Statistical analysis plan: the effectiveness of public support for high-potential businesses

Programme

This analysis concerns the Innovation Vouchers Programme (IVP). Below we briefly outline background information on this programme.

Innovation Vouchers Programme (IVP):

- IVP was an RCT run in 2015 (across 3 waves). The purpose of the programme was to encourage SMEs to work with external knowledge providers, with the goal that this would lead to increased knowledge and capacity within those businesses and in turn to increased innovation.
- Participants were randomly assigned to the control or treatment group. Participants in the control group were not allocated to receive an innovation voucher (£5,000 value), while participants in the treatment group were allocated to receive a voucher.
- There were 1,463 participating SMEs, including 356 in the control group and 1,107 in the treatment group (roughly ¼ allocated to the control group and ¾ to the treatment group). Places in the treatment group were allocated based on available budget.
- Participants were subject to eligibility checks, which reduced the number of participants from 2,149 to 1,463.
- Contributors to IVP included Innovate UK and the Innovation Growth Lab (IGL).

Research Questions

In our analysis, we aim to answer the following research questions:

	Research Question
Primary	RQ1**:** Can public programmes provide support that increases businesses growth?
Secondary	RQ1a**:** To what extent do estimates of impact from the programme vary by the evaluation methodology that is applied?
	RQ1b**:** How long might it take to observe impacts on business performance and over what period might they be sustained?

Variables

Datasets:

To complete our analysis, we are using data from the following sources:

- Innovate UK (IUK)
- Longitudinal Business Database (LBD)
- Office for National Statistics (ONS)
- Intellectual Property Office (IPO)

Manipulated variables:

The manipulated variables in our analysis are:

- For the experimental analysis of IVP, assignment to the treatment group, when compared to the control group
- For the quasi-experimental analysis of IVP, assignment to the treatment group, when compared to a group of businesses not participating in the experiment with similar characteristics

Primary and secondary outcomes:

All outcomes are measured in relation to the programme year, defined as the year of programme participation: 2015

The **primary outcomes** for our analysis are computed as follows, using data from the ONS Longitudinal Business Database (LBD):

- Cumulative turnover, the sum of annual turnover between the programme year and the programme year + 4, in £0,000. Turnover for years in which a business is marked as "inactive" will be counted as zero. Each annual turnover value will be adjusted for inflation by using the GDP deflator value in that year.
- Cumulative employment ("job years"), defined as the sum of people employed by the company between the programme year and the programme year + 4. Employment for years in which a business is marked as "inactive" will be counted as zero.
- A proxy measure of productivity, defined as turnover per employee in the programme year + 4. If employment is reported as zero or is missing in a particular year, we will consider productivity that year to be zero.

The **secondary outcomes** for our analysis are computed as follows, using data from the LBD, Intellectual Property Office (IPO) and Innovate UK:

- Turnover in the programme year + 4, in £0,000.
- Employment in the programme year + 4.
- Survival, defined as whether the business is marked as "active" in the LBD in the programme year + 4, as a binary measure (0 or 1).
- Number of patents, defined as the sum of patents issued to the business between the programme year + 1 and 2018.
- Awards of R&D funding between the programme year + 1 and the programme year + 4, defined as the sum of grants received from Innovate UK, in £0,000.

Covariates:

The covariates (control variables) in the analysis include pre-intervention values of outcome variables, and company characteristics in the programme year.

The following variables will be taken directly from the data sources:

- Turnover in each year prior to the programme, in £0,000 (adjusted for inflation).
- Employment in each year prior to the programme, in number of employees.
- Business sector in the programme year, one of 15 categories corresponding to the primary business activity (from SIC classification).
- Region: one of 12 UK nations and regions.

The following variables will be computed:

- Productivity in the year prior to the programme, defined as the turnover per employee.
- Logarithm of company age in the programme year, defined as the number of years since incorporation of the business in the programme year.
- Multiplant status, defined as whether or not a company had multiple locations (i.e. more than one reporting unit) in the programme year, a binary measure (0 or 1).
- Patent history, defined as the number of patents issued to the company between 2011 and the programme year.
- Trademark history, defined as the number of trademarks issued to the company between 2011 and the programme year.
- Registered design history, defined as the number of registered designs issued to the company between 2011 and the programme year.
- Urban or rural status, defined using the 2011 NSPL Output Area classification:[1]
    - (1) large urban (A1-B1 in categories used in England and Wales, 1 in Scottish categories)
    - (2) other urban (C1-C2 in categories used in England and Wales, 2 in Scottish categories)
    - (3) rural (all other categories)
    - (4) none (for businesses registered in Northern Ireland)

[1] This variable is not available for businesses that have their registered address in Northern Ireland.The distinction between the urban and rural categories is defined differently in England and Wales to Scotland: for this reason, it is important to include the binary indicator variable of whether the business's registered address is in Scotland in any model in which the urban/rural location is included.

