

Using RCTs as a research method for SME policy research: The UK experience

ERC Research Paper 66

April 2018

Using RCTs as a research method for SME policy research: The UK experience

Stephen Roper

Enterprise Research Centre and Warwick Business School
stephen.roper@wbs.ac.uk

The Enterprise Research Centre is an independent research centre which focusses on SME growth and productivity. ERC is a partnership between Warwick Business School, Aston Business School, Imperial College Business School, Strathclyde Business School, Birmingham Business School and Queen's University School of Management. The Centre is funded by the Economic and Social Research Council (ESRC); Department for Business, Energy & Industrial Strategy (BEIS); Innovate UK, the British Business Bank and Intellectual Property Office. The support of the funders is acknowledged. The views expressed in this report are those of the authors and do not necessarily represent those of the funders.

ABSTRACT

RCTs have a number of theoretical advantages over more standard quasi-experimental or econometric research methodologies, particularly the avoidance of selection bias. In the UK, budgetary pressures have created a renewed interest in evidence-based policy-making and one aspect of this has been the implementation of a series of randomised controlled trials (RCTs) to investigate the effectiveness of different types of SME support. These focus on the value of external business advice, innovation support, leadership coaching, management and leadership training and the effectiveness of alternative forms of communicating with firms. We briefly describe the trials and identify the operational and tactical issues which arise in the implementation of research experiments based on an RCT approach. Experience from medical trials also suggests the value of replication and synthesis, epitomised in the Cochrane Reviews, suggesting the value of repeated trials.

Key words: Randomised Control Trial; SME policy; UK.

Acknowledgements

Valuable comments were provided by Prof Mark Hart, ERC and Aston Business School and James Phipps, Innovation Growth Lab, Nesta. Opinions expressed are those of the author alone.

CONTENTS

ABSTRACT	3
1. INTRODUCTION	5
2. RCTS AS A RESEARCH APPROACH TO ENTERPRISE POLICY ...	8
3. POLICY RCTS IN PRACTICE – THE UK EXPERIENCE	12
3.1 Creative Credits	13
3.2 Growth Vouchers	16
3.3 The Small Business Charter (SBC) Growth Vouchers Programme	18
3.3 The Growth Impact Pilot (GIP)	19
3.4 Innovation Vouchers.....	20
3.5 Messaging Trials.....	21
4. DISCUSSION	22
5. CONCLUSIONS	24
REFERENCES	26

1. INTRODUCTION

When David Laws the new Chief Secretary to the UK Treasury first arrived at his desk after the 2010 general election he found a one line note from his predecessor Liam Byrne. It said: 'Dear chief secretary, I'm afraid to tell you there's no money left'.¹ By 2012 the position had changed relatively little with George Osborne the UK Chancellor of the Exchequer quoted as saying: 'The British Government has run out of money because all the money was spent in the good years. The money and the investment and the jobs need to come from the private sector'.² The implications of the resulting period of 'austerity' in the UK have been significant, with real terms reductions in many areas of public spending, and a re-evaluation of the effectiveness of many policy interventions. One consequence has been a renewed interest in 'What works' and the evidence base which underpins policy intervention. In terms of the science and research budgets this is evident in an emphasis on 'impact', i.e. 'the demonstrable contribution that excellent research makes to society and the economy'.³ New research organisations have also been created specifically oriented towards providing government with information on policy effectiveness in specific policy arenas. 'The initiative is based on the principle that good decision-making should be informed by the best available evidence on both what works and what does not work. What Works is a world first: it's the first time any government has taken a national approach to prioritising the use of evidence in decision-making'.⁴ Ten 'What Works Centres' have been established focussing on topics including: health and social care, educational achievement, crime reduction, early years intervention, local economic growth and improved quality of life for older people

¹ Source: Owen, P (2010) 'Ex-Treasury secretary Liam Byrne's note to his successor: there's no money left', The Guardian, 17th May 2010. Accessed online 3/10/2014 at www.theguardian.com.

² Source: Morgan, R (2012) 'George Osborne: UK has run out of money', The Telegraph, 26th February 2012. Accessed online 3/10/2014 at www.telegraph.co.uk.

³ Source: Economic and Social Research Council. Accessed on line 3/10/2014 at www.esrc.ac.uk/funding-and-guidance/impact-toolkit/what-how-and-why/what-is-research-impact.aspx.

⁴ Source: <https://www.gov.uk/guidance/what-works-network>. Accessed 3/12/2017

The renewed focus on evidence of policy effectiveness has also led to increased interest in the value of experimental evaluation methods involving randomised control trials (RCTs). Although unusual in terms of industrial policy, such experimental methods of policy evaluation are well established in medicine, social policy and development economics. Burtless (1995, p. 63), for example, cites Greenberg and Schroder (1991) who 'identified more than 90 separate field trials involving a wide range of distinctive research areas including health insurance, prisoner rehabilitation, labour supply, worker training and housing subsidies'. Banerjee and Duflo (2008, p. 32) also describe the 'recent surge in experimental work' in development economics. In the context of small business policy evaluation, Potter and Storey (2007), however, provide an extensive review of best practice in OECD countries without any mention of either the application or potential for experimental methods. Similarly, UK government guidance on industrial policy evaluation focuses entirely on non-experimental ex post evaluation approaches (BIS 2009). Related observations might be made in relation to the evaluation of R&D and innovation policies: despite the increasing importance of evaluation as part of the process of development of technology policy, evaluation approaches have until recently remained almost universally ex post and non-experimental (Laredo 1997).

Over the period since 2010 the UK five publicly funded RCT projects have been undertaken in the UK designed to test aspects of small business and innovation policy. The first of these was the Creative Credits project which ran from 2009-11 and tested the value of vouchers to link SMEs to creative enterprises (Bakhshi et al. 2015). The second, and largest experiment - the 'Growth Vouchers' programme - was ran from 2013-15 with a budget of around \$45m, and aimed to evaluate the effectiveness of external advice in promoting small business growth and the comparative value of an online and face-to-face diagnostic. Five different types of business advice were tested in a project involving 20,000 SMEs. Third, the 'Growth Impact Pilot (GIP)' - related to a \$300m support and mentoring scheme called the Growth Accelerator - ran from 2014-16 and aimed to assess the impact of business mentoring. This was a smaller experiment involving around 600 SMEs within a tight turnover band. Fourth, 'Small Business Charter Growth Vouchers Trial' was a \$4.5m project intended to explore the value of business school-SME leadership and management education which was due to

start in late 2014. Finally, the Innovation Vouchers RCT involving around 800 treated firms is a more recent trial due to complete in early 2018. In addition to these business support trials there have been a number of communication trials with a focus on behavioural change, many with support from the Innovation Growth Lab (IGL)⁵.

