completely assess several key outcomes of the NIGMS SBIR/STTR program. Our evaluation demonstrated that the NIGMS program performed above baseline expectations along all dimensions, and comparably to other federal agency SBIR/STTR grant programs. In addition, we show that the use of extant data increasingly is a viable, less expensive, and more reliable approach to gathering data for evaluation studies.

Keywords

Innovation; Database; Research evaluation; Research policy; Small Business Innovation Research; U.S. federal government

Publisher’s Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declarations of interest: None
1.0 Introduction

The National Institute of General Medical Sciences (NIGMS) supports the fourth-largest Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) grant programs at the National Institutes of Health (NIH). As part of its role in ensuring effective investment of taxpayer resources, NIGMS conducted an evaluation of its SBIR/STTR grants since inception of the program. This retrospective evaluation, the first that the Institute has conducted on these programs, focused on the role that the SBIR/STTR program had in promoting innovation, commercialization, and survival for small businesses that participated.

The SBIR program was established by Congress in 1982 and the STTR program was established (along with an SBIR reauthorization) in 1992 to strengthen the role of small, innovative firms in federally-supported research and development (R&D). The overarching mission of the SBIR/STTR programs is “to support scientific excellence and technological innovation through the investment of federal research funds in critical American priorities to build a strong national economy.” Among the major goals established in pursuit of this mission are stimulating technological innovation and increasing private-sector commercialization of innovations derived from federal R&D (U.S. Small Business Administration, 2017a). In the 1992 reauthorization of the SBIR program, Congress emphasized the program’s goal of commercialization, increasing the priority of this goal (Archibald & Finifter, 2000).

Both SBIR and STTR are three-phase programs. In Phase I, a six-month (SBIR) to one-year (STTR) grant or contract is awarded to a small business (and any partners) to establish technical merit, feasibility, and commercial potential of the proposed R&D effort. If Phase I is successful, up to two years of additional funding may be provided in Phase II to continue the effort begun in Phase I. Phase III, not funded by the SBIR/STTR program, is for the small business to pursue commercialization of the results of Phases I and II.

The SBIR and STTR programs are funded through set-asides of each federal agency’s extramural research budget. The NIH has the largest SBIR/STTR program among civilian agencies (U.S. Small Business Administration, 2017b). Of the 27 institutes and centers that make up the NIH, NIGMS has the fourth-largest SBIR/STTR grant program (NIGMS does not fund SBIR/STTR contracts) (National Institutes of Health, 2017). The size of the set-asides used to fund the SBIR/STTR programs has grown with each Congressional

---

1 Abbreviations used in this article:
D&B: Dun and Bradstreet
DUNS: Data Universal Numbering System
ExPORTER: Exported Portfolio Online Reporting Tools: Expenditures and Results
IMPAC: Information for Management Planning Analysis and Coordination
NASA: National Aeronautics and Space Administration
NIGMS: National Institute of General Medical Sciences
NIH: National Institutes of Health
PARDI: Portfolio Analysis and Reporting Data Infrastructure
RePORTER: Research Portfolio Online Reporting Tools: Expenditures and Results
SBA: Small Business Administration
SBIR: Small Business Innovation Research
STTR: Small Business Technology Transfer
USPTO: United States Patent and Trademark Office
reauthorization of the programs. What began in 1983 as a small $639,000 program at NIGMS had increased to become a major $82.5 million investment by the Institute by 2017. From the programs’ inception through fiscal year 2017, NIGMS SBIR/STTR funding totaled more than $1 billion.\(^2\) Along with substantial growth in the program has come increased interest in evaluating what the return on this investment has been.

As noted in previous studies of the SBIR/STTR Programs, establishing clear evidence of successful outcomes is challenging (National Research Council Committee for Capitalizing on Science, Technology, and Innovation, 2009a). There has been little in the way of systematic collection of outcomes attributable to SBIR/STTR grants and contracts. As a result, most studies have relied on proxy measures such as successful transition to Phase II funding (National Academies of Sciences, Engineering, and Medicine, 2016a), success in obtaining Phase III funding (National Research Council Committee for Capitalizing on Science, Technology, and Innovation, 2009b), and patents applied for and received that acknowledge support from SBIR/STTR funding (Edison, 2014). The primary source of such information has been surveys of past recipients of SBIR/STTR funding, which in most cases had very low response rates (Onken, Aragon & Calcagno, 2019; National Institutes of Health, Office of Extramural Research, 2003).

An alternative to surveys that offers the potential for more comprehensive and reliable information on program outcomes is the mining of existing publicly available datasets containing identifying information that facilitates linkages among records from disparate sources. Linked datasets, whose increasing use has been fueled by the greater availability of public datasets and improvements in our ability to merge them, are a critical component of current efforts to strengthen evidence-based decision making in both the private sector and the federal government (Commission on Evidence-Based Policymaking, 2017; Morrel-Samuels, Francis, & Shucard, 2009). In this paper, instead of surveys we rely on existing databases as the primary source of information to examine the outcomes of NIGMS SBIR/STTR funding. To assess the returns on NIGMS’s investment, we use several measures of innovation, commercialization, and survival including awarded patents, forward patent citations, and firm status. Each of these indicators is described in more detail in section 2.0, and comparisons to other SBIR/STTR and small business data are present in section 3.0.

