



ARTICLE

DOI: 10.1057/s41599-018-0069-9

OPEN

# Why are there (almost) no randomised controlled trial-based evaluations of business support programmes?

Margaret Dalziel<sup>1</sup>

**ABSTRACT** Based on the achievements of randomised controlled trials (RCTs) in medicine, and the need for effective government interventions in support of business, some have advocated for the use of RCTs in the evaluation of business support programmes (BSPs). Notwithstanding these recommendations, the use of RCTs in the evaluation of BSPs has been resisted by (almost) all. Policy makers and managers are correct in their reluctance to undertake RCT-based evaluations for four reasons. First, while RCTs require the random allocation of support, judicious programmes select firms on the basis of potential and amenability to support. Second, while RCTs require treatments that exhibit low variability, the most effective BSPs draw upon substantive knowledge to provide support that is customised. Third, BSPs aim to produce outliers—firms whose performance is exceptional. When outliers are present, very large samples will be required to produce reliable results. Finally, an RCT may not yield a meaningful contribution to knowledge. The strength of an RCT is its ability to estimate the magnitude of the treatment effect under controlled conditions. But where much depends on the nature of participants and circumstances, we seek evidence of what works, for whom, in which circumstances, and why.

<sup>1</sup>University of Waterloo, Waterloo, Ontario, Canada. Correspondence and requests for materials should be addressed to M.D. (email: [mdalziel@uwaterloo.ca](mailto:mdalziel@uwaterloo.ca))

The randomised controlled trial (RCT) is a very beautiful technique, of wide applicability, but as with everything else there are snags (Cochrane, 1972, p 22).

## Introduction

Heralded as the gold-standard in the evaluation of medical interventions, the use of RCTs is increasingly being advocated as a means of evaluating business support programmes (most recently by Bakhshi et al., 2015; Banerjee et al., 2015; previously by Burtless, 1995; Storey, 1998). The Economist reports that the percentage of National Bureau of Economic Research working papers that mention RCTs in the abstract has increased five-fold over the last 10 years (Economist, 2016). And several organisations and initiatives that promote, fund, and disseminate the results of RCTs have been launched including: the Innovation Growth Lab and the What Works Network in the UK; and the Abdul Latif Jameel Poverty Action Lab (J-PAL), the International Initiative for Impact Evaluation, and the American Economic Association's RCT Registry in the US.

Notwithstanding these recommendations, researchers report that there has not yet been an RCT-based evaluation of a programme that supports R&D or innovation (Crespi et al., 2011; De Blasio et al., 2014). While there have been a number of experimental evaluations of modest microfinance and business training programmes, mostly in developing countries (e.g., Banerjee et al., 2015; Crépon et al., 2015; Fairlie et al., 2015), I am aware of only one RCT-based evaluation of a business support programme (BSP) in a developed country, and in that case the programme was designed specifically to support the RCT evaluation (Bakhshi et al., 2015).

In this paper, I address the question of why there are (almost) no RCT-based evaluations of BSPs in developed countries. I begin by presenting the arguments for government interventions in support of business and describing the nature and extent of support. The next section reviews the literature on the strengths and limitations of RCTs, which has been developed by economists and evaluators in fields such as public health, education, and international development, where RCTs are prevalent. The following section focuses on the case of business support and identifies the features of business support programme design that make evaluation challenging and the use of RCTs rare. The final sections consider the implications for policy makers and evaluators, and conclude.

## Why and how governments support businesses

Governments invest large sums in BSPs with a view to accelerating economic development and the benefits it provides in terms of jobs and prosperity. As these funds are raised through taxation, this transfer of wealth from the citizenry to the corporate sector needs to be justified, particularly because, in many cases, the citizens providing the funds have fewer resources than the business people who benefit most directly from the support (Lazonick and Mazzucato, 2013).

Both the economic and innovation system rationales for government support of business are based on the premise that such support, in the fullness of time, yields benefits to society that exceed the costs. Economists believe that the invisible hand of the market is generally best able to allocate resources efficiently and that government intervention needs to be justified in terms of market failure. Positive externalities—benefits to third parties such as knowledge spillovers—are one form of market failure. As a consequence of knowledge spillovers, the society-wide returns to investments in R&D are likely to exceed the private returns to the focal firm, leading to under-investments in R&D from society's perspective (Hall et al., 2010; Jaffe et al., 1993).

Government investments in programmes that support R&D respond to this form of market failure by inducing firms to invest in R&D. Other forms of market failure provide the justification for government investments in other forms of business support such as clean technologies and new ventures. Negative externalities, costs to third parties such as pollution, explain why governments invest in environmental regulations and support clean technology companies (Jaffe et al., 2005); information asymmetries and liquidity constraints explain why governments support ventures (Freel, 2007).

Innovation—the design and deployment of new products, processes, and services—is an emergent process that involves a diverse set of actors, resources, and institutions (Anderson et al., 2014; Dougherty, 2016). Innovation scholars argue that the state plays a creative role in supporting innovation and economic development that goes beyond correcting market failures, to encompass the generation and pursuit of opportunities that may be too uncertain to attract private sector investment (Lazonick and Mazzucato, 2013). Examples include the important role of the US government in nurturing nascent technologies and industries such as the internet, nanotechnology, and electric vehicles (Fuchs, 2010; Mazzucato, 2015). Indeed, most of the significant innovations of the past few decades have resulted from collaborative efforts involving industry, university, and government actors (Block and Keller, 2015). There are also social arguments for business support, such as the provision of jobs for disadvantaged individuals, support for economic development in challenged regions, or assistance for declining industries (Uchitelle, 2017). Such interventions increase the welfare of the supported groups, but may not lead to increases in innovation, productivity, or overall net social welfare.

