Catalogue no. 11-633-X — No. 063
ISSN 2371-3429
ISBN 978-0-660-99770-4

Analytical Studies: Methods and References

Applied Quantitative Impact Analysis: A Review of Contemporary Methods for Evaluating Programs Serving Businesses

by Manassé Drabo, Landry Kuate and Michael Willox

Release date: May 27, 2026

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, the Agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Applied Quantitative Impact Analysis: A Review of Contemporary Methods for Evaluating Programs Serving Businesses

by

Manassé Drabo, Landry Kuate and Michael Willox

Economic and Social Analysis and Modelling Division, Analytical Studies and Modelling Branch, Statistics Canada

11-633 063

2026003

ISSN 2371-3429
ISBN 978-0-660-99770-4

May 27, 2026

Analytical Studies: Methods and References

Papers in this series provide background discussions of the methods used to develop data for economic, health and social analytical studies at Statistics Canada. They are intended to provide readers with information on the statistical methods, standards and definitions used to develop databases for research purposes. All papers in this series have undergone peer and institutional review to ensure that they conform to Statistics Canada's mandate and adhere to generally accepted standards of good professional practice.

The papers can be downloaded free at www.statcan.gc.ca.

Table of contents

Abstract

This article presents a methodological framework for evaluating the economic impacts of public programs, applied to agencies like Crown corporations (e.g., Business Development Bank of Canada [BDC], Farm Credit Canada and Export Development Canada). Using difference-in-differences, event studies and survival analysis, it estimates quasi-causal effects on outcomes such as sales, employment and business longevity. The paper outlines each method's theory, implementation and data needs, emphasizing the value of Statistics Canada's linked datasets. It also discusses key assumptions with a case study on BDC illustrating practical application. The work promotes collaboration with Statistics Canada for evidence-based policy evaluation.

Keywords: difference-in-differences, event study, survival analysis, impact assessment, policy analysis, firm performance, public program effectiveness.

1. Introduction

Statistics Canada has a long history of collaborating with other government departments and Crown corporations to assist them in evaluating the impact of their policies and programs (e.g., unemployment insurance and tax competitiveness) with the objective of providing Canadians with unbiased, accurate and relevant information about the issues that affect them most. Recently, Statistics Canada has taken on a larger role, assisting Canadian organizations, primarily federal government agencies and Crown corporations, in measuring the impact of their services on the clients they serve. Impact evaluation support as an analytical service has grown rapidly in the last decade because of Statistics Canada's continued improvements in integrating its extensive and diverse microdata assets with partner organizations' unique client information in a secure environment that protects the confidentiality of individuals' and organizations' information.

This article provides an overview of the process that Statistics Canada follows to support organizations in conducting impact analysis and describes the commonly used statistical methods employed. To understand the types of information that Statistics Canada includes in its reports, results from a recent study performed for the Business Development Bank of Canada (BDC), a federally funded Crown corporation, are presented. The BDC programs provide financial services primarily to Canadian small and medium-sized enterprises (SMEs). In general, their services are aimed at enhancing sales, employment and longevity, particularly when business conditions are most challenging, like during the recession of 2008–2009 or the COVID-19 pandemic.

The article is structured as follows: Section 2 describes the study context and outlines key arguments for the methodological approach rationale. Section 3 presents the methodological framework and details data requirements, Section 4 discusses limitations, Section 5 applies the framework to Crown corporation programs and Section 6 concludes the paper.

2. Program evaluation context and methodological rationale

Governments increasingly seek to evaluate the effectiveness of programs delivered through public departments and private firms, aiming to ensure accountability, optimize resource allocation and improve outcomes. However, rigorous causal evaluation often faces practical constraints—especially in real-world policy environments where randomization is infeasible, data are fragmented and interventions are rolled out unevenly. Moreover, technical and financial constraints prevent conducting an evaluation before implementation. In many cases, programs are evaluated after their implementation, which can be very challenging (Frenette, Chan, & Handler, 2025). In this context, agencies and some government departments have engaged in collaboration with Statistics Canada. As the national statistical agency, Statistics Canada offers secure access to high-quality administrative data, longitudinal records and standardized metrics that enable robust, scalable and policy-relevant econometric analysis. Their infrastructure supports empirical work that balances methodological rigour with operational feasibility.

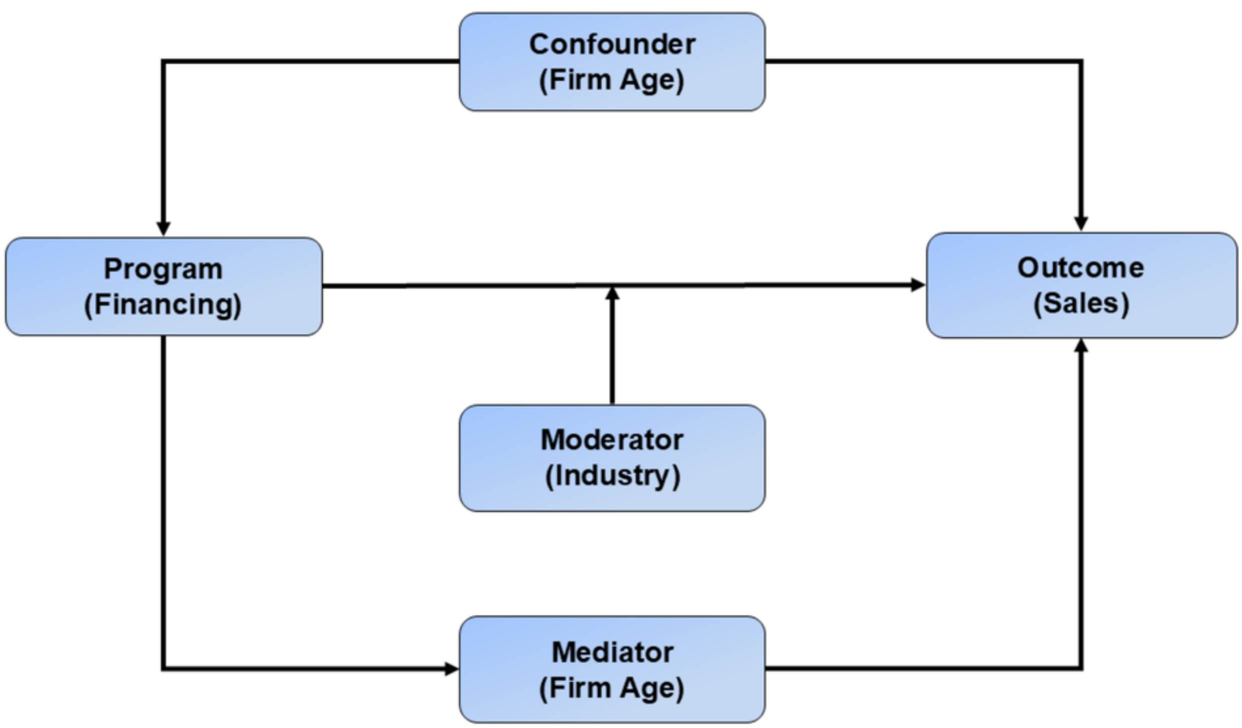
When assessing whether a program improves business performance, it's important to understand how various external factors can influence the relationship between the program and changes in outcomes (e.g., employment and sales). In impact analysis, variables such as confounders, mediators and moderators play distinct roles (Figure 1). Confounders are variables (e.g., firm age, size and managerial experience) that influence both a business's decision to participate in a program (treatment assignment) and its final outcomes. Confounders are important to consider because they can create the illusion of an effect where none exists. For instance, a business receiving a loan may exhibit sales growth, but this could stem from favourable market conditions rather than the program itself (i.e., the loan, which is the "intervention"). When market conditions help or harm young and old businesses differently, for example, confounders like firm age can obscure the true impact of the program.

Mediators, like confounders, influence outcomes; however, they are best described as an output of the program that, in turn, causes a change in the outcome. They do not influence businesses' decisions to participate in the program that provides a "treatment" like a loan. Consequently, they are described as lying along the causal path. For example, a loan may increase a firm's marketing capacity, which then drives higher sales; the marketing investment is the mediator. Last, moderators influence the strength of the effect without influencing the decision to participate in the program. For instance, a program may benefit tech firms more than manufacturers, making industry type a moderator.[1]

Understanding these relationships is important because evaluating the causal impact of such programs is complex because of confounding factors (e.g., differences in firm characteristics, industry dynamics or macroeconomic trends) which can obscure whether observed outcomes result from the programs themselves. Consequently, to provide reliable evidence for policy makers, stakeholders and funding agencies, robust methodologies are essential to isolate these program effects from other influences on outcomes.

Distinguishing between confounders, moderators and mediators is not just a technical detail—it is essential for understanding how programs generate change. Identifying confounders allows analysts to isolate the program's effect from external influences, while recognizing mediators helps to reveal the mechanisms through which a program operates. Similarly, knowing which moderators positively or negatively affect the strength of the impact allows analysts to assess for whom or under what conditions a program is most effective.

Figure 1
Causal pathways between programs and outcomes influenced by external factors: Confounders, mediators, and moderators with an illustrative example



Source: Author's calculations.

These distinctions enable more accurate estimates of direct effects (those attributable solely to the program) and indirect effects (those channelled through intermediate variables). Moreover, program managers are not passive participants in this process. Their detailed knowledge about

1. More technically, a mediator lies on the causal pathway, while a moderator interacts with the treatment to change the magnitude of the effect on the outcome.

how programs are designed, delivered and targeted is essential to identifying the relevant influencing factors and to properly structuring an impact evaluation. By documenting and sharing information on eligibility criteria, implementation practices and expected pathways of change, federal agencies and organizations strengthen the validity of causal analysis and contribute directly to the success of evidence-based evaluation.

While approaches using randomized controlled trials (RCTs), regression discontinuity designs (RDDs) and instrumental variable (IV) are considered among the best for causal inference, they are not always suitable for program evaluations. RCTs require random assignments, which may be ethically or politically unacceptable in public service delivery. RDDs depend on sharp eligibility thresholds that many programs lack, and IV methods hinge on finding valid instruments—often elusive in complex institutional settings. Moreover, these methods typically focus on narrow populations or short-term effects, limiting their generalizability and policy relevance. By contrast, quasi-experimental designs like difference-in-differences (DiD), event studies (ESs) and survival analysis (SA) offer flexible frameworks that can accommodate staggered implementation, heterogeneous treatment effects and long-term dynamics.

This article discusses these three advanced quantitative methods—DiD, ESs and SA—to assess and isolate the quasi-causal effects of federal agencies and organizations by comparing specific outcomes of client businesses that participate in or receive support from the agencies' programs. The outcomes of these clients, often referred to as "the treated" or members of the "treatment group" are compared with the outcomes of a comparison group composed of non-clients over time to provide a robust alternative to descriptive statistics.

The objectives of this article are threefold. First, they describe the theoretical foundations, practical implementation and data requirements of these methods in a clear, accessible manner. Second, they demonstrate their applicability to a wide range of policy evaluations beyond Crown corporations. Third, they illustrate their use in evaluating Crown corporation programs, with detailed results for the BDC showing significant sales and employment growth. In doing so, this article builds on the recent work by Frenette, Chan, & Handler (2025) on the opportunity of leveraging Statistics Canada data for impact analysis and outlines methodological framework details. By demonstrating the feasibility and value of these methods, the article aims to foster a culture of accountability and continuous improvement in public program delivery.

3. Methodological framework and data requirements

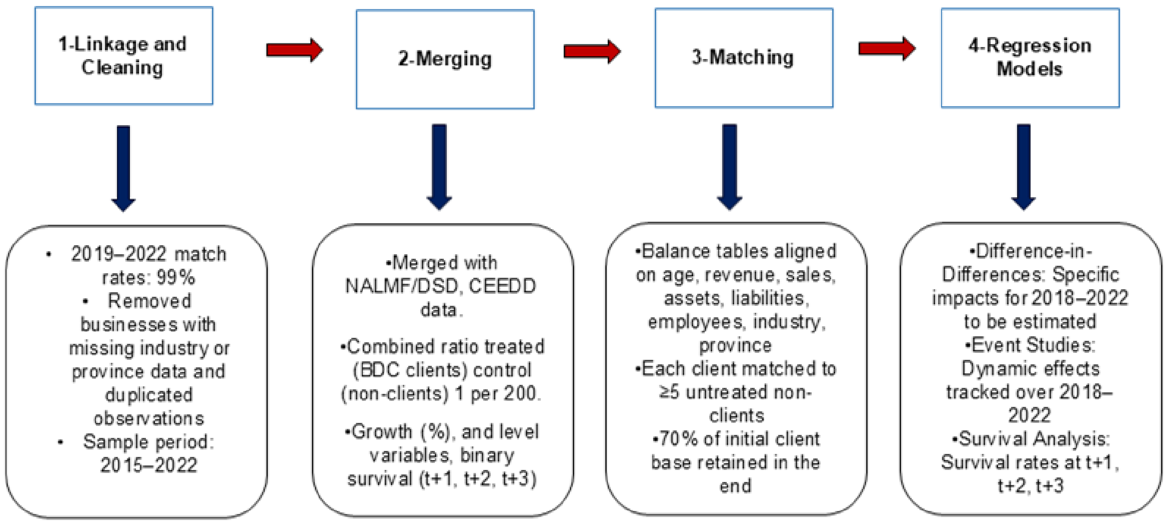
Evaluating federal programs requires methods and data that can help isolate quasi-causal effects from external factors, such as firm characteristics, industry trends or macroeconomic shocks. Statistics Canada's Economic and Social Analysis and Modelling Division has developed a robust analytical setting that can be summarized in four steps, as described in Figure 2.

Step 1 consists of linking the client or program data to Statistics Canada's Business Register (BR).[2] The resulting linked dataset is then cleaned by removing firms with missing information required for linking, duplicate information or low-quality linkage to the BR.[3] In step 2, the linked data from step 1 are merged with other Statistics Canada business datasets such as the National Accounts Longitudinal Microdata File (NALMF) and the Diversity and Skills Database (DSD),

2. Microdata linkage is conducted in accordance with Statistics Canada's Directive on Microdata Linkage.
3. Linkage quality was evaluated using a weighted scoring system that prioritized name similarity, geographic consistency and other identifiers such as phone number and address. Matches were classified into five categories: A (excellent) for near-perfect matches, B (very good) for strong matches, C (good) for moderate matches, D (acceptable) for minimal confidence and E (use with caution) for weak matches. The best match for each record was flagged, though alternative matches were retained for review.

components of the Canadian Employer–Employee Dynamics Database (CEEDD). A brief description of these data is provided in Section A1 of the appendix.

Figure 2
Statistics Canada methodological framework for impact studies analysis, illustrative example



Source: Statistics Canada, Economic and Social Analysis and Modelling Division.

The selection of firms that compose the comparison group for analysis is undertaken in step 3 by matching treated (clients) with untreated firms with similar characteristics that do not receive the product or participate in the program. Step 4 of the methodology describes the regression models that rely on three complementary econometric approaches—DiD, ESs and SA—to assess the impact of services on business outcomes, including sales, employment and longevity. Additional details on the regression model are provided in Section A2 of the appendix. The remainder of this section describes each step in detail.

Examples of results from the BDC's financing programs for SMEs are used to illustrate the information and guidance on interpreting results that organizations typically receive in an impact analysis report. In most cases, Statistics Canada's impact analysis reports focus on providing program managers the tools and knowledge to understand the results and convey them to their stakeholders as required. This gives the agency control over how to disseminate the information in these reports and any accompanying statistical output.

3.1 Linkage and cleaning

The effectiveness of DiD, ESs and SA hinges on high-quality, longitudinal data. The CEEDD is a comprehensive, linked administrative dataset that captures the evolving relationships between employers and employees across the Canadian economy.

Agencies requesting impact evaluations typically provide Statistics Canada with business client data, allowing clients to be identified in the CEEDD. The agency's client data may also include information about the nature of the services their clients received. For example, the BDC evaluation included data on clients receiving financing only, advisory services only or both, with precise start years to define treatment periods. These client data are linked to the BR, a comprehensive database of Canadian businesses capturing annual data on revenue, employment, industry (using North American Industry Classification System codes), firm size (e.g., number of employees and revenue bands) and location (e.g., province and postal code). Once client businesses are identified in the BR, their corresponding information can be located in the CEEDD, as described in more detail in Section 3.2.