### 2.0 Indicators and Data Sources

#### 2.1 NIH Grants

Records of NIGMS support for the SBIR and STTR programs were drawn from NIH’s Information for Management Planning Analysis and Coordination (IMPAC) II database, an internal database of grant applications and awards maintained by NIH’s Office of Electronic Research Administration. The IMPAC II database provided information on grantee firms and their Phase I and Phase II funding. While we used an internal NIH database as our source data, a public version of the database is available (National Institutes of Health, 2018a; National Institutes of Health, 2018b), along with additional data we retrieved from NIH’s Research Portfolio Online Reporting Tools website (National Institutes of Health, 2018c).

\(^2\)These figures include funds for new Phase I and Phase II awards as well as noncompeting grant renewals.

*Eval Program Plann.* Author manuscript; available in PMC 2020 December 01.
2.2 Patent Awards

An invention patent is often an early step in the path to commercialization of a product and the ready availability of patent data has contributed to their widespread use to measure innovation. Previous research has shown patent activity to be an adequate surrogate for more nuanced measures of innovation (Arcs & Audretsch, 1989), and patents are frequently a good predictor of technological and economic performance measured in other ways (de Rassenfosse & van Pottelsberge de la Potterie, 2009; Griliches, 1998; Hagedoorn & Cloots, 2003; Keller & Holland, 1982).

To assess the patent activity of SBIR/STTR grantees, we relied on two sources:

1. Acknowledgments of federal grant support are found in the Government Interests section of awarded patents, available in the United States Patent and Trademark Office (USPTO) Full-Text and Image Database (U.S. Patent and Trademark Office, 2016a). Text mining was used to search the Government Interests section of awarded patents for references to NIGMS SBIR/STTR support and extract grant numbers that were then matched against NIH grant records.

2. Awarded patents reported by NIGMS SBIR/STTR grantees through the federal Interagency Edison (iEdison) reporting system. iEdison helps federal government grantees and contractors comply with the Bayh-Dole Act regulations requiring government-funded inventions be reported to the federal agency that made the grant or contract award. NIH makes its iEdison records for patent awards publicly available in its RePORTER database of grant awards (National Institutes of Health, 2018a) and as a bulk download through ExPORTER (National Institutes of Health, 2018b).

We also obtained ancillary information on NIGMS-supported patents from the USPTO’s PatentsView database (U.S. Patent and Trademark Office, 2016b). PatentsView provided useful information such as disambiguated patent assignees.

Our data were limited to awarded patents only. Including patent applications would have provided a more complete picture of the patenting activities of former grantees. However, we were unsuccessful in finding a method to reliably identify patent applications resulting from NIGMS funding. Even in applications for the awarded patents in our data, funding acknowledgments were often missing.

2.3 Downstream Patents

A second-order effect of SBIR/STTR funding can be observed in the “forward citation” of patents: the citation of SBIR/STTR-funded patents in subsequent downstream patent applications. While the number of forward citations a patent receives has been used in many previous studies to measure economic value or an invention’s impact, several investigators have noted the limitations of this measure (Jaffe & de Rassenfosse, 2017). For example, earlier patents are cited for a variety of reasons, making their interpretation ambiguous (e.g., unlike the USPTO, the European Patent Office categorizes their patent citations, including “X” and “Y” citations used to show lack of innovation). Also, Gambardella, Harhoff and
Verspagen (2008) found that forward citations account for a small percentage of variance in the ultimate economic value of patents.

While there are limitations, Gambardella et al. also found that patent citations correlated significantly with economic value and was the best predictor among several alternatives they examined. Previous research has found forward citations to be related to a variety of long-term outcomes (Lanjouw & Schankerman, 1999) including commercial success as measured by stock market value (Trajtenberg, 1990) and sales in the pharmaceutical industry (Guo, Hu, Zheng, & Wang, 2013). A summary of several patent citation validation studies can be found in Jaffe & de Rassenfosse (2017).

In many of these studies, a distinction is made between a self-citation (citation of a patent held by the same firm) and a citation in other firms’ patents. Self-citations may have a different relationship to the private value of a patent or the broader social impact of an invention than citations by other firms (for example, see Hall, Jaffe, and Trajtenberg, 2005). For this reason, in this study we exclude self-citations in analyses of forward citations.

We obtained information on forward citations of patents using the NIH’s Portfolio Analysis and Reporting Data Infrastructure (PARDI), a non-public NIH database that combines NIH IMPAC II grants data with publication and patent records, including the Clarivate Analytics Derwent World Patents Index®.

### 2.4 Business Survival (Mergers, Acquisitions, and Bankruptcies)

The survival of an SBIR/STTR grantee as a “going concern” provides indirect evidence of the firm’s success in commercializing its products and services. The merger or acquisition of a former grantee also is an indication that the firm was successful in developing intellectual property or products of value. We used both publicly available and proprietary databases to determine each NIGMS SBIR/STTR-supported firm’s current status: active, merged (or acquired), or inactive. Three sources were used for firm status: OpenCorporates (OpenCorporates, 2018), Crunchbase (Crunchbase, 2018), and a unique dataset purchased from Dun & Bradstreet (D&B; Dun & Bradstreet, 2018). OpenCorporates is a database aggregator that combines information about companies from a variety of sources including government websites and application programming interfaces, publicly available datasets, and through Freedom of Information requests. Crunchbase is a crowd-sourced service that provides information on companies, investments, industry trends, and news about public and private companies. D&B is a well-known provider of commercial data and analytics.