Business support takes a variety of forms. This paper considers grants and loans to companies and knowledge-based support delivered by third-party innovation intermediaries (alternative forms of direct support) and tax credits (indirect support). The Organisation for Economic Co-operation and Development (OECD) reports on direct and indirect support for industrial research and development (R&D). Direct support for industrial R&D totaled approximately \$43.7 billion across OECD and major economies in 2015, the most recent estimate (OECD, 2017a). Indirect support—R&D tax credits—is estimated at \$44.5 billion (OECD, 2015). In addition to providing support for R&D, governments provide support for new ventures, by sponsoring business incubators and accelerators, of which there are now estimated to be 7000 globally (INBIA, 2017), and support to small and medium sized enterprises (SMEs), largely through loan guarantees (OECD, 2017b). In addition, governments provide support for regions, by sponsoring economic development organisations and by providing economic incentives that entice firms to grow in, or relocate to, the sponsoring jurisdiction. New data estimate the value of these economic incentives at \$45 billion for 2015, for the US alone (Bartik, 2017). Governments provide further support to businesses through favourable environmental, employment, and accounting regulations (Lazonick and Mazzucato, 2013; Murphy et al., 2016).

## RCT strengths and limitations

Early articulations of the merits of RCTs were presented by American psychologist Donald Campbell and by Scottish doctor

Archibald (Archie) Cochrane. Campbell was noted for his work on methodology and his perspectives on the importance of experimentation in public policy (Campbell and Stanley, 1963). Conscious of the anecdotal nature of the basis for medical diagnosis and intervention, and of the moral and practical requirements for widespread, low cost, and effective health services, Cochrane made a case for effective and efficient interventions where effectiveness was tested through RCTs (Cochrane, 1972). While there can be no doubt that RCTs have made significant contributions to the advancement of knowledge and human welfare, the promotion of RCTs by some has led to a counter-vailing literature on their limitations. The contributors to this literature reject the notion that RCTs represent a methodological 'gold standard' (Cartwright, 2010), and regard the push for experimental evaluations as normative and reductionist, and insensitive to the importance of context and diversity in methodologies for social science research (Morrell, 2008; Pawson, 2013). In the following I review the strengths and limitations of RCTs.

**Strengths.** A significant challenge in evaluating the effectiveness of an intervention is the determination of what would have happened in the absence of the intervention. RCTs address this question of the counterfactual by ensuring that the subjects that receive the intervention are as much like the subjects that do not receive the intervention as possible. When the study sample is large and representative of the population of interest, RCTs produce a valid and reliable estimate of the Average Treatment Effect (ATE), the expected effect of the treatment on a member of the population. The similarity of treatment and control groups eliminates explanations for differential outcomes that are unrelated to the intervention, resulting in high internal validity (Deaton, 2010; Oakley, 1999).

But internal validity can be compromised by outlier responses to treatment, that is, subjects that respond exceptionally well to treatment (Deaton and Cartwright, 2016). This is an important caveat when contemplating the use of RCTs for the evaluation of business support programmes, because factors that are difficult to measure, such as managerial capabilities, may explain the responsiveness of firms to treatment. The outlier participants who perform exceptionally well may be the basis upon which a positive ATE for the programme is claimed if they are part of the treatment group, or the basis upon which a negative or null ATE results if they are part of the control group. Deaton and Cartwright (2016) use a simulation to show that a sample size of 1000 is required to compensate for a single positive outlier sufficiently well to have a better than 94% chance estimating the correct sign of the ATE.

If the sample is representative of the population of interest then external validity, or the degree to which the experimental results are replicable, will be high as well. But this is difficult to achieve because properly conducted RCTs make significant demands on subjects, making it difficult to choose representative trial settings (Campbell and Stanley, 1963; Deaton, 2010; Greenberg et al., 1999; Sanson-Fisher et al., 2007). Sanson-Fisher et al. (2007) describe a case where only 16 of 228 sites agreed to participate in a population health study. Clearly the 16 sites that agreed to participate in the study were in some respects different from the 212 sites that refused to participate, and so the external validity of the study is compromised. Where meta-analyses of multiple controlled studies are conducted, and where individual study results are in agreement, or where differences can be explained, such analyses provide evidence of effectiveness across multiple settings.

There are circumstances where the random allocation of treatment amongst subjects is, or is perceived as, the fairest approach. These circumstances include the common situation where resources are limited and treatment is desired by many. For example, in California, the state elected to randomly allocate film tax credits amongst eligible firms when the demand for such tax credits exceeded available resources (Weatherford, 2016). Similarly, international development interventions are sometimes allocated to randomly selected communities, in part in the interests of experimental evaluations, and in part because randomly allocating support may be perceived as fair.

**Limitations.** Experimental evaluations that focus on the estimation of the ATE have been criticized for distracting from efforts to understand why and how interventions produce outcomes. As well, they may inhibit learning, make high demands on subjects, and be costly. In the following I consider the limitations of RCTs.

*Narrowness of scope.* The most severe criticism of RCTs is that they are narrow in scope and rely on complementary investigations to provide a theoretical and practical context for their measurement results (Cartwright, 2010). Victora et al., (2004) distinguish between the statistical probability of an effect, the plausibility of an effect, for which evidence of the causal pathway is necessary, and the adequacy of an effect, whether or not the effect is of sufficient magnitude to be meaningful in substantive, rather than only in statistical, terms. RCTs address only the statistical probability of an effect. Pawson and Tilley (1997) conclude that '30 years of project evaluation in sociology, education, and criminology was largely unsuccessful because it focused on whether projects worked instead of on why they worked'.

The problem may be that social scientists who advocate a reliance on RCTs for establishing causality in social interventions may have misunderstood their role in medical research. Biomedical interventions are subjected to multiple investigations during their long gestation. For example, during the 17–24 years that elapse between investments in biomedical research and the commercialization of new molecular entities, potential new therapies are subjected to laboratory and animal testing, testing on small numbers of human beings, and finally clinical trials that employ RCTs and large samples (Toole, 2012). The causal pathways that connect the intervention to the effect are generally investigated prior to conducting the RCTs. The RCTs are used to determine effects in different population subgroups, not to determine whether or not there's an effect, or to establish the mechanisms that produce effects (Nixon, 2015). Biomedical researchers recognise that RCTs establish only the probability of effects and employ multiple methodologies to fully investigate how, why, under what conditions, and for whom interventions produce effects.