Statistical Analysis

Below, we outline the steps for our statistical analysis.

Primary research questions

The primary research question is RQ1.

RQ1: Did the Innovation Vouchers Programme increase business growth?

In all analyses for RQ1, we will first run a bivariate model including only the manipulated variable (treatment/control group, or participant/comparison group) as a predictor, followed by multivariate models containing the covariates. The multivariate models will be treated as our definitive results. We will report the estimates for the predictors, the estimates for the covariates, the p-values (raw and after adjustment for multiple hypothesis testing), and confidence intervals. When using probit models, instead of the estimated coefficient itself, we will report the estimated marginal effect at the control group mean.

With numeric/continuous outcome measures, the estimates derived from the models including covariates (model 2 in each case) will be considered as the definitive estimates. With binary outcomes, the estimates from the probit models will be considered as definitive, but the size of the marginal effects will be checked against the OLS estimates.

To accompany the statistical models, we will produce a series of line plots depicting the outcome measures over time, depicting the treatment and control groups (for the experimental analysis) or the participants and comparison group (for quasi-experimental analysis) with confidence intervals shown as error bars.

Primary analysis

Our primary analysis of RQ1 will measure the impact of participation in the three programmes on business growth and productivity. For IVP, this will be accomplished by means of an experimental analysis comparing the control and treatment groups.

Experimental analysis:

For our *experimental analysis* of RQ1, we will measure differences in the key long-term outcomes - cumulative turnover, cumulative employment, and productivity - between businesses in the treatment group and businesses in the control group in IVP.

The analysis will estimate the intention-to-treat (ITT) effect, comparing all businesses assigned to the control group with all businesses assigned to the treatment group.

The continuous outcomes of the policy experiment will be tested using regression models of the following two forms:

$$Y_i = \alpha + \beta T_i + \varepsilon_i \quad (1)$$

$$Y_i = \alpha + \beta T_i + \gamma X_i + \varepsilon_i \quad (2)$$

where, for each individual i, Y_i is the dependent variable, T_i is an indicator variable defined to be equal to 1 if respondent i is in the treatment arm being tested and zero if the respondent is in the control group against which that treatment is being compared, X_i is a matrix of the covariates listed above (see Variables), and ε_i is a random error term.

A secondary binary outcome (business survival) will additionally be tested with Probit models of the following two forms:

$$P(Y = 1|T_i) = \Phi(\beta T_i + \varepsilon_i) \quad (1)$$

$$P(Y = 1|T_i, \gamma X_i) = \Phi(\beta T_i + \gamma X_i + \varepsilon_i) \quad (2)$$

where, for each individual i, $P()$ is the probability of the outcome, $Y = 1$ is the binary response variable representing company survival, T_i is an indicator which is equal to 1 if respondent i is in the treatment arm being tested and zero if the respondent is in the control group against which that treatment is being compared, X_i is a matrix of covariates, Φ is the cumulative distribution function (CDF) of the standard normal distribution, and ε_i is a random error term.

Secondary analysis

Comparison of regions and urban/rural locations:

In our secondary analysis of RQ1, we will first assess the primary outcomes - cumulative turnover, cumulative employment, and productivity - in different regions and locations in the UK. For the analyses above, we will complete an analysis for different regions in the UK, and for rural and urban locations.

For the analysis of the different regions, we will derive an estimate for the impact of the intervention in each of the 9 English regions and for each of the three other UK nations. Nation/region will be included as a factor in the statistical models, such that estimates will be produced for each region (level).

We will also assess each of the primary outcomes by location, using our three urban/rural categories. For the analysis of urban and rural locations, we will derive an estimate for the impact of the intervention in urban areas (when compared to rural and small areas).