A combination of firm name, address, key personnel, and DUNS numbers were used to ensure accurate matches among these data sources and the NIH grant records. In ambiguous cases (e.g., firms with the same name, missing DUNS numbers, and different addresses), in-depth manual internet searches and reviews of NIH grant files were performed in an attempt to create accurate matches.

Business survival is likely to be a function of how long ago the firm had first received NIGMS support (a potential surrogate for firm age) so, for analyses of business status, firms were stratified into 5-year cohorts based on their first year of NIGMS support.
3.0 Results

3.1 Transitions to Phase II Award

A successful transition to Phase II funding is the first step toward commercializing the results of SBIR/STTR-funded R&D. Funding for Phase II is contingent on Phase I results that demonstrate technical feasibility and commercialization potential. Of the 1,925 Phase I projects funded through fiscal year 2012, 699 (36.2 percent) had made a successful transition to Phase II funding by fiscal year 2015. The performance benchmark established by the Small Business Administration (SBA) for transition rate is 25 percent (U.S. Small Business Administration, 2017c). When considering the top 25 businesses receiving the highest number of NIGMS Phase II awards, the transition rate was 68.9 percent. While the differences in fields of research, funding mechanisms, and other characteristics make any direct comparisons challenging, a similar analysis of National Aeronautics and Space Administration (NASA) SBIR/STTR grantees from 2005 to 2012 found a transition rate of 56.4 percent (National Academies of Sciences, Engineering, and Medicine, 2016a).

3.1 Patent Awards

Figure 1 shows the number of patents citing one or more NIGMS SBIR/STTR grants in the USPTO and iEdison databases. Text mining of the Government Interests section of awarded patents identified 299 patents citing NIGMS SBIR/STTR grant support. The iEdison database also contained 141 patents citing support from these programs. A total of 371 different patents were reported in at least one of these databases. However, there have been no strong enforcement mechanisms available to ensure that federal support is acknowledged in either the Government Interests section of patents or in iEdison. As a result, many patents reported through iEdison did not contain a funding acknowledgement in the corresponding patent’s Government Interests section. Likewise, many references to NIGMS grants in Government Interests sections were not found in iEdison.

Underreporting occurs in both sources, suggesting that there is an even larger universe of patents supported by the NIGMS SBIR/STTR program that includes patents reported in neither database. Among patents known to have been supported by an NIGMS SBIR/STTR grant (i.e., those officially reported through iEdison), only 49 percent contained a grant citation in the Government Interests section of the patent. Assuming that this underreporting of 51 percent is similar among patents not reported to iEdison, the 299 patents found in USPTO represent a subset of a larger universe of at least 610 patents supported by the NIGMS SBIR/STTR program (299 × 100/51 = 610).3 Of this estimated universe, the 371 patents identified through either source represent a sample of 61 percent.

Figure 2 shows that the cumulative percentage of SBIR/STTR grants awarded through fiscal year 2008 that have resulted in one or more patents is about 9 percent (more recent cohorts of Phase I awards would not have had time to generate patents by 2015). Note that this includes only known patents. If the known patents represent 61 percent of a larger universe

---

3Note that this universe will be even larger if the underreporting of Government Interests is even greater among those patents not reported through iEdison. It is not unreasonable to assume that patents not reported through iEdison are even more likely to fail to acknowledge Government Interests in the patent application.
of patents, we might extrapolate that approximately 15 percent of SBIR/STTR grants have produced at least one patent.

Patent activity is significantly higher for projects that successfully transitioned to a Phase II award relative to other Phase I awardees in the same year (Figure 3). Logistic regression showed that a Phase II transition increased the odds of obtaining a patent by a factor of 5.35 over the odds of patenting with only Phase I support. Approximately 18 percent of Phase II awards received an acknowledgment in one or more patents, exceeding the SBIR performance benchmark of 15 percent (U.S. Small Business Administration, 2017c). If corrected for underreporting found in the USPTO and iEdison databases, the true percentage may be as high as 30 percent.

Previous evaluations that used surveys to analyze patent activity found that “between 35 and 45 percent of all companies with SBIR awards… developed sufficient technical knowledge to be worth the time and expense of a patent application (and award) …” (National Research Council, 2009). Four recent surveys of the Phase II SBIR/STTR programs at the NIH (National Academies of Sciences, Engineering, and Medicine, 2015), NASA (National Academies of Sciences, Engineering, and Medicine, 2016a), the Department of Energy (National Academies of Sciences, Engineering, and Medicine, 2016b), and the Department of Defense (National Research Council, 2014) found an average of 42.1 percent of Phase II projects resulted in at least one patent. Considering potential bias due to item non-response in these surveys, we adjusted these findings assuming that non-responsive businesses have not produced a patent. When this adjustment was performed, the average patenting rate for Phase II awards was estimated to be 27.5 percent in these previous studies.

The cumulative number of known patents generated by the NIGMS SBIR/STTR programs is compared to the cumulative program costs in Figure 4. Figure 5 shows that the average SBIR/STTR investment leveraged in patent generation has remained constant at about $3 million dollars per patent for the past decade, reaching a minimum of $2.6 million in 2015 (unadjusted for underreporting). However, when plotted as constant 1983 dollars the funding per patent has been decreasing since 1998. If our estimate of the true number of patents resulting from NIGMS SBIR/STTR funding (after adjusting for underreporting) were accurate, the average investment per patent would be only $1.6 million.