To some extent the criticism that RCTs examine only probably of an effect can be lodged against a range of methodologies for estimating an ATE. Quasi-experimental methodologies such as instrumental variables, difference-in-difference estimation, and propensity score matching, like RCTs, focus on the estimation of an ATE at the expense of investigating how, and for which subgroups, outcomes are significant. The difference is that because quasi-experimental designs are 'after-the-fact' evaluations (Jaffe, 2002), alternative samples can be constructed for additional analyses without compromising the design of the evaluation. This is not the case for RCTs. In an RCT, subjects are randomly allocated to treatment and control groups prior to the intervention. If, for example, ex post subgroup analysis is performed, where subgroups have not been established through random

allocation, then this *ex post* analysis does benefit from the experimental design (Deaton, 2010).

Of course, RCTs can be complemented with non-experimental studies of why programmes succeed or fail, for whom, and under what circumstances (Oakley et al., 2006). And studies can be designed explicitly to evaluate mechanisms. For example, rather than trying to estimate the effect of an educational program, Duflo et al. (2012) use an RCT to examine the procedural question of whether or not financial incentives reduce teacher absenteeism. Public health researchers have developed a range of techniques, including multi-arm studies and factorial trials, to augment the explanatory power of RCTs (Bonell et al., 2012).

*Negative effects on learning.* While some authors feel that RCTs are indispensable to learning and rigorous evaluation, even in fields where they difficult to conduct (Oliver et al., 2002; Walwyn and Wessely, 2005), others lament the effect that RCTs have on the ability of programme designers and managers to learn. At the portfolio level, a focus on RCTs as a preferred evaluation methodology means that programmes that are amenable to RCTs are more likely to be evaluated than those that are not (Ravallion, 2015). The standardisation of treatments that is inherent to an RCT prevents more micro level experimentation and learning, constraining the administration of programmes, and subordinating programme managers to the needs of evaluators (Perrin, 2002). Writing on the use of RCTs in population health, Sanson-Fisher et al. (2007) observe that given the need for flexible, broad, and complex interventions, a focus on those that can be tested by RCTs may threaten the development and evaluation of innovative interventions with potentially significant public health consequences.

*High demands made on subjects.* RCTs make high demands on subjects that go beyond the risks they would face by knowingly subjecting themselves to an experimental intervention. This is because a subject in an RCT doesn't know whether or not they will receive the treatment, and so they are less well equipped to deal with unanticipated effects and consequences. As observed by Campbell and Stanley, [The experimental design] 'is so demanding of cooperation on the part of respondents or subjects so as to end up with research done only on captive audiences rather than the general citizen of whom one would wish to speak' (Campbell and Stanley, 1963). RCTs are typically conducted in situations where subjects are vulnerable: medical trials are conducted on sick patients, trials in international development on the poor, trials in education on students, in training on the unemployed, and in studies of recidivism on the recently incarcerated. In a review of experimental social science, Greenberg et al. report that most RCTs in economics have been carried out by rich people on poor people and that 'the scarcity of experiments involving the middle and upper class is extraordinary' (1999, p 159).

*Costs.* As observed by Heckman and Smith (1995) RCTs, like all high-quality evaluations, can be expensive. The head of the World Bank Development Impact Evaluation Division estimates the average cost of an impact evaluation in the field of international development at \$500,000 (Nature, 2015). Sanson-Fisher et al. (2007) cite the costs of four RCTs in the field of population health (US\$2.5 million to US\$45 million), where the costs can be especially high if multiple pilot settings are used with an interest in increasing the generalizability of results. But costs are lower in fields such as education, and there may be opportunities for cost savings. Not all programmes are ready or suitable for evaluation, in some cases because they have not been running long enough to be operating smoothly or to have generated results, and in other

cases because their record keeping is inadequate, or because they are not sufficiently large (Petticrew et al., 2012). In such cases, low cost performance assessments should precede or supplant more costly approaches such as RCTs (Epstein and Klerman, 2012).

### The evaluation of Business Support Programmes

In the following, I consider the implications for evaluation of alternative modes of business support, focusing on the feasibility and advisability of experimental evaluations.

**Direct support—grants and loans.** Grants and loans range from large investments in single firms, which are often politically motivated and not subject to rigorous evaluation, to programmes that distribute funds to suitably qualified firms, typically on a competitive basis. One of the largest and longest running grant programmes is the Small Business Innovation Research (SBIR) programme in the US. Founded in 1982 to stimulate technological innovation and help meet federal R&D needs, the programme requires US federal agencies with extramural R&D budgets of more than \$100 million to allocate a share of their R&D budget, currently 3.2%, to SBIR programmes (SBIR, 2014). Programmes are highly competitive—approximately 13% of SBIR applicants are selected for Phase I funding, and approximately 50% of those are awarded Phase II funding (SBIR, 2017). Since 1982, SBIR programmes have awarded 112,500 grants worth a total of \$26.9 billion. Japan, the UK, and the Netherlands have developed SBIR-type programmes (OECD, 2010).

Being awarded a competitive grant has an impact on a company that persists beyond the lifetime of the award. Investing in early stage companies is risky because returns are subject to technological, market, and managerial uncertainties. It has been shown that firms that receive R&D grants from competitive programmes are more likely to receive downstream equity investments than firms that do not receive grants, possibly due to the quality signal of the award (Feldman and Kelley, 2006; Meuleman and De Maeseeneire, 2012). The process of selecting companies for grants may also have an effect on company behaviours as programmes may provide feedback to both successful and unsuccessful applicants. Even where specific feedback is not provided, the acceptance or rejection itself is informative and may prompt firms to change their behaviours.

There have been three highly regarded evaluations of the SBIR programme. The first two evaluations produced conflicting results. Employing a matched sample approach, Lerner, (2000) found that in regions with substantial venture capital activity, firms that receive SBIR funding experience greater growth in revenues and employment, and are more likely to attract venture capital financing, than firms that do not receive funding. Using an instrumental variable to account for observed and unobserved differences between firms that did and did not receive SBIR funding, Wallsten, (2000) found that the differences in employment growth and R&D investments between funded and unfunded firms disappear, suggesting that SBIR funds 'crowd out' the private investments in R&D.

