Chilean Journal of Statistics
Volume 16, Number 1, June 2025, Pages 1–21
DOI: 10.32372/ChJS.16-01-01

Causal Inference and Econometrics
Research Article

A statistical methodology for assessing the impact of the Financial Inclusion 90 Program under limited compliance

Ernesto San Martín[1,2,3*], María Soledad Graell[4], Carolina Maldonado[5,8], Macarena Borquez[6], and Iván Canales[7]

[1]Millennium Nucleus on Intergenerational Mobility: From Modelling to Policy (MOVI), Santiago, Chile
[2]Faculty of Mathematics, Pontificia Universidad Católica de Chile, Santiago, Chile
[3]LIDAM/CORE, Université Catholique de Louvain, Louvain, Belgium
[4]Gerencia de Sostenibilidad, Caja de Compensación Los Andes, Santiago, Chile
[5]Subgerencia de Clientes, Caja de Compensación Los Andes, Santiago, Chile
[6]Gerencia de Marketing, Caja de Compensación Los Andes, Santiago, Chile
[7]Chapter Advance Analytics, Caja de Compensación Los Andes, Santiago, Chile
[8]Tapp, Caja de Compensación Los Andes, Santiago, Chile

(Received: 18 October 2024 · Accepted in final form: 21 May 2025)

Abstract

Financial literacy and education have become pivotal on the global policy agenda, as reflected in initiatives such as the Organisation for Economic Co-operation and Development / International Network on Financial Education High-Level Principles on National Strategies for Financial Education. To improve financial well-being, Caja de Compensación Los Andes and Tapp implemented the Financial Inclusion I90 Program (I90 Program) in Los Muermos, Chile. However, challenges related to participant consent and compliance within the treatment-control design introduced biases that compromised the study's internal validity. This article presents an alternative methodology for impact evaluation based on partial-identification techniques to address these uncertainties and biases. By modeling a range of plausible counterfactual behaviors, the proposed approach delineates partial-identification regions for key probabilities, thereby quantifying the extent of uncertainty in the findings. Furthermore, it emphasizes that impact evaluation should inform policy decisions —consistent with Neyman's concept of inductive behavior— rather than merely predict outcomes in similar contexts. Ultimately, the method offers a framework on evaluation by incorporating policymakers' beliefs and explicitly acknowledging inherent uncertainties. By quantifying potential risks through partial-identification regions, it enables more informed and flexible policy decisions based on a realistic appraisal of implementation challenges.

Keywords: Causal inference · Impact evaluation · Inductive behavior · Non-compliance · Partial identification.

Mathematics Subject Classification: Primary 62F99 · Secondary 62C12, 62P20

*Corresponding author. Email: esanmar@uc.cl (E. San Martin)

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

1. Introduction

1.1 *Context*

Financial literacy and education have emerged as key priorities on the global policy agenda, as underscored by the High-Level Principles on National Strategies for Financial Education developed by the Organisation for Economic Co-operation and Development/International Network on Financial Education (OECD/INFE) and endorsed by the Leaders of the Group of Twenty (G20) in 2012 [1]. Reliable data highlight the need for financial education, helping to identify the demographic groups most in need and offering insights into where progress is occurring and where further efforts are required. Longitudinal measures facilitate the tracking of progress, allowing policymakers to adapt and enhance strategies accordingly. In addition, cross-country comparisons of key financial literacy indicators provide opportunities to identify best practices, replicate successful policies, and address common challenges through shared solutions.

1.2 *Stakeholders and context of the Financial Inclusion I90 Program*

Aligned with this global initiative, Caja de Compensación Los Andes and Tapp launched the Financial Inclusion I90 Program (I90 Program) as part of their ongoing commitment to improving financial well-being. Caja Los Andes is the largest compensation fund in Chile, with over 72 years of history and more than 4 million members. As a non-profit organization, it forms part of the national social security system and supports the state in delivering statutory benefits, including family allowances and the management of medical leave. Beyond these legal mandates, it also provides a broad range of additional services aimed at improving social well-being, particularly in the areas of health, education, recreation, and financial support. In line with its mission to promote financial inclusion, Caja Los Andes created Tapp, a subsidiary that offers a free and easily accessible prepaid card, specifically designed to serve individuals underserved by the traditional financial system.

The program was implemented in Los Muermos, a commune in the Los Lagos Region of southern Chile, approximately 1,000 kilometers south of Santiago, the capital. Los Muermos has a population of around 17,000 people and a density of about 11 inhabitants per square kilometer. Its municipality, governed by a mayor from the Unión Demócrata Independiente (UDI), played a key role in facilitating the intervention.

Key participating organizations included the Ilustre Municipalidad de Los Muermos, BNP Paribas Cardif, Mastercard, Pontificia Universidad Católica de Chile, Universidad de Los Lagos, Destácame, and Sencillito. BNP Paribas Cardif, the insurance arm of the French banking group BNP Paribas, specializes in protection and savings products. Mastercard is a global leader in payment technologies, facilitating electronic transactions worldwide. Destácame is a Chilean fintech offering financial education and access to credit through personalized credit reports. Sencillito is a Chilean platform enabling bill payments online or in physical locations. The academic partners were the Pontificia Universidad Católica de Chile —ranked 93rd globally and 2nd in Latin America according to QS 2025— and Universidad de Los Lagos, a public university ranked in the 201-250 range in Latin America.

To enable causal comparison, Maullín was selected as the control commune. Maullín is geographically close to Los Muermos and shares a similar rural profile and political leadership (its mayor is an independent supported by the UDI). These similarities, along with the need to avoid indirect exposure to the intervention in Los Muermos (such as through public marketing), justified the choice of an external but comparable commune as control. The Universidad de Los Lagos was responsible for delivering the educational content, while the Pontificia Universidad Católica de Chile led the program's impact evaluation.

1.3 *Educational activities and evaluation instruments*

Several instruments were used to monitor and evaluate this pilot program, including the measurement of the financial well-being index established by the OECD [2] through a financial literacy questionnaire (Appendix A, levels of participation in various activities managed by Caja Los Andes and Tapp, participant satisfaction and evaluation surveys, as well as the number of account openings. The activities implemented were the following:

- In-person workshops —Topics included family budgeting, accounting, sales, responsible debt management, fraud awareness, savings, money management, and financial products.
- Talks at educational institutions —Tailored presentations aimed at increasing financial awareness and knowledge among students.
- Educational meals —Informative sessions combined with meals to create a supportive and engaging learning environment.
- Educational content dissemination —Distribution of educational materials via WhatsApp, email, and phone to ensure continuous engagement.
- Broadcasting educational capsules on local television —Delivery of concise, targeted financial education content to reach a broader audience.
- Broadcasting educational messages on local radio stations —Use of radio as an accessible medium to disseminate financial education within the community.

It is important to note that the financial literacy component of the questionnaire aligns with the OECD/INFE definition of financial literacy, which is described as: “A combination of awareness, knowledge, skills, attitudes, and behaviors necessary to make sound financial decisions and ultimately achieve individual financial well-being” [1].

The questions themselves are primarily drawn from existing surveys, all of which were validated and endorsed by OECD/INFE experts. These questions represent best practices in the measurement of financial literacy and inclusion. Since its first implementation in 2010 as part of the OECD’s inaugural international survey on financial literacy and inclusion, the questionnaire has been used to assess financial literacy levels across diverse populations. In 2015-2016, approximately 40 countries and economies participated in an international survey on adult financial literacy skills, using data collected through this toolkit. The results were published for an initial set of countries in [1].

The selection of the Los Muermos commune was based on its alignment with key program criteria: a rural setting, a population of fewer than twenty thousand inhabitants, and relative isolation from urban centers. The population was segmented into groups based on their roles within the community: 20% women, 55% small and medium-sized enterprises (SMEs) and dependent employees, and 25% elderly individuals.