For the *experimental analysis* of IVP, the outcomes will be tested with models of the following two forms:

$$Y_i = \alpha + \beta T_i + \gamma Z + \beta_2 (T_i * Z) + \varepsilon_i \quad (1)$$

$$Y_i = \alpha + \beta T_i + \gamma Z + \gamma X_i + \beta_2 (T_i * Z) + \varepsilon_i \quad (2)$$

where, for each individual i, Y_i is the dependent variable, T_i is an indicator variable defined to be equal to 1 if respondent i is in the treatment arm being tested and zero if the respondent is in the control group against which that treatment is being compared, Z is an indicator variable for the region or location (rural or urban), X_i is a matrix of covariates which includes Z, and ε_i is a random error term. $\beta_2(T_i * Z)$ is the interaction between the treatment arm and the region or location.

Alternative and additional outcomes:

Next, we will measure business growth using *alternative* and *additional* outcome measurements, including turnover and employment in the outcome year, business survival, patents and R&D funding. As in the primary analysis, for IVP this will be an experimental analysis. As turnover and employment are continuous outcomes, we will use OLS regression models as described above. As business survival is a binary outcome, we will use Probit models.

Uptake of innovation vouchers:

We will also apply an instrumental variable (IV) approach to measure the impact of uptake of the voucher on primary outcomes of interest (cumulative turnover, cumulative employment, productivity). The reason for this additional analysis is the relatively low uptake of vouchers among companies assigned to the treatment group (66.5%), which may result in a difference between ITT and LATE estimates.

When estimating the effect of being assigned to the treatment group on cumulative turnover, we can consider treatment status as an *instrument* because it affects cumulative turnover only through redemption of the voucher. If treatment assignment and subsequent cumulative turnover are correlated, this provides evidence that redemption of the voucher causes changes in cumulative turnover.

To measure the effect of redemption of the voucher on the outcomes of interest for the Innovation Vouchers Programme, we will apply an instrumental variables approach using the Two Stage Least Squares (TSLS) estimator approach.

In Stage 1, we will estimate the relationship between random assignment and redemption of the voucher (the dependent variable):

$$Z_i = \alpha + \beta T_i + \gamma X_i + \varepsilon_i \qquad (1)$$

where, for each individual i, Z_i is an instrumental variable defined to be equal to 1 if respondent i has redeemed the innovation voucher and zero if the respondent did not redeem the voucher, T_i is an indicator variable defined to be equal to 1 if respondent i is in the treatment arm being tested and zero if the respondent is in the control group against which that treatment is being compared, X_i is a matrix of covariates, and ε_i is a random error term.

In Stage 2, we will measure the relationship between assignment to treatment group and outcomes, using predicted values from Stage 1. This will be tested using regression models of the following two forms:

$$Y_i = \alpha + \beta\hat{T}_i + \varepsilon_i \quad (2)$$

$$Y_i = \alpha + \beta\hat{T}_i + \gamma X_i + \varepsilon_i \quad (3)$$

where, for each individual i, Y_i is the outcome variable, $\hat{T}_i$ is the prediction redemption of the voucher derived from the first stage, X_i is a matrix of covariates, and ε_i is a random error term.

Secondary research questions

The secondary research questions include RQ1a and RQ1b.

RQ1a: Do estimates vary based on methodology?

In all analyses for RQ1a, we will first run a bivariate model including only the manipulated variable as a predictor, followed by multivariate models containing the covariates. The multivariate models will be treated as our definitive results. We will report the estimates for the predictors, the estimates for the covariates, the p-values (raw and adjusted), and confidence intervals.

Quasi-experimental analysis:

For our analysis of RQ1a, we will complete a *quasi-experimental analysis* of IVP. The results of this analysis will be compared to the experimental analysis completed for RQ1.

This analysis will estimate the intention-to-treat (ITT), comparing all businesses that applied to participate in the programme, and met the eligibility criteria, with a comparison group of similar businesses identified from the LBD through matching. Businesses in the control group for IVP will be excluded from this analysis.

The outcomes of the programmes will be tested using regression models of the following two forms:

$$Y_i = \alpha + \beta C_i + \varepsilon_i \quad (1)$$

$$Y_i = \alpha + \beta C_i + \gamma X_i + \varepsilon_i \quad (2)$$

where, for each individual i, Y_i is the dependent variable, C_i is an indicator variable defined to be equal to 1 if respondent i is in the participant group and zero if the respondent is in the comparison group against which that participant group is being compared, X_i is a matrix of covariates, and ε_i is a random error term.

RQ1b: How does impact change over time?

In all analyses for RQ1b, we will first run a bivariate model including only the manipulated variable as a predictor, followed by multivariate models containing the covariates. The multivariate models will be treated as our definitive results. We will report the estimates for the predictors, the estimates for the covariates, the p-values (raw and adjusted), and confidence intervals.

To answer RQ1b, we will evaluate outcomes in each year after the programme is delivered (programme year). For IVP, this will be accomplished by means of an experimental analysis.

Experimental analysis:

For our *experimental analysis* of RQ1b, we will measure the impact of IVP after 1 year, 2 years, 3 years, 4 years, 5 years, 6 years, and 7 years, comparing the treatment group to the control group. The outcome year will be included in the statistical models as a factor, such that estimates will be produced for each year (level).

The outcomes of the programmes will be tested using regression models of the following two forms, repeated for each outcome year t:

$$Y_{it} = \alpha_t + \beta_t T_{it} + \varepsilon_{it} \quad (1)$$

$$Y_{it} = \alpha_t + \beta_t T_{it} + \gamma_t X_{it} + \varepsilon_{it} \quad (2)$$

where, for each individual i, Y_{it} is the dependent variable (outcome) in year t, T_{it} is an indicator variable defined to be equal to 1 if respondent i is in the treatment arm being tested and zero if the respondent is in the control group against which that treatment is being compared, X_{it} is a matrix of covariates, and ε_{it} is a random error term.

For the continuous outcomes of cumulative turnover, cumulative employment and productivity, we will use OLS regression models. We will first use bivariate models, and then multivariate models with control variables.

Pre-Analysis

Assembly of Datasets

Ingest datasets:

In order to join our datasets to the Longitudinal Business Database, we assembled the ingest dataset for IVP.

First, we selected the variables from the programme data (from IUK) that related to information provided by applicants, and details of the experimental condition or funding status of those applicants.

Next, we combined the programme datasets with information from multiple additional sources, including:

- Company registration numbers (CRNs), postcodes and dates of incorporation from Companies House
- UK nation or region and urban/rural classification from the National Statistics Postcode Lookup (NSPL) database
- IP outcomes including patents, registered designs, and trademarks from the Intellectual Property Office (IPO) data
- Details of any additional R&D funding obtained from IUK

Combination of these data sources was achieved by first identifying the unique company registration number (CRN) for each company in Companies House. To determine the correct CRNs, we searched Companies House automatically, using the API. The outputs - the first page of search results for each company name in the programme dataset for IVP - were saved. We then used an assignment procedure to find the most likely match between the search results and the companies listed in each of the programmes, as follows:

1. Matching CRNs: the CRN provided by the company is the same as the search result
2. Close CRNs: the CRN provided by the company is very close to the search result (with a Levenshtein distance < 3)
3. Postcode match / Locality match: the postcode provided by the company match the search result
4. For all other companies, the first search result is assigned, and an additional flag is added for those results with the lowest Levenshtein distance

For companies falling into category 4, we undertook additional manual checks to determine if the CRN assigned is indeed the correct company. Two researchers performed these checks independently and without consulting one another, to achieve inter-rater reliability. The results were then compared, and any disagreements resolved through discussion. Some companies are likely to be correctly identified, but are lacking a high degree of certainty - in these cases, we add an additional flag that the companies have a “low confidence”
Only companies for which we have a sufficient degree of confidence in the CRN will be included in the analysis, so companies that cannot be identified will be excluded from the ingest datasets. Note that some companies were not incorporated at the time the support programmes were delivered, but were later incorporated.

After identifying the CRNs, we again used the Companies House API to collect any Standard Industrial Classification (SIC) codes provided by the companies. SIC codes represent different industries. The codes - which have 4-5 numeric digits - have 731 possible values. For our purposes, we group these into general sector categories (A-U), using the first digit of the SIC code. SIC codes and the corresponding sectors are listed on the Companies House website, and we have adapted this list into a spreadsheet which is then used to assign sectors.

Following the SIC codes, we add information from the NSPL dataset, including the region and the 2011 Output Area classification. The data is combined by means of the postcode associated with the company's registered address on Companies House. The Output Areas are used to generate the three urban/rural categories used in our analysis.

Next, we add in information about intellectual property, including any patents, trademarks and registered designs. These outputs are combined as outcome variables (following programme participation) and as covariates (prior to programme participation).

The resulting ingest datasets contain the *minimum* number of variables necessary to undertake our intended analyses, in line with ONS policy. The IVP ingest dataset contains 15 variables.

Data requests:

The Longitudinal Business Database (LBD) is a dataset with information about individual businesses (see LBD variables). Due to its sensitive nature, associated with the potential to identify individuals, the LBD is only available via the Secure Research Service (SRS) in the ONS.

To access the LBD, we submitted a project application on the ONS Research Accreditation Service which contains specific details of the intended analyses, datasets to be ingested into the SRS, and ethical considerations. Once the application was approved, we shared the Innovation Vouchers Programme data with ONS for the purpose of ingestion into the SRS and linking between our datasets and the LBD.

All outputs from the analysis must be approved by the ONS prior to export. No numeric, graphical, or qualitative results that can be used to identify an individual business will be allowed outside the SRS. Only accredited individuals are permitted to work with data in the SRS.

Data linking:

Datasets have been linked by ONS at the time of ingesting the programme data into the SRS. The ONS matched company registration numbers (CRNs) to ONS enterprise references (entrefs).

We will report the extent to which our ingest datasets are linked with the LBD. We will report the total number of companies linked via CRN, and the number for which no matches were found. We will also report the total number of matches for the variables used to compute our outcome measures and covariates: turnover, employees, local units, SIC codes.

Prediction Survey

Prior to the statistical analysis, we elicited predictions about the impacts of the three innovation programmes on long-term outcomes for participating businesses. The respondents include analysts and policy and programme specialists at Innovate UK, the Department of Business and Trade, and the Innovation Growth Lab, who have experience with innovation programmes and policies, but were not necessarily involved in any of the three programmes considered in our evaluation. We also surveyed members of the IGL network and participants on the Social Science Prediction Platform, who have experience answering prediction surveys, but are very unlikely to be familiar with business support programmes.

The goal of the prediction measurements is to allow construction of a prior distribution, to be used in Bayesian models for the legacy evaluations project (see Bayesian analysis).

The prediction survey involves three steps:

1. Provision of a brief description of the programmes
2. Explanation of the outcome measures
3. Elicitation of estimates

Initial Ingest

The initial ingest was an opportunity for us to evaluate the feasibility of the analysis presented in this plan. During this step, we ingested and examined only data regarding the features and outcomes of the control group for IVP.

We completed power calculations to determine the magnitude of differences between the treatment and control groups that will be required in order to detect differences between them. We then compared the MDES values to a benchmark value to assess the feasibility of the intended analyses.

Power calculations:

We completed power calculations to determine the largest difference in outcomes that will achieve sufficient statistical power to limit the likelihood of Type-II error. As the power calculations used only control group data, we assumed an equal number of businesses in the treatment and control groups.

Since we are testing changes over time, we completed this calculation for each of the primary outcomes: cumulative turnover, cumulative employment, and productivity. We first calculated the SD for each of the primary outcomes, and then the absolute (percentage) difference required, and the MDES, to achieve sufficient statistical power.

We reported the:

- Assumptions of our power analysis (e.g. power of 80%, alpha of 5%)
- Number of observations in the control groups

- Baseline/reference value and SD for the outcome variable
- Difference in mean value (% distance from baseline/reference) needed for the result to be detectable under our assumptions.

After completing our primary analysis and secondary analysis, we compared the observed difference between the treatment and control groups to the values calculated as necessary to achieve sufficient statistical power.

Covariates and timescale:

Another key aspect of the initial ingest was to finalise the selection of covariates to be included in the statistical models for both the primary and secondary research questions. Some characteristics of businesses have a large number of possible values (e.g. sector) which increases the degrees of freedom in our statistical models. During the initial ingest, we used data from the control group only to measure the distribution of values for these variables, both in the outcome years and in subsequent years.

We also determined the feasibility of extending the timescale for outcomes of interest, by measuring their variation following the COVID-19 pandemic (i.e. outcomes after 2019/20). This allowed us to determine how far to extend our analysis of RQ1b.

Finally, we recorded the number of missing values in the outcome data and covariates, and any outliers, to help with refining this analysis plan.

Balance tests

As a first step in our experimental analysis, we will complete balance tests by computing the mean values of all baseline characteristics, and testing for joint significance of these characteristics in predicting treatment status.

We will first review our datasets by describing key characteristics of the treatment group, control group, and entire sample for IVP. These characteristics, as described above, include categorical variables (such as region) and continuous variables (such as company age).

First we will compute mean values for all baseline characteristics, and then conduct an F-test for joint significance of these characteristics in predicting treatment status.

We will report the:

- Number of companies in each condition
- Percentage of companies in different categories in base year: business size, sector, urban/rural
- Mean (average) values of continuous or binary variables in base year: turnover, number of employees, productivity, company age, number of plants, number of locations
- Standardised difference of variables in T and C in base year: turnover, number of employees, productivity, company age, number of plants, number of locations
- Results of F-test for joint significance: F-statistic, degrees of freedom, p-value,

To accompany this, we will create density plots of the outcome measures and covariates.

Matching

For our quasi-experimental analyses, we will compare programme participants to a comparison group, which will be constructed by means of matching.

Using the database of SMEs in the LBD, we will use a logistic model with the binary outcome of participation in the programme, and the following characteristics in the programme year as matching variables:

- Log of business age in programme year
- Sector in programme year (determined with SIC code in LBD)
- UK region in programme year (identical to LBD region if 1 reporting unit, otherwise "multiple")
- Turnover in each year prior to the programme year
- Number of employees in each year prior to the programme year
- Multiplant status

If we are given permission to ingest a large dataset of IP outputs to the SRS, we will also include the following matching variables:

- Patents granted prior to the programme year
- Trademarks granted prior to the programme year
- Registered designs granted prior to the programme year

Potential matches will be drawn from the whole population of businesses in the LBD that did not participate in or apply to IVP (i.e. do not appear in the programme dataset). Businesses in the IVP treatment group will be matched to businesses with similar characteristics in the programme year.

We will use two general approaches: propensity score matching (PSM) and coarsened exact matching (CEM). PSM will be the primary matching approach. It is a statistical method that matches treated units to control units based on their estimated probability (propensity score) of receiving treatment, calculated using the observed covariates listed above. Our preferred PSM method is many-to-one nearest neighbour matching (NNM), in which multiple control units can be matched to a single treated unit, each control unit is only used once, and a fixed ratio of control to treated units is specified. We will test different ratios (e.g. 4:1 and 5:1) to identify the optimal ratio. The advantage of many-to-one NNM, when compared to one-to-one NNM, is that it usually reduces variance and increases precision in treatment estimates.

We will also employ calipers in our NNM, in which a maximum allowable distance (caliper) is specified, beyond which matches are rejected. Calipers have been found to be an effective addition to NNM because they can increase balance and reduce bias (Austin, 2013; Austin & Stuart 2015). In line with Austin (2011), we will first attempt to use a caliper of 0.2 of the standard deviation of the logit of the propensity score, a value found to minimize mean squared error in estimated treatment effects. However, we will also explore other calipers, such as a more lenient 0.25 and a more restrictive 0.1. To compute the maximum allowable distance, we will first calculate the propensity scores, apply a logit transformation, calculate the standard deviation of the logit propensity score, and multiply by the caliper (e.g. 0.2). Testing multiple options will allow us to explore the tradeoffs of wider calipers (more matches but potentially worse balance) and narrower calipers (fewer matches and better balance).

As an alternative approach, we will also use coarsened exact matching (CEM) and report the results from the resulting analysis in our supplementary findings. CEM involves (temporarily) reducing continuous variables into categories, before implementing exact matching. We will

test different degrees of coarseness to assess the influence of our choices on balance and the number of matched units.

We will assess the quality of the matches from our PSM and CEM approaches with balance tests, and will select the best matching procedure prior to running our quasi-experimental analysis (see below). With both PSM and CEM, there is a risk that treatment units remain unmatched – and dropping any units due to matching would violate the ITT principle. In other words, we will find matches for as many of the units in the treatment group as possible, to preserve the ITT estimation in our statistical analysis. Therefore, our first priority is to ensure all or almost all treatment group units are matched. Our second priority is to maximise balance between the treatment and control (comparison) groups.

If our PSM approaches result in more than 2% of the intervention/treatment group units being unmatched, we will consider other options, such as NNM with replacement, or the omission of some matching characteristics. When considering balance, post-matching std mean differences for each of the matching variables should be between -.1 and .1, and post-matching variance rations for each of the matching variables should be between 0.5 and 2.

To test the matching approaches, and implement the optimal matching procedure, we will use the MatchIt package in R.

Balance tests (matching):
To assess the balance between the programme participants and the matched comparison group, we will use a series of comparisons generated by functions in MatchIt (see example here).

The function `summary.matchit()` displays information including standardised mean differences, variance ratios, and empirical CDF statistics. We will also generate Love plots to summarise the quality of the matches, using `plot.summary.matchit()`.
To select which PSM and CEM method to apply, we will select the approach with the best overall balance across all three programmes, and for which all treatment units are matched in each programme.

Descriptive statistics

Prior to the full statistical analysis, we will generate descriptive statistics for IVP, including observable characteristics of the participating and non-participating businesses and the long-term outcomes of interest (mean, minimum, maximum, and SD of values, alongside the sample sizes of full observations). We will conform to the ONS rules about identifying information, such that these descriptive statistics can be exported and included in our results.

For the prediction survey, we will report the estimates and degrees of confidence by the respondent group, as well as the expected responses of others. The estimates will also be presented using visualisations such as boxplots (for relative measures) and forest plots (for point estimates).

Correction for multiple comparisons

Our analysis includes many statistical models. The primary analysis relates to the research question RQ1, and comprises 3 statistical models for IVP.

The secondary analysis includes secondary research questions RQ1a, RQ1b, as well as secondary analysis for RQ1.

To address concerns about increased Type-I error due to multiple hypothesis testing, we will use a correction for the family-wise error rate (FWER). This correction will relate to both the primary and secondary analyses. We will correct for the FWER following Romano and Wolf (2005), using the `wildrwolf` package.

Following Guess et al. (2023), we will apply an adjustment to the primary analysis only, comprising 3 hypotheses for IVP. Second, we will apply the adjustment for the primary and secondary analyses for each research question.

We will report adjusted p-values alongside the model results for all outcomes.

LBD variables

C. List of variables

Variable Name	Variable Description
year	Year of IDBR snapshot
quarter	Quarter of IDBR snapshot
entref	IDBR enterprise reference
ruref	IDBR reporting unit reference
wowref	IDBR enterprise group reference
luref	IDBR local unit reference. 8 digits + checkletter
wow_ultfoc	Ultimate Parent reference provided by Dun & Bradstreet
ent_employment	Enterprise employment
ent_legalstatus	1-digit code representing the legal status of the business
ent_sic2007	Enterprise SIC 2007 code
ent_sic92	Enterprise SIC 1992 code
ent_turnover	Enterprise turnover reported for a business in thousands
ent_active_type	LBD activity marker for enterprises
ru_employment	Reporting unit employment
ru_region	Reporting unit Region
ru_rusic2007	Reporting unit SIC 2007 code
ru_rusic92	Reporting unit SIC 2003 code
ru_turnover	Reporting unit turnover
lu_employment	Local unit Employment
lu_region	Local unit region
lu_sic2007	Local unit SIC 2007 code
lu_sic92	Local unit SIC 2003 code
ru_active_type	LBD activity marker for reporting units
lu_active_type	LBD activity marker for local units
dummy_ent	A dummy enterprise indicator variable

Pre-registration of previous analyses on short term programme effects

Prior to the current project on long-term effects of the Innovation Vouchers programme, short-term effects of the same programme and same RCT were analyzed in a separate project. The project was pre-registered on the AEA RCT registry, with ID AEARCTR-0001556.

The preregistration outlined hypotheses relating to collaboration, innovation activities and outputs, and business performance. The following indicators were identified: percent of innovation activities conducted with the help of external partners, percent of firm's turnover spent on innovation, percent of turnover coming from new or improved products or services, and business turnover. Data was to be collected by means of surveys.

The findings of this analysis were published in Kleine, M., Heite, J., & Huber, L. R. (2022). Subsidized R&D collaboration: The causal effect of innovation vouchers on innovation outcomes. *Research Policy*, *51*(6), 104515. https://doi.org/10.1016/j.respol.2022.104515