ENHANCING DEFENSE ACQUISITION PROCESSES WITH EVIDENCED-BASED ANALYSIS: AN ILLUSTRATIVE CASE USING DOD'S SMALL BUSINESS INNOVATION RESEARCH (SBIR) PROGRAM

Major Toby Edison and Thomas R Edison*

ABSTRACT. This paper proposes and demonstrates that experimental and quasi-experimental evaluation methods can be applied to parts of the defense acquisition system providing evidence of program effectiveness. Specific example presented is a quasi-experimental evaluation of the Department of Defense (DOD) Small Business Innovation Research (SBIR) program. This paper demonstrates that quasi-experimental methods can be used to evaluate certain aspects of the DOD acquisition system and provides policy analysts with new tools to meet Congressional requirements for acquisition system evaluation.

The Office of Management and Budget (OMB), the Government Accountability Office (GAO), and the House Armed Services Committee (HASC) unanimously agree that the DOD does not objectively measure the performance and effectiveness of its acquisition system.

The paper recommends that more quasi-experimental studies be conducted and actual experimental studies be executed. These methods can help the DOD overcome the well documented deficiency in evaluating the effectiveness of its acquisition systems.

---

* Major Toby Edison, PhD, Dr. Thomas R Edison, DBA, Researcher: Major Toby Edison is a Professor of Program Management for Defense Acquisition University (DAU), West Region, Los Angeles, CA. Toby received his doctorate at the RAND Graduate School in Public Policy Analysis. His thesis was an evaluation of the DOD SBIR (Small Business Innovation Research) program. He has a MBA from UC Irvine. He has held several key acquisition positions within the US Air Force in program management while stationed at Hanscom AFB, MASS. and at Los Angeles AFB, CA. Among the programs he has been actively involved in were the Space Radar and Joint STARS programs. He is currently on active duty with the US Air Force deployed overseas. He is married with two children. Major Toby Edison, PhD, Professor of Program Management, Defense Acquisition University 222 N. Sepulveda Blvd, Suite 1220. El Segundo, CA 90245-5659 toby.edison@dau.mil Ph: 310-606-5924 Fax: 310-606-5925

MOTIVATIONAL QUOTES

The following quotes are provided as background for establishing the need for enhancing DOD's ability to conduct evidence-based analyses:

Findings.-The Congress finds that-
(1) waste and inefficiency in Federal programs undermine the confidence of the American people in the Government and reduces the Federal Government's ability to address adequately vital public needs;
(2) Federal managers are seriously disadvantaged in their efforts to improve program efficiency and effectiveness, because of insufficient articulation of program goals and inadequate information on program performance; and
(3) Congressional policymaking, spending decisions and program oversight are seriously handicapped by insufficient attention to program performance and results.
-- Introduction to the Government Performance and Results Act of 1993

(Sec. 5403) Directs each federal agency required to participate in the SBIR or STTR program to: (1) develop metrics evaluating the effectiveness and benefit of the program which are scientifically based, reflect the agency's mission, and include factors relating to the economic impact of the programs;
(2) conduct an annual evaluation of their SBIR and STTR programs using such metrics; and
(3) report each evaluation's results to the Administrator and the small business committees.
--Public Law 111-84, signed by President Obama on October 28, 2009; authorizes National Defense for FY2010, and specifically authorizes the DOD SBIR/STTR Programs through September 30, 2010

"The Panel began with the question of how well the defense acquisition system is doing in delivering value to the warfighter and

the taxpayer. For the most part, the Panel found that there is currently no objective way to answer this question. For most categories of acquisition, only anecdotal information exists about instances where the system either performed well, or poorly. Even where real performance metrics currently exist, they do not fully address the question. The Panel strongly believes that the defense acquisition system should have a performance management structure in place that allows the Department's senior leaders to identify and correct problems in the system, and reinforce and reward success."
--House Armed Services Committee Panel on Defense Acquisition Reform Findings and Recommendations, March 23, 2010

INTRODUCTION

Evaluating the effectiveness of any government program is difficult. Data on the program's output is often hard to obtain, selection into the program is usually not random and few programs are structured to facilitate the application of causal effects analysis. The Department of Defense (DOD) Small Business Innovation Research (SBIR) program is one such government program. Evaluating the effectiveness of the DOD SBIR program is required by Congress who direct each federal agency to "develop metrics evaluating the effectiveness and benefit of the program which are scientifically based, reflect the agency's mission, and include factors relating to the economic impact of the programs." Despite this legal requirement and nearly 30 years of running the SBIR program neither DOD administrators, nor policy analysts evaluating the program know whether the program is actually effective in supporting the DOD R&D mission by transitioning and integrating SBIR funded technologies into DOD weapon systems. In their assessments, the Government Accountability Office (GAO) and the Office of Management and Budget (OMB), found that the effectiveness of the DOD SBIR program had not been demonstrated (GAO, 2005) (OMB, 2005). The SBIR program is not alone in the DOD for its lack of objectives evidence of its performance effectiveness.