After the linkage, the data are cleaned to remove firms with missing relevant information (e.g., enterprise identifier, industry or province and duplicate observations).

3.2 Merging

For DiD and ESs, agencies need longitudinal outcome data for treated and control groups, covering key performance indicators like sales and export values (in Canadian dollars) and employment. Using a business identifier found in the BR, Statistics Canada merges the resulting dataset in step 1 to one or more of the business databases in the CEEDD (see Liu & Zhang, 2025, for a detailed description of the CEEDD and its structure). The CEEDD's primary database for impact analysis at Statistics Canada is the NALMF. It covers incorporated enterprises with employees but excludes unincorporated businesses without staff. For analyses involving sole proprietorships, data from the Financial Declaration File and T2 Schedule 50 may be required. Another important database is the DSD, which captures demographic information on business owners and employees for the following dimensions: sex, age, marital status, immigrant status, disability status, Indigenous peoples and racialized groups.[4]

Once the datasets are integrated, variables of interest are constructed to support impact analysis. These include treatment indicators based on program participation timing and intensity; firm-level covariates such as size, industry and growth trends; and outcome variables like employment growth, survival and productivity metrics. Where applicable, firm characteristics are also aggregated or lagged to minimize endogeneity. Workforce-level variables from the DSD enable the incorporation of equity- and diversity-relevant dimensions, further enriching the causal interpretation of program impacts.

3.3 Matching

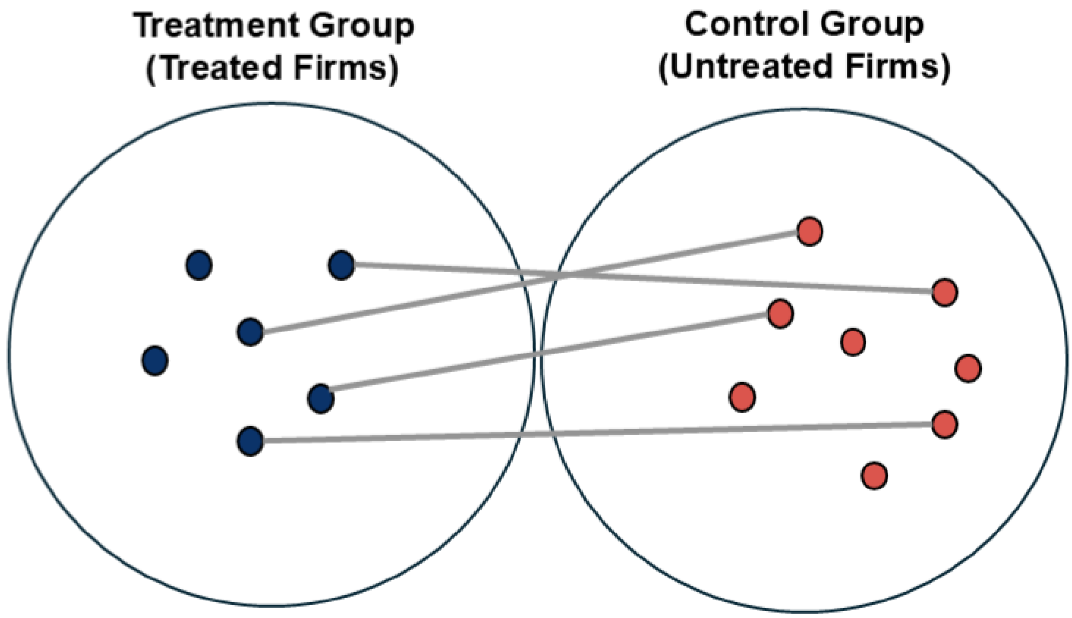
To test whether a program (e.g., government loan or grant) causes a business to grow or survive longer, it's important to compare participating or treated firms with similar firms that didn't receive the support—a counterfactual group. This is where matching comes in. Matching means finding untreated businesses that look as similar as possible to the treated ones based on characteristics like industry, size, revenue, location and number of employees. By making the groups comparable, one can reduce the influence of external factors, such as firm age or market conditions, which could otherwise distort the results. Matching methods are commonly used in program evaluation to build comparison groups that closely resemble the group receiving the intervention (Stuart, 2010). Matching estimators include propensity score matching (PSM), nearest neighbours matching, caliper matching, kernel matching (KM) among others. Recent impact study analyses have focused on the KM estimator, which consists of weighting untreated firms based on their similarity to treated ones using a kernel function. The KM estimator uses all untreated firms and requires a careful bandwidth selection for its efficiency.

Figure 3 shows a one-to-one matching process used in impact analysis, such as PSM, to construct a comparable control group. The large circle on the left represents the treated group (firms that received the program or intervention, shown as blue dots), and the large circle on the right represents the untreated or control group (firms that did not receive the program or intervention, shown as red dots). A grey line connecting a blue dot to a red dot indicates a successful match, where the two firms share highly similar observable pre-treatment characteristics (e.g., size, age and industry). The goal is to ensure that the only systematic difference between the matched firms is the treatment itself, isolating the treatment's causal effect. As shown, not every firm in the image is matched. Some treated and untreated firms remain unmatched (unconnected dots) and were excluded from the analysis because no sufficiently

4. For more information on how Indigenous peoples and racialized groups are captured in the CEEDD, see Gueye, Lafrance-Cooke and Oyarzun (2022); Gueye (2023); and Gueye (forthcoming).

similar counterpart could be found, which is a common outcome in real-world matching procedures.

Figure 3
Matching based on observable firm characteristics



Source: Statistics Canada.

While matching estimators can enhance quasi-causal inference in DiD analysis by improving covariate balance and reducing bias from observable confounders, their use introduces several critical concerns that may outweigh these benefits. Among these concerns is the risk of violating the parallel trends assumption—matching aligns groups on observed characteristics but cannot account for unobserved, time-varying factors that may drive differential trends. Additionally, matching often reduces sample size and statistical power by discarding unmatched units, and results can be highly sensitive to the choice of matching algorithms, leading to instability in estimates. These limitations underscore the importance of rigorous diagnostics, sensitivity checks and robust inference methods when integrating matching into DiD frameworks.

Applying matching estimators allows comparison between matched treated firms and the resulting comparable matching untreated firms. This comparison helps isolate the program's true effect. For this reason, it is important to track businesses before and after they receive support to see how their outcomes change over time. For example, using data from the CEEDD to track business activity, SA looks at whether businesses are still active a few years after receiving support.

3.4 Regression models

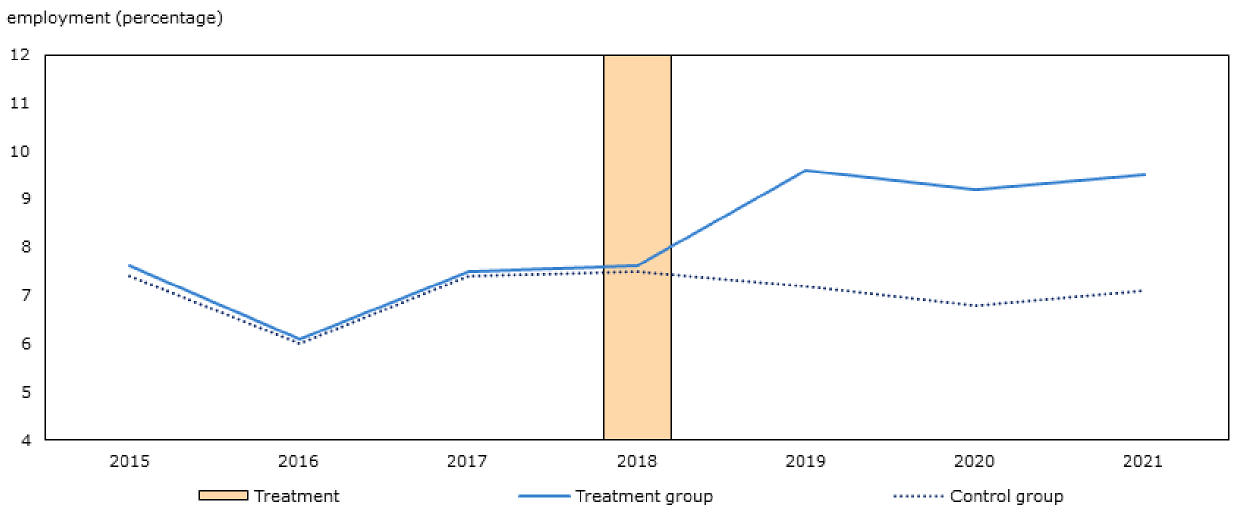
This article discusses three commonly used methods for impact analysis to measure the quasi-causal effects of federal agencies and organizations programs: DiD, ESs and SA. Each approach compares outcomes for treated firms to a carefully matched group of similar untreated firms, helping to isolate the true impact of the program. Their shared use of regression and matching makes them both accessible and powerful tools for policy evaluation. A summary comparison between methods is provided in Table 1 of Section A3 of the appendix.