In each case the RCT principle is simple. Firms are invited to apply for a support scheme and then once they have applied they are randomly allocated to a control group (who do not receive the support) and a treatment group (who do receive support). In large enough samples randomisation will eliminate any selection bias issues allowing subsequent differences in performance between the treatment and control groups to be attributed to the support measure. The methodological and practical advantages of experimental and non-experimental research methods have been widely debated in the context of social policy interventions (Burtless 1995; Heckman and Smith 1995; Bratberg, Grasdahl, and Risa 2002; Deaton 2010; Banerjee and Duflo 2011). For example, experimental methods based on randomised allocation have – at least in theory – the advantage of transparency and may be more convincing to policy-makers than the results of more complex econometric evaluation approaches (Burtless 1995). Implementing experimental approaches may, however, lead to other specific biases (Heckman and Smith 1995), while small-scale experimental studies may fail to replicate the likely macro-impacts of a scheme which is implemented at national level – that is, they lack external validity (Garfinkel, Manski, and Michalopoulos 1990). Perhaps the key advantage of experimental approaches, however, and the central issue with non-experimental evaluation approaches, relates to potential selection biases. In terms of policy for innovation or small and medium enterprises (SMEs), selection biases might arise where a policy initiative seeks to back winners or is focussed on firms which have an established track record of growth or innovation. In this sense the receipt of public support may itself need to be treated as endogenous (Garcia and Mohnen 2010).

In this paper we provide an overview of the value of RCTs for SME policy research and, drawing on the UK experience, identify implementation issues which may offset any theoretical advantages of an experimental approach.

⁵ See <http://www.innovationgrowthlab.org/blog/making-innovation-and-growth-policy-more-experimental-first-year-igl>.

2. RCTS AS A RESEARCH APPROACH TO ENTERPRISE POLICY

Potter and Storey (2007) identify five reasons why industrial policy evaluation might be undertaken: to establish the impact of industrial policy; to inform the allocation of funding to alternative policy measures; to demonstrate value for money; to stimulate debate about forms of public intervention; and, to contribute to improvements in the design and administration of policy interventions. In each case the problem of causal inference is the same, i.e. that the treated and non-treated outcomes for any single firm are never observed (Holland 1986). The analytical problem this raises is how to estimate the counterfactual and the difference between the actual realised outcomes and the potential outcomes if no treatment had been administered. Ideally, the substitute for the unobserved (untreated) outcome needs to meet two criteria: (i) it should be observable to the researcher; and, (ii) it should be an 'internally' valid substitute for the set of untreated outcomes. Internal validity in this sense requires that 'the only difference between the member of the control group and the member of the treated group corresponds to the fact that the latter is treated and the first one is not' (Reiner 2011, p. 18).

More comprehensively, Imbens and Wooldridge (2009) outline three situations which describe the allocation of subjects to a control and treatment group. The first, and simplest, is the classical experimental situation of randomised allocation in which allocation is unrelated to outcomes. The second allocation mechanism – 'un-confounded allocation' – occurs where assignment is independent of outcomes but may be related to subject characteristics. Here, where the assignment mechanism is either observable or discoverable, sampling and/or statistical approaches can be used to minimise any systematic differences between the characteristics of the treatment and control groups and provide a valid estimate of treatment effects (Burtless 1995). In practice, research approaches to evaluating SME policy vary in the sophistication of their approach to un-confoundedness. Potter and Storey (2007), for example, cite evaluation studies which compare the performance of treated firms with control groups of 'typical' firms (Lambrecht and Pirnay 2005), and studies which use 'matched' control groups based on treatment and control groups with similar baseline

characteristics (Lerner 1999). However, despite careful matching or selection of control groups, the potential remains for bias in terms of the background characteristics of the two groups (Bratberg, Grasdahl, and Risa 2002).

This has led to the development and widespread application of a third group of econometric research approaches which can ‘control’ ex post for potential selection biases by either implicitly or explicitly modelling the probability that a firm will be in the treatment rather than the control group, and then estimating the impact of the treatment ‘controlling’ for any selection biases (Bratberg, Grasdahl, and Risa 2002; Imbens and Wooldridge 2009). This type of evaluation, allowing for potential selection biases, has been emphasised by the OECD as best practice in the context of SME support measures (Potter and Storey 2007), and has been widely used in the UK for the evaluation of supports for small business growth (Hart et al. 2008; Mole et al. 2008; Mole et al. 2008; Mole et al. 2011) and innovation (Foreman-Peck 2013)

Implementing this type of non-experimental approach, however, often involves making assumptions about the underlying causal process which determines the allocation of firms to the treatment and non-treated groups (Burtless 1995). Specific approaches may also pose challenges in terms of identification, requiring, for example, the use of a variable or group of variables which influence allocation but which have no influence on subsequent outcomes.⁶ Early comparisons of non-experimental econometric estimators with experimental estimators tended to favour experimental estimators (Heckman and Hotz, 1989), and highlighted the sensitivity of the results of non-experimental studies to the choice of econometric approach (Bratberg, Grasdahl, and Risa 2002). This led Burtless (1995) to conclude that: ‘the classical experimental estimator still has a major advantage over non-experimental estimators for users who care about the statistical precision of the estimates they use. But the more important advantage

⁶ Arguably this issue may be more difficult to deal when the units of analysis are firms rather than individuals. For example, factors which are difficult to observe such as management and leadership quality may influence performance and link to selection. In education or employment policy experiments, individuals’ educational or work histories may be easier to observe.

is that the validity of the experimental estimator depends upon assumptions that are ordinarily much easier to evaluate – and to believe’ (Burtless 1995, p. 73)⁷.