The indeterminate effectiveness of the relatively small SBIR program is just one case of the DOD generally not examining its acquisition

processes. Congress finds that the Department of Defense acquisition system does not routinely use objective methods to measure and improve its functions. Specifically, on March 23, 2010, the House Armed Services Committee on Defense Acquisition Reform concluded that there is no objective way to determine "how well the defense acquisition system is doing in delivering value to the warfighter." (HASC, 2010) Congress has officially required evidence-based policy administration by all Federal Agencies since 1993 through the Government Performance and Results Act (GPRA). The GAO finds fault with the DODs implementation of the GPRA, finding serious flaws in the DOD's Program Management business processes which are responsible for managing DOD acquisition. Specifically the GAO cites, that the DOD's plan to improve program management "lacked basic information, such as identifying specific business areas and key elements, such as goals, objectives, and performance measures." (GAO, 2010) There is ample evidence that DOD's measurement of its acquisition processes needs improvement. Unfortunately for many of the complex and unique acquisition processes that the DOD manages, instituting suitable performance measures has proved difficult. This paper shows that performance measurement tools do exist for one small piece of the defense acquisition portfolio, the DOD SBIR program.

This paper proposes a methodology for measuring the performance of the DOD SBIR program using quasi-experimental methods from the broader program evaluation literature. The paper describes the DOD SBIR program. It then describes the basics of the DOD SBIR program and examines two key biases in past DOD SBIR program evaluations that have confounded researchers: response bias and selection bias. The paper documents strategies to mitigate these biases using quasi-experimental methods that have been used in other program evaluations. Next, the paper illustrates that a better evaluation of the DOD SBIR program is possible if better methods are applied to existing data. The paper concludes with suggestions for strengthening the evaluation of the SBIR program with better data collection methods and with randomization. With evidence that better evaluations of defense acquisition processes are possible, the paper concludes with suggestions for further evidence-based research

efforts.

DESCRIPTION OF THE DOD SBIR PROGRAM AND BIASES IN PAST EVALUATIONS

Congress requires that all federal agencies with extramural R&D budgets in excess of $100M, including the Department of Defense, set aside 2.5% of their R&D budget for the SBIR program. The broad purpose of the program is to provide contracts to qualifying small businesses to support each agency's research mission, and to commercialize the funded technologies. In 2010, the SBIR program represents about 1% of the $108B that the Department of Defense spends on procurement. Congress sets the emphasis of the program with the following four goals: 1) to stimulate technological innovation; 2) use small businesses to meet federal R&D needs; 3) foster participation by disadvantaged businesses; and 4) increase private sector commercialization of federally funded research. (OSADBU, 2007). Congress has placed increased emphasis on the goal of enhancing private sector commercialization of DOD SBIR funded projects.

The law also requires the participating federal agencies to structure their SBIR programs with three-phases, with specific funding ceilings for each phase. Phase I funds up to $100K for a 6-month feasibility study competitively awarded to firms. Phase II is the principal R&D phase, which awards up to $750K over 18 months to the most promising Phase I submissions. Phase III is the commercialization phase, which is the period when firms sell their mature technologies to interested parties—often DOD prime contractors or program offices. No pre-allocated SBIR program funds support Phase III commercialization; however, if a topic reaches Phase III, the firm can be awarded a contract for that technology immediately, without competition. The design of the SBIR Phases is intended to transition the most promising technologies from the thousands of ideas of the participating small contractors into fielded technologies.

Within the constraints of the program, Congress offers freedom for the agencies to manage the SBIR program to fit the R&D strategies of

the participating agencies, which are important to understand in order to evaluate the program. Each agency has many noteworthy organizational innovations for managing a large dollar R&D program without explicit overhead that is required to award contracts and grants in relatively small dollar amounts. The 2008 DOD annual report to Congress on the SBIR program highlights some of these challenges. In 2008 the DOD solicited proposals for nearly 1,000 topics, for which they processed over 12,000 proposals, ultimately awarding about two SBIR contracts per topic. In order to manage this administrative workload, the DOD manages the process online—publishing two or three SBIR solicitations a year online, requiring proposers to register with the DOD SBIR program with their unique federal contractor identification number and to submit their proposals online. These online contract management tools will be shown later to be invaluable for measuring the program effectiveness.