Given the limitations and disappointing results of these evaluations, it is tempting to envision a definitive experimental evaluation of the SBIR programme. But it would be difficult to conduct such an evaluation without diminishing the effects of the programme in the process. First, the random allocation of funds, even amongst a subset of highly-qualified applicants, may result in the funding of less meritorious applicants, and reduced outcomes. Also, randomisation would diminish the signaling effect of the award to downstream equity investors, and would diminish the learning opportunities associated with the acceptance–rejection feedback. And because outcomes are

expected to be skewed, with a small number of companies responsible for most of the impact, large samples will be required for reliable results (Deaton and Cartwright, 2016). So the most recent and rigorous evaluation of SBIR uses a regression discontinuity design rather than an RCT (Howell, 2017). Based on data from over 5000 applicants to the programme between 1983 and 2013, Howell, like Lerner, finds that winning an SBIR award makes it more likely that a firm will attract venture capital financing and increase revenues, especially where the firm is young and operates in an emerging sector.

There have been several recent RCT-based evaluations of grant programmes in which the amount of funding awarded is low. These include evaluations of microcredit programmes in Hyderabad and Morocco, and the Creative Credits programme in the UK. In all cases the outcomes were modest. The programme in Hyderabad distributed loans of \$200 and found that after 1.5 years recipients were no more likely than non-recipients to own a business or start a new business, although they were more likely to invest in existing businesses (Banerjee et al., 2015). The programme in Morocco found that recipients experienced no gain in income or consumption relative to non-recipients, although households identified as more likely to borrow saw a rise in investment assets used for self-employment activities (Crépon et al., 2015). The Creative Credits programme saw the random allocation of £4000 vouchers for consulting services to qualifying SMEs, finding effects on innovation after 6 months, which disappeared after 12 months (Bakhshi et al., 2015). Collectively, these evaluations show that small amounts of funding do not bring about transformative effects, at least not when they are randomly allocated.

**R&D tax credits.** One of the most widely used tools for supporting innovation is the R&D tax credit, used by 29 of 35 OECD countries (OECD, 2017a). Because R&D tax credits are widely used and the intervention is easily measured, there is an extensive literature on their effects. In terms of input additionality, studies show that for every \$1 of forgone tax revenues, firms spend approximately an additional \$1 on R&D (Köhler et al., 2012). Studies of the impact of R&D tax credits on firm outputs report an increased probability of new product development, but no impact on economic performance (Czarnitzki et al., 2011; Köhler et al., 2012). The high cost of R&D tax credits, combined with the evidence of their marginal impacts, has focused attention on where their impact is greatest. A review of the literature suggests that R&D tax incentives be targeted at firms that are more highly innovative and industry sectors with a higher propensity to conduct R&D (Palazzi, 2011).

Given the fact that the number of firms receiving R&D tax credits is large, and that the treatment is quantifiable, it would seem that R&D tax credits would constitute an ideal setting for an RCT. But there has yet to be an RCT-based evaluation of R&D tax credits (De Blasio et al., 2014; Bronzini and Iachini, 2014). The random allocation of R&D tax credits is not advisable because R&D projects typically unfold over multiple years and there is often a significant lag between the time a firm invests in R&D, and the time when it is able to capitalise on that investment. Unpredictability in the allocation of R&D tax credits, as would be required by their random allocation, will diminish a firm's ability to plan its R&D spending, and may have a negative effect on its R&D behaviour or performance.

Also, the random allocation of financial resources for businesses, in the interests of rigorous evaluation, is likely politically infeasible. Business leaders, who may have financial resources and relationships with decision makers at their disposal, would campaign vigorously against the random allocation of a

resource on which they had come to depend. Jurisdictions compete with one another for large and growing businesses that offer jobs, and the random allocation of R&D tax credits might be enough to cause business leaders to threaten to relocate to a more 'business-friendly' jurisdiction. Politicians are likely to find the views of business leaders more persuasive than those of evaluators (Andrews, 2017; Cairney, 2016).

**Support delivered by third-party innovation intermediaries.**

Funding provided by governments to third-party innovation intermediaries is used to provide knowledge-based services, sometimes offered in combination with funding, to client companies. Knowledge-based services include R&D services, access to specialised equipment, and business services including training and advice on sales and marketing, intellectual property, exporting, etc. (Howells, 2006). Programmes are necessarily small and limited to firms in specific regions or industries, or to firms at specific stages of development or with certain technological requirements. Only a fraction of the firms in a country conduct R&D and of those, few will choose to avail themselves of the services of an innovation intermediary (Beise and Stahl, 1999). This is in part because the services of such organisations rely on substantive capabilities in specific areas that will not be useful for all companies.

Research has shown that firms with high absorptive capacity are more likely to collaborate (Eom and Lee, 2010) and that collaborators sometimes develop long-term relationships that allow the firm and the intermediary to work together on problems that exhibit greater 'information gaps' over time (Izushi, 2003). Relationships of greater intensity are associated with greater outcomes (Autio et al., 2008; Barge-Gil and Modrego, 2011). Even for new ventures, generic support is of little value and successful programmes rely on a complex combination of human capital, networks and experience, which must be built over time (Hochberg, 2016). Here too, firms with greater absorptive capacity are more likely to receive advisory services and coachable entrepreneurs experience greater impacts (Cumming and Fischer, 2012).

As a consequence of the importance of the characteristics of the firms being served, many programmes are highly selective and admit only the most promising firms. Such programmes aim to combine selection and treatment effects for superior outcomes. Czarnitzki and Delanote (2012) show that companies identified as young innovative companies (YIC), that is, small, young companies whose R&D expenses exceed 15% of their revenues, grow more quickly than other new, technology-based firms that do not meet the YIC requirement. The YIC programme offered by Business Finland is an example of a programme that selects such companies for support, in the hopes that with support, they will grow large enough to have a significant effect on the national economy. To be eligible for the programme, the new venture must exhibit strong growth motivation, competent management, and a compelling competitive advantage. The programme admits approximately 30 companies a year and provides them with bridging services and up to about €1 million in funding (Autio and Ranniko, 2016). A comparison of the sales of companies treated by the YIC programme and a control group of unsuccessful applicants to the programme shows that 3 years after the intervention, treated companies have mean sales of over €2.5 million, while the control group has mean sales of about €1 million, both well above typical mean sales for young companies. Controlling for selection effects shows that the estimated treatment effect of the programme is an increase in sales of 130% after 3 years (Autio and Ranniko, 2016). Clearly an evaluation technique that demands the random assignment of

treatments would be unsuitable for the evaluation of the Business Finland YIC programme and other programmes that select suitable companies for treatment.