Thus, in non-experimental evaluations either un-confounded or confounded allocation – where allocation has some dependence on potential outcomes – can have profound implications for a researcher’s ability to obtain reliable estimates of policy treatment effects. In experimental approaches, by contrast, random assignment to the treatment and control group should avoid such biases and allow more accurate estimation of treatment effects (Burtless 1995). A number of potential implementation issues arise, however, even with experimental designs involving randomised allocation (Reiner 2011). First, in small samples randomised allocation may fail to eliminate differences in the characteristics of firms in the treatment and control groups, influencing the internal validity of the experiment⁸. A second threat to internal validity in experimental studies highlighted by Heckman and Smith (1995) is the potential for substitution bias where members of a control group are able to substitute alternative forms of support for the focal treatment effect. If significant, this contamination may undermine the estimate of the treatment derived from the difference in outcomes between the treatment and control groups. In the context of an industrial policy initiative where the subjects are firms rather than individuals it is also possible to envisage a related ‘signalling bias’ where firms in the treatment group are able to attract additional investment because they are in the treated group⁹. Third, and a factor which proves important in UK experiments, is attrition with firms in the treatment and control groups dropping out of the trial during the impact period (Bakhshi et al. 2015).

⁷ More recently, however, developments in propensity scoring, matching estimators and instrumental variables provide alternative statistical approaches to dealing with un-confoundedness in different contexts (Imbens and Wooldridge 2009).

⁸ As a result Bruhn and McKenzie (2009) suggest that in small samples (30-100 observations) the similarity of treatment and control groups is better where pair-wise matching, stratification or re-randomisation approaches are used rather than simple random allocation. For larger sample sizes (>300), however, their simulation analysis suggests that the choice of randomisation approach is a much less significant factor.

⁹ Meuleman and Maeseneire (2012), for example, find that Belgian SMEs which obtained an R&D subsidy were better able to access long-term debt than other firms. From a policy standpoint this is a desirable element of the policy impact but from a purely experimental point of view this may be seen as contamination which influences the scale of any measured treatment effect.

Finally, it is not clear how readily the results of any single policy experiment involving a small proportion of a target population can be generalised to the entire target population, i.e. whether the results of small scale policy experiments are ‘externally’ valid. Two issues here arise in both individual RCTs and other types of pilot evaluation projects: First, Garfinkel et al. (1990), discuss the ‘macro’ or society-wide effects which would occur if an intervention was implemented at macro-level but which may not be replicated in small scale studies. These include: policy effects on the economic equilibrium; widespread diffusion of information about the programme; and, social interaction and norm formation which might influence programme participation or other related decisions. Second, external validity may also be reduced if the characteristics of those firms participating in a programme differ significantly from the wider target group (Burtless 1995). Applicants for innovation support measures, for example, may be more strongly innovation-oriented than firms in general, a capability which has been linked to above average business performance (Rosenbusch, Brinckmann, and Bausch 2011). One further issue arises because of the nature of RCTs. External validity may also be impacted by ‘randomisation bias’, the idea that the adoption of random allocation itself might either induce or discourage some types of firms from applying for a scheme (Heckman and Smith 1995).

Issues related to the validity (either external or internal) of individual trials have suggested the potential value of triangulation or meta-analytical reviews across a number of RCTs. In medicine this approach is evident in the Cochrane Library which has established clear protocols for the conduct of individual RCTs and the undertaking of ‘reviews’ or meta-analyses of individual RCTs. Two recent Cochrane reviews illustrate this approach. One Cochrane review considering treatments for frozen shoulder synthesised the results from nineteen separate RCTs (Page et al. 2014). Another considered three RCTs in a treatment for early stage throat cancer (Yuan et al. 2014). In each case the reviews compare and contrast the results from the studies and identify generalised findings based on the principle that any one study may either reflect methodological or operational anomalies which may influence the results. In the context of medicine, however, interventions are primarily mechanical: a treatment which works with one human being is very likely to work with another independent of the context in which they are living. In terms of support programmes for firms, however, the situation is

arguably more complex given the heterogeneity of small firms, the social and interactive nature of business, variety in the way support measures are operated and the importance (and diversity) of the contexts in which firms operate. Internal and external validity may therefore be *more* difficult to maintain in industrial policy RCTs than in medicine. Arguably this means that standards of evidence should also be higher, requiring consistent evidence from multiple RCTs from different contexts, before robust conclusions should be drawn¹⁰.

3. POLICY RCTS IN PRACTICE – THE UK EXPERIENCE

The UK has a relatively long history of conventional SME policy evaluation, referenced widely in OECD best practice guidelines (Potter and Storey 2007) despite considerable concerns about the value of many evaluation studies¹¹. The value of evaluation evidence for policy-upgrading and design is therefore widely accepted within public service, and interest has been reinforced by current budgetary pressures and the policy choices this has enforced. Internationally too, the ‘movement for evidence based policy’ has re-emphasised the value of evidence in policy design and development. Lunn (2013, p. 99), for example, quotes the US Coalition for Evidence-Based policy as follows:

‘... to increase government effectiveness through the use of rigorous evidence about what works’ [because] ‘... a concerted government effort to build the number of these proven interventions, and spur their widespread use, could bring rapid progress to social policy similar to that which transformed medicine’.

Particularly interesting here is the analogy with medicine and the suggested parallels to the meta-analysis and synthesis in the Cochrane Reviews discussed earlier.

¹⁰ Innovation and management studies come closest to this approach in meta-analyses (Mueller, Rosenbusch, and Bausch 2013; Rosenbusch, Brinckmann, and Bausch 2011). See also the IGL database which also provides data on multiple trials: <http://www.innovationgrowthlab.org/igl-database>

¹¹ See for example: <https://www.nao.org.uk/report/evaluation-government/>.

Five publicly-funded enterprise policy RCTs have either been completed or are currently taking place in the UK each of which tests a different aspect of public support for SMEs. In addition, there have been a number of randomised communication trials designed to explore the potential to ‘nudge’ SME behaviours. We discuss each in turn.