The average NIGMS investment for each patent is distinct from the average total R&D cost of a patented invention, as SBIR/STTR firms use other sources of funding for R&D activities not captured in this analysis. These data simply allow comparisons between the level of programmatic input (the NIGMS funding) to one measure of R&D output (patents). For example, on a dollar-for-dollar basis, the NIGMS SBIR/STTR programs have generated approximately twice the number of patents as NIGMS’s other research grant programs. Whereas the SBIR/STTR programs currently represent about 3.1 percent of the NIGMS research grant budget (and an even smaller percentage in earlier years), these programs have generated 6.0 percent of the patents that we were able to link to NIGMS support.

Comparisons to other published per-patent R&D costs are difficult for several other reasons. There are few such studies— mostly single-sector studies or reports by industrial
associations (Neuhaeusler, Frietsch, Mund & Eckl, 2017)—and estimates vary greatly by technology field. For example, the National Science Foundation’s National Center for Science and Engineering Statistics estimated patent costs in 2008 ranged from $3.2 million in the medical equipment and supplies industry, to $6.4 million in scientific research and development services, to $20.5 million for pharmaceuticals and medicines. For all small businesses (regardless of technology sector), R&D spending per patent was $4.5 million (Shackelford, 2013). Based on annual SV Wissenschaftsstatistik (German R&D survey data) collected on behalf of the German Federal Ministry of Education and Research, Neuhaeusler et al. calculated 2009 business R&D expenditures per patent in Germany for a variety of technological fields. In the fields of research typically supported by NIGMS funding, per patent costs in Germany were (in 2009 U.S. dollars, ascending order): medical technology ($866,000), analysis of biological materials ($1.1 million), chemical engineering ($2.0 million), organic fine chemistry ($2.1 million), biotechnology ($6.4 million), and pharmaceuticals ($20.1 million) (Neuhaeusler, et al., 2017). The averaged patent costs in these six fields is $5.4 million invested per patent.

3.2. Downstream Patents

Table 1 shows the number of SBIR- and STTR-supported patents that were cited by downstream patents. When self-citations (citations in which the patent assignees of the citing and cited patents are the same) are removed, 69.3 percent of SBIR- and STTR-generated patents were cited by an average of 19.9 subsequent patents. (Note that self-citations have not been corrected for any transfer of ownership of patents over time. This may have resulted in an underestimate of the amount of self-citation.) A recent report of patent activity associated with NIH-sponsored research grants found that the average NIH-supported patent awarded in 2006 through 2008 was cited an average of 13.9 times and that all NIGMS-supported patents (which would have included the SBIR/STTR patents) were cited an average of 14.6 times (Kalutkiewicz & Ehman, 2017).

The 5,114 citations of SBIR/STTR-supported patents shown in Table 1 is a duplicated count that includes instances of multiple SBIR/STTR-supported patents by the same citing patent. When these duplicates were removed, there were 3,776 unique patents that cited one or more SBIR/STTR supported patents, resulting in an average of 11 unique downstream patents for each SBIR/STTR patent.

3.3 Business Survival

In fiscal years 1983–2015, NIGMS supported a total of 1,196 different companies. Table 2 shows the number of these that were found in the three business data sources we used. Firm status was available for 893 (75 percent) in OpenCorporates, 298 (25 percent) in Crunchbase, and 1,108 (93 percent) in D&B. Status was available for 1,160 firms (97 percent) in at least one of these data sources, and status was available from at least two sources for 900 firms (75 percent). The firms for which information was available from multiple sources were recipients of 85 percent of all NIGMS SBIR/STTR funding.

For the 900 firms found in at least two of the business data sources, concordant information was found for 592 of the firms (66 percent), which we considered “confirmed” statuses.4

4 Onken et al. Page 8 Eval Program Plann. Author manuscript; available in PMC 2020 December 01.
Extensive internet research was performed on the remaining firms to resolve discrepancies in
the statuses reported from the three data sources and locate businesses for which no
information was available. We were able to confirm the status of an additional 344 firms,
resulting in a database of 936 firms (78 percent of all grantee firms) with a confirmed status.

For the 224 firms reported by at least one of the business data sources and whose status
could not be confirmed through internet research, we assumed their status distribution to be
the same as the distribution of firms with confirmed status, conditioned on receiving NIGMS
funding in the same 5-year period and having the same reported status from the three
databases.

Of the 36 firms not found in any of the three data sources, status couldn’t be resolved for 30
firms. These were assumed to be inactive businesses.

Combining the confirmed and imputed statuses, a total of 756 firms (63 percent) were
estimated to be active, 218 (18 percent) had merged or been acquired, 189 (16 percent) were
no longer active, and status remained unknown for 33 (3 percent). Figure 6 shows the status
for each 5-year cohort of firms based on their first year of NIGMS support. Among the
oldest grantees, about 40 percent are estimated to be active, 25 percent have merged or been
acquired, and 35 percent are inactive.

Survival curves of active firms and those active, acquired, or merged with another firm, are
shown in Figure 7 and compared to a survival curve of all small businesses in the U.S.
provided by the SBA (U.S. Small Business Administration, 2014). The survival of small
businesses follows an exponential decay curve and similar exponential decay curves were fit
to the NIGMS grantee data. Additional data from the Bureau of Labor Statistics suggest that
survival in the professional, scientific, and technical services industry does not differ
markedly from the survival rate of all private firms (Bureau of Labor Statistics, 2017).

4.0 Discussion

The results of this study included several measures of innovation, commercialization, and
survival of businesses that were recipients of funding from NIGMS. In this discussion, we
focus on the holistic findings as they relate to the evaluation of the NIGMS SBIR/STTR
portfolio.