RCTs have been used to evaluate knowledge-based services to aspiring entrepreneurs, mostly in developing countries. For example, Calderon et al. (2013) used a RCT design to evaluate the effect of business training on female entrepreneurs in rural Mexico. The 165 treated entrepreneurs received six 8-h modules of training and 30-page 'textbooks' to accompany each model, a more substantive treatment than was offered by many previous experimental business training programmes in developing countries. The authors find that those assigned to treatment earn higher profits, have larger revenues, serve a greater number of clients, are more likely to use formal accounting techniques, are more likely to be registered with the government, and that the effects are sustained through the medium-term (Calderon et al., 2013). Martinez et al. (2016) evaluate the effect of a programme that combines business training and funding for entrepreneurs in Chile. They find that the programme increases employment in the short run through self-employment, and in the long run through wage work. Fairlie et al. (2015) conduct an experimental evaluation of an entrepreneurship training programme in the US. They find that those receiving the modest amount of training (the cost of which was estimated at \$1321) are significantly more likely to own a business after 6 months, but that after 18 or 60 months there are no effects of having participated in the programme.

In summary, RCTs have shown that standard knowledge-based services may be effective for entrepreneurs in developing countries with little business knowledge. But such programmes are unlikely to be very effective in advanced countries where the demand for relevant, customised knowledge is greater. And there is a high opportunity cost to denying continued support in cases where support delivery involves knowledge-intensive relationships that have been developed over multiple years between highly qualified support providers and recipients.

## Implications

RCTs are most appropriate for the evaluation of policy interventions in general, and BSPs in particular, when:

**1. A meaningful and cost-effective contribution to knowledge is envisioned.** This may occur when it is not known whether or not an intervention, or a feature of an intervention, is likely to have an effect or the magnitude of the effect; the envisioned RCT is likely to advance understanding regarding the presence or magnitude of the effect; and the benefits of the knowledge generated are likely to exceed the costs, broadly understood, of the measurement exercise (Petticrew et al., 2012). For BSPs, these three criteria will be difficult to meet. In other cases, judgment will be required to determine the degree to which previous evaluations are pertinent in determining whether or not an effect is likely, whether or not past estimates of its magnitude are useful, and whether or not the presence or magnitude of an effect is likely to be measurable (Pearce et al., 2015).

**2. The treatment exhibits low variability.** Treatments need not be uniform, but there must be sufficiently low-variability in treatments for it to be clear what is, and what is not, a treatment. This rules out interventions that are highly customised and treatments that build on past interventions the effects of which may still be unfolding, both of which are common in the case of business support. Some, but not all, are of the view that RCTs are problematic in psychiatry for similar reasons (Walwyn and Wessely, 2005). In fields where it is expected, or at least desirable, that actual treatments adhere to protocol, techniques have been

developed to measure and address deviations (Mowbray et al., 2003). Fidelity to key intervention processes and functions is more important than fidelity to activities (Bonell et al., 2012) and research has shown that both fidelity and adaptation are important (Durlak and Dupre, 2008).

**3. The random allocation of treatments does not severely diminish their effectiveness.** As has been observed by many, human beings are purposeful and will adjust their behaviours in response to changing conditions. When treatments are randomly allocated, this may have an effect on the population of subjects that seek treatment. This is problematic for interventions whose outcomes depend on the selection of subjects, and for treatments that are ineffective if their availability is not predictable. Also, subjects must be similar with respect to amenability to treatment. Where amenability to treatments is highly variable, outcomes will be highly skewed, and very large samples will be required to produce reliable results (Deaton and Cartwright, 2016).

**Implications for policymakers.** For policymakers there is an important precursor to determining how to evaluate a BSP and that is establishing the rationale for the programme. This involves identifying the economic or innovation systems logic that justifies the intervention, the selection of the targets of the intervention, and the causal pathway, or theory of change, that leads from intervention to outcome of interest.

Some programmes are designed to be general, while others are designed to be selective. General programmes serve all eligible companies, while selective programmes serve companies that are selected for support on the basis of demanding qualification requirements that few companies can attain, or on the basis of a competitive process. The justification for selective programmes is that the economy and citizens benefit disproportionately from highly successful firms that provide employment opportunities, tax revenues, and innovation system leadership. Policy makers must decide whether they want to offer only general business support or whether they want to engage in selective business support. The choice has implications for evaluation, as selective programs cannot be evaluated using RCTs.

A second implication is that evaluators are not passive observers. Habicht et al., (1999, p 14) warn that 'key individuals in donor or international agencies, as well as the evaluators themselves, may have been trained to regard probability assessments [RCTs] as the gold standard and fail to understand that this approach is seldom mandatory or even feasible for the routine evaluation of programme effectiveness'. In addition to methodological preferences, evaluators may ascribe to an overly rational view of governance. As Pearce and Raman observe, 'There is a danger that the current UK government's interest in RCTs is driven not by their methodological suitability, but because they lend themselves to a model of governance that values context-free quantification and benchmarking' (Pearce and Raman, 2014, p 398).

A final comment concerns the political feasibility of RCT-based evaluations of BSPs. Even admirers of RCTs concede that they are likely politically infeasible for the evaluation of BSPs (Jaffe, 2002; Storey, 1998). The reason for this is that, unlike the vulnerable populations that are typically the subjects of RCTs (Greenberg et al., 1999), business people have voice. More than other members of society, they are independent and, in many cases, influential. This does not stop them from arguing vociferously for government support, but it makes it all but impossible for governments to distribute that support without consideration for the concerns of the would-be recipients.

**Table 1 Why there are (almost) no RCTs of BSPs****Problems with using RCTs for evaluating BSPs**

1. Random allocation of support. Randomly allocating support will greatly diminish the effectiveness of support as the capabilities of companies is a key determinant of outcomes.

2. Standardised treatments. Standardised treatments are problematic in cases where effectiveness depends on interventions designed to address specific challenges employing specialised knowledge and long-term generative relationships.

3. Reliability of results. BSPs aim to produce outliers—firms whose performance is exceptional. When outliers are present, very large samples will be required to produce reliable results.

4. Contribution to knowledge. The benefits will rarely outweigh the costs. RCTs are best for determining the presence and size of an effect under controlled conditions. Business outcomes are heavily influenced by external factors and so the presence and size of an effect will vary greatly across implementations.

**Exceptions and alternatives**

It may be possible to randomise some support for ventures of modest potential—microcredit and business training interventions have been evaluated using RCTs.

Purely financial support is intrinsically standardised. It may be possible to standardise some knowledge-based support in terms of key intervention processes and functions without unduly compromising the quality of support.

Econometric and theory-based qualitative evaluations are the alternatives to RCTs.

Evaluations should consider the mechanisms by which effects are produced and how effects vary across treated subjects and implementations.

**Implications for evaluators.** Evaluators should begin by understanding the programme's rationale, client base, and expected impacts. Programmes need to be ready for evaluations, especially evaluations that are costly and time-consuming (Epstein and Klerman, 2012; Smith, 2004). Also, the intervention must be of a magnitude that is sufficient to generate a discernable effect and the measurement of impact must be sufficiently proximate to the phenomenon to be able to capture that effect (Bonell et al., 2011). Where programmes are light and, e.g., provide general information, networking opportunities, or modest amounts of funding, there can be little expectation of a measurable effect on company performance. And regardless of the research design, evaluators need to be sensitive to the appropriateness of their assumptions and the limitations of their data and methodology.

RCTs may be appropriate where there is no effectiveness penalty to low-variability treatments that are randomly allocated. Where this is not the case, evaluators can use techniques such as natural experiments and quasi-experimental techniques as alternatives to RCTs. Natural experiments leverage naturally occurring and unanticipated changes in policy or circumstances to introduce variability in treatments. Quasi-experimental techniques include regression discontinuity design, instrumental variables, difference-in-differences estimation, and propensity score matching, all of which use econometrics to statistically isolate treatment effects (Hottenrott et al., 2017). Also, there are recent extensions to quasi-experimental techniques that extend their reach and internal validity, allowing evaluators to address a broader range of situations and produce more reliable estimates of treatment effects.

**Conclusion**

A premise of medical interventions is that we want every human being to be healthy, or as healthy as our knowledge, resources, and generosity allows. The same is not true of government interventions in support of businesses. Most people are not concerned about the health of very many companies, and in many cases, would prefer it if the health and power of some companies were diminished. For dis-interested citizens, companies are but the means to desired ends: needed or coveted products and services, pride in local or national successes, jobs for employees, training for students, financial returns for investors, taxes for governments, philanthropic support for local or national sports teams, innovative business ecosystems, etc. We want

companies to survive and grow, not for intrinsic reasons, but because companies are instrumental to the ends we seek. The implication is that business support programmes are and should be designed to achieve the greatest effects, not to help all companies. The primary explanation for why there are (almost) no RCT-based evaluations of government interventions in support of business, is that the requirement for the random allocation of support makes it impossible to select companies for support based on factors that are known to increase the likelihood of successful outcomes (Table 1).

Received: 3 April 2017 Accepted: 18 January 2018

Published online: 06 February 2018

**References**

- Anderson N, Potočník K, Zhou J (2014) Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. *J Manag* 40(5):1297–1333
- Andrews L (2017) How can we demonstrate the public value of evidence-based policy making when government ministers declare that the people 'have had enough of experts'? *Palgrave Commun* 3(1):11
- Autio E, Rannikko H (2016) Retaining winners: can policy boost high-growth entrepreneurship? *Res Policy* 45(1):42–55
- Autio E, Kanninen S, Gustafsson R (2008) First- and second-order additionality and learning outcomes in collaborative R&D programs. *Res Policy* 37(1):59–76
- Bakhshi H, Edwards JS, Roper S, Scully J, Shaw D, Morley L et al. (2015) Assessing an experimental approach to industrial policy evaluation: applying RCT to the case of creative credits. *Res Policy* 44(8):1462–1472
- Banerjee A, Duflo E, Glennerster R, Kinnan C (2015) The miracle of microfinance? Evidence from a randomised evaluation. *Am Econ J Appl Econ* 7(1):22–53
- Barge-Gil A, Modrego A (2011) The impact of research and technology organizations on firm competitiveness. Measurement and determinants. *J Technol Transf* 36(1):61–83
- Bartik TJ (2017) A new panel database on business incentives for economic development offered by state and local governments in the United States. Prepared for the Pew Charitable Trusts
- Beise M, Stahl H (1999) Public research and industrial innovations in Germany. *Res Policy* 28(4):397–422
- Block FL, Keller MR (2015) State of innovation: the US government's role in technology development. Routledge, New York
- Bonell C, Fletcher A, Morton M, Lorenc T, Moore L (2012) Realist randomised controlled trials: a new approach to evaluating complex public health interventions. *Social Sci Med* 75(12):2299–2306
- Bonell CP, Hargreaves J, Cousens S, Ross D, Hayes R, Petticrew M et al. (2011) Alternatives to randomisation in the evaluation of public health interventions: design challenges and solutions. *J Epidemiol Commun Health* 65(7):582–587

- Bronzini R, Iachini E (2014) Are incentives for R&D effective? Evidence from a regression discontinuity approach. *Am Econ J: Econ Policy* 6(4):100–134
- Burtless G (1995) The case for randomised field trials in economic and policy research. *J Econ Perspect* 9(2):63–84
- Cairney P (2016) *The politics of evidence-based policy making*. Springer, Berlin
- Calderon G, Cunha JM, De Giorgi G (2013) *Business literacy and development: evidence from a randomised controlled trial in rural Mexico*. Barcelona Graduate School of Economics Working Paper Series
- Campbell DT, Stanley JC (1963) *Experimental and quasi-experimental designs for research*. Houghton Mifflin Company, Boston
- Cartwright N (2010) What are randomised controlled trials good for? *Philos Stud* 147(1):59–70
- Cochrane AL (1972) Effectiveness and efficiency: random reflections on health services. The Nuffield Provincial Hospital Trust
- Crépon B, Devoto F, Dufo E, Parienté W (2015) Estimating the impact of microcredit on those who take it up: evidence from a randomised experiment in Morocco. *Am Econ J: Appl Econ* 7(1):123–150
- Crespi G, Maffioli A, Mohnen P, Vázquez G (2011) *Evaluating the impact of science, technology and innovation programs: a methodological toolkit*. Inter-American Development Bank, Washington
- Cumming DJ, Fischer E (2012) Publicly funded business advisory services and entrepreneurial outcomes. *Res Policy* 41(2):467–481
- Czarnitzki D, Delanote J (2012) Young innovative companies: the new high-growth firms *Ind Corp Chang* 22(5):1315–1340
- Czarnitzki D, Hanel P, Rosa JM (2011) Evaluating the impact of R&D tax credits on innovation: a microeconomic study on Canadian firms. *Res Policy* 40(2):217–229
- De Blasio G, Fantino D, Pellegrini G (2014) Evaluating the impact of innovation incentives: evidence from an unexpected shortage of funds. *Ind Corp Change* 24(6):1285–1314
- Deaton A (2010) Instruments, randomization, and learning about development. *J Econ Lit* 48(2):424–455
- Deaton A, Cartwright N (2016) *Understanding and misunderstanding randomised controlled trials*. National Bureau of Economic Research Working Paper No. 22595
- Dougherty D (2016) *Taking advantage of emergence: productively innovating in complex innovation systems*. Oxford University Press, New York
- Dufo E, Hanna R, Rya SP (2012) Incentives work: getting teachers to come to school. *Am Econ Rev* 102(4):1241–1278
- Durlak JA, DuPre EP (2008) Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am J Community Psychol* 1(3-4):327–350
- Economist (2016) Economists are prone to fads, and the latest is machine learning. <http://www.economist.com/news/finance-and-economics/21710800-big-data-have-led-latest-craze-economic-research-economists-are-prone>. Accessed 20 Mar 2017
- Eom B, Lee K (2010) Determinants of industry–academy linkages and, their impact on firm performance: The case of Korea as a latecomer in knowledge industrialization. *Res Policy* 39(5):625–639
- Epstein D, Klerman JA (2012) When is a program ready for rigorous impact evaluation? The role of a falsifiable logic model. *Eval Rev* 36(5):375–401
- Fairlie RW, Karlan D, Zinman J (2015) Behind the GATE experiment: Evidence on effects of and rationales for subsidized entrepreneurship training. *Am Econ J Econ Policy* 7(2):125–161
- Feldman MP, Kelley MR (2006) The ex ante assessment of knowledge spillovers: Government R&D policy, economic incentives and private firm behavior. *Res Policy* 35(10):1509–1521
- Freel MS (2007) Are small innovators credit rationed? *Small Bus Econ* 28(1):23–35
- Fuchs ER (2010) Rethinking the role of the state in technology development: DARPA and the case for embedded network governance. *Res Policy* 39(9):1133–1147
- Greenberg D, Shroder M, Onstott M (1999) The social experiment market. *J Econ Perspect* 13(3):157–172
- Habicht JP, Victora CG, Vaughan JP (1999) Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *Int J Epidemiol* 28(1):10–18
- Hall BH, Mairesse J, Mohnen P (2010) Measuring the returns to R&D. *Handb Econ Innov* 2:1033–1082
- Heckman JJ, Smith JA (1995) Assessing the case for social experiments. *J Econ Perspect* 9(2):85–110
- Hochberg YV (2016) Accelerating entrepreneurs and ecosystems: the seed accelerator model. *Innov Policy Econ* 16(1):25–51
- Hottenrott H, Lopes-Bento C, Veugelers R (2017) Direct and cross scheme effects in a research and development. *Res Policy* 46(6):1118–1132
- Howell T (2017) Financing innovation: evidence from R&D grants. *Am Econ Rev* 107(4):1136–1164
- Howells J (2006) Intermediation and the role of intermediaries in innovation. *Res Policy* 35(5):715–728
- INBIA (International Business Incubator Association) (2017) *Business incubators*. Originally available at <https://inbia.org/>; Now available at [https://en.wikipedia.org/wiki/Business\\_incubator](https://en.wikipedia.org/wiki/Business_incubator). Accessed 17 Jan 2018
- Izushi H (2003) Impact of the length of relationships upon the use of research institutes by SMEs. *Res Policy* 32(5):771–788
- Jaffe AB (2002) Building programme evaluation into the design of public research-support programmes. *Oxford Rev Econ Policy* 18(1):22–34
- Jaffe AB, Newell RG, Stavins RN (2005) A tale of two market failures: technology and environmental policy. *Ecol Econ* 54(2):164–174
- Jaffe AB, Trajtenberg M, Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *Q J Econ* 108(3):577–598
- Köhler C, Laredo P, Rammer C (2012) The impact and effectiveness of fiscal incentives for R&D. Compendium of evidence on the effectiveness of innovation policy. Part of the Compendium of Evidence on the Effectiveness of Innovation Policy Intervention Project led by the Manchester Institute of Innovation Research (MIOIR), University of Manchester, UK
- Lazonick W, Mazzucato M (2013) The risk-reward nexus in the innovation-inequality relationship: who takes the risks? Who gets the rewards? *Ind Corp Change* 22(4):1093–1128
- Lerner J (2000) The government as venture capitalist: the long-run impact of the SBIR program. *J Priv Equity* 3(2):55–78
- Mazzucato M (2015) *The entrepreneurial state: debunking public vs. private sector myths*. Anthem Press, New York
- Martinez CA, Ruiz-Tagle J and Esteban Puentes (2016) *Microentrepreneurship support programme in Chile*. [https://www.researchgate.net/publication/43294822\\_Micro-Entrepreneurship\\_Support\\_Programme\\_in\\_Chile\\_Impact\\_Evaluation](https://www.researchgate.net/publication/43294822_Micro-Entrepreneurship_Support_Programme_in_Chile_Impact_Evaluation). Accessed 11 Dec 2016
- Meuleman M, De Maeseineire W (2012) Do R&D subsidies affect SMEs' access to external financing? *Res Policy* 41(3):580–591
- Morrell K (2008) The narrative of 'evidence-based' management: a polemic. *J Manag Stud* 45(3):613–635
- Mowbray CT, Holter MC, Teague GB, Bybee D (2003) Fidelity criteria: development, measurement, and validation. *Am J Eval* 24(3):315–340
- Murphy G, Siedschlag I, McQuinn J (2016) Employment protection and industry innovation. *Ind Corp Change* 26(3):379–398
- Nature (2015) Can randomised trials eliminate global poverty? <http://www.nature.com/news/can-randomised-trials-eliminate-global-poverty-1.18176>. Accessed 17 Jan 2018
- Nixon J (2015) The big debate: RCTs and realist evaluation. University of Leeds. [https://www.youtube.com/watch?v=m0I5O2v\\_8MM](https://www.youtube.com/watch?v=m0I5O2v_8MM) The relevant passages occur at 12:55-14:22, 58:57-1:00:08, and 1:11:49-1:13:48. Accessed 17 Jan 2018
- Oakley A (1999) Paradigm wars: some thoughts on a personal and public trajectory. *Int J Social Res Methodol* 2(3):247–254
- Oakley A, Strange V, Bonell C, Allen E, Stephenson J (2006) Process evaluation in randomised controlled trials of complex interventions. *BMJ* 332(7538):413–416
- OECD (2010) *Public procurement programmes for small firms – SBIR like programmes*. [www.oecd.org/innovation/policyplatform/48136807.pdf](http://www.oecd.org/innovation/policyplatform/48136807.pdf). Accessed 17 Jan 2018
- OECD (2015) *OECD science, technology and industry scoreboard 2015: innovation for growth and society*. OECD Publishing, Paris
- OECD (2017a) *Main science and technology indicators, volume 2016 issue 2, table 36—percentage of business enterprise expenditure on R&D (BERD) financed by government*. OECD Publishing, Paris
- OECD (2017b) *Financing SMEs and entrepreneurs 2016: an OECD Scoreboard*. OECD Publishing, Paris
- Oliver P, Piachaud J, Done J, Regan A, Cooray S, Tyrer P (2002) Difficulties in conducting a randomised controlled trial of health service interventions in intellectual disability: implications for evidence-based practice. *J Intellect Disabil Res* 46(4):340–345
- Palazzi P (2011) *Taxation and innovation*. OECD Taxation Working Papers Number 9. <http://dx.doi.org/10.1787/22235558>
- Pawson R (2013) *The science of evaluation: a realist manifesto*. Sage, UK
- Pawson R, Tilley N (1997) *Realistic evaluation*. Sage, UK
- Pearce W, Raman S (2014) The new randomised controlled trials (RCT) movement in public policy: challenges of epistemic governance. *Policy Sci* 47(4):387–402
- Pearce W, Raman S, Turner A (2015) Randomised trials in context: practical problems and social aspects of evidence-based medicine and policy. *Trials* 16(1):394
- Perrin B (2002) How to—and how not to—evaluate innovation. *Evaluation* 8(1):13–28
- Petticrew M, Chalabi Z, Jones DR (2012) To RCT or not to RCT: deciding when 'more evidence is needed' for public health policy and practice. *J Epidemiol Community Health* 66(5):391–396
- Ravallion M (2015) *The World Bank: why it is still needed and why it still disappoints*. Center for Global Development Working Paper No. 400. Social

- Science Reseach Network <http://dx.doi.org/10.2139/ssrn.2623133> Accessed 17 Jan 2018
- Sanson-Fisher RW, Bonevski B, Green LW, D'Este C (2007) Limitations of the randomised controlled trial in evaluating population-based health interventions. *Am J Prev Med* 33(2):155–161
- SBIR (2014) SBIR Policy Directive. <https://www.sbir.gov/about/about-sbir#sbir-mission>. Accessed 17 Jan 2018
- SBIR (2017) Proposal success rate of all 2013 SBIR Phase 1 Awards (Phase 1 awards/Phase 1 proposals received). SBIR-STTR. <https://www.sbir.gov/node/736115>. Accessed 17 Jan 2018
- Smith J (2004) Evaluating local economic development policies: theory and practice. In: Nolan A, Wong G (eds) *Evaluating local economic and employment development: how to assess what works among programmes and policies*. Organisation for Economic Cooperation and Development, p 287–332. Paris
- Storey DJ (1998) *Six steps to heaven: evaluating the impact of public policies to support small businesses in developed economies*. Centre for Small & Medium Sized Enterprises, Warwick Business School. UK
- Toole AA (2012) The impact of public basic research on industrial innovation: evidence from the pharmaceutical industry. *Res Policy* 41(1):1–12
- Uchitelle L (2017) *Making it: why manufacturing still matters*. The New Press, New York
- Victoria CG, Habicht JP, Bryce J (2004) Evidence-based public health: moving beyond randomised trials. *Am J Public Health* 94(3):400–405
- Wallsten SJ (2000) The effects of government-industry R&D programs on private R&D: the case of the Small Business Innovation Research program *Rand J Econ* 31(1):82–100
- Walwyn R, Wessely S (2005) RCTs in psychiatry: challenges and the future. *Epidemiol Psychiatr Sci* 14(3):127–131
- Weatherford B, Sisney J (2016) California's first Film Tax Credit Programme. Legislative Analyst's Office. [www.lao.ca.gov/reports/2016/3502/First-Film-Tax-Credit-Prog-092916.pdf](http://www.lao.ca.gov/reports/2016/3502/First-Film-Tax-Credit-Prog-092916.pdf). Accessed 14 Mar 2017

## Acknowledgements

The author thanks the Menton campus of Sciences Po for kindly hosting her 2017 academic leave.

## Additional information

**Competing interests:** The author is both an academic and a principal with an evaluation consultancy. The author declares no competing financial interests.

**Reprints and permission** information is available online at <http://www.nature.com/reprints>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018