3.1 Creative Credits

The UK’s first industrial policy experiment focussed on a policy instrument called ‘Creative Credits’, an innovation voucher programme designed to stimulate new innovative partnerships between SMEs and local creative service providers (Bakhshi et al. 2011)¹². The scheme was designed to help SMEs to benefit from the advantages of open innovation, in terms of stimulating creativity, enhancing product quality and generating potential reputational benefits which signal the quality of firms’ innovation activities¹³. Implementing open innovation partnerships poses particular challenges for resource-constrained SMEs, however (van de Vrande et al. 2009).

The Creative Credits policy experiment was conducted in the Manchester City Region of North-West England between September 2009 and October 2010. Creative Credits were promoted and marketed through a number of media channels, business networks and tele-marketing¹⁴. Over two thousand firms made some form of enquiry about the scheme with a total of 672 SMEs making eligible applications¹⁵. On receipt, applications from SMEs were first checked for eligibility¹⁶, Creative Credits were then randomly distributed across applicant firms. In total one hundred and fifty Creative Credits were ‘awarded’. Each

¹² Supporting evidence is found in Bakhshi (2009), Stam, de Jong and Marlet (2008) (2008) and Muller, Rammer and Truby (2009) which all report that businesses which make proportionately greater use of creative services introduce more innovations.

¹³ New innovation partnerships may also provide SMEs with access to networks which create commercial opportunities, and allow them to search their technological environment in a more systematic fashion, resulting in improved access to technology developed elsewhere (Powell 1998; Niosi 2003).

¹⁴ In promoting the scheme care was taken to minimise any selection biases. For example, companies tele-marketing the scheme used random methods of identifying which SMEs to call to promote the scheme (Bakhshi et al. 2013, p. 43).

¹⁵ This was equivalent to around 1:8 of the eligible business population (Bakhshi et al. 2011, p. 12).

¹⁶ Eligibility details are outlined in detail in Bakhshi et al. (2013). They included criteria relating to location, size and legal status.

Creative Credit had a face value of £4,000 with recipient firms also required to contribute a minimum of £1,000 to the cost of the project. Subsequent to the award, SMEs were encouraged to identify a new creative partner and develop a collaborative project proposal¹⁷, with all but two Creative Credits being used. Once a partnership was formed, projects were required to be completed within five months.

Data collection for the Creative Credits policy experiment comprised four sequential surveys of the treatment and control groups¹⁸. Over the course of the four surveys, and despite the payment of small cash incentives to firms to encourage continued participation in the data collection, significant attrition in response was experienced. By Survey 4, response numbers in the control group had fallen to 157, 52.2 per cent of those firms responding to Survey 1. In the treatment group attrition was less severe, with 78.0 per cent of Survey 1 respondents also responding to Survey 4. A key issue in terms of the internal validity of the experiment was whether the characteristics of the respondents to all of the four surveys and those firms which dropped out were similar, or whether attrition was systematically related to some respondent characteristic. Comparing the starting characteristics (from Survey 1) of stayers and those firms which did drop out, however, suggested no systematic differences to the initial control and treatment groups. This suggests that despite significant attrition this aspect of the internal validity of the experiment was maintained (Bakhshi et al. 2013).

The surveys also suggested that the award of a Creative Credit increased the probability that firms went ahead with their project within the five months since

¹⁷ To help with this process a web-based marketplace – a Creatives Gallery – of eligible creative firms was designed and made available to all eligible SMEs in the treatment and control groups. The aim of creating the online Gallery was to explore the potential for a minimal brokerage model and reduce the burden of administrative costs of the pilot project. SMEs were not allowed to work with creative companies which they had previously worked with, however.

¹⁸ Survey 1 was a baseline survey undertaken at the time firms were allocated to the treatment and control groups. Survey 2 was undertaken at a point just after the firms in the treatment group had completed their projects. Surveys 3 and 4 undertaken six month and twelve months later focussed on output, behavioural and network additionality.

the allocation of the Creative Credits by around 84 per cent¹⁹, a level of project additionality very similar to that reported for the Dutch, Swiss and Austrian innovation voucher schemes (Cornet, Vroomen, and van der Steeg 2006; Good and Tiefenthaler 2011)²⁰. Six months after the end of the treatment, firms in the treatment group were significantly more likely to be undertaking product/service and process innovation, and had (at the 10 per cent level) a significantly more positive distribution of sales growth rates than firms in the control group. These positive output additionality effects were short-lived, however, with no significant differences between the treatment and control groups after 12 months (Bakhshi et al. 2013)²¹.

In the Creative Credits experiment randomisation worked relatively well despite sample sizes – particularly for the treatment group - being relatively small (Bruhn and McKenzie, 2009). Attrition too led to no significant bias in the characteristics of those respondents which continued to respond to later surveys. Both helped to maintain internal validity. Another aspect of internal validity is the appropriateness of the timescale over which policy effects are measured and additionality interpreted. Behavioural or network benefits from measures such as Creative Credits, for example, may take some time to translate both into new innovation and innovation outcomes (Cunningham, Gök, and Laredo 2013). In the Creative Credits experiment measuring outcomes at several points helped to avoid potentially misleading inferences (Hewitt-Dundas and Roper 2011). To examine external validity the characteristics of Creative Credit applicants were compared to the population of eligible firms using a small ancillary survey of non-applicant firms (Bakhshi et al. 2011). This suggested that applicants were similar in size to the eligible population but were less likely to be exporters, more likely to have a high proportion of graduate employees and more likely to have engaged

¹⁹ Of the 301 firms in the control group which responded to our baseline survey, 36 firms (12 per cent) went ahead anyway with their projects. Among the group of 150 firms which were assigned Creative Credits 144 (96.0 per cent) actually commissioned projects.

²⁰ We modelled project additionality using both simple OLS and as a robustness check a treatment model allowing for potential sample selection. As expected, selection effects proved insignificant. As part of the same exercise we also investigated whether the small cash incentives provided to firms to help encourage survey response had biased these results. No evidence of any significant bias could be identified. Details of these models and those referred to later in this section can be found in Bakhshi et al. (2013).