Innovation

As mentioned in Section 2.2, the generation of patents from SBIR/STTR grantees was the
primary proxy used as a measure of innovation. Results suggested that the SBIR/STTR
portfolio had a higher patenting rate than that of other grants administered by NIGMS. In the
case of Phase II grant recipients, patenting activity exceeded SBA benchmarks, suggesting
that the program is meeting its targets. However, when comparing these rates against those

Lack of concordance was due in large part to ambiguities created by multiple records in the OC database for firms having different
statuses in multiple U.S. jurisdictions.

Confirmatory information included press releases, news articles, and what appeared to be active corporate websites with dates.
from evaluations of other SBIR/STTR programs, our analysis had lower unadjusted values. This discrepancy highlights a key difference between our approach and survey-based approaches.

Non-response bias can result in an overestimation of patenting activity on surveys, as businesses may be less likely to respond to survey questions related to patenting if they have not demonstrated success in patenting. The use of extant data allows for full coverage of all firms, but similarly suffers from missing data. In our case, multiple overlapping sources of patent data allowed for an estimate of true values that adjust for missing data in each source. Adjusting the reported outcomes for item non-response (in surveys) and missing data (in patent databases) resulted in similar estimates of patenting rates, suggesting that: a) there is concurrent validity in the extant data used in this study, and b) the NIGMS SBIR/STTR portfolio performs similarly to other agency programs.

Our approach to identifying patents is a more conservative one, designed to ensure that the attribution to NIGMS funding occurs only when NIGMS has been directly acknowledged by the inventor(s). While Bayh-Dole Act regulations require government-funded inventions be reported to the federal agency that made the grant or contract award, enforcement of these regulations can be a challenge (Rai & Sampat, 2012; U.S. Archive). In future work, various matching techniques will be utilized to determine if additional NIGMS-funded patents exist yet not identified by this current approach. We surmise that NIGMS funding is only one of many sources of funding for these firms, and we are also pursuing other data sources that will accurately capture levels of funding from other sources, such as venture capital, other federal and non-federal funding, and even other SBIR/STTR grants from other agencies.

When considering the cost efficiency of the NIGMS SBIR/STTR program, it was found that the average amount of federal funding needed to produce a patent was between $1.6 million and $2.6 million. In this regard, the NIGMS program has been successful, given that other studies estimate R&D spending of $3.2 to $6.4 million per patent in the medical equipment and scientific research business sectors, areas relevant to the NIGMS mission. While it’s possible additional funding sources were used to produce the NIGMS-supported patents, it remains that the return on the public investment is greater than the return on investments from all sources.

Overall, the NIGMS SBIR/STTR portfolio demonstrated innovation activity that generally met baseline standards and had performance comparable to other evaluations of SBIR/STTR grant programs across government.

**Commercialization**

While no direct measure of commercialization activities for SBIR/STTR grantees was collected, successful transitions to Phase II funding, patent activity, and downstream patent citations indicate a firm’s ability to contribute to further commercialization of the work supported by its grants. Patenting activity appears to be similar to other SBIR/STTR portfolios, suggesting a similar level of commercialization activity for NIGMS grantees. When considering downstream patent citations, however, some differences begin to appear.
As mentioned earlier, approximately 70 percent of the patents that we analyzed had downstream patent citations that did not originate from the original patent assignees. Thus, a large fraction of patents generated by SBIR/STTR grantees are being referenced in subsequent patent applications. On average, these patents are being cited by 19.9 subsequent patents, notably higher than the average values for NIGMS or NIH supported patents (14.6 and 13.9, respectively). These additional downstream patent citations represent continued investment in similar technologies and innovation. Assuming each patent required an additional $3.2 million in R&D investment (Shackelford, 2013), this would constitute $12.1 billion in follow-on R&D activity resulting from work supported by $1.1 billion in initial investments. This is a larger return on funding than seen by Kalutkiewicz and Ehman (2017), who observed the equivalent of approximately $6 billion in downstream R&D activity for every $1 billion invested in NIH research grant funding despite using a higher estimate of the R&D costs per downstream patent. This finding suggests a notable contribution of the SBIR/STTR portfolio towards downstream economic activity, even though it does not directly measure other commercialization and licensing activity of the patented technologies.

In addition to patent-related metrics, a firm’s transition from Phase I to Phase II SBIR/STTR funding was considered as a proxy for both survival and the ability of a firm to further commercialization efforts. Along these lines, grantees in the NIGMS portfolio had a higher transition rate (36.2 percent) than the SBA-established baseline of 25 percent. The most prolific grantees transitioned at rates nearing 70 percent, which exceeded similar analyses of NASA SBIR/STTR programs (National Academies of Sciences, Engineering, and Medicine, 2016a). This suggests that grantees were meeting or exceeding expectations in demonstrating the viability of the technologies at early stages such that further investment was warranted for continued commercialization efforts.

These indicators, while not direct evidence of commercialization of technologies in the marketplace, point to the fact that the NIGMS SBIR/STTR program has been successful in developing the commercial potential for the technologies supported by the grants. A fuller characterization of the nature of these technological contributions is one area for future research. For example, the nature of backward citations in SBIR/STTR patents may predict the extent to which these inventions are built on in the future (see, for example, Ahuja & Lampert, 2001, and Kelley, Ali, & Zahra, 2013).