3.4.1 Difference-in-differences

One of the most effective ways to assess whether a business support program—such as a loan, grant or advisory service—causes meaningful changes in outcomes like employment or sales is through a method known as difference-in-differences (DiD). This approach compares how these outcomes evolve over time between two groups: firms that received the program (the treatment group) and similar firms that did not (the control group) (Angrist & Pischke, 2009). The logic is straightforward: if both groups were on similar paths before the program, any consistent difference that emerges afterward may be attributed to the program. Chart 1 shows a hypothetical example of employment levels before and after the treatment year, 2018. If the treated firms begin to outperform the control firms only after receiving support, the divergence may indicate the size of the program's effect on the clients' employment levels. Throughout the article, the impact is called the average treatment effect on the treated (ATT) rather than other treatment effect estimates such as the intention-to-treat or marginal treatment effects.

For this comparison to be valid, the treatment and control groups must be comparable before the program begins. This is where matching plays an important role. By selecting non-client firms that closely resemble client firms in terms of size, industry, region and other characteristics, the analysis ensures both groups were operating under similar conditions before the treatment. This process helps eliminate the possibility that pre-existing differences, which are unrelated to the program, are driving the observed outcomes.

Chart 1
Difference-in-differences for a program aimed at increasing employment, illustrative example



Source: Statistics Canada.

A foundational assumption of DiD is the parallel trends assumption. It asserts that in the absence of the program, both groups would have continued to follow the same time path as before the treatment. Testing the parallel trends assumption in a DiD framework is inherently limited because the counterfactual—what would have happened to the treated group in the absence of the intervention—cannot be observed. While examining pre-intervention trends between treated and control groups can offer suggestive evidence, it does not constitute a definitive test. Similar trends before treatment may indicate plausibility, but they don't guarantee that those trends would have continued post-treatment without intervention.

This limitation highlights the importance of having a long and stable pre-intervention period. The more extended and consistent the similarity in trends, the stronger the case for assuming parallel trajectories would have persisted. However, researchers must remain cautious—even well-aligned pre-treatment paths cannot fully validate the assumption, and any divergence of post-treatment could reflect changes in underlying differences rather than treatment effects. Thus, while pre-trend analysis is useful, it should be treated as supportive—not conclusive—evidence.

Chart 1 shows a case where pre-treatment trends are parallel in the employment levels. In this example, the gap in the employment levels that emerges during the post-treatment period between the treatment and control groups may represent the impact of the program or ATT. This method is particularly useful because it accounts for both time trends and baseline differences in employment for the treated and untreated firms, unlike descriptive comparisons, which often ignore these factors.

By combining DiD with matching and using detailed firm-level data, such as payroll records or corporate tax filings, it is possible to isolate an estimated effect of a program with greater confidence. This approach not only improves the credibility of evaluation findings but also provides decision makers with more accurate information about how their programs affect their clients.

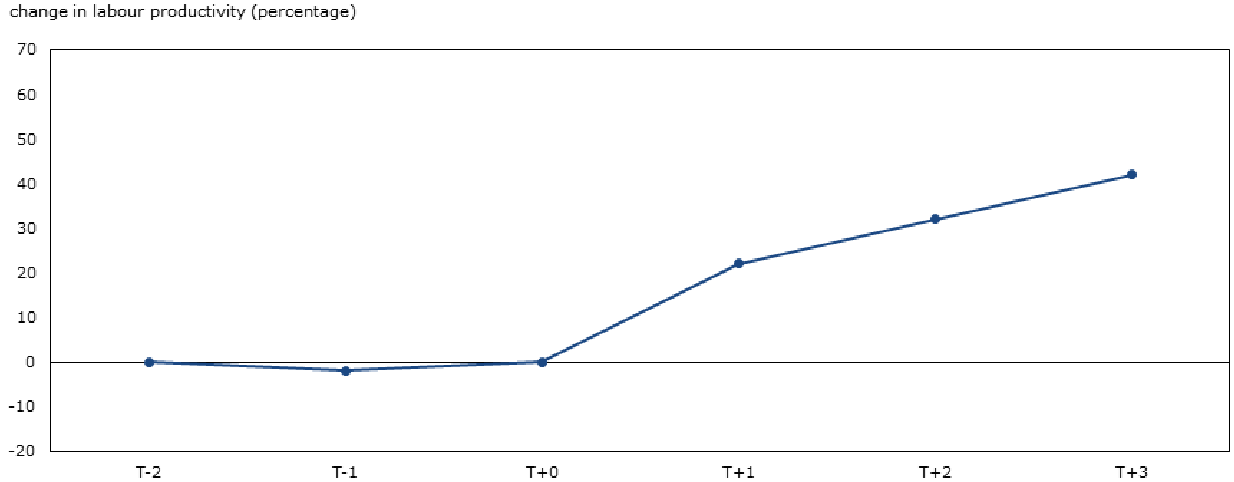
3.4.2 Event studies

ESs extend the DiD approach by showing how the effects of a program develop over time (Khandker, Koolwal, & Samad, 2010). Instead of comparing average outcomes before and after a program, ESs track changes in each year leading up to and following the initial treatment. This produces a detailed timeline of how business outcomes—such as labour productivity, sales or employment—respond to the program. Chart 2 illustrates this approach by plotting the average change in labour productivity for treated firms, relative to a baseline period. The horizontal axis shows time in relative terms, where t is the year that the clients begin participating in the program. The years before the program (t–2, t–1) serve as a pre-treatment check, while the years after the program (t+1, t+2, t+3) reveal how treatment effects unfold over time.

As with DiD, the credibility of an event study depends on comparing treated firms with similar untreated firms—identified through matching on characteristics like size, industry and region. This ensures both groups faced similar conditions before the program, allowing the analysis to isolate changes likely driven by the program itself.

ESs also provide valuable insights into the timing and duration of program impacts. For example, Chart 2 shows no measurable improvement in productivity when the program begins (t+0), but large and growing gains emerge in the years that follow. This pattern suggests that the program's benefits accumulate over time, possibly as firms adapt or reinvest. Such dynamic effects would be missed by simpler methods like descriptive statistics, which typically report only a single average outcome. Moreover, because ESs display results for each period, they help analysts identify whether effects fade, plateau or continue to rise.

Chart 2
Event study measuring programs' impacts on changes in labour productivity relative to a control group



Source: Statistics Canada.

Conducting ESs requires data spanning several years before and after firms begin participation in the program. While this allows for a good understanding of impact trends, it also introduces challenges. For instance, some firms may exit the dataset in later years, either because they shut down or because their data are no longer reported. This attrition can bias results if the firms that drop out are systematically different. To address this, analysts rely on large samples and careful data cleaning and may apply techniques like inverse probability weighting to adjust for dropouts.[5]

3.4.3 Survival analysis

SA is another way to assess whether a business program helps firms succeed over time, especially during difficult economic periods. In this context, "survival" simply refers to whether a business is still active one, two or three years after receiving program support. Instead of using complex statistical models, this approach focuses on a more straightforward question: Are firms that were part of the program more likely to stay in business than similar firms that were not?

To answer that question, the method compares survival rates between treated firms (those that received support) and a matched group of untreated firms that are similar in size, industry, region and other characteristics. Matching helps ensure that differences in survival rates can be linked to the program itself rather than other factors. This approach fits naturally within the same DiD framework used to study outcomes like sales or employment, making it easier for agencies already familiar with those methods to apply it to business survival.

Firms are considered alive if they continue reporting positive values for revenue and employment. If that information disappears, the firm is considered as having exited the market. The method then compares the percentage of firms still operating after one, two or three years between the treated and control groups. For example, if 80% of treated firms are still active two years later, compared with 70% of their matched peers, the results suggest, but do not necessarily prove, that the program increased the survival rate by 10 percentage points. It is important to note that the key assumption here is different from that of a DiD analysis. Because all firms start in a single state (existence), the parallel trends assumption is not applicable. Instead, the most important assumption is that the reasons for a firm dropping out of the dataset are not related to the program itself, a concept known as "uninformative censoring."

5. Inverse probability weighting is a method used to adjust for differences in how likely firms are to participate in the program. It builds on the matching process by giving more weight to firms that look similar but were less likely to be treated, helping to reduce bias from firms dropping out.

While this method doesn't capture exactly how long each firm survives or why some close or die earlier than others, it provides a clear and easily communicated snapshot of program impact. It is especially helpful for evaluating how programs support business resilience in times of crisis, like during a recession or the COVID-19 pandemic. Although more advanced survival models exist, this simplified approach offers agencies a practical and intuitive way to assess whether their programs are helping firms stay in business.

4. Limitations

4.1 Limitations of difference-in-differences using two-way fixed effects

Did relies heavily on the parallel trend's assumption, which posits that treated and control groups would have followed similar outcome trajectories (e.g., sales or employment) in the absence of the program. However, this assumption is not directly testable because one never observes the counterfactual trend for the treated group. Even if pre-treatment trends appear similar, that doesn't guarantee they would have remained parallel post-treatment. If the assumption is violated—because of unobserved confounders, policy anticipation or differential shocks—the DiD estimate will be biased.

Other notable limitations include :

(i) heterogeneous treatment timing: when units receive treatment at different times, standard DiD (especially two-way fixed effects [TWFE] models) can produce misleading estimates caused by negative weighting or contamination from already-treated units

(ii) time-varying confounders: if factors that influence the outcome evolve differently across groups, they can bias the estimated treatment effect

(iii) limited pre-treatment data: without a long pre-treatment period, it's difficult to assess whether trends are truly parallel

(iv) spillover effects: if treatment affects control units (e.g., through markets or networks), the control group is no longer a valid counterfactual.

ESs are sensitive to survivor bias, where firms that close drop out of the sample, potentially inflating later-year coefficients. Clients with higher survival rates attributable to program support may show larger sales differences in later years, reflecting survival rather than direct program effects. Data attrition, as firms exit, reduces sample size and widens confidence intervals, particularly for long horizons. It can be mitigated by using a large sample, but attrition remains a concern. Unlike descriptive statistics, which avoid attrition by focusing on a single point in time, ESs provide evidence of how the influence that programs have on outcomes evolves over time.

The DiD-based SA, which uses a binary survival outcome (alive or dead at t+1, t+2 or t+3), is limited to comparing whether treated firms are more likely to survive than firms in the matched control group. This approach provides estimates of the probability of survival based on program participation at specific points in time after the treatment. However, it does not model the full timing of exit or the probability of survival over time as traditional survival models do. In other words, this approach does not estimate the changing risk of exit as time progresses. For example, it can show that clients have a 10-percentage-point higher probability of being alive at some point after the initial treatment but not measure the duration of survival before closure or the probability of firms surviving over multiple subsequent years.

An important consideration when interpreting findings from analyses using DiD and ES is their external validity, or the extent to which the results can be generalized to firms other than just those that received treatment. Research designs, like DiD and ES, use matching methods to construct control groups, specifically to estimate the ATT. This approach improves internal validity by comparing firms that participated in the program with a set of observably similar firms that did not.

However, this focus on treated firms limits the generalizability of the results. The firms that participated in the program are, by definition, not a random sample of all businesses; they may be more motivated, face unique challenges, or possess specific growth potential that made them eligible or likely to apply. Because matched control groups are constructed to mirror the treatment group, the findings represent the programs' effect on particular sets of firms (clients) under a specific set of economic conditions. Consequently, caution should be exercised when extrapolating estimates of ATT to predict the effects on firms with different characteristics or in different economic contexts.

4.2 Alternative to two-way fixed effects: The staggered difference-in-difference estimator

The Callaway and Sant'Anna (2021) estimator represent an advancement for DiD analyses in staggered adoption settings, where units become treated at different times. Rather than relying on a single TWFE regression, their approach decomposes the problem into a series of simpler 2×2 DiD comparisons: each group-time ATT (g, t) is estimated separately for a specific cohort (group g first treated at time g) and post-treatment period t, using only valid comparisons (typically never-treated or not-yet-treated units as controls). These group-time effects are then aggregated into summary parameters (overall ATT, dynamic/event-study effects or calendar effects) using explicit, researcher-chosen weights.

Compared with traditional TWFE models, this methodology offers several key advantages. First, it avoids the well-documented problems of TWFE in staggered designs, such as negative weighting of treatment effects and contamination from already-treated units serving as controls, which can lead to biased or uninterpretable estimates even under parallel trends. Second, it provides clear, interpretable parameters that directly accommodate treatment effect heterogeneity across cohorts and over time. Third, it supports conditional parallel trends (after covariates) through flexible methods like inverse probability weighting, regression adjustment or doubly robust estimation, and allows the use of not-yet-treated units as additional controls for better comparability.[6]

However, the approach is not without drawbacks. It can be less statistically efficient than some alternatives in certain settings, requiring larger samples for precise estimates, particularly when many cohorts exist or pre-treatment periods are limited. Small cohort sizes may lead to omitted or imprecise group-time effects. Additionally, like all DiD methods, it still relies on a (conditional) parallel trends assumption that remains fundamentally untestable, though pre-trends tests are more straightforward.

This estimator is becoming a standard in applied econometrics for program evaluation and is currently used in ongoing impact analysis projects at Statistics Canada to provide more robust causal evidence in staggered treatment contexts. However, this focus on treated firms limits the generalizability of the results. The firms that participated in the program are, by definition, not a random sample of all businesses; they may be more motivated, face unique challenges, or possess specific growth potential that made them eligible or likely to apply. Because matched control groups are constructed to mirror the treatment group, the findings represent the programs' effect on sets of firms (clients) under a specific set of economic conditions. Consequently, caution should be exercised when extrapolating estimates of ATT to predict the effects on firms with different characteristics or in different economic contexts.

6. As noted by Callaway and Sant'Anna (2021), the use of not-yet-treated units is motivated by identification rather than by the number of available controls. Even when many untreated units exist, only those that remain untreated at time t provide valid counterfactuals for the treated cohort. We clarify this point here for completeness.

5. Practical application: Evaluation of support programs

Consider a program that supports SMEs across Canada by offering financial assistance (grants or scholarships), loans or loan guarantees, as well as advice or training sessions. The program administrator wants to know whether these programs help these businesses grow, for example, by creating more jobs or increasing sales. To find out, one would look at businesses that received assistance in a particular year, say 2018, and track them from 2015 (before the assistance) to 2021 (a few years after). This would allow us to compare what happened before, during and after the assistance. As mentioned in the previous sections, the analysis would compare companies that received support with similar companies that did not receive support to highlight the real effects of the program. One would also consider factors such as treatment timing (when firms receive their initial treatment), the company age, size, location and sector of activity to avoid being misled by external factors.

Statistics Canada generally compares the list of businesses assisted by the program to its database of Canadian businesses. In most studies, approximately 75 to 90 percent of firms receiving program assistance are successfully matched with one or more firms that did not participate in the program. The percentage of successfully matched program participants depends on the quality of the program's business data (unique identification, address and industry sector) and the variables of interest. Matching is performed using a KM method, with each business receiving assistance being matched with at least two similar businesses that did not receive assistance, based on criteria such as age, revenue, number of employees and location.

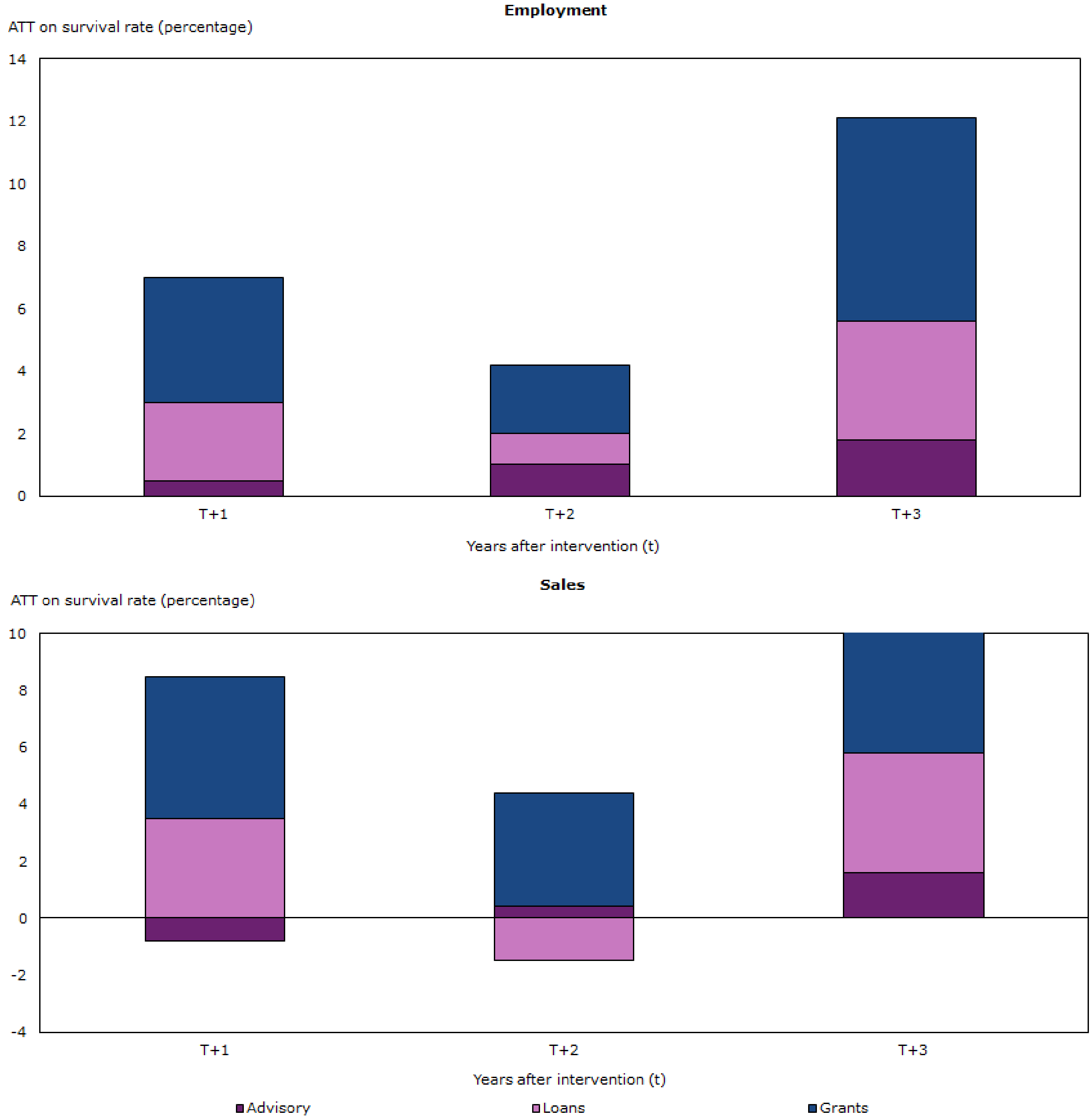
Estimates of the ATT for different treatments are not strictly comparable when treatment effects are estimated separately for each type of treatment, and the matched sample of firms included in each of the analyses is not identical. For example, a group of firms that chose to receive a grant might be fundamentally different (e.g., larger or more productive) from the group of firms that chose a loan or an advisory service. Because each ATT is tailored to a unique set of treated firms, directly comparing the changes in outcomes like employment or sales across the three treatments doesn't precisely measure, for example, if grants are generally better than loans.

Despite this technical limitation, comparing the estimated ATTs across programs still provides substantial utility, especially when the differences in matched samples are relatively small. For instance, comparing results highlights the relative strengths of treatments individually. Comparing ATTs can be particularly helpful for guiding future research designs, especially when the objective is to build a deeper understanding of multi-treatment interactions.[7] In addition, identifying differences in program effectiveness across various periods, often reveals heterogeneity in the firms most sensitive to each intervention. While these comparisons must be made with caution, their insights often allow program administrators to optimize the mix of available supports and target specific interventions more effectively at different types of businesses.

While impact analysis generally tends to follow the method described in Callaway and Sant'Anna (2021), these examples of impact studies conducted by the Economic and Social Analysis and Modelling Division are based on DiD, ES and SA models that include a TWFE specification. Regardless of the specification used, the following impact study results provide a useful example of how impact analysis findings are interpreted and conveyed to a broader audience.

7. It is highly recommended that organizations new to quantitative impact analysis start small by evaluating individual programs as distinct treatments before engaging in more complex causal inference designs, like those that estimate treatment effects for program interactions or groups of programs.

Chart 3
Difference in difference in employment for clients in 2020 compared with non-clients



Note: This figure shows average treatment effects on the treated (ATT) at t+1, t+2 and t+3 for employment and sales, by service type. Estimates control for firm characteristics and time effects. Bars are overlaid, not stacked. Each bar represents the size of the treatment effect starting from zero, with smaller bars plotted in front of larger ones.
Source: Authors' calculations.

Chart 3 shows that in terms of employment, grants appear to be the most effective treatment, with clear gains in the first year and the strongest impact in the final year, with effects rising from around four percentage points at the outset to more than six points after three years. Loans are also associated with gains in employment, although their impact is more moderate and temporarily weakens during the second year, falling from around two and a half points in the first year to less than four points thereafter.[8] Advisory services bring only marginal improvements, generally less than two points over the entire period.

The sales pattern is similar, but with consistently larger effects associated with grants, which start at nearly five points, decline slightly in 2020 and rebound strongly in the third year. Loans fluctuated more sharply, with an initial increase followed by a temporary decline and then a recovery, while advisory services remained modest throughout the period, never exceeding two points.

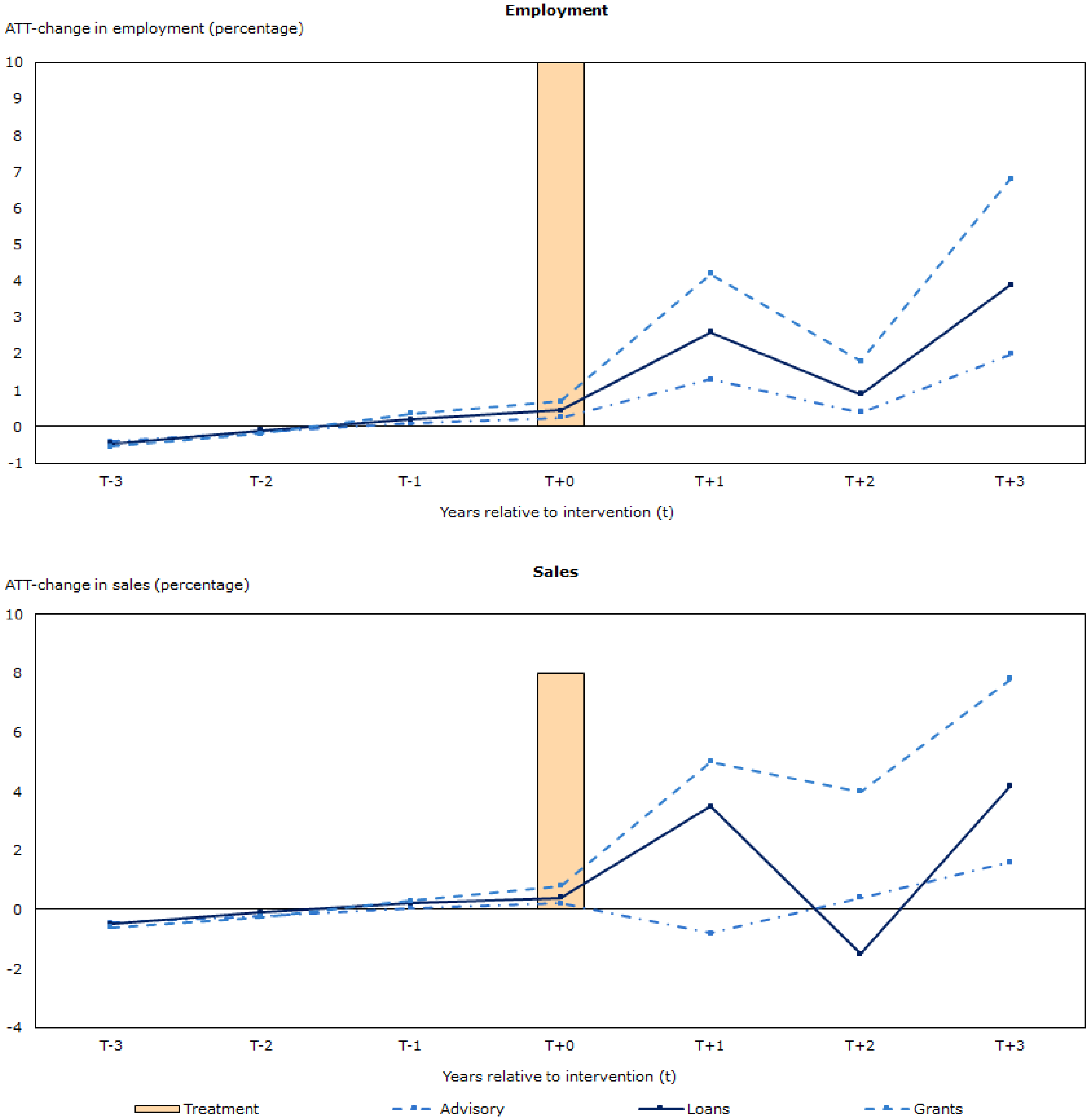
The DiD summarizes the average effect over different periods and is easy to interpret for policy making, but it does not reveal how quickly the effects manifest themselves, whether they accumulate, and how they respond to shocks over time. Chart 4 presents the event study estimates that answer these timing questions while providing a practical diagnosis for the parallel trend hypothesis.

The pretreatment coefficients from t−3 to t+0 remain close to zero with only minor fluctuations, which suggests that treated firms and their matched counterparts followed similar paths before the intervention. After treatment, the trajectories diverge in a way that confirms the trends observed in the DiD results and reveals a richer dynamic. In terms of employment, grants have the most pronounced effect. Employment increased in client firms compared with the control group in the first year, declined in 2020, and rebounded in the third year, from just over four to nearly seven percentage points higher. Employment in firms receiving loans follows a more moderate trajectory, with gains in the first year compared with the control group, a slight slowdown in the second year and a recovery in the third year. Employment in firms receiving consulting services shows more modest but positive changes throughout the period compared with the control group, never exceeding two points.

In terms of sales, the overall trend is similar, although firms receiving loans experienced a more pronounced decline relative to the control group during the pandemic before recovering. The results of this event study complement the DiD estimates by confirming that previous trends are likely to be parallel and illustrating how the program's impacts vary over time relative to the initial treatment period, including the temporary decline during the pandemic and the subsequent rebound, particularly for grants.

8. This dip at t+2 is consistent with a retooling or J-curve hypothesis. If loans are primarily used to finance machinery and equipment, the disruptions from installation and learning curves can temporarily depress sales while labour is retained or reallocated for implementation, followed by a rebound in t+3 as efficiency gains are realized. Testing the J-curve hypothesis would require additional analysis.

Chart 4
Event study measuring the impact of programs on employment changes among clients compared with non-clients

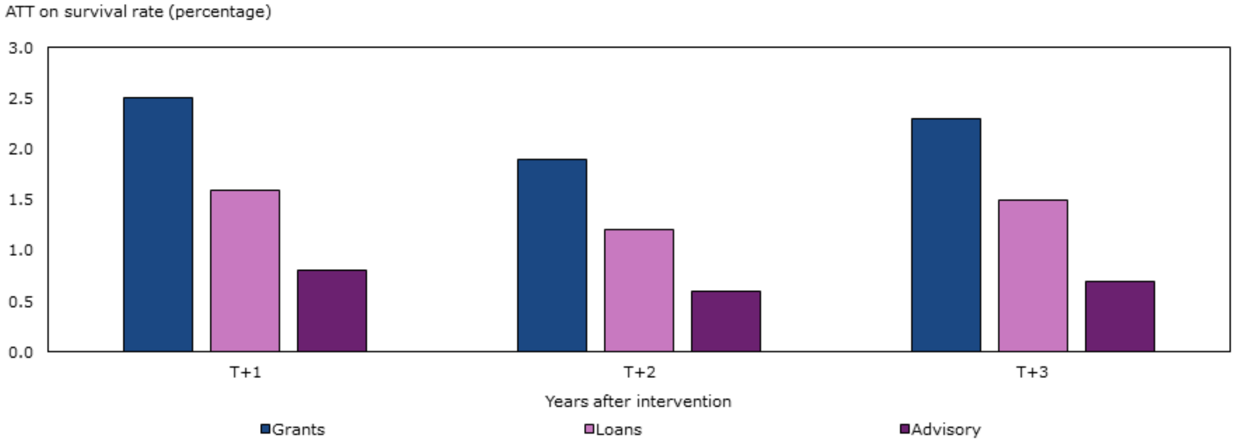


Note: This figure displays dynamic treated (ATT) estimates from t−3 to t+3 for employment and sales. Pre-treatment coefficients diagnose parallel trends, post treatment trajectories show persistence and timing.
Source: Authors' calculations.

Chart 5 presents the ATT for firm survival, which is defined as a firm remaining active and reporting positive employment or income in the data at each period. These ATT estimates reflect the incremental benefit compared with matched non-clients, meaning the values are modest and should not be confused with gross survival rates. Firms receiving grants offer the greatest incremental benefit, improving survival by about two to three percentage points over a three-year period, with a slight decline in 2020. Firms receiving loans offer more modest but still significant gains, generally in the range of one to one and a half percentage points, while firms receiving counselling services provide only marginal improvements over the control group, generally less than one percentage point.

These trends reflect the findings of the DiD study and the event study: the more pronounced effects of grants in terms of performance translate into greater benefits in terms of firm survival, the effects of loans remain positive but are somewhat sensitive to shocks, and the effect of advisory services are more muted.

Chart 5
Survival analysis (one year later) in terms of employment among clients in 2020 compared to non-clients



Note: This figure reports a survival indicator (active if positive employment or revenue) for treated (ATT) clients relative to matched non-clients, by service type, at t+1, t+2 and t+3.
Source: Authors' calculations.

These estimates are quasi-causal and rely on the assumption of parallel trends, which cannot be verified directly. Matching reduces bias related to observable differences but does not eliminate differences unobserved in the data or latent heterogeneity. Attrition and data limitations may also affect the results, particularly for later years. Despite these caveats, the analysis demonstrates the value of linked administrative data and robust econometric methods for evidence-based program evaluation.

6. Conclusion

Organizations are increasingly seeking to evaluate program effectiveness to ensure accountability, optimize resource allocation and improve outcomes. Statistics Canada works with outside organizations (for example BDC, EDC and FCC) to apply impact analysis frameworks to evaluate how their services affect client outcomes.[9]

It is important to note that impact analysis serves as one input to evaluation and that rigorous causal estimation often faces practical constraints, particularly when randomization is not possible and programs are evaluated after implementation. This study shows how a well-designed evaluation framework, combining the DiD method, ESs and SA, can be applied to measure the economic impact of programs. The results illustrate how multiple quantitative methods can complement each other. The DiD method summarizes average effects at key periods, ESs validate hypotheses and reveal dynamic patterns, and SA links performance gains to business continuity. Together, they offer a robust complement to other approaches, such as targeted interviews or case studies that help isolate program impacts from confounding factors. While these models cannot eliminate unobserved heterogeneity, they provide a practical and credible basis for supporting evidence-based evaluation.

9. Ratté, S. (2025). Measuring BDC's impact on clients during the pandemic study. Business Development Bank of Canada.
Ratté, S. (2022). Measuring BDC's impact on clients (2014–2018) (ST-CLIENTIMPACT-E2206). Business Development Bank of Canada.

Authors

Manassé Drabo, Landry Kuate and Michael Willox are with the Economic and Social Analysis and Modelling Division, Analytical Studies and Modelling Branch, at Statistics Canada.

Appendix

A1 Understanding of Statistics Canada data

Primary datasets used for evaluation federal government programs serving the businesses include the following:

- Business Register (BR): Statistics Canada's continuously maintained central repository of baseline information on businesses and institutions operating in Canada.
- Canadian Employer–Employee Dynamic Database (CEEDD): A set of administrative files linked and developed by the Economic Analysis Division. The main files used for quantitative impact assessment include the
  - National Accounts Longitudinal Microdata File (NALMF)
  - Diversity and Skills Database (DSD).

Business Register

- The BR is a repository of information reflecting the Canadian business population and exists primarily to supply frames for all economic surveys in Statistics Canada.
  - Includes corporations, unincorporated businesses (self-employed), sole proprietors, government entities, non-profit organizations, partnerships and financial funds.

- The BR is updated by various sources:
  - files from Canada Revenue Agency such as the T2, T1, GST, PD7 and T4
  - data derived from other Statistics Canada surveys or other sources
  - data collected directly from survey respondents.

- Some variables used for the projects come directly from the BR, such as industry (North American Industry Classification System) and geography.

What data should programs collect to ensure quality linkages to the Business Register?

- All data that can uniquely identify businesses should be collected.
  - Names and addresses alone don't necessarily identify a business uniquely.

- The Business Number is the gold standard for linking a business to the BR.
- Ideally, without a Business Number, several names and addresses including postal codes and provinces or territories (e.g., legal and operating names and addresses) are required to link the business with the proper BR entity.
- Other useful variables to link businesses to the proper BR entity include email addresses, web sites (URLs), phone numbers, and geocodes such as latitude and longitude coordinates.

National Accounts and Longitudinal Microdata Files

The most extensively used database in the CEEDD for impact analysis at Statistics Canada is the National Accounts and Longitudinal Microdata Files (NALMF), which includes corporate income tax returns (T2) containing many of the details found in corporate income statements and balance sheets; payroll data (T4, PD7) related to employment and labour income; and sales information from which GST/HST is calculated. Notably, the NALMF provides information at the enterprise level and does not cover the set of unincorporated enterprises without employees. Therefore, if the agency's clients include sole proprietorships, information from the Financial Declaration File and the T2 Schedule 50 may be needed for their impact studies analyses.

Diversity and Skills Database

Using information from the CEEDD, the DSD was developed to allow researchers to analyze business performance and evaluate programs in a fulsome fashion. The DSD covers the 2005-to-2021 period (at the time of the current article) and can provide demographic information on business owners and employees for the following dimensions:

- sex
- age
- marital, immigrant and disability status
- Indigenous peoples (business owners only)
- racialized groups (business owners only)
- Sex of C-Suite and Board of Directors for publicly traded (2025/26 project)
- information on businesses majority-owned—if more than 50% of its shares belong to individuals who identify as in a specific group.

A2. Additional details on regression models

Since DiD is a foundational method within the impact study's analytical framework, this section expands its implementation and extensions. The ES approach generalizes DiD by allowing for dynamic treatment effects over time, while SA adapts the DiD strategy to contexts where the outcome of interest is the survival rate—treating it as the dependent variable in assessing intervention effects.

DiD is used to estimate the average treatment effect on treated (ATT). The sample firms included matched treatment and comparison groups. If the matching process is successful, the only characteristic that differs between the matched treatment and comparison groups will be the treatment.

A simple standard linear regression DiD estimator (Equation 1) includes binary indicators for treatment (T) and a post-treatment period (P) to capture differences in treatment and control groups and the covariates X that may confound the interpretation of the interaction parameter TP, which represents the average treatment effect on the treated.

$$Y = \beta_0 + \beta_1 T + \beta_2 P + \beta_3 TP + \beta_4 X + \varepsilon \quad (1)$$

Interpreting coefficients: Coefficients representing the treatment effect are interpreted in percentage change in the outcome of interest in treated groups relative to the control groups. However, when variables have been transformed using logarithm forms or an Inverse Hyperbolic Sine (HIS) function, coefficients should be transformed to derive the percentage change using the following formula:[10]

$$\%\Delta = (\exp(\beta) - 1) \times 100 \quad (2)$$

Here, $\%\Delta$ represents the coefficient β as a percentage change where β is the measure of ATT. For example, if the coefficient for the outcome number of employees equals 0.2, the percentage change would be 22.1%, not 22%. For coefficients with small values, the difference between the percentage change and the coefficient is small, but it increases exponentially with the value of the coefficient.

A3. Comparison of methods

Table 1

10. Bellemare and Wichman (2020) show that the method to calculate the percentage change in the outcome of interest is the same regardless of whether the explanatory variables in a standard DiD regression are transformed using a HIS or a log transformation.

Feature	Difference-in-Differences (DiD)	Event Study	Survival Analysis
Core Question it Answers	What was the overall impact of a program by comparing the change in an outcome for the treatment group to the change for the control group?	How did an outcome evolve in the periods just before and after a specific event or policy was introduced?	How did a program affect the timing of an event, such as how long it took for something to happen (e.g., a business to fail or succeed)?
Simple Analogy	You compare the growth in sales at a store that ran a new ad campaign vs. a similar store that didn't. You look at the difference in their sales growth, not just their final sales numbers.	You track a firm's profits for several years *before*, *during*, and *after* they start exporting. The focus is on the pattern of change around the moment exporting began.	You test two program participants and non-participants to see which group has more active firms. The goal is to see if one group is active consistently for more years than the other.
What it Compares	The change in outcome before and after the program for the treatment group vs. the control group.	The outcome in multiple time periods leading up to and following a single event for the group that experienced it.	The time until a specific event occurs for the treatment group vs. the control group.
Key Assumption	Parallel Trends: You must assume that if the program never happened, the two groups would have continued to have similar trends in their outcomes.	No Anticipation: You assume the outcome wasn't already changing in anticipation of the event, and that no other major event happened at the exact same time.	Independent Censoring: You must assume that the reasons you stop observing a subject (e.g., the study ends) are not related to the event of interest itself.
Example Use Case	Did a new government grant increase the average annual revenue of businesses that received it, compared to those that didn't?	How did a new trade policy announcement affect business investment levels in the quarters immediately before and after it took effect?	Did a mentorship program for startups boost their chance of survival?
What the Result Looks Like	Typically, a single number that represents the program's estimated impact (e.g., "The program increased revenue by an average of $10,000").	A graph showing the impact in each period around the event, allowing you to see if the effect was immediate, grew over time, or faded.	A "survival curve" graph showing the percentage of businesses still operating over time. You compare the curve of the treated group to the control group.

References

Angrist, J. D., & Pischke, J.-S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton University Press.

Callaway, B., & Sant'Anna, P. H. C. (2021). *Difference-in-Differences with multiple time periods*. Journal of Econometrics, 225 (2), 200–230.

Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2010). Handbook on impact evaluation: Quantitative methods and practices. World Bank Publications.

Frenette, M., Chan, W., & Handler, T. (2025). Leveraging Statistics Canada data integration opportunities for program evaluation. *Economic and Social Reports, Statistics Canada, Catalogue no. 36-28-0001*.

Hamit-Haggar, M. (2024). Assessing the economic impact of Accelerated Growth Service program participants, 2017 to 2019. *Reports on Special Business Projects, Statistics Canada, Catalogue no. 18-001-X*.

Rosa, J., Nikzad, R., Demers, F., & Poghosyan, T. (2024). Impact assessment of business innovation and growth support programs on firm performance in Canada using the CDM model. *Reports on Special Business Projects, Statistics Canada, Catalogue no. 18-001-X.*

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. Statistical Science, 25(1), 1–21.