²¹ Robustness checks using multivariate models allowing for potential selection effects confirmed these results (Bakhshi et al. 2013).

in prior innovation (Bakhshi et al. 2011, pp. 15-16). Issues were therefore evident with the external validity of the experiment.

3.2 Growth Vouchers²²

Announced originally in 2013 the Growth Vouchers RCT was the largest RCT that the UK has run and focused on testing the value of five different types of business advice for SME growth (Mole et al. 2008). The experiment aimed to provide subsidised business advice to around 20,000 small businesses at a total programme cost of around \$45m. Each business applying for the programme received either a face-to-face diagnostic interview with a consultant or undertook an on-line diagnostic and was then randomly allocated a Growth Voucher. Each Growth Voucher was worth up to £2,000 (just over \$3,000) and could be used to cover half the cost of business advice in one of five areas: raising finance and managing cash flow; recruiting and developing staff; improving leadership and management skills; marketing, attracting and keeping customers; and making the most of digital technology.

The Growth Voucher experiment focused on three research questions: Did firms given a Growth Voucher out-perform those that did not get one? Did businesses which undertook an on-line assessment outperform those which undertake a face-to-face needs assessment? And, which of the five different types of business advice create the greatest return (BIS 2014). The treatment phase of the experiment began in January 2014 and ended in March 2015. Around 1:6 firms receiving the diagnostic were allocated to a control group. Follow-up surveys were planned for 2-3 years after the treatment phase with longer-term follow-up intended through matching with administrative data.

No quantitative assessment of long-term treatment effects from the whole sample of Growth Voucher recipients has yet been published. An early evaluation published in 2016 does, however, highlight some of the challenges in the implementation of the programme and its short-term impacts. This report relates

²² See <https://www.gov.uk/government/collections/growth-vouchers-programme>. Accessed 3/12/2017.

to a cohort of firms that entered the programme between January to August 2014 (BIS 2016). The evaluation suggested that randomisation worked well with very similar treatment and control groups, suggesting strong internal validity at that point. Only around a third of firms in the treatment group went on to purchase business advice with their vouchers compared to a planning assumption of 50 per cent. The three most common reasons firms in the treatment group gave for not using their Growth Voucher were, an inability to find a suitable supplier (40%), insufficient funds to meet the match-funding requirement (36%) and lack of time to locate suitable suppliers (32%) (BIS 2016). The lower level of voucher use than anticipated could have created difficulties in terms of internal validity, and the original power calculations for the trial (Cabinet Office Behavioural Insights Team 2014). However, as average spending by those using a voucher was higher than anticipated this offset the lower level of voucher use in terms of calculating the treatment effects. Other issues arose linked to the complex design of the trial and the uneven take-up of the five areas of business advice. This reduced the power of some elements of the trial while strengthening others. Other results from this early evaluation were largely positive with voucher recipients/users more likely to increase skills, undertake business and marketing planning, and increase turnover than firms in the control group.

Other implementation issues also arose during the Growth Vouchers project, with the initial recruitment of companies slower than anticipated, although eventually recruitment significantly exceeded the original targets (BIS 2014). Some concerns were also noted around the variability of the initial face-to-face diagnostic or needs assessment – with some consultants more thorough than others. This variability in one of the treatments delivered suggests the need for a cautious approach to inference. Other concerns both from business advisors and firms focussed on the randomised nature of the trial itself and its potential to generate biases in the type of SMEs applying for the scheme (Heckman and Smith 1995). One business advisor commented:

'It might make sense from a research point of view but from a just common sense, normal, human-being point of view, telling someone, 'It would be really helpful for you to get one of these Growth Vouchers and let's talk

about how you could use it, but at the end of the day I'm just going to hit a button and you might get it or you might not' - people think that's a bit mad. ... These are real people with real businesses. This is not a game.' (BIS 2014, p. 30)

Concerns were evident too on the part of companies, one commenting:

'Why go through all that if it's just a lottery. I wouldn't have done it if I'd known', (BIS 2014, p. 54)

And similarly:

'I was aware that not everyone would get one but it wasn't made clear to me till afterwards that it was a randomised process. I wouldn't have bothered applying if I'd known it was randomised I would have had no interest in that'. (BIS 2014, p. 32).

3.3 The Small Business Charter (SBC) Growth Vouchers Programme

The Small Business Charter Growth Voucher Programme RCT aimed to test whether 'firms who receive subsidised leadership and management education from Business Schools perform better than those who have received no such support'²³. The focus of this RCT was therefore rather different to that of the Growth Voucher Programme (business advice) and the GIP (coaching). Like the main Growth Voucher programme, however, the SBC Growth Voucher Programme was intended to operate on a co-funding basis with government contributing a £2000 Growth Voucher per business and the business providing matching funding. The treatment (leadership and management education) provided was intended to include ten management development workshops in addition to mentoring support and networking events. Firms were to be recruited between October 2014 and March 2015, with the treatment lasting around six months and then firms were to be tracked through 2015 and 2016 to calibrate improvements in performance. Six UK business schools were included in the launch announcement with the aim of recruiting up to 150 firms each to

²³ See <https://www.gov.uk/government/publications/small-business-charter-growth-voucher-programme-terms-and-conditions>. Accessed 3/12/2017.

participate in a diagnostic workshop²⁴. Half of the participants would then be randomly allocated a Growth Voucher with the other half constituting the control group.

Recruitment to the trial proved limited making the full RCT difficult to operationalise although tracking of participants continues. In part, this was due to concerns by many of the 18 English business school partners, who had recently been awarded the Small Business Charter, about the potential reputational risks arising when they were required to inform businesses that they had not been awarded a Growth Voucher. As a result only 6 of the 18 business schools decided to engage with the initiative and participate in the management development workshops and networking events. As in the Growth Voucher evaluation, take-up among those awarded a growth voucher was also limited due to firms' reluctance to contribute to the costs of the management development workshops.