**Firm Survival**

Key to the long-term viability of the technologies developed through the SBIR/STTR program, the continued survival of a firm was assessed through both Phase II transitions and data on the continued operation or acquisition of grantee firms beyond the period of funding by NIGMS. As shown earlier, grantees transitioned to Phase II funding at a rate higher than those found in other studies of SBIR/STTR programs, including the SBA baseline of 25 percent, suggesting that firms with SBIR/STTR support exceed expectations in terms of survival beyond the initial phase of the program. However, this only indicates short-term survival, and does not provide a measure of whether the firm remained in business beyond the grant period.
Identifying firm survival beyond the grant period is not feasible using only NIH records, given that typical progress reporting requirements stop with the end of a project’s grant funding. As a result, establishing firm status requires external sources, and missing data in these sources can be the result of a firm being acquired, shuttered, or simply not being easily found. Despite this, and the fact that our approach did not include any direct outreach to firms, we were able to obtain some information on firm status data for 97 percent of grantees. Data were missing in all the sources we used, but the merging of multiple sources allowed for both increased coverage and cross-validation of findings. Our finding that 63 percent of firms were estimated to still be operating in some capacity, and that their time-dependent survival was higher than typical SBA data on the rate of small business closures, suggested that the businesses supported by NIGMS through its SBIR/STTR programs were in operation longer than typical firms. Further information on the survival of businesses within the biomedical R&D sector would be needed for a direct assessment of the impact of the SBIR/STTR program, but the current data suggest that the program does identify grantees who are likely to be successful and helps maintain the vibrancy of these small businesses, in keeping with the mission of the program.

4.1 Lessons Learned—The use of extant data requires a careful assessment of the quality of the information contained in the available datasets. The data in extant sources often are collected for purposes other than research and they may not meet the same standards of rigor and quality typical in primary data collected for research. The integration of multiple data sources for the same information, such as our use of three alternative sources to determine business survival, along with rigorous methods to resolve conflicts among them, can increase the reliability of the final dataset. When there is agreement among several data sources, more labor-intensive, in-depth research can be devoted to resolving ambiguous cases.

Similarly, care must be taken in evaluating the quality of matches between records in diverse data sources. In this study, reliable matches—for example, between a firm’s grant records and its records in business intelligence databases—were ensured by using multiple criteria to ensure a successful match. Again, ambiguous cases can be identified, and in-depth research targeted to resolving them.

4.2 Conclusions—This study was unique in its quantification of outcomes using linked data, drawn largely from publicly-available sources. Most previous studies of the SBIR/STTR program have relied on surveys of former grantees to gather outcome information. Such surveys are costly to administer, and response rates have typically been very low, making non-response bias a potential threat to the validity of those results. The availability of large publicly-available datasets continues to grow and methods of text mining and linking databases continue to improve (see, for example, Dornbusch, Schmoch, Schulze, & Bethke, 2012; Maraut & Martinez, 2014; Penuel & Means, 2011). As a result, the use of extant data increasingly becomes a viable, less expensive, and more reliable approach to gathering data for evaluation studies.

Using linkages between NIGMS grant funding records and other business databases, we were able to document more completely several key outcomes of NIGMS SBIR/STTR...
funding and more firmly establish the unique contributions of these programs within the NIGMS research grant portfolio. By all measures examined in this study, when compared to other available benchmarks, the NIGMS SBIR/STTR programs have been successful in meeting the purposes for which they were established.

Acknowledgments:

We thank the following people whose valuable suggestions contributed to the preparation of the manuscript: Drs. Jon R. Lorsch, Judith H. Greenberg, and William A. Gern.

Funding: This work was supported in part by contract HHSN316201300006W/HHSN26300001 from the National Institute of General Medical Sciences (NIGMS) to Medical Science & Computing, LLC.

Author Biographies:

James Onken is President of Research Enterprise Analytics (REA), specializing in the evaluation of research programs and the analysis of federal research policies. Prior to founding REA, he worked for the National Institutes of Health for 27 years, holding senior positions performing program analysis and developing the NIH data infrastructure for research portfolio analysis.

Andrew C. Miklos is Data Collection and Analysis Section Chief in the Office of Program Planning, Analysis, and Evaluation (OPAE) at NIGMS, where he informs NIGMS grants policy decisions with analyses combining administrative records with other data sources. His background is in biochemistry and biophysics, having formerly worked in NMR research on the thermodynamics of protein folding and binding.

Travis F. Dorsey is an analyst for Medical Science & Computing, LLC working in the Office of Program Planning, Analysis, and Evaluation (OPAE) at NIGMS, where he informs NIGMS grants policy decisions with analyses combining administrative records with other data sources. His background is in economics and political economy, having formerly worked in the financial sector performing data analytics.

Richard Aragon is chief of the OPAE, where he advises on the Institute’s strategic planning activities, analyzes and evaluates its research and training programs, and serves as its legislative liaison. Prior to joining NIGMS, Richard held two positions in the Office of the Assistant Secretary for Financial Resources within the Office of the HHS Secretary: directing the Division of Program Integrity Assessment, Integration, and Oversight and the Division of Outreach, Communications, and Training.

Anna Maria Calcagno is the deputy chief of the OPAE. Prior to joining the NIGMS, she served as the data analytics team lead within the Office of Program Integrity Coordination at Department of Health and Human Services where she worked on one of the HHS Secretary’s Strategic Priorities. Dr. Calcagno’s over decade-long career of federal service began as a health science administrator at the National Cancer Institute (NCI).
Bibliography

Ahuja G, & Lampert CM (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. Strategic Management Journal, 22, 521–543.