The program aimed to empower individuals and promote behavioral change while providing digital tools to support informed and responsible financial decision-making, ultimately improving their financial well-being. It followed the OECD methodological framework, which focuses on three components: knowledge, attitude, and behavior. The objective was to positively impact the financial-health index of key segments within the Los Muermos commune, with an expected reach of approximately 60% of the population.

The following interventions were implemented within this methodological framework:

- Knowledge —Local mass communication through radio, television, outdoor advertising, interviews, public activations, and large-scale events.
- Attitude —Educational initiatives delivered via both local and digital media, as well as in-person formats, including educational capsules, WhatsApp groups, workshops, community activations, and ambassador programs.
- behavior —Promotion of access to and usage of financial products such as loans, savings accounts, insurance, and prepaid cards.

The results of this pilot program provide insights into the extent to which the financial-health index of individuals in a locality can be improved. They also help to identify the most effective strategies for potential implementation in other areas.

1.4 *Quasi-experimental design*

The I90 Program employed a two-cluster quasi-experimental design. The Los Muermos commune was designated as the treated cluster, while the Maullín commune served as the control cluster. The intervention was assigned at the cluster (commune) level, whereas financial-health index outcomes were measured at the individual level. Because only two clusters exist, internal validity relies on the exchangeability of Los Muermos and Maullín prior to the intervention.

In the original design, participants from both communes were to be selected randomly: surveyors would visit randomly chosen households and administer the questionnaire to at least one resident. This approach aimed to establish a baseline measurement of the financial-health index prior to the program implementation in Los Muermos.

As previously mentioned, a key component of the program involved providing personalized training sessions to selected participants in Los Muermos. Upon completion of these sessions, both treated participants and those in the control group were to be reassessed to determine their financial-health index score.

This design not only enabled a comparison of changes in the financial-health index between the two communes, but also aimed to quantify the magnitude of those changes. While such a design is generally considered stronger than purely observational studies, its internal validity rests on the assumption that the two communes are exchangeable; for further discussion, see [3].

However, two issues arose that ultimately undermined the presumed superiority of the chosen design: not all contacted individuals consented to participate, and among those who did consent, only a small proportion fully complied with the intervention. Table 1 summarizes the raw absolute figures, from which we observe the following:

- In Los Muermos:
  - 58% of those contacted agreed to participate in the baseline measurement;
  - of these, 41% remained for the endline measurement;
  - among this 41%, only 8% attended the training sessions, while 92% did not comply with the intervention.
- In Maullín:
  - 42% of those contacted agreed to participate in the baseline measurement;
  - of these, 39% remained for the endline measurement.

Table 1: Participant flow by location: contacted, consented, endline, and trained.

n	Los Muermos	Maullín
$n_{\text{contacted}}$	862	738
$n_{\text{consented}}$	500	310
n_{endline}	204	120
n_{trained}	17	-

1.5 *Limitations of the experimental design*

Following [4, 5], we paraphrase the definition of a randomized experiment as follows:

> Let random samples of individuals be drawn from the population of interest and assigned to treatment groups. Let all members of a treatment group receive the same treatment, and suppose that each subject complies with the assigned treatment. Then the distribution of outcomes experienced by the members of a treatment group will be the same (up to random sampling error) as would be observed if the treatment in question were administered to the entire population.

Let Y_t denote the potential outcome for an individual under treatment $t \in \{0,1\}$ ($t=1$ treated, $t=0$ control). Let $\boldsymbol{x} \in \mathbb{R}^p$ be a vector of covariates that characterise the individual, and let $z \in \{0,1\}$ record the realised treatment assignment.

With two randomly assigned clusters, we adopt the conditional independence $(Y_0, Y_1) \perp z \mid \boldsymbol{x}$ and invoke the "stable unit treatment value assumption"; hence the observed outcome equals Y_t whenever $z=t$ and there is no interference between individuals. Although strong with only two clusters, this assumption provides a theoretical benchmark for the discussion that follows. These conditions imply the expression given by

$$P\big(Y_t \mid \boldsymbol{x}, z=1\big) = P\big(Y_t \mid \boldsymbol{x}, z=0\big) = P\big(Y_t \mid \boldsymbol{x}\big), \quad t \in \{0,1\}. \tag{1}$$

In the absence of non-consent, the data would identify $P(Y_1 \mid \boldsymbol{x})$ and $P(Y_0 \mid \boldsymbol{x})$.

If Y_t is binary, with $Y_t = 1$ indicating a favourable outcome, treatment 1 is socially preferable whenever

$$P\big(Y_1 = 1 \mid \boldsymbol{x}\big) \; > \; P\big(Y_0 = 1 \mid \boldsymbol{x}\big). \tag{2}$$

Because only two clusters were randomised, the events $\{Y_1 = 1\}$ and $\{Y_0 = 1\}$ refer to the sets $\{m \in M\colon Y_1(m) = 1\}$ and $\{m \in M\colon Y_0(m) = 1\}$, respectively. Therefore, the inequality in (2) reflects a group-level treatment comparison, which characterizes a two-cluster quasi-experimental design.

This formulation is intended to guide policy: in an ideal setting, one would offer treatment 1 as broadly as feasible to replicate the experimental conditions. In democratic contexts, however, individuals may decline participation, and fiscal constraints may limit coverage. As a result, full compliance is rarely attainable [6, 7], and partial uptake weakens the external and internal validity of the experimental findings.

In the present study, as detailed in Subsection 1.4, individuals in both Los Muermos and Maullín had to provide informed consent before participating. Let $c \in \{0,1\}$ indicate consent ($c=1$ for consent, $c=0$ otherwise). Because outcomes are unobserved for non-consenters, neither $P(Y_1 \mid \boldsymbol{x})$ nor $P(Y_0 \mid \boldsymbol{x})$ is point-identified, and the inequality expressed in (2) cannot be directly tested. By the law of total probability, we have that

$$P(Y_t \mid \boldsymbol{x}) = P(Y_t \mid \boldsymbol{x}, c=1)P(c=1 \mid \boldsymbol{x}) + P(Y_t \mid \boldsymbol{x}, c=0)P(c=0 \mid \boldsymbol{x}), \quad t \in \{0,1\}.$$

The data reveal $P(c=i \mid \boldsymbol{x})$ for $i=0,1$ and $P(Y_t \mid \boldsymbol{x}, c=1)$, but not $P(Y_t \mid \boldsymbol{x}, c=0)$; the latter term embodies the fundamental missing-data problem induced by non-consent.

1.6 *The ignorability condition in observational studies*

The limitations discussed above motivate a shift in strategy for policy evaluation. One of the most widely adopted strategies relies on the so-called ignorability condition. While this strategy is based on a probabilistic assumption similar to the one stated in (1), its practical implementation differs: instead of selecting a population with shared characteristics x and then randomly assigning treatment $t=1$ or control $t=0$ —thereby constructing a two-cluster quasi-experimental design— an observational study compares two already formed subpopulations: those under treatment and those under control. The goal is to identify a set of covariates $\widetilde{\boldsymbol{x}}$ such that

$$Y_t \perp z \mid \widetilde{\boldsymbol{x}}, \quad t \in \{0,1\} \tag{3}$$

can reasonably be assumed.

Formally, for any random variables V, W, S, the condition $V \perp W \mid S$ means that, for all bounded measurable functions f, we have that $\mathrm{E}(f(V) \mid W, S) = \mathrm{E}(f(V) \mid S)$. A rigorous treatment of conditional independence and its role in statistical inference can be found in [8]. Additional references on the ignorability condition include [9, 10, 11], among others.

Given the formal similarity between the expressions presented in (1) and (3), ignorability can be viewed as turning an observational study into a local two-cluster quasi-experiment, conditional on $\widetilde{\boldsymbol{x}}$. Thus, the analyst's task is to select covariates for which the conditional-independence assumption is credible. Achieving balance in $P(z = 1 \mid \widetilde{\boldsymbol{x}})$ is a by-product rather than the primary goal.