3.3 The Growth Impact Pilot (GIP)

While the Growth Vouchers experiment focused on the value of different types of business advice, the Growth Impact Pilot experiment focused on testing the impact of business leadership coaching taken alongside leadership and management training. The GIP was closely related to the UK's national Growth Accelerator programme which provided structured coaching for the leadership teams of SMEs with significant growth potential (OECD 2012). The main Growth Accelerator aimed to assist 26,000 firms over a three-year period (2014-16), with public investment in the scheme approaching \$250m (£200m). In the main Growth Accelerator programme firms undertook an initial face-to-face diagnostic and then a period of structured coaching before being able to access subsidised leadership and management training. In the GIP the ordering of support measures was reversed with firms offered subsidised leadership and management training at the outset – as an incentive to participate in the

²⁴ See <https://smallbusinesscharter.org/small-business-charter-launches-growth-voucher-programme-for-small-businesses/>. Accessed: 4/12/2017.

programme - and then the randomised possibility of subsequent leadership coaching²⁵.

The GIP was a significantly smaller experiment than the Growth Vouchers Programme with the aim of recruiting 600 firms (i.e. 300 control and 300 treatment) over the April 2014 - March 2015 period. Follow up surveys were then to be undertaken through 2015 and 2016. Recruitment to the GIP trial was slower than anticipated, raising questions about internal validity, although target numbers were finally achieved. Again, as in the Growth Voucher programme and Small Business Charter (SBC) Growth Vouchers Programme there was some uncertainty on the part of those delivering the programme about reputational damage given the randomised nature of the support. No evaluation or impact reports have yet been published on the GIP but longer-term tracking of participating firms using administrative data is on-going.

3.4 Innovation Vouchers

Innovate UK started an Innovation Vouchers programme in 2012 to provide support for SMEs' innovation activities. The programme provides SMEs from all sectors financial support of up to £5000 for engaging the services of experts from academia, research and technology organisations or the private sector to pursue a certain innovative project within the firm. The Innovation Vouchers programme has been organised around fixed application dates, and the RCT evaluation of the programme focuses on support provided during the 2016-2018 period. Recent updates to the registered Trial Protocol²⁶ suggest the original target numbers (2100 firms) have been scaled back to 800 firms in the treatment group and 200 firms in the control group. The Trial Protocol suggests four hypotheses:

- Hypothesis 1: Innovation vouchers have a positive effect on a firm's collaboration with external partners.
- Hypothesis 2: Innovation vouchers have a positive effect on a firm's innovation activities.

²⁵ See <https://www.apm.org.uk/sites/default/files/growth%20impact%20pilot.pdf>. Accessed: 3/12/2017.

²⁶ See <https://www.socialscienceregistry.org/trials/1556>.

- Hypothesis 3: Innovation vouchers have a positive effect on a firm's innovation output.
- Hypothesis 4: Innovation vouchers have a positive effect on a firm's business performance indicators.

No reports or analysis on the outcomes of the Innovation Vouchers trial have yet been published (December 2017).

3.5 Messaging Trials²⁷

Alongside the experiments described earlier designed to impact on business performance **** experiments have been undertaken to explore the potential value of different messaging strategies. The first such trial aimed to encourage those registering for a UK government business mentoring initiative (Get Mentoring) to go on to the next step of the accreditation process and complete a training programme. Alternative e-mail messages were randomly sent to those registering for the scheme the success of the messaging measured by subsequent take-up of the training programme. Content which was emotionally engaging was found to have a significant impact with important gender differences also identified in individuals' responses. The process of testing and steadily improving messaging contributed to the programme surpassing its target of training 15,000 new mentors by December 2012.

A second series of messaging trials were undertaken as part of the promotion of the Growth Voucher trial outlined earlier. Trials tested simple tweaks to improve engagement with an email newsletter providing details of available business support. The results were positive:

'... opening rates increased by 3.9 per cent just by inviting businesses to reflect on their growth ambitions in the subject line ("Realise your hopes and aspirations"), but highlighting that support was "free" completely undermined this'. ... Overall, the email led to 9,000 additional applications to the Growth Vouchers Programme, and provided BEIS with

²⁷ This section draws heavily on the account by James Phipps, Nesta, on the UK experience of messaging trials. See <https://www.nesta.org.uk/blog/taking-first-steps-business-policy-experiments>

*valuable information on how we could most effectively communicate with small firms*²⁸.

4. DISCUSSION

In 2007 Jonathan Potter and David Storey described six alternative research approaches to SME policy evaluation (Potter and Storey 2007). The sixth – and most complex – approach in their typology relates to econometric approaches which control for selection bias. Arguably, however, RCTs – which avoid, rather than control for, selection bias offer a seventh alternative research approach.

The use of this research approach for examining enterprise policy remains in its infancy, however, while the experience of the UK highlights some of the operational and logistic difficulties with this approach. For example, the challenges to internal validity are substantial. In industrial policy experiments, randomisation, for example, cannot be double blind as in medical trials and, in some cases, may be subject to small sample biases (Bruhn and McKenzie 2009). The face-to-face nature of many types of business support activity may also result in the heterogeneity of the treatment where business advisors vary in quality or motivation (Mole et al. 2008). Contamination may also be evident where firms allocated to a control group seek to substitute other support for the focal treatment, or where the treatment itself acts as a signal of firm quality (Kleer 2010; Meuleman and De Maeseneire 2012). These issues pose significant challenges for experimental design, challenges reinforced by uncertainty over impact periods and the potential for attrition in follow-up surveys (Bakhshi et al. 2013).

Challenges also remain in terms of external validity, i.e. the extent to which the results of any experiment can be generalised to the broader population of SMEs or non-applicants. The question of whether applicant group for any scheme is typical of all firms is an issue in any evaluation and in the estimation of treatment effects. Specific issues arise in experiments, however, where - as in the case of Growth Vouchers – firms may be discouraged from applying to the scheme due to the lottery element (Heckman and Smith 1995). Other, perhaps weaker, firms

²⁸ See <https://www.nesta.org.uk/blog/taking-first-steps-business-policy-experimentation>.

may of course be encouraged to apply by the random allocation element of the scheme and they assess the chance of success to be greater than schemes where applications are subject to peer review. Neither the inducement or deterrent effect of randomisation seems likely in healthcare trials, providing another illustration of the more complex nature of industrial policy experiments. A related issue – not discussed in the literature as far as we are aware – is the potential impact of randomisation on the supply-side of policy experiments, and the decision by delivery partners whether or not to participate in an experiment. Randomisation may be seen, for example, as raising ethical issues for delivery partners who might be keen to assist all client firms²⁹. It may also pose reputational risks if delivery partners were seen by the media or general population as ‘experimenting with firms’. This issue arose in the case of the Growth Voucher and GIP trials and was one of the reasons for the scaling back of the Small Business Charter Growth Voucher experiment.

With the exception of the Growth Voucher experiment the other UK policy experiments also remain relatively small and outcomes will therefore reflect the geographical, market, industry and network context within which the experiments take place (van der Duin et al. 2006). The Creative Credits experiment, for example, was explicitly targeted at SMEs in the Manchester region of North West England due to a perceived deficiency in the level of co-ordination between SMEs and designers in the region (MIER 2009). The Creative Credits project was also undertaken in the immediate aftermath of the financial crisis also potentially impacting on outcomes. Indeed one firm commented on their Creative Credit project *‘it was useful at the time, but because of the recession and all the rest of it, I don’t believe we’ve really seen the benefit yes, but I suspect we will do’*. Another firm remarked *‘... commerce in the last months has been affected by the recession so it’s difficult to know what whether the Creative Credit has helped deflect further shrinkage in trade than would have been suffered and what impact it would have had in normal circumstances’*. As with potential randomisation biases the extent of contextual influences on outcomes is, in practice, difficult to anticipate and quantify. Replication across diverse contexts is therefore important.

²⁹ This issue can be offset through the design of the experiment. The Growth Voucher trial offered all participants some support for example.

Overcoming these challenges in as cost-effective manner as possible suggests an experimentation strategy based on a small number of repeated experiments where each experiment is just large enough to maintain internal validity. Repetition – as in the Cochrane Reviews cited earlier – would provide confirmatory evidence and may help, in particular, to avoid false positives or negatives. Ideally, repeated experiments would also be undertaken in contexts which vary significantly, enhancing the generalisability of any consistent policy effects.

5. CONCLUSIONS

The UK experience suggests that randomised control trials do represent a feasible research approach for industrial policy instruments, particularly for support measures which are novel, uniform in terms of the treatment provided, and which are targeted at a relatively large group of firms. With care, internal validity can be maintained in individual experiments, and replication can help to overcome contextual challenges to external validity. This is important as without such policy experiments – whether RCTs or more conventional pilot projects – there is an increased risk of inappropriate or ineffective policy interventions. In terms of business support, the financial stakes are considerable, with one recent estimate being that around €152bn pa is spent on business support across Europe, €9.8bn in the UK³⁰.

As always with evaluation studies, however, the value is only derived ex post as policy interventions are either modified, extended or curtailed. Appropriately interpreting the evidence provided by RCTs (and other small scale policy trials) remains important (Lunn 2013). Two concerns arise in particular. First, at their best, policy experiments test the effectiveness of a rather specific treatment on a particular group of firms. For example, the GIP experiment focused on leadership coaching. Its results may provide good evidence on the impact of leadership coaching for a particular size group of UK small firms. Beyond this target group, however, inferences about the more general value of coaching will only be

³⁰ See <http://www.innovationgrowthlab.org/blog/much-%E2%82%AC152-billion-spent-on-business-supporting-businesses-does-it-work>.

tentative, and unwarranted inferences should be avoided. Indeed, changing the delivery mechanism, resource-base or eligibility criteria may either enhance or reduce the effectiveness of leadership coaching. Second, individual RCT and more conventional policy experiments are inevitably context specific – a more serious issue than in medical trials which generally relate to specific bio-chemical mechanisms– and the generalisability of evidence from individual experiments needs to be carefully considered.

REFERENCES

Bakhshi, H. 2009. Creative Supply Chain Linkages and Innovation: Do the Creative Industries Stimulate Business Innovation in the Wider Economy? *Innovation: Management, Policy & Practice* **11:169-189**.

Bakhshi, H.; J. Edwards; S. Roper; J. Scully; and D. Shaw. 2011. Bridging the divide: Using innovation vouchers to link SMEs and creative service providers – first results from a UK mixed-methods evaluation. In *Evaluating innovation policy, methods and applications*, ed. EUNIP. Florence.

Bakhshi, H.; J. Edwards; S. Roper; J. Scully; D. Shaw; L. Morley; and N. Rathbone. 2013. Creative Credits - a randomized controlled industrial policy experiment. London Nesta.

Bakhshi, H.; J.S. Edwards; S. Roper; J. Scully; D. Shaw; L. Morley; and N. Rathbone. 2015. Assessing an experimental approach to industrial policy evaluation: Applying RCT plus to the case of Creative Credits. *Research Policy* **44:1462-1472**.

Banerjee, A.V. and E. Duflo. 2008. Title. NBER working paper 14467.

———. 2011. Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty. *Public Affairs*.

BIS. 2009. RDA Evaluation: Practical Guidance on implementing the Impact Evaluation Framework, ed. I.a.S. Department for Business. London

———. 2014. Growth Vouchers Programme - formative evaluation report ed. I.a.S. Department for Business. London.

———. 2016. Growth Vouchers Programme Evaluation Cohort 1: Impact at six months. In *Research Paper* London BIS.

Bratberg, E.; A. Grasdal; and A.E. Risa. 2002. Evaluating social policy by experimental and nonexperimental methods. *Scandinavian Journal of Economics* **104:147-171**.

Bruhn, M. and D. McKenzie. 2009. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic*

Journal-Applied Economics **1:200-232**.

Burtless, G. 1995. The case for randomised field trials in economic and policy research *Journal of economic perspectives* **9:63-84**.

Cabinet Office Behavioural Insights Team. 2014. GROWTH VOUCHERS PROGRAMME - Trial Protocol (full)

. London Cabinet Office.

Cornet, M.; B. Vroomen; and M. van der Steeg. 2006. Do innovation vouchers help SMEs to cross the bridge towards science? In *No 58 CBP Discussion Paper* CBP.

Cunningham, P.; A. Gök; and P. Laredo. 2013. The Impact of Direct Support to R&D and Innovation in Firms. Compendium of Evidence on the Effectiveness of Innovation Policy. London: NESTA.

Deaton, A. 2010. Instruments, Randomization and Learning about Development. *Journal Of Economic Literature* **48:424-455**.

Foreman-Peck, J. 2013. Effectiveness and efficiency of SME innovation policy. *Small Business Economics* **41:55-70**.

Garcia, A. and P. Mohnen. 2010. Impact of government support on R&D and Innovation In *UNU-Merit Working Paper 2010-034*.

Garfinkel, I.; C.F. Manski; and C. Michalopoulos. 1990. Title. Institute for research on poverty, discussion papers Dp 922-90, University of Wisconsin-Madison.

Good, B. and B. Tiefenthaler. 2011. Innovation voucher- small is beautiful. *Plattform - Forschungs- und Technologieevaluierung*:**45-53**.

Greenberg, D. and M. Shroder. 1991. *Digest of the social experiments*. Madison: Institute for research on poverty, University of Wisconsin.

Hart, M.; N. Driffield; S. Roper; and K. Mole. 2008. Evaluation of Regional Selective Assistance and its successor, Selective Finance for Investment in England (SFIE). *BERR Occasional Paper No. 2 March 2008*.

Heckman, J.J. and J.A. Smith. 1995. ASSESSING THE CASE FOR SOCIAL EXPERIMENTS. *Journal of economic perspectives* **9:85-110**.

Hewitt-Dundas, N. and S. Roper. 2011. Creating advantage in peripheral regions: the role of publicly funded R&D centres. *Research Policy* **40:832-**

841.

Holland, P. 1986. Statistics and causal inference *Journal of the American Statistical Association*:**945-960**.

Imbens, G.W. and J.M. Wooldridge. 2009. Recent Developments in the Econometrics of Program Evaluation. *Journal Of Economic Literature* **47:5-86**.

Kleer, R. 2010. Government R&D subsidies as a signal for private investors. *Research Policy* **39:1361-1374**.

Lambrecht, J. and F. Pirnay. 2005. An evaluation of public support measures for private external consultancies to SMEs in the Walloon region of Belgium. *Entrepreneurship And Regional Development* **17:89-108**.

Laredo, P. 1997. Evaluation in France: A decade of experience. In *Policy evaluation in innovation and technology: towards best practices*, ed. OECD. Paris.

Lerner, J. 1999. The government as venture capitalist: the long-run impact of the SBIR program. *Journal of Business* **72:285-318**.

Lunn, P. 2013. Researchers imply, policymakers infer: the relationship between evidence and policy. *Journal of the statistical and social inquiry society of Ireland* **XLII:96-106**.

Meuleman, M. and W. De Maeseneire. 2012. Do R&D subsidies affect SMEs' access to external financing? *Research Policy* **41:580-591**.

MIER. 2009. Innovation, Trade and Connectivity report. In *Manchester Independent Economic Review*.

Mole, K.; M. Hart; S. Roper; and S. Saal. 2008. Assessing the Effectiveness of Business Support Services in England: Evidence from a Theory Based Evaluation. *International small business journal*
Forthcoming

Mole, K.F.; M. Hart; S. Roper; and D. Saal. 2008. Differential Gains from Business Link Support and Advice: A Treatment Effects Approach. *Environment and Planning C* **26:316-334**.

Mole, K.F.; M. Hart; S. Roper; and D.S. Saal. 2011. Broader or deeper? Exploring the most effective intervention profile for public small business

- support. *Environment and Planning A* **43:87-105**.
- Mueller, V.; N. Rosenbusch; and A. Bausch. 2013. Success Patterns of Exploratory and Exploitative Innovation: A Meta-Analysis of the Influence of Institutional Factors. *Journal Of Management* **39:1606-1636**.
- Muller, K.; C. Rammer; and J. Truby. 2009. The Role of Creative Industries in Innovation. *Innovation: Management, Policy & Practice* **11:148-168**.
- OECD. 2012. An International Benchmark analysis of public programmes for high growth firms'. . Paris.
- Page, M.J.; S. Green; S. Kramer; R.V. Johnston; B. McBain; and R. Buchbinder. 2014. Electrotherapy modalities for adhesive capsulitis (frozen shoulder). *Cochrane Database of Systematic Reviews* **10**.
- Potter, J. and D. Storey. 2007. OECD Framework for the Evaluation of SME and Entrepreneurship Policy and Programmes. Paris OECD.
- Reiner, C. 2011. Evaluating innovation policies by chance? The case for randomised R&D-programme evaluation. *Plattform - Forschungs- und Technologieevaluierung* **37:15-28**.
- Rosenbusch, N.; J. Brinckmann; and A. Bausch. 2011. Is innovation always beneficial? A meta-analysis of the relationship between innovation and performance in SMEs. *Journal of Business Venturing* **26:441-457**.
- Stam, E.; J.P. de Jong; and G. Marlet. 2008. Creative Industries in the Netherlands: Structure, development, innovativeness and the effects on urban growth. *Geografiska Annaler. Series B, Human Geography* **90:119-132**.
- van de Vrande, V.; J.P.J. de Jong; W. Vanhaverbeke; and M. de Rochemont. 2009. Open innovation in SMEs: trends, motives and management challenges. *Technovation* **29:423-437**.
- van der Duin, P.; R. Ortt; D. Hartmann; and G. Berkhout. 2006. Innovation in Context. In *Managing Technology and Innovation* ed. R.M. Verburg; J.R. Ortt; and W.M. Dicke. Abingdon Routledge
- Yuan, Y.; X. Zeng; Y. Hu; T. Xie; and Y. Zhao. 2014. Omentoplasty for oesophagostomy after oesophagectomy. *Cochrane Database of*

Systematic Reviews **10.**



Centre Manager
Enterprise Research Centre
Warwick Business School
Coventry CV4 7AL
CentreManager@enterpriseresearch.ac.uk

Centre Manager
Enterprise Research Centre
Aston Business School
Birmingham, B4 7ET
CentreManager@enterpriseresearch.ac.uk