Archibald R, & Finifter D (2000). Evaluation of the Department of Defense Small Business Innovation Research Program and the Fast Track Initiative: A Balanced Approach. In N. R. Council, The Small Business Innovation Research Program: An Assessment of the Department of Defense Fast Track Initiative Washington, DC: National Academy Press.

Arcs ZJ, & Audretsch DB (1989). Patents Innovative Activity. Eastern Economic Journal, XV(4), 373–376.

Bureau of Labor Statistics. (2017). Establishment Age and Survival Data. Retrieved November 29, 2017, from Business Employment Dynamics: https://www.bls.gov/bdm/bdmage.htm Commission on Evidence-Based Policymaking. (2017). The Promise of Evidence-Based Policymaking Washington DC: The Commission on Evidence-Based Policymaking.

Crunchbase. (2018). Who We Are Retrieved February 15, 2018, from https://about.crunchbase.com/about-us/de

Rassenfosse G, & van Pottelsberge de la Potterie B. (2009). A Policy Insight into the R&D Patent Relationship. Research Policy, 779–792.

Dornbusch F, Schmoch U, Schulze N, & Bethke N (2012). Identification of university-based patents: A new large-scale approach. Research Evaluation, 22(1), 52–63. doi: 10.1093/reseval/rvs033

Dun & Bradstreet. (2018). Our Data Retrieved February 15, 2018, from http://www.dnb.com/about-us/our-data.html

Edison T (2014). Analyzing Patents Generated by SBIR Firms. Proceedings of the Eleventh Annual Acquisition Research Symposium Graduate School of Business and Public Policy, Naval Postgraduate School.

Gambardella A, Harhoff D, & Verspagen B (2008). The value of European patents. European Management Review, 5, 60–84.

Griliches Z (1998). Patent Statistics as Economic Indicators: A Survey. In Griliches Z, R&D and Productivity: The Econometric Evidence (pp. 287–343). Chicago, IL: University of Chicago Press.

Guo Y, Hu Y, Zheng M, & Wang Y (2013). Patent indicators: a window to pharmaceutical market success. Expert Opinion on Therapeutic Patents, 23(7), 765–771. doi: 10.1517/13543776.2013.792806 [PubMed: 23611022]

Hagedoorn J, & Clooet M (2003). Measuring Innovative Performance. Is there an advantage in using multiple indicators. Research Policy, 32, 1365–1379.

Hall BH, Jaffe A, & Trajtenberg M (2005) Market value and patent citations. RAND Journal of Economics, 36(1), 16–38.

Jaffe AB & de Rassenfosse G (2017). Patent citation data in social science research: Overview and best practices. Journal of the Association for Information Science and Technology, 68(6), 1360–1374.

Kalutkiewicz MJ, & Ehman RL (2017). Patents, Pasteur, and Productivity New York: Manhattan Institute for Policy Research.

Keller RT, & Holland WE (1982). The Measurement of Performance among R&D Professional Employees: A Longitudinal Analysis. IEEE Transactions of Engineering Management, 29, 54–58.

Kelley DJ; Ali A, & Zahra SA (2013). Where do breakthroughs come from? Characteristics of high-potential inventions. Journal of Product Innovation Management, 30, 1212–1226.

Landau JO, & Schankerman M (1999). The Quality of Ideas: Measuring Innovation with Multiple Indicators National Bureau of Economic Research, NBER Working Paper No. 7345 Retrieved May 8, 2017, from http://www.nber.org/papers/w7345.pdf

Maraut S, & Martinez C (2014). Identifying author-inventors from Spain: methods and a first insight into results. Scientometrics, 101(1), 445–476.

Morrel-Samuels P, Francis E, & Shucard S (2009). Merged Datasets: An Analytic Tool for Evidence-Based Management. California Management Review, 52(1), 120–139.

National Academies of Sciences, Engineering, and Medicine. (2015). SBIR/STTR at the National Institutes of Health Washington, DC.
National Academies of Sciences, Engineering, and Medicine. (2016a). SBIR at NASA Washington, DC: The National Academies Press.

National Academies of Sciences, Engineering, and Medicine. (2016b). SBIR/STTR at the Department of Energy Washington, DC: The National Academies Press. doi:10.17226/23406

[dataset] National Institutes of Health. (2017). Funding Facts Retrieved April 5, 2017, from https://report.nih.gov/fundingfacts/fundingfacts.aspx

[dataset] National Institutes of Health. (2018a). NIH Research Portfolio Online Reporting Tools: Expenditures and Results (RePORTER) Retrieved February 15, 2018, from https://projectreporter.nih.gov/reporter.cfm

[dataset] National Institutes of Health. (2018b). NIH Exported Portfolio Online Reporting Tools: Expenditures and Results (ExPORTER) Retrieved February 15, 2018, from https://exporter.nih.gov/ExPORTER_Catalog.aspx

National Institutes of Health. (2018c). NIH Research Portfolio Online Reporting Tools Retrieved February 15, 2018, from https://report.nih.gov/

National Institutes of Health, Office of Extramural Research. (2003). National Survey to Evaluate the NIH SBIR Program: Final Report Bethesda: U.S. Department of Health and Human Services.

National Research Council. (2009). An Assessment of the SBIR Program at the National Aeronautics and Space Administration Washington, DC: The National Academies Press. doi:10.17226/12441

National Research Council. (2014). SBIR at the Department of Defense Washington, DC: The National Academies Press. doi:10.17226/18821

National Research Council, Committee for Capitalizing on Science, Technology, and Innovation. (2009a). An Assessment of the Small Business Innovation Research Program, Policy and Global Affairs (Wessner CW, Ed.) Washington, DC: National Academies Press.

National Research Council, Committee for Capitalizing on Science, Technology, and Innovation. (2009b). Venture Funding and the NIH SBIR Program Washington, DC: National Academies Press.

Neuhaeusler P, Frietsch R, Mund C, & Eckl V (2017). Identifying the Technology Profiles of R&D Performing Firms — A Matching of R&D and Patent Data. International Journal of Innovation and Technology Management, 14, 1–30.

Onken J, Aragon R, & Calcagno AM (2019). Geographically-Related outcomes of U.S. funding for small business research and development: Results of the research grant programs of a component of the National Institutes of Health. Evaluation and Program Planning, 77 In press.

OpenCorporates. (2018). About OpenCorporates Retrieved February 15, 2018, from https://opencorporates.com/info/about

Penuel WR & Means B (2011). Using large-scale databases in evaluation: Advances, opportunities, and challenges. American Journal of Evaluation, 32(1), 118–133. doi: 10.1177/1098214010388268

Rai AK & Sampat BN (2012). Accountability in patenting of federally funded research. Nature Biotechnology, 30, 953–956.

Shackelford B (2013). InfoBrief: One in Five U.S. Businesses with R&D Applied for a U.S. Patent in 2008 Arlington: National Center for Science Engineering and Statistics.

Trajtenberg M (1990). A penny for your quotes: patent citations and the value of innovations. The RAND Journal of Economics, 21(1), 172–187. U.S. Archive. From https://ia800203.us.archive.org/0/items/underreportingfe00unit/underreportingfe00unit.pdf

[dataset] U.S. Patent and Trademark Office. (2016a). Patent Full-Text Databases Retrieved August 22, 2016, from http://patft.uspto.gov/netehtml/PTO/index.html

[dataset] U.S. Patent and Trademark Office. (2016b). PatentsView Retrieved November 23, 2016, from http://www.patentsview.org

U.S. Small Business Administration. (2014). Frequently Asked Questions Retrieved from https://www.sba.gov/sites/default/files/FAQ_March_2014_0.pdf

U.S. Small Business Administration. (2017a). About SBIR Retrieved April 5, 2017, from SBIR-STTR: America’s Seed Fund: https://www.sbir.gov/about/about-sbir#sbir-mission
[dataset] U.S. Small Business Administration. (2017b). SBIR-STTR Award List Retrieved April 5, 2017, from https://www.sbir.gov/sbirsearch/award/all

U.S. Small Business Administration. (2017c). Performance Benchmark Requirements for Phase I Retrieved April 5, 2017, from SBIR-STTR: America’s Seed Fund: https://www.sbir.gov/performance-benchmarks
Highlights:

- Extant database linkages were successfully used to assess an SBIR/STTR program
- The NIGMS SBIR/STTR program performed above expectations along all dimensions
- The use of extant data increasingly can be a cost-effective approach in evaluation
Figure 1.
Venn diagram of NIGMS SBIR/STTR-supported patents found in the USPTO and iEdison databases.
Figure 2.
Cumulative patenting activity of NIGMS SBIR/STTR-funded grants. The dashed portion of the line includes recent projects that are unlikely to have had time to generate patents by 2015.
Figure 3.
Cumulative patenting activity of NIGMS SBIR/STTR-funded grants that received only Phase I support and those that made the transition to Phase II. The dashed portion of the line includes recent projects that are unlikely to have had time to generate patents by 2015.
Figure 4.
Cumulative NIGMS SBIR/STTR funding and number of supported patents.
Figure 5.
Average NIGMS SBIR/STTR program funding per patent generated. Real and constant 1983 dollars.
Figure 6.
Business status for five-year cohorts of NIGMS grantees, where cohorts are based on first year of NIGMS support.
Figure 7.
Percent of NIGMS-supported firms still active as a function of time from first NIGMS award and survival of all small businesses from birth year. Dashed lines indicate best-fitting exponential decay curves.
Table 1.
Number of patents associated with SBIR and STTR grants, the number cited by other patents, and the average number of citations each patent received.

|                     | SBIR | STTR | Total SBIR/STTR |
|---------------------|------|------|-----------------|
| Total Patents       | 330  | 41   | 371             |
| Cited Patents       | 248  | 30   | 278             |
| Citations           | 5,170| 914  | 6,084           |
| Excluding Self-Citations |
| Cited Patents       | 231  | 26   | 257             |
| Citations           | 4,249| 865  | 5,114           |
| % Cited             | 70.0%| 63.4%| 69.3%           |
| Avg Citations       | 18.4 | 33.3 | 19.9            |
Table 2.
Number of NIGMS SBIR/STTR grantees found in the OpenCorporates, Crunchbase, and Dun and Bradstreet databases.

| Number of Sources | Firms | Total GM Funding   | Percent of GM Funding |
|-------------------|-------|--------------------|-----------------------|
| 0                 | 36    | $10,346,947        | 1.1%                  |
| 1                 | 260   | $130,351,040       | 14.0%                 |
| 2                 | 662   | $484,139,514       | 51.9%                 |
| 3                 | 238   | $307,634,028       | 33.0%                 |
| Total             | 1,196 | $932,471,529       | 100.0%                |