In what sense might the ignorability assumption be inadequate for evaluating a public policy? A first reaction is to ask whether the condition is plausible in the application at hand. While legitimate, this perspective is insufficient because it overlooks the ultimate purpose of policy evaluation. Our aim is not merely to predict what would happen if treatment $t = 1$ were implemented; rather, it is to influence the policymaker's will so that actions bring about the desired outcomes. This view is grounded in Neyman's concept of inductive behavior [12] to be discussed later.

Assumptions should be read as guidance for action –premises on which the policymaker may base interventions. Then, what course of action is implied by ignorability? Because the condition casts the observational study as a local two-cluster quasi-experiment, it effectively prescribes identifying the subpopulation characterised by $\widetilde{\boldsymbol{x}}$ and, in principle, administering treatment $t = 1$ there. While feasible, this is only one option. Partial-identification strategies, by combining observations with assumptions, can point to alternative actions. Developing such alternatives is precisely one of the objectives of the present work.

1.7 *Objective of the article*

In this article, we propose an alternative methodology for evaluating the potential impact of the I90 Program. First, we model the evaluation problem using partial identification techniques, as introduced by Manski; see, for example, [5] and the references therein. This approach allows us to account for uncertainties stemming from different sources of bias —such as informed consent and non-compliance — and to incorporate various counterfactual behaviors relevant to the context.

Second, we argue that evaluating the impact of a program differs fundamentally from predicting what would happen in similar contexts or among new participants once the program is implemented. Instead, our focus is on the implementer of the program: the evaluation results derived under different counterfactual assumptions are meant to inform and influence the implementer's decisions —guiding not only whether to implement the program, but also how to act in light of its outcomes. This perspective is aligned with the concept of inductive behavior, as introduced in [12].

The rest of this article is organized as follows. Section 2 presents the results observed during the program's implementation. In Section 3, we introduce the concepts of partial identification and inductive behavior. In Section 4, the evaluation of the I90 Program under the proposed framework is developed. Section 5 provides a discussion of our results, whereas Section 6 states our conclusions.

2. The observed impact of the Financial Inclusion I90 Program

2.1 *Notation and basic facts*

We begin by establishing the notation used to describe both the design structure and attrition issues, as well as to analyse the observed impact of the program on the financial-health index.

Let M denote the population of interest, that is, the set of all contacted individuals. For each $m \in M$, we record the following:

- $c(m) = 1$ if the individual consented to participate (and was therefore measured at baseline); $c(m) = 0$ otherwise.
- $z(m) = 1$ if the individual belongs to the treated cluster (Los Muermos); $z(m) = 0$ if the individual belongs to the control cluster (Maullín).
- $\zeta(m) = 1$ if the individual attended all financial training sessions; $\zeta(m) = 0$ otherwise.
- $w(m) = 1$ if the individual was measured at endline; $w(m) = 0$ otherwise.
- $y_1(m)$ is the financial-health index measured at baseline.
- $y_2(m)$ is the financial-health index measured at endline.

In relation to empirical probabilities, the raw counts in Table 1 imply that $P(c = 1) = 0.58$ (Los Muermos), $P(c = 1) = 0.42$ (Maullín), $P(w = 1 \mid c = 1, z = 1) = 0.41$, and $P(\zeta = 1 \mid c = 1, z = 1, w = 1) = 0.08$.

About sub-populations, using these probabilities, we define three mutually exclusive groups analyzed throughout the article as follows:

- Population 1 (Maullín, measured at baseline & endline) stated as

$$\{m \in M \colon c(m) = 1, z(m) = 0, w(m) = 1\}.$$

- Population 2 (Los Muermos, city-level intervention only) presented as

$$\{m \in M \colon c(m) = 1, z(m) = 1, w(m) = 1, \zeta(m) = 0\}.$$

- Population 3 (Los Muermos, city-level intervention and training) established as

$$\{m \in M \colon c(m) = 1, z(m) = 1, w(m) = 1, \zeta(m) = 1\}.$$

In Table 2, missing outcomes (y_1 or y_2) are indicated by "–". For instance, the endline score y_2 is observed in Los Muermos only when $c = 1,\ z = 1,\ w = 1$, irrespective of training attendance.

Table 2: Participant selection and attrition.

	Consent to participate	Baseline score	Treatment assignment	Compliance with training	Compliance with endline	Endline score
Los Muermos (intervention)	$c = 1$	y_1	$z = 1$	$\zeta = 1$	$w = 1$	y_2
				$\zeta = 0$	$w = 1$	y_2
				$\zeta = 0$	$w = 0$	–
	$c = 0$	–	–	–	–	–
Maullín (control)	$c = 1$	y_1	$z = 0$	–	$w = 1$	y_2
	$c = 1$	y_1	$z = 0$	–	$w = 0$	–
	$c = 0$	–	–	–	–	–

2.2 *Observed results*

To gauge the I90 Program's potential impact on the financial-health index, we contrast the baseline and end-line empirical cumulative distribution functions. As a descriptive measure of separation we report the Kolmogorov–Smirnov (KS) statistic, that is the supremum distance between the two empirical cumulative distribution functions. Because the same individuals are observed at both waves, the samples are paired rather than independent, and the index is discrete with many ties; the usual reference distribution of the two-sample KS statistic is therefore only approximate and tends to be conservative.

Accordingly, we treat the resulting p-values merely as gauges of dissimilarity, not as formal tests of significance. (With only two clusters, any cluster-level correlation would further widen the true sampling distribution.) Under this framing, the null hypothesis is that the baseline and end-line distributions coincide.

Figure 1 shows the empirical curves. Using the KS statistic, we find the following:

- For Population 1 —Assuming the null hypothesis (equality of baseline and endline distributions) is true, the p-value is 0.428. That is, such similarity between distributions is quite likely under the null.
- For Population 2 —The p-value is 0.2352, indicating that such similarity is still reasonably expected under the null hypothesis.
- For Population 3 —The p-value is 0.074. In this case, the observed similarity is less likely under the null, suggesting potential effects of the intervention.

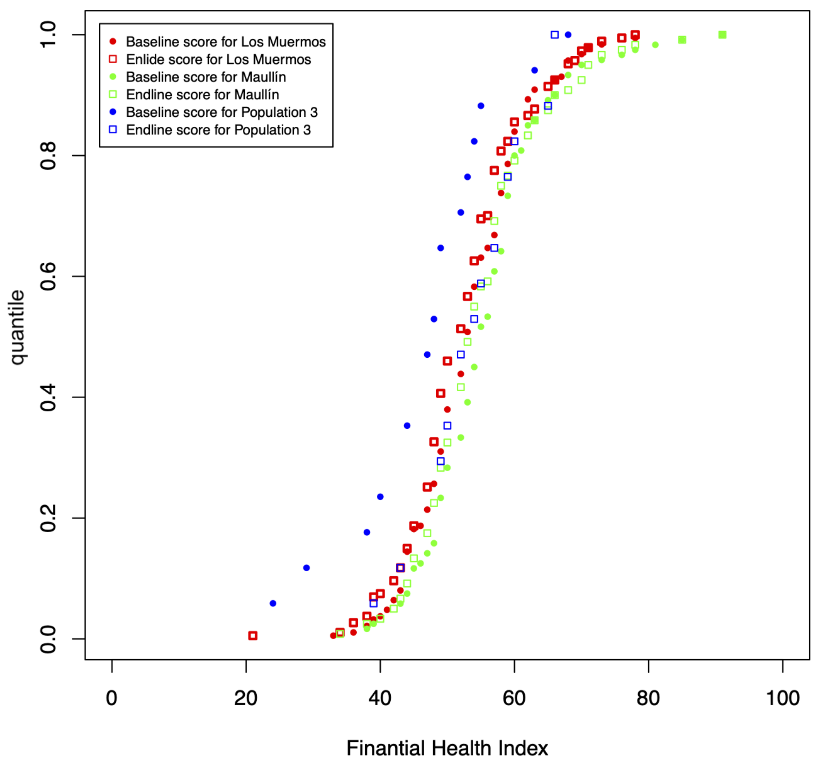

Figure 1: Empirical distribution functions of the financial-health index at baseline and endline.

The third result, combined with the fact that the endline distribution for Population 3 is visibly shifted to the right relative to the baseline, is consistent with a possible improvement in the perceived financial-health index among those who received both the city-wide intervention and the training sessions. It is also worth noting that the baseline distribution for Population 3 was initially worse than those of Populations 1 and 2. However, following the implementation of the I90 Program, Population 3 not only appears to improve but also reaches the level of the other groups, whereas Populations 1 and 2 display no noticeable change. Because the KS test does not account for non-response, self-selection into training, or the psychometric limitations of the index, this pattern should be viewed as suggestive rather than conclusive evidence of causal improvement.

A brief comment is warranted on the interpretation of hypothesis testing in this context. A convention is to fix a threshold —typically 0.05— such that if the p-value exceeds this threshold, the null hypothesis is not rejected. This convention, associated with Neyman, requires the threshold to be defined before observing the data, as part of a pre-specified decision rule. Another approach, associated with the later Fisher, treats the p-value as a property of the data, not a decision boundary, and uses it as a continuous measure of evidence against the null. For detailed discussions on these perspectives and the conceptual confusion surrounding them [13, 14, 15]. In our case, since the goal is causal inference rather than formal hypothesis testing, we prefer to report the p-value as a descriptive summary of how similar or different the empirical distributions actually are, under the assumption that the null hypothesis is true.

2.3 *Mobility of the financial health index*

The results above suggest that the increase in the financial-health index is more pronounced in Population 3 compared to Populations 1 and 2, where the change appears negligible. However, interpreting the difference $y_2 - y_1$ as a gain may be misleading, because it only captures the absolute difference between the two values, without accounting for their positions on the scale. Indeed, the quantity $y_2 - y_1$ satisfies identities such as

$$y_2 - y_1 = (y_2 - c) - (y_1 - c) = (y_2 + c) - (y_1 + c), \quad \text{for any } c > 0,$$

which shows that the same numerical gain can occur at very different points on the scale. For instance, moving from 2 to 4 and from 6 to 8 both represent a two-point increase, yet the substantive implications may diverge sharply.

The raw difference therefore captures magnitude but ignores starting and ending positions. A two-point rise from a low baseline can signal a much deeper transformation than the same rise from an already favourable position. Because the underlying index is essentially ordinal –constructed from summed Likert items– treating it as if it possessed true interval properties would be misleading [16].

To obtain more interpretable evidence, we analyse the conditional distribution of ranks at endline given baseline ranks. This rank-mobility approach, inspired by intergenerational-mobility studies [17], highlights upward or downward movement relative to the whole distribution rather than relying on raw score differences.

Operationally, we define four rank categories based on quartiles stated as

$$\begin{array}{l|l}
r_1 = 1 \text{ if } y_1 \leq q_1(0.25); & r_2 = 1 \text{ if } y_2 \leq q_2(0.25); \\
r_1 = 2 \text{ if } q_1(0.25) < y_1 \leq q_1(0.5); & r_2 = 2 \text{ if } q_2(0.25) < y_2 \leq q_2(0.5); \\
r_1 = 3 \text{ if } q_1(0.5) < y_1 \leq q_1(0.75); & r_2 = 3 \text{ if } q_2(0.5) < y_2 \leq q_2(0.75); \\
r_1 = 4 \text{ if } y_1 > q_1(0.75); & r_2 = 4 \text{ if } y_2 > q_2(0.75),
\end{array}$$

where q_1 and q_2 denote the quantiles of the baseline and endline distributions, respectively.

To assess the effect of treatment intensity, we directly compare Populations 2 and 3. We consider the effect to be positive (respectively, negative) if the probability of upward mobility —that is, moving from a lower rank $r_1 = l$ to a higher rank $r_2 = k$, with $k > l$— is greater (respectively, smaller) among those who received both the city intervention and financial training, compared to those who received only the city intervention. That is, we interpret the treatment as having a positive effect if

$$P(r_2 = k \mid r_1 = l, c = 1, z = 1, \zeta = 1, w = 1) > P(r_2 = k \mid r_1 = l, c = 1, z = 1, \zeta = 0, w = 1),$$

and as negative if

$$P(r_2 = k \mid r_1 = l, c = 1, z = 1, \zeta = 1, w = 1) < P(r_2 = k \mid r_1 = l, c = 1, z = 1, \zeta = 0, w = 1)$$

or as null if the two are equal.

Tables 3 and 4 show the empirical mobility matrices for Populations 2 and 3, respectively. Focusing on the upper triangle of each table —that is, the cells where $r_2 > r_1$— we observe that upward mobility is more pronounced in Population 3. This suggests that the combined treatment (city intervention and financial training) had a stronger effect than the city intervention alone.

Table 3: Mobility of financial-health index for Population 2 (city intervention only).

r_1	$P(r_2 = 1 \mid r_1 = l)$	$P(r_2 = 2 \mid r_1 = l)$	$P(r_2 = 3 \mid r_1 = l)$	$P(r_2 = 4 \mid r_1 = l)$
$l = 1$	0.489	0.298	0.191	0.021
$l = 2$	0.327	0.286	0.327	0.061
$l = 3$	0.163	0.286	0.306	0.245
$l = 4$	0.024	0.119	0.286	0.571

Table 4: Mobility of financial-health index for Population 3 (city intervention + training).

r_1	$P(r_2 = 1 \mid r_1 = l)$	$P(r_2 = 2 \mid r_1 = l)$	$P(r_2 = 3 \mid r_1 = l)$	$P(r_2 = 4 \mid r_1 = l)$
$l = 1$	0.000	0.167	0.500	0.333
$l = 2$	0.000	0.000	0.333	0.667
$l = 3$	0.000	0.250	0.250	0.500
$l = 4$	0.000	0.000	0.250	0.750

3. Partial identification and inductive behavior

3.1 *Context*

We now introduce the components of the proposed methodology for program evaluation. The original design of the I90 Program followed a treated-versus-control group framework with random selection, which theoretically allows the extrapolation of results to other communes similar to Los Muermos. However, as discussed in Subsection 1.5, the presence of non-consent to participate induces an identification problem that renders it impossible to apply the traditional framework of a randomized experiment to evaluate the program's impact. Furthermore, among participants in Los Muermos who remained in the study through endline, only 8% complied with the financial training sessions.

These limitations raise the following question:

> "How can we assess the impact of the Financial Inclusion I90 Program when the compliance rate is low, recognizing that an impact evaluation is distinct from simply predicting the program's eventual effects?"

In the sections that follow, we propose a methodology that explicitly incorporates the uncertainties associated with this assessment, with the goal of informing policymaker decisions by providing a structured basis for selecting among different possible courses of action. Our approach combines the perspective of partial identification [5, 18] with the statistical concept of inductive behavior, as developed in [12, 19, 20].

3.2 *Partial identification*

To clarify the notion of partial identification, we begin with its rationale. Our interest lies in the distribution of the financial-health index at endline among participants in Los Muermos who consented to participate, denoted $P(y_2 \mid c = 1, z = 1)$. However, not all of these individuals complied with the endline measurement, and thus this distribution cannot be directly observed. The only observable distribution is that of participants who both consented and were measured at endline, given by $P(y_2 \mid c = 1, z = 1, w = 1)$.

In econometric terms, this observable distribution is said to be identified. The question is: how does the distribution of interest $P(y_2 \mid c = 1, z = 1)$ relate to the identified distribution? By the law of total probability, we have

$$\begin{aligned} P(y_2 \mid c = 1, z = 1) = {} & P(y_2 \mid c = 1, z = 1, w = 1)\, P(w = 1 \mid c = 1, z = 1) \\ & + P(y_2 \mid c = 1, z = 1, w = 0)\, P(w = 0 \mid c = 1, z = 1). \end{aligned}$$

In this decomposition, the conditional probability $P(y_2 \mid c = 1, z = 1, w = 0)$ —that is, the distribution among individuals who consented but were not measured at endline— is completely unknown. As a result, the distribution of interest $P(y_2 \mid c = 1, z = 1)$ is unidentified in the standard sense. Nevertheless, it can be partially identified. Since probabilities lie between 0 and 1, the unknown component can be bounded as follows:

- If $P(y_2 \mid c = 1, z = 1, w = 0) = 0$, then $P(y_2 \mid c = 1, z = 1) \geq P(y_2 \mid c = 1, z = 1, w = 1)P(w = 1 \mid c = 1, z = 1)$, which defines the lower bound of the distribution.
- If $P(y_2 \mid c = 1, z = 1, w = 0) = 1$, then $P(y_2 \mid c = 1, z = 1) \leq P(y_2 \mid c = 1, z = 1, w = 1)P(w = 1 \mid c = 1, z = 1) + P(w = 0 \mid c = 1, z = 1)$, which defines the upper bound of the distribution.

Thus, we identify a region within which all possible distributions $P(y_2 \mid c = 1, z = 1)$ that are compatible with the observed data must lie. This region is called the partial identification region. Figure 2 illustrates this concept. The region also enables us to quantify the uncertainty induced by participants who did not comply with the endline measurement, which directly affects the identifiability of the target distribution $P(y_2 \mid c = 1, z = 1)$.

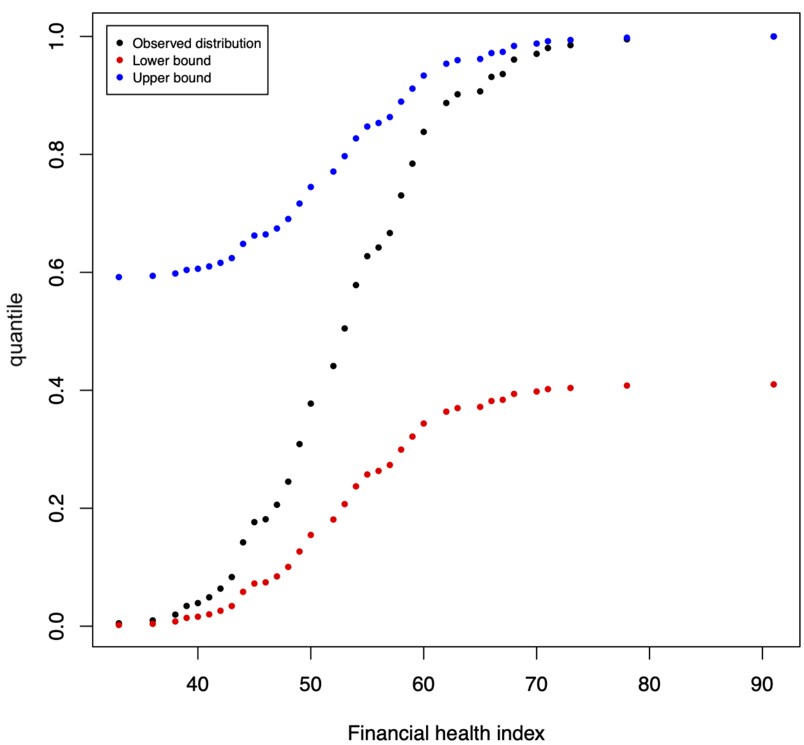

Figure 2: Partial identification region of the endline distribution among participants in Los Muermos who consented to participate.

In our case, this uncertainty is reflected in the width of the interval between the lower and upper bounds, which equals $P(w = 0 \mid c = 1, z = 1)$, the proportion of participants not measured at endline. According to the data, this proportion is 59%. The shaded region in Figure 2 illustrates the ambiguity introduced by non-response, accommodating multiple plausible shapes for the unobserved component of the distribution.

In general terms, "Partial identification means that the available evidence and maintained assumptions imply a bound on treatment response but not a precise value. (Formally, a quantity is partially identified if the evidence and assumptions imply that the quantity lies in some informative set of values but do not enable one to determine its precise value. The set of feasible values is called the identification region)" [21, p. 2065]. In the context of program evaluation, the strength of partial identification analysis lies in the fact that "wide bounds add transparency to [...] causal inferences, because they make clear how much information is 'present' in the data relative to how much information is supplemented by causal identification or parametric assumptions)" [22, p. 5].

The partial identification perspective can be traced back to seminal contributions by [23, 24]. More recently, this framework has gained traction in econometrics, particularly for addressing challenges such as missing data and measurement error [25, 26, 27].

Another empirical challenge addressed through this approach involves the identification of a joint distribution when only the marginal distributions are known —a problem considered in [28, 29, 30]. The field of political science has also adopted the partial identification approach in causal inference, as shown in [31, 32, 33]. In epidemiology, the methodology has been applied to a range of problems involving causal inference as well [22, 34].

3.3 *Inductive behavior*

In [20], Neyman described the final phase of scientific inquiry in three steps: (i) the visualization of several possible sets of hypotheses relevant to the phenomena under study, (ii) deductions from these sets of hypotheses, and (iii) an act of will, or a decision to adopt a particular course of action — possibly guided by a deliberate stance toward the various sets of hypotheses identified in step (i) [20, p. 10]. According to Neyman, this final step is not a matter of logical reasoning but rather an expression of volition. The first two steps allow for the derivation of universally valid formulas (obtained via deductive logic), which serve as normative regulators of belief [20, p. 15].

This framework constitutes the essence of the concept of inductive behavior, which reflects the "recognition that the purpose of every piece of serious research is to provide grounds for the selection of one of several contemplated courses of action" [20, p. 16]. In other words, scientific findings influence the behavior of a client —or, in our case, a policymaker— prompting them to act as though the world functions according to the scientific explanation provided.

As Neyman himself stated in [19]: "with many phenomena, certain permanencies appear quite stable. This creates the habit of regulating our actions regarding some observed events by referring to the permanencies which, at the particular moment, seem to be established. This is what we call inductive behavior" [19, p. 1].

4. Evaluation of the Financial Inclusion I90 Program

4.1 *Context*

Every impact evaluation faces the inherent challenge known as the "fundamental problem of causal inference" [35]. This arises from the impossibility of observing the same statistical unit under both treatment and control conditions.

More precisely, define $r_2(1)$ as the rank in the endline distribution of the financial-health index that an individual would attain if exposed to both the city-wide intervention and the financial training sessions. Similarly, let $r_2(0)$ denote the rank that the same individual would attain at endline if exposed only to the city intervention.

Focusing on participants in Los Muermos who consented to participate and were measured at endline (that is, the set $\{m \in M\colon c(m) = 1, z(m) = 1, w(m) = 1\}$), the evaluation aims to compare $P(r_2(1) = k \mid r_1 = l)$ and $P(r_2(0) = k \mid r_1 = l)$, for $k > l$. For brevity, we omit the conditioning on $c = 1, z = 1, w = 1$ in what follows. These conditional probabilities are decomposed using the law of total probability as

$$\begin{aligned} P(r_2(1) = k \mid r_1 = l) = {} & P(r_2(1) = k \mid r_1 = l, \zeta = 1)P(\zeta = 1 \mid r_1 = l) \\ & + P(r_2(1) = k \mid r_1 = l, \zeta = 0)P(\zeta = 0 \mid r_1 = l), \qquad (4) \\ P(r_2(0) = k \mid r_1 = l) = {} & P(r_2(0) = k \mid r_1 = l, \zeta = 1)P(\zeta = 1 \mid r_1 = l) \\ & + P(r_2(0) = k \mid r_1 = l, \zeta = 0)P(\zeta = 0 \mid r_1 = l), \qquad (5) \end{aligned}$$

where:

- $P(r_2(1) = k \mid r_1 = l, \zeta = 1)$ is the probability of endline rank k if both interventions are received, among those who actually received both and had baseline rank l;
- $P(r_2(1) = k \mid r_1 = l, \zeta = 0)$ is the same as above, but for those who received only the city intervention;
- $P(r_2(0) = k \mid r_1 = l, \zeta = 1)$ is the probability of endline rank k if only the city intervention were received, among those who actually received both;
- $P(r_2(0) = k \mid r_1 = l, \zeta = 0)$ is the same as above, among those who actually received only the city intervention; and
- $P(\zeta = 1 \mid r_1 = l)$ is the proportion of participants who attended the training sessions, among those with baseline rank l.

Among these, the terms $P(r_2(1) = k \mid r_1 = l, \zeta = 1)$, $P(r_2(0) = k \mid r_1 = l, \zeta = 0)$, and $P(\zeta = 1 \mid r_1 = l)$ are identified from the data, since they refer to observed subgroups. In contrast, the cross-terms —$P(r_2(1) = k \mid r_1 = l, \zeta = 0)$ and $P(r_2(0) = k \mid r_1 = l, \zeta = 1)$— remain unidentified due to the Fundamental Problem of Causal Inference: they refer to counterfactual scenarios not observed in the data. Therefore, to evaluate the program, one must introduce additional assumptions in order to partially identify the target probabilities $P(r_2(1) = k \mid r_1 = l)$ and $P(r_2(0) = k \mid r_1 = l)$ for $k > l$. In the following, we explore three plausible scenarios, each yielding partial identification regions for these probabilities. These regions, in turn, may guide policymakers in their decision-making process by explicitly quantifying the uncertainty introduced by the fundamental limits of causal inference.

4.2 *Scenario 1: Rational individual choice*

Scenario 1 can be described as follows: the implementer of the I90 Program assumes that residents of Los Muermos make rational decisions regarding participation in the financial training sessions. In particular, those who choose to attend the sessions believe that their financial-health index would deteriorate if they did not, whereas those who choose not to attend believe they would benefit even without the training.

In essence, this scenario assumes that individuals self-select into treatment based on their own expectations, aiming for what they perceive to be the best possible outcome. Probabilistically, for all $k > l$, this assumption is expressed as

$$\begin{aligned} P(r_2(1) = k \mid r_1 = l, \zeta = 0) &\leq P(r_2(0) = k \mid r_1 = l, \zeta = 0), \\ P(r_2(1) = k \mid r_1 = l, \zeta = 1) &\geq P(r_2(0) = k \mid r_1 = l, \zeta = 1). \qquad (6) \end{aligned}$$

The first inequality states that, for those who did not attend the training sessions ($\zeta = 0$), their endline outcome would not have improved if they had attended. The second states that, for those who did attend ($\zeta = 1$), their outcome would have been worse had they not attended.

Combining these aspects with the decompositions stated in (4) and (5), we obtain the partial identification regions, for $k > l$, given by

$$P(r_2 = k \mid r_1 = l, \zeta = 1)P(\zeta = 1 \mid r_1 = l) \leq P(r_2(1) = k \mid r_1 = l) \leq P(r_2 = k \mid r_1 = l);$$
$$(7)$$
$$P(r_2 = k \mid r_1 = l, \zeta = 0)P(\zeta = 0 \mid r_1 = l) \leq P(r_2(0) = k \mid r_1 = l) \leq P(r_2 = k \mid r_1 = l).$$

For formal details of this result, see Appendix B. Note that $P(r_2 = k \mid r_1 = l)$ represents the overall empirical probability of a transition from baseline rank l to endline rank k, regardless of training participation.

Table 5 presents the resulting identification bounds. Under this scenario, if individuals are assumed to have acted optimally regarding training participation, then both extremes —universal training and no training— would lead to lower financial mobility rates than those actually observed. In other words, letting each individual decide whether to participate could produce better outcomes than enforcing a uniform policy. Consequently, if the policymaker accepts this rational-choice view, the outcomes observed under self-selection may already reflect an optimal balance, and policy interventions that constrain individual choice might yield inferior results.

Table 5: Partial identification regions under Scenario 1 (rational individual choice).

Transition (baseline → endline)	All participants receive training		No participants receive training	
	Lower	Upper	Lower	Upper
From rank 1 to rank 2	2.0%	28.0%	26.0%	28.0%
From rank 2 to rank 3	2.0%	33.0%	31.0%	33.0%
From rank 1 to rank 3	2.0%	23.0%	22.0%	24.0%
From rank 3 to rank 4	4.0%	26.0%	23.0%	26.0%
From rank 2 to rank 4	4.0%	10.0%	6.0%	10.0%
From rank 1 to rank 4	4.0%	6.0%	2.0%	6.0%

4.3 *Scenario 2: Optimistic policymaker*

Scenario 2 can be described as follows: the implementer of the I90 Program is optimistic about the comprehensive treatment, believing that attending the training sessions is generally preferable to not attending. One indication of such optimism is the decision to implement a program that had not previously been tested in Chile.

In probabilistic terms, this optimism is expressed as

$$P(r_2(1) = k \mid r_1 = l, \zeta = 0) \geq P(r_2(0) = k \mid r_1 = l, \zeta = 0);$$
$$P(r_2(1) = k \mid r_1 = l, \zeta = 1) \geq P(r_2(0) = k \mid r_1 = l, \zeta = 1),$$

for all $k > l$.

The first inequality states that, for participants who actually received only the city intervention, their outcomes would have improved had they also attended the training sessions. The second inequality states that for those who did receive both components, outcomes would have been worse under city intervention alone.

Combining these assumptions with the decompositions presented in (4) and (5), we obtain the identification regions, for $k > l$, formulated as

$$\begin{aligned} P(r_2 = k \mid r_1 = l) &\leq P(r_2(1) = k \mid r_1 = l) \\ &\leq P(r_2 = k \mid r_1 = l, \zeta = 1)P(\zeta = 1 \mid r_1 = l) \\ &\quad + P(\zeta = 0 \mid r_1 = l); \end{aligned} \tag{8}$$

$$P(r_2 = k \mid r_1 = l, \zeta = 0)P(\zeta = 0 \mid r_1 = l) \leq P(r_2(0) = k \mid r_1 = l) \leq P(r_2 = k \mid r_1 = l).$$

For a formal details of this result, see Appendix B. As before, $P(r_2 = k \mid r_1 = l)$ denotes the empirical transition probability from rank l to rank k, regardless of training status.

Based on the results in Table 6, we observe that if the decision is to implement both the city intervention and the training sessions, the financial mobility rates (first column of percentages in the table) would likely improve. Thus, if the policymaker proceeds with offering training sessions to all, the results are expected to surpass those observed in Los Muermos. In fact, they may even outperform the outcomes under the status quo, where only some individuals opted into training.

Consequently, if the policymaker adopts an optimistic stance —believing that training participation is universally beneficial— and actively encourages participation (for example, via incentives or persuasion), the outcomes would likely exceed those already observed. In this view, comprehensive implementation is preferable to self-selection.

Table 6: Partial identification regions under Scenario 2 (optimistic policymaker).

Transition (baseline → endline)	All participants receive training		No participants receive training	
	Lower	Upper	Lower	Upper
From rank 1 to rank 2	35.0%	91.0%	34.0%	35.0%
From rank 2 to rank 3	33.0%	96.0%	31.0%	33.0%
From rank 1 to rank 3	24.0%	91.0%	22.0%	24.0%
From rank 3 to rank 4	26.0%	96.0%	23.0%	26.0%
From rank 2 to rank 4	10.0%	98.0%	6.0%	10.0%
From rank 1 to rank 4	6.0%	93.0%	2.0%	6.0%

4.4 *Scenario 3: Less intensive intervention is better*

Scenario 3 can be described as follows: the program implementer believes that the city-level intervention alone may be more effective than complementing it with financial training sessions. This scenario may be particularly relevant when considering whether it is worthwhile to invest additional resources in implementing the training component.

In probabilistic terms, this assumption is expressed as

$$\begin{aligned} P(r_2(1) = k \mid r_1 = l, \zeta = 0) &\leq P(r_2(0) = k \mid r_1 = l, \zeta = 0); \\ P(r_2(1) = k \mid r_1 = l, \zeta = 1) &\leq P(r_2(0) = k \mid r_1 = l, \zeta = 1), \end{aligned}$$

for all $k > l$.

The first inequality suggests that participants who received only the city intervention would have fared worse if they had also attended training. The second implies that even those who received both components would have performed better had they received only the city-level intervention.

Combining these assumptions with the decompositions formulated in (4) and (5), we derive the identification regions, for $k > l$, stated as

$$
\begin{aligned}
P(r_2 = k \mid r_1 = l, \zeta = 1)P(\zeta = 1 \mid r_1 = l) &\leq P(r_2(1) = k \mid r_1 = l) \leq P(r_2 = k \mid r_1 = l); \\
P(r_2 = k \mid r_1 = l) &\leq P(r_2(0) = k \mid r_1 = l) \\
&\leq P(r_2 = k \mid r_1 = l, \zeta = 0)P(\zeta = 0 \mid r_1 = l) \\
&\quad + P(\zeta = 1 \mid r_1 = l). \qquad (9)
\end{aligned}
$$

For formal details of this result, see Appendix B. Again, $P(r_2 = k \mid r_1 = l)$ denotes the empirical probability of transitioning from rank l to rank k, regardless of training status.

Table 7 summarizes the resulting partial identification regions. Under this pessimistic assumption, the implementation of training sessions is viewed as counterproductive, and eliminating them would lead to improved financial mobility outcomes. That is, if the implementer truly believes that the city intervention alone is superior, then acting accordingly —by omitting the training— would lead to higher transition probabilities.

Table 7: Partial identification regions under Scenario 3 (less intensive intervention).

Transition (baseline → endline)	All participants receive training		No participants receive training	
	Lower	Upper	Lower	Upper
From rank 1 to rank 2	2.0%	35.0%	35.0%	45.0%
From rank 2 to rank 3	2.0%	33.0%	33.0%	37.0%
From rank 1 to rank 3	2.0%	23.0%	24.0%	33.0%
From rank 3 to rank 4	4.0%	26.0%	26.0%	30.0%
From rank 2 to rank 4	4.0%	10.0%	10.0%	12.0%
From rank 1 to rank 4	4.0%	6.0%	6.0%	14.0%

5. Discussion

The implementation of a public policy or program ultimately rests on the discretionary decision of the policymaker. While this observation may seem self-evident, it is crucial to explicitly incorporate this dependency into the impact evaluation process. Moreover, the decision to implement a policy inherently involves a degree of risk which, if appropriately quantified, can offer valuable insights to inform the policymaker's course of action.

This article proposes a framework for evaluating the impact of a policy or program that explicitly accounts for both elements. We begin by addressing the fundamental identification problem in impact evaluation: the impossibility of observing the same statistical unit simultaneously under both treatment (or intervention) and control conditions.

Rather than relying on conventional ignorability assumptions, we adopt a partial identification approach to the relevant probabilities. This strategy highlights that the identification problem does not admit a unique solution; instead, it produces at least three plausible alternatives. Each of these alternatives corresponds to what the policymaker is willing to assume —either about the behavior of individuals (Scenario 1) or about their own policy stance (Scenarios 2 and 3).

Each scenario yields partial identification regions for the probabilities of interest, $P(r_2(1) = k \mid r_1 = l)$ and $P(r_2(0) = k \mid r_1 = l)$. These regions not only capture all plausible values under the assumptions of each scenario, but also quantify the uncertainty involved: wider intervals signal greater ambiguity. This representation offers a practical way for policymakers to assess the robustness of the evaluation and the degree of uncertainty associated with each course of action.

Faced with such an evaluation framework, different policymakers may interpret and respond to the findings differently. In the case study presented, three scenarios are examined. For instance, under Scenario 1, the policymaker assumes that individuals make optimal choices regarding whether to attend the training sessions. In that case, the mobility observed in Los Muermos would not necessarily improve through universal enforcement of the intervention. Importantly, the evaluation is conducted with respect to the actual target population, not some external or hypothetical group. As such, the results are not predictions for a different commune, but rather tools to inform the policymaker's stance on implementing the intervention, while still allowing individuals to opt in or out of training.

This approach aligns with Neyman's concept of inductive behavior, whereby scientific knowledge influences action through the identification of stable patterns in empirical phenomena.

6. Conclusions

This study proposed a methodological framework for policy evaluation based on partial identification, without imposing distributional assumptions on the data-generating process. Such assumptions are often implausible in causal inference settings involving observational data and limited compliance.

As discussed previously, our conclusions are specific to the populations studied. Because policy evaluation differs from statistical prediction, we did not compute standard errors in the conventional sense. Instead, uncertainty is conveyed by the width of the partial-identification regions, which quantify structural ambiguity arising from unobserved counterfactuals. These bounds reflect structural uncertainty only, and do not account for sampling variability; if desired, that component could be examined with a cluster-aware bootstrap, but our focus here is on identification limits rather than frequentist inference.

Consistent with Neyman's concept of inductive behavior, we interpret these evaluations as tools to support decision-making under uncertainty, where the final course of action depends ultimately on the will of the policymaker.

Although a rightward shift in the endline distribution was observed for the subgroup that complied with training, partial-identification analysis shows that this pattern is compatible with a wide range of unobserved outcomes; it cannot, therefore, be taken as definitive evidence that the programme improved financial health. Moreover, the index is a subjective composite whose psychometric properties in this population remain to be validated, so any apparent gains should be interpreted with caution.

Two main limitations must be acknowledged. First, subgroup sample sizes are modest. A second implementation of the intervention was considered but proved financially infeasible. Second, the statistical and conceptual approach adopted here departs from mainstream techniques, introducing alternative interpretations —particularly regarding the meaning of "evaluation".

Note that our intention is not to replace existing methods but to provide a complementary perspective, reminding policymakers that multiple valid paths may coexist, and that scientific results should inform—rather than prescribe—policy action.

This perspective opens avenues for interdisciplinary research. A promising direction is to work directly with policymakers through focus groups to understand how they interpret different partial-identification scenarios, echoing recent work on heuristics and elite judgment [36]. Future developments may include multicentre designs, Bayesian borrowing of strength, cost-effectiveness analyses, and the incorporation of behavioural financial metrics, all of which could enhance the practical utility of partial-identification frameworks in policy evaluation.

Appendices

Appendix A: Financial health questionnaire

The financial health questionnaire consists of two sets of questions: the first set pertains to the respondent's current situation, while the second focuses on the occurrence of specific financial events.

The first set is introduced with the question: How well do the following statements describe you or your situation? The items are the following:

- I can cope with a significant unexpected expense.
- I am securing my financial future.
- Due to my financial situation, I believe I will never have the things I want in life.
- I can enjoy life because of the way I manage my money.
- I am barely getting by financially.
- I am concerned that the money I have or save will not last.

Each item is rated on a Likert scale with the following response categories: Not at all; Very little; Somewhat; Very well; Completely.