As highlighted in this paper's introductory quote from the 2009 re-authorization of the SBIR program, Congress requires the program administrators to develop metrics on the program's effectiveness. The DOD has created a metric called the Commercialization Achievement Index. This index is not deemed sufficient to measure the program's effectiveness (OMB, 2005). Across all federal SBIR programs, since its inception the effectiveness of the SBIR program (to increase commercialization) has never been evaluated and this applies more broadly than to the specific DOD program (GAO, 2005). Among the specific reasons the GAO cites for this lack of objective evaluation are the lack of an agreed upon measure of effectiveness for commercialization and lack of reliable data on the program. Published evaluations of the SBIR program typically suffer from two common issues identified in the broader literature on program evaluation: selection bias and response bias.

The key aspects of past DOD SBIR program evaluations that are presumed to cause bias are the fact that evaluations must be performed after selection and with self-reported survey data. Response bias affects program evaluations that rely on surveys because it is presumed that program participants' over-report the output resulting from the program. Participants have an incentive to

attribute more benefit from program participation in a survey so that the program will continue to receive funding and the participant continue to receive the benefits of the program. Selection bias is the presumption that program administrators are not selecting program participants at random. Specifically, selection bias invalidated after-the-fact evaluations because it is assumed that more capable participants are selected at a higher rate and that these firms, in the absence of the program, are more productive. In the case of the DOD SBIR program analyzed in this paper, winning firms were bigger, older, and more experienced defense contractors and as a result had more non-SBIR defense contracts before and after winning a SBIR award.

An ideal experiment of the SBIR program would randomly assign SBIR program treatment on a population of firms qualifying for the SBIR program and see if the treated firms have more future defense contracts than untreated firms. Such an experiment has not been conducted, which motivates the example in this paper, estimating the treatment effect for winning a DOD SBIR award with after-the-fact evaluation methods and non-survey data.

STRATEGIES TO MITIGATE BIASES

To perform an effectiveness evaluation on the DOD SBIR program, this paper builds a data set based on 2003 SBIR applications. To control for response bias, the applications were matched to the defense contract database rather than to survey data. The analysis uses after-the-fact quasi-experimental models to control for selection bias, which have been shown to approximate the results of a randomized controlled trial under certain assumptions.

The program evaluation literature documents that the least biased program evaluations rely on a neutral source of outcome data (i.e. not reported by administrators or participants), have pre- treatment and post treatment observations, contain many characteristics of the participants and collect data on the treated population and a representative control population. The data set created for this analysis uses defense contract award data as the outcome of interest.

The contract award data are an output of the defense accounting process represented by the DD Form 350, which documents and publishes every contract award greater than $25K. The DOD identifies each contract awardee with a unique contractor identification number which can be linked electronically to other data bases the DOD maintains. This paper links to the DOD's Central Contractor Registry(CCR) and the DOD SBIR program's database of SBIR applications to capture firm characteristics in the database. The characteristics of each firm are important to after-the-fact program evaluations, because researchers can explain some of the variation in program effectiveness by correlating program outcomes with firm characteristics. For example, larger firms might win more defense contracts dollar simply because they have the capacity to take on more DOD funded work, regardless of whether they won a SBIR award. The DOD SBIR program's database of SBIR applications captured information on all firms that applied for the DOD SBIR program by year of application and identifies the firm's proposal that won an award. These pieces of information enabled the identification of a treatment population which applied for and won a SBIR award in a given year and a control population of firms that applied for but did not win an award. Creating a comparable control group with distinguishing characteristics is the crucial ingredient identified by program evaluation literature to controlling selection bias.

To control for selection bias the current program evaluation literature suggests using doubly robust estimation (DRE) methods to estimate the relationship between winning a SBIR award and future defense contract dollars. As the name implies, researchers use two methods to estimate a treatment effect. The first method prescribed is propensity score matching (PSM) which uses the observable covariates of the firms to create balanced treatment and control population. The second method prescribed is to perform a statistical estimation of the treatment effect that uses the characteristics of the firms to explain variation in future defense contracts (usually a regression with controls model). By combining two different estimation strategies, researchers have two chances to build the correct model. According to DRE theory, this approach will estimate a consistent treatment effect even if only one of the models is correct.

The characteristic of double robustness is achieved in after-the-fact program evaluations when the estimation from the PSM model and the statistical model are consistent in magnitude and significance. Under ideal conditions and with enough descriptive data, by applying these methods, a better estimate of the treatment effect from winning a SBIR award on future defense contract dollars is possible.

A NAÏVE ESTIMATE OF SBIR TREATMENT

In order to show why using a balanced treatment and control population are better than using raw data, this paper begins with a naïve estimate of the DOD SBIR program's treatment effect. Researchers with a treatment and control group typically estimate a treatment effect with differences in differences estimate. The first difference is calculated by subtracting the outcome observed before treatment and after treatment for each group. The second difference is equal to the difference in treatment between treated and non-treated observations.

A differences in differences is not the same as a typical program evaluation report based on a survey. A survey based estimate can only report the average raw output data on the treated group. For example, the National Academies of Sciences reports the average raw survey response to estimate sales generated by SBIR funded research to be $1.3M per SBIR project (Wessner, 2007). This average survey response is not a differences in differences because it does not compare the results to non-treated observations. Because the dataset created for this paper identifies winners and losers, it can be used to estimate a naïve differences in differences. Naïve means that that selection bias is not controlled.

The dataset used for this estimation is based on the entire population of DOD SBIR applicants in 2003 obtained from the Department of Defense SBIR administrative website. From the population of 2003 applications, a subset of 1460 firms who also applied in 2004 and who had a contractor identification number in the Central Contractor

Registry was identified as the population of interest. The DOD SBIR administrative database identifies 687 of these firms as winning a 2003 SBIR contract, with 773 applying for but not winning in 2003. These 1460 firms were matched with their contractor identification numbers to the form DD350 database maintained by the Department of Defense Directorate for Information Operations and Reports. The DD350 contains all contract actions greater than $25K organized by year and by individual contractor identification number.

Using the SBIR application data-set, the first difference between average total non-SBIR defense contract dollars won in 2004 minus the 2003 total **Δ04-03** is $651K for the average winner and $203K for the average loser (see Table 1). The second difference, the average treatment difference between winners and losers, is $448K. This naive treatment effect is assumed to be affected by selection bias.

Group/Year	2003	2004	Δ04-03
Winners	1,430	2,081	651
Losers	456	659	203
ΔW-L	974	1,422	**$448K**

TABLE 1: NAIVE DIFFERENCES IN DIFFERENCES

The effect of selection bias is presumably the cause of the SBIR winners having on average
$974K more contracts than losers in 2003, and $1.4M more contracts in 2004. Because winners have more contracts to start out with and firms with more past contracts will probably win more future contracts before and after winning in 2003, it is impossible to isolate the effect of winning the SBIR award in 2003. To improve on this naïve estimate more advanced statistical techniques are needed.

EVIDENCE OF A SBIR TREATMENT EFFECT

The naïve treatment effect estimate can be improved by using the characteristics of firms to explain some of their variation in treatment. The characteristics are used two ways to control variation. The first method to control variation using firm characteristics is to use an algorithm to balance the characteristics of the treatment and control populations. The balancing algorithm will discard outlying observations so that the treatment and control populations will be theoretically identical to a randomized controlled trial population. The second method to produce a better estimate of treatment effect using firm characteristics uses the firm characteristics to explain variation in the outcome. For example, by using a pre-treatment observation of defense contracts before a firms wins a SBIR contract, some of the variation in the post-treatment contract award amounts can be explained.

Applying these two methods to the dataset build for this paper can better estimate a treatment effect for the DOD SBIR program. This research method is described by Ho, Imai, King, and Stuart (2007) as doubly robust estimation. Double robust estimation protocols prescribe balancing populations and then using statistical methods to estimate the treatment effect. Analyses in Ho, Imai, King and Stuart (2007) show consistency between the results of RCT studies analyzed with DRE methods. Their analyses support the conclusion that estimates of causal treatment effects can be produced by DRE methods if researchers properly balance the treatment and control groups or researcher apply the correct statistical model. Their analyses based on thousands of different population balancing assumptions and statistical models with data from randomized controlled trials supports the conclusion that if the average treatment effect estimated with balanced treatment and control groups is consistent with the estimated treatment effect from another statistical model (such as a regression model) then the DRE estimate can be considered a causal estimate.

The model demonstrated estimates the future average increase in non-SBIR defense contracts for firms winning a 2003 DOD SBIR award. The key parameter of interest is the correlation between winning a 2003 SBIR award and non-SBIR defense contracts in 2004.

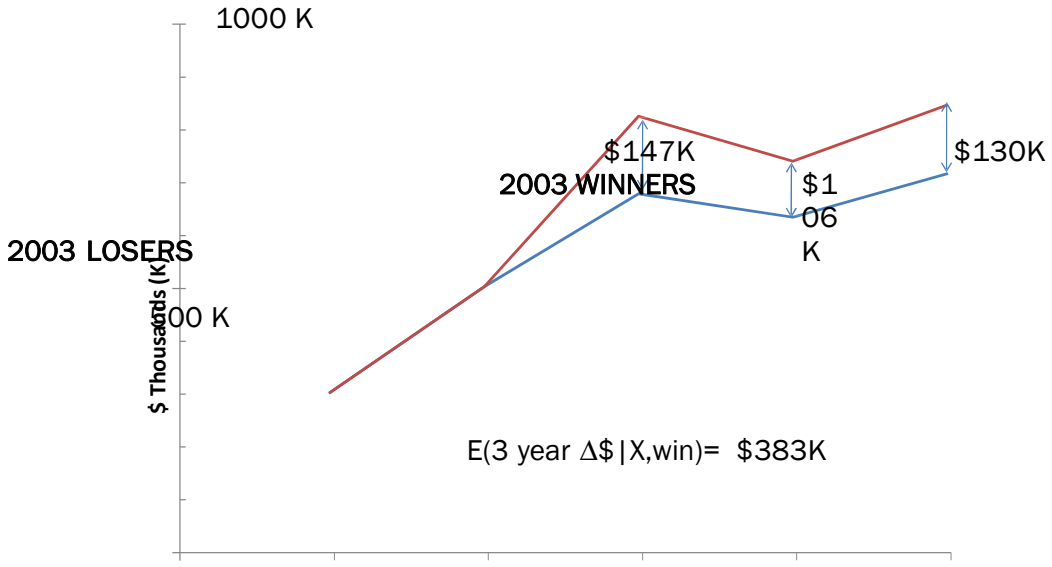
The control variables are total non-SBIR contracts in 2002, total SBIR contracts in 2002, the firms' first contract year, the number of employees in 2003, whether the firm won a defense contract as a sub-contractor in 2003, the number of topics submitted in 2003, and the total number of past Phase I or II awards.

The populations are balanced using the Coarsened Exact Matching protocols described by Iacus, King, & Porro (2008). The balanced population retains 534 firms that won in 2003 and 681 losing firms for a 83% post matching retention rate. As an example of the improvement in post- matching balance, the raw population had a different in 2002 non-SBIR contracts of $925K, the matched population $58K.

The doubly robust estimation model estimates a $147K treatment effect, with confidence level of greater than 99%. Based on this estimate, there is empirical support that the SBIR program increases defense contracts in 2004 for firms winning SBIR contracts in 2003.

The estimation that the DOD SBIR program does significantly increase non-SBIR defense contracts one year after award might be missing delayed effects two or three years after award. A three year commercialization time horizon is supported by surveys on the self-reported commercialization outcomes related to the SBIR program by the National Academies of Science (Wessner, 2007) and contract award analysis by RAND (Held, 2006), both of which find that the majority of commercialization activity occurs 3 years after a SBIR award. A doubly robust estimation is used to estimate several treatment effects for the non-SBIR DOD contracts won by firms in 2005 and 2006 who won a 2003 DOD SBIR award. The doubly robust estimated treatment effect for the 2005 non-SBIR contract dollar difference is $106K; the 2006 difference is $130K. Both estimates are statistically significant at the greater than 99% confidence level. These estimations of a lagged treatment effect support a conclusion that for the average firm, winning a DOD SBIR award puts a company on a sustained path towards winning more future DOD contract dollars than had they not won.

Winning a DOD SBIR award appears to put winning firms on a path of higher non-SBIR defense contract award dollars. Figure 1 illustrates that for the period between 2004 and 2006 firms which applied for and won a 2003 SBIR contract won an average of $383K more defense contracts than a matching set of firms who applied for but did not win a 2003 DOD SBIR award. The DOD SBIR program appears to be effective at increasing commercialization of SBIR funded technologies through defense contracts.



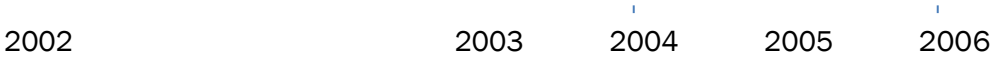


FIGURE 1: THREE YEAR ESTIMATED TREATMENT EFFECT OF WINNING A 2003 SBIR AWARD.

The Department of Defense explicitly acknowledges that access to new technology and a strong industrial base are crucial to United States national security (OSD, 2010). The evidence presented in this paper suggest that the DOD SBIR program may be providing both access to new technologies and broadening the industrial base by transitioning new technologies developed by small businesses into defense programs through defense contracts. The evidence that firms winning SBIR contracts increase their future sales to the DOD at a higher rate than had they not won supports the belief that the DOD SBIR program contributes to the DOD mission. Prior to this analysis the DOD emphasized without proof that they used the DOD SBIR program to support mission oriented research needs rather than to increase private sector commercialization. With proof that the commercialization path from SBIR funded R&D into standard defense prime contracts may be enhanced, the DOD can fulfill their GPRA requirement to demonstrate the effectiveness of their administration of the program and support the DOD preference against private sector commercialization.

HOW TO IMPROVE DOD SBIR PROGRAM EVALUATION

This paper's analysis is motivated by a literature review of the SBIR program, which contains numerous government reports, policies and regulations requiring better evaluations of the DOD SBIR program. Most of the policy responses to the need for better evaluation, such as the DOD- developed Commercialization Achievement Index, and the surveys conducted by the GAO and National Academies of Science, fall short of actually providing data for better evaluation because the data collected is incomplete, presumably subject to response bias and does not collect data on treated and untreated populations. By using the already-existing defense contract database, this paper showed that there exists a data source free from self-reported survey response bias to evaluate the program. Additionally, by using econometric methods to control for selection bias, this paper provides policy makers with one example of the ability to evaluate one key aspect of the program.

The policy recommendations on how to improve evaluation will increase the number of studies on the program, allow researchers to explore more evidence of SBIR research output, and improve the policy recommendations of the program evaluations. This paper motivates three possible policy implementations the DOD can use to improve the evaluation of the DOD SBIR program. The first is to make the DOD SBIR administrative data accessible to more researchers. The second would be to build automated links to the applying SBIR firms to other innovation proxies – most specifically, the US Patent database, the iEdison database, and technical publication databases. Finally, to more conclusively evaluate the DOD SBIR program, some form of Randomized Control Trials will need to be implemented. The enormous number of topics and applicants makes the DOD SBIR a good candidate to implement RCT's to evaluate the effectiveness of the SBIR program.

Evaluation Recommendation 1: Make SBIR Administrative Data Available To Researchers

The first recommendation to improve evaluation of the SBIR program, making administrative data more accessible to researchers, is a low cost, easily implementable policy change with potential for significant payback. As already documented in the review of SBIR evaluations, one of the consistent themes of all past SBIR program evaluations is

the lack of reliable, consistent data and the resulting lack of conclusive studies about the program's effectiveness. Additionally, the broader literature on R&D evaluations in general suffers from the same problems: lack of reliable data and a resultant dearth of conclusive evaluations on R&D programs. Opening the wealth of already- existing data collected by the DOD SBIR program to policy analysts would be an enormous step towards improving collective knowledge about how effective R&D subsidy programs really are.

One specific example of data that is available to program administrators but not to program evaluators is the proposal evaluation scores used to award SBIR contracts. If these scores were made available to researchers, then researchers could use those scores to better match firms in propensity score models or to control for variation in outcome. Importantly, since the DOD SBIR program is probably already collecting this information for administrative purposes in electronic formats and making the data available to administrators via the internet, the cost to make the data accessible to R&D policy researchers would be minimal. The payback for making this data available to research policy analysts that have spent decades trying to determine the efficacy of R&D policies with nearly zero reliable data is potentially significant. Policy makers could have more fact-based studies to improve policy to meet the spirit and intent of the Government Performance and Results Act.

Evaluation Recommendation 2: Link SBIR Funding To More Innovation Proxy Data Sources

The second policy recommendation to improve the evaluation of the DOD SBIR program is to enable automated matching of SBIR administrative data to other sources of innovation output data such as patent data, innovation tracking databases, sales data, venture capital funding, or technical publication data. Per US law, any SBIR participant is mandated to report to the government the details of any inventions or patents generated from the program. Unfortunately, the reporting is often decentralized, and the data collected is not easily linked to the actual source of funding. There are certainly more research outputs than just increased DOD sales tracked through the defense contracting database that could be used to measure the impact of the DOD SBIR program. Examples of potentially useful data sources are the US Patent and Trademark Database, technical paper databases, databases of firms such as COMPUSTAT, HOOVERS or

DUNS, venture capital tracking databases, initial public offering databases, merger databases, or Internal Revenue Service data. Currently automated linking of SBIR participant data to another data source is not possible because not all of the databases can be linked using contractor identification numbers or DUNS numbers. The lack of a common standard firm identifier leaves researchers with the option of trying to match research inputs to output based on firm names, which contain tremendous variation in spelling within and across databases. The SBIR program could require firms to include their DUNS number in the already-required government interest statements for patents generated by SBIR funds. For matching technical publications, the SBIR program could require firms to report SBIR-generated technical publications with full citations in future application packages. Since SBIR application packages are submitted electronically, the government can begin to understand the impact of the SBIR program on the body of technical knowledge through patent disclosure analysis and technical publication analysis.

The most expedient link to establish might be the link between SBIR funding and the interagency Edison (iEdison) database maintained by the National Institutes of Health. This database was created to fulfill the statutory requirement for federally funded researchers to report inventions and patents developed with federal funds. Currently it collects data from some, but not all, DOD research organizations. DOD SBIR policies could be modified to require winning firms to report inventions and patents through this database, and to require the inclusion of the funding contract number and the correct contractor identification number.

A final suggestion to improve tracking of SBIR output activity would be to require proposing firms to submit their tax identifier number to conclusively link SBIR funding to actual growth in revenue. Since all firms winning SBIR awards must be US companies, this policy intervention would cover the entire population of awardees. Moreover, since the IRS reports on income are legally required to be accurate and are subject to the possibility of auditing, the validity of the sales and revenue data will be substantially more accurate than the data self-reported in surveys. Another strength of this source of data would be that the study population could be expanded beyond the non-representative sample of survey respondents to include potentially all SBIR applicants.

The strengthening of the links between DOD SBIR program data

sources and data sources on innovation proxies will greatly improve the quality and quantity of analyses possible on the program. If any of these policy recommendations improve evaluating the link between innovation subsidies to innovation output, a new era of R&D policy evaluation can begin and better R&D policies can be created.

Evaluation Recommendation 3: Implement Limited Randomized Control Trials For Improved Evaluations

The final suggestion for improving evaluation of the SBIR program is to continue to apply and refine research methods proven to mitigate biases, including using randomized controlled trails. The Government Performance and Results Act requires all agencies to strive towards evidence based policy implementation. The gold standard research method to provide conclusive evidence of program effectiveness would be to conduct a randomized control trial by randomizing some aspects of the contract awards. Of all the R&D subsidy and small business programs and the program evaluations reviewed for this paper, the SBIR program might be the most conducive to incorporating randomization to improve evaluation.

One practical suggestion to implement an RCT would be to select a subset of some of the topic awards with a random process. Since each topic receives around 15 applications, the suggestion would be to identify the 5 highest rated applications, randomly select the winner from those 5 applications, and track the relative performance of the firms that received the award and those who did not. There is a possibility that this type of experiment could be double blind because the firms would never know if they received the award due to random assignment and the program managers actually managing the SBIR contract could be kept blind to the actual award decision. The DOD SBIR program is an ideal candidate for incorporating some aspect of an RCT to evaluate the program. There are hundreds of topics each year, thousands of applicants, the research budget is by its very nature discretionary (not on a program's critical path, nor vital for national security), and the firms can be tracked over time.

In lieu of the opportunity to perform an RCT, researchers should continue to apply the propensity score and doubly robust estimation methods to SBIR administrative data. These after- the-fact estimation protocols could be improved if the actual evaluation scores were made available to researchers. If the evaluation scores were made available, researchers could use the scores to better match firms with

balancing algorithms. Researchers could use the proposal evaluation scores in regression models to explain more variation in the outcomes of interest.

Current best practices in developmental economics have adopted RCT's (Rodrik & Rosenzweig, 2009). The focus of developmental economics (on improving the lives of the citizens of poor nations through interventions such as micro-financing, distributing anti-mosquito nets, improving immunizations and improving potable water supplies) by its nature makes it a much humbler and moderately funded field than national R&D policy analysis. Rodrik & Rosenzweig (2009) note that in the field of development economics: "Randomized controlled trials (RCTs), in which randomly-selected subpopulations are selected for an intervention and then outcomes are compared across the treated and untreated populations, have been used to evaluate the causal effects of specific programs (e.g., cash transfers, subsidies to medical inputs), delivery mechanisms (e.g., kinds of financial products), and, less pervasively, to obtain evidence on fundamental behavioral assumptions that underlie models used to justify policy–e.g., adverse selection." If policy administrators can adopt RCT methods to determine the best way to deliver developmental economics policy interventions, then the better-funded, higher-profile field of R&D policy analysis should be able to muster the resources and institutional will necessary to implement limited RCT studies to better understand the efficacy of the $1B+ DOD SBIR program.

Bottom line: Policy makers should seriously consider incorporating randomization into the DOD SBIR program to improve the evaluation of the program and to demonstrate how to build evaluation tools into other government programs. These three suggestions could help revolutionize the way the SBIR program is evaluated and offer a wider variety of answers to the policy questions. With more data available, better links to research output and actual experimental results, the artifacts of the DOD SBIR program that actually work best can be understood, refined and applied as best practices across the DOD and Federal government. With better analyses, policy makers can use facts to craft and administer better policies. This paper has provided a small sample of the research possible if evaluation data and tools are improved. If any form of these recommendations is adopted, the DOD SBIR program would be better evaluated.

FURTHER EVIDENCE-BASED ACQUISITION POLICY ANALYSES

The program evaluation tools demonstrated in this paper highlight that it is possible to evaluate the effectiveness of some aspects of defense acquisition systems. The after-the-fact tools demonstrated in this paper and the suggestion to implement randomized controlled trials can be applied to other areas of the defense acquisition system to provide policy makers evidence of how well policy changes perform. Specifically there are policy changes enacted by the Weapon System Acquisition Reform Act (P.L.111) and the National Defense Authorization Act for FY 2009 (P.L. 110-417) that are worthy of consideration for evaluation with experimental and quasi-experimental methods. Some examples of the policy recommendations that might be suited for experimental analysis are: the emphasis on competition, the requirement for prototyping, the implementation of earned value management and the increase in the number of acquisition professionals.

For example, estimating the effectiveness of maintaining competition throughout the acquisition lifecycle could be part of a randomized trial or could be analyzed using quasi- experimental methods. For an RCT, policy makers could randomly pick which current program would be required to implement competition in technology development, prototyping and production. Analysts could estimate the effect of competition by measuring the difference in cost changes and schedule delays on the programs with and without competition. If randomization of competition requirements is infeasible, after-the-fact analyses could estimate the effect of competition on cost and schedule. The evaluator could use the characteristics of the different programs (weapon type, joint program, service of program office, year of program initiation), along with an identifier on whether they had competition or not, to build treatment and control groups and to explain other variations in program outcomes.

CONCLUSION

Congress is re-emphasizing its direction to the DOD to improve the evaluation methodologies for the defense acquisition system. This paper highlights that for some aspects of the defense acquisition system quasi-experimental methods can be applied and do provide evidence to estimate program efficacy. This paper recommends that DOD policy makers build more experimental and quasi-experimental links into the current DOD SBIR program to improve the evidence

available to acquisition policy makers. Based on this demonstration, policy makers should consider broadening the application of these methods beyond the SBIR program to additional acquisition system aspects that can be analyzed with experimental and quasi-experimental models.

REFERENCES

DIOR. (n.d.). *DOD Personnel &Procurement Statistics*. Retrieved from http://siadapp.dmdc.osd.mil/.

GAO. (2010, March 25). *DOD Approach to Business Transformation.* Retrieved March 23, 2010, from GAO: http://www.gao.gov/highrisk/risks/dod-management/business_transformation.php.

GAO. (2005). *Observations on the Small Business Innovation Research Program: Statement for the Record of Anu K. Mittal, Director Natural Resources and Environment Team.* Washington D.C: Government Accountability Office.

GAO. (2009). *Reports on the Government Performance and Results Act.* Washington DC.: Government Accountability Office.

HASC. (2010). *House Armed Services Committee Panel on Defense Acquisition Reform, Findings and Recommendations .* Washington D.C.

Held, B. E. (2006). *Evaluation and Recommendations for Improvement of the Department of Defense Small Business Innovation Research Program.* RAND: Santa Monica.

Ho, D., Imai, K., King, G., & Stuart, a. E. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis , 15* (3), 199-236.

House Armed Services Committee Panel on Defense Acquisition Reform . (2010). *House Armed Services Committee Panel on Defense Acquisition Reform Findings and Recommendations.* Washington D.C.

Iacus, S., King, G., & Porro, G. (2008). *CEM: A Matching Method for Observational Data in the Social Sciences.* Retrieved November 15, 2009, from www.agrocampus-ouest.fr/math/useR- 2009/.../pdf/Porro+Iacus+King.pdf.

National Defense Authorization Act for FY 2009. (2009). *P.L. 110-417* .

OMB. (2005). *OMB Program Assessment Rating Tool (Part): DOD Small Business Innovation Research Program.* Office of Management and Budget: Office of Management and Budget.

OSADBU. (2007). DOD Annual Report Summary. Office of Small and Disadvantaged Business Utilization.

OSADBU. (2010, Feb 26). *DOD SBIR Annual Report.* Retrieved Feb 26, 2010, from www.dodsbir.com.

OSD. (2010). *Quadrennial Defense Review Report.* Washington D.C.

Rodrik, D., & Rosenzweig, M. (2009). Development Policy and Development Economics: An Introduction*. In D. R. Rosenzweig, *Handbook of Development Economics, vol. 5.* North-Holland.

Weapon System Acquisition Reform Act . (2009). *P.L.111* .

Wessner, C. W. (Ed.). (2007). *An Assessment of the Small Business Innovation Research Program at the*
Department of Defense. Washington, D.C.: National Academies Press.