The second set is introduced with the question: How often do the following situations occur? The items are the following:

- Purchasing a gift for a wedding, birthday, or other occasion would impose a burden on my monthly finances.
- I have money left over at the end of the month.
- I am behind on my finances.
- My finances control my life.

Each of these is answered using the following Likert scale: Never; Almost never; Sometimes; Often; Always.

Appendix B: Technical details

We derive here the partial identification region stated in (7). Starting from the decomposition formulated in (4), we write

$$\begin{aligned} P(r_2(1)=k \mid r_1=l) &= P(r_2(1)=k \mid r_1=l,\zeta=1)P(\zeta=1 \mid r_1=l) \\ &\quad +P(r_2(1)=k \mid r_1=l,\zeta=0)P(\zeta=0 \mid r_1=l) \\ &\leq P(r_2(1)=k \mid r_1=l,\zeta=1)P(\zeta=1 \mid r_1=l) \\ &\quad +P(r_2(0)=k \mid r_1=l,\zeta=0)P(\zeta=0 \mid r_1=l) \\ &= P(r_2=k \mid r_1=l,\zeta=1)P(\zeta=1 \mid r_1=l) \\ &\quad +P(r_2=k \mid r_1=l,\zeta=0)P(\zeta=0 \mid r_1=l) \\ &= P(r_2=k \mid r_1=l), \end{aligned}$$

where we have used the fact that when conditioning on the actual treatment received, $r_2(1)=r_2$ for those with $\zeta=1$, and $r_2(0)=r_2$ for those with $\zeta=0$.

Since $P(r_2(1)=k \mid r_1=l,\zeta=0)\geq 0$, it follows that

$$P(r_2(1)=k \mid r_1=l) \geq P(r_2(1)=k \mid r_1=l,\zeta=1)P(\zeta=1 \mid r_1=l),$$

establishing the lower bound of the first inequality defined in (7).

Similarly, from the expressions given in (5) and (6), we have

$$\begin{aligned}
P(r_2(0)=k \mid r_1=l) &= P(r_2(0)=k \mid r_1=l,\zeta=1)P(\zeta=1 \mid r_1=l)\\
&\quad + P(r_2(0)=k \mid r_1=l,\zeta=0)P(\zeta=0 \mid r_1=l)\\
&\leq P(r_2(1)=k \mid r_1=l,\zeta=1)P(\zeta=1 \mid r_1=l)\\
&\quad + P(r_2(0)=k \mid r_1=l,\zeta=0)P(\zeta=0 \mid r_1=l)\\
&= P(r_2=k \mid r_1=l,\zeta=0)P(\zeta=1 \mid r_1=l)\\
&\quad + P(r_2=k \mid r_1=l,\zeta=0)P(\zeta=0 \mid r_1=l)\\
&= P(r_2=k \mid r_1=l).
\end{aligned}$$

Since $P(r_2(0)=k \mid r_1=l,\zeta=1) \geq 0$, it follows that $P(r_2(0)=k \mid r_1=l) \geq P(r_2(0)=k \mid r_1=l,\zeta=0)P(\zeta=0 \mid r_1=l)$, which gives the lower bound for the second inequality stated in (7). The partial identification regions presented in (8) and (9) follow by similar arguments.

Statements

Acknowledgements

The authors would like to thank the Editors-in-Chief and the two anonymous reviewers for their valuable comments and suggestions, which substantially improved the quality of this article.

Author contributions

Conceptualization: E. San Martin, M.S. Graell, M. Borquez; data curation: I. Canales, C. Maldonado; formal analysis: E. San Martin, I. Canales; investigation: E. San Martin, M.S. Graell, C. Maldonado, M. Borquez, I. Canales; methodology: E. San Martin, I. Canales, C. Maldonado; software: E. San Martin; supervision: E. San Martin, M.S. Graell; validation: E. San Martin, M.S. Graell, M. Borquez, C. Maldonado; visualization: E. San Martin; writing original draft: E. San Martin; writing review and editing: E. San Martin. All authors have read and agreed to the published version of the article.

Conflicts of interest

The authors declare no conflict of interest. The conclusions presented in this article do not necessarily reflect the institutional views of Caja de Compensación Los Andes.

Data and code availability

The data and computational code are available from the authors upon request.

Declaration on the use of artificial intelligence (AI) technologies

The authors declare that no generative AI was used in the preparation of this article.

Funding

This research was supported by a grant from Caja de Compensación Los Andes. The first author was also partially funded by the Millennium Nucleus on Intergenerational Mobility: From Modelling to Policy (MOVI), grant NCS2021072, of the Millennium Science Initiative, which is a program of the Chilean National Agency for Research and Development (ANID) under the Ministry of Science, Technology, Knowledge and Innovation.

References

[1] Atkinson, A. and Messy, F.A., 2016. OECD/INFE international survey of adult financial literacy competencies. Technical Report, OECD, Paris, France. DOI: 10.1787/28b3a9c1-en

[2] Kempson, E., 2009. Framework for the development of financial literacy baseline surveys: A first international comparative analysis. OECD, Paris, France.

[3] Chandler, J., Cumpston, M., Li, T., Page, M., and Welch, V., 2019. Cochrane Handbook for Systematic Reviews of Interventions. Wiley, Hoboken, NJ, US.

[4] Fisher, R.A., 1935. The Design of Experiments. Oliver and Boyd, London, UK.

[5] Manski, C.F., 2007. Identification for Prediction and Decision. Harvard University Press, Cambridge, MA, US.

[6] Manski, C.F., 1997. The mixing problem in program evaluation. The Review of Economic Studies, 64, 537–553. DOI: 10.2307/2971730

[7] Pepper, J.V., 2003. Using experiments to evaluate performance standards: What do welfare-to-work demonstrations reveal to welfare reformers? The Journal of Human Resources, 38, 860–880. DOI: 10.2307/1558782

[8] Florens, J.P., Mouchart, M., and Rolin, J.M., 1995. Elements of Bayesian Statistics. Marcel Dekker, New York, NY, US.

[9] Rosenbaum, P.R. and Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika, 70, 41–55. DOI: 10.1093/biomet/70.1.41

[10] Imbens, G.W., 2000. The role of the propensity score in estimating dose-response functions. Biometrika, 87, 706–710. DOI: 10.1093/biomet/87.3.706

[11] Hernán, M.A. and Robins, J.M., 2020. Causal Inference: What If. Chapman and Hall/CRC, Boca Raton, FL, US.

[12] Neyman, J., 1938. L'estimation statistique traitée comme un problème classique de probabilité. Actualités Scientifiques et Industrielles, 739, 25–57.

[13] Gigerenzer, G., 2004. Mindless statistics. The Journal of Socio-Economics, 33, 587–606. DOI: 10.1016/j.socec.2004.09.033

[14] Gigerenzer, G., 2014. The superego, the ego, and the id in statistical reasoning. Keren, H.J. and Lewis, C. (editors). A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues, pp. 311–340. Psychology Press, Hove, UK.

[15] Haller, H. and Krauss, S., 2002. Misinterpretations of significance: A problem students share with their teachers. Methods of Psychological Research, 7, 1–20. DOI: 10.5283/EPUB.34338

[16] Krantz, D., Luce, R.D., Suppes, P., and Tversky, A., 1971. Foundations of Measurement. Vol. I: Additive and Polynomial Representations. Academic Press, New York, NY, US.

[17] Cholli, N.A. and Durlauf, S.N., 2022. Intergenerational mobility. Technical Report No. w30099, National Bureau of Economic Research, Cambridge, MA, US. DOI: 10.3386/w29760

[18] Manski, C.F., 2003. Partial identification of probability distributions. Springer, New York, NY, US.

[19] Neyman, J., 1950. First Course in Probability and Statistics. Holt, Rinehart and Winston, New York, NY, US.
[20] Neyman, J., 1957. "Inductive behavior" as a basic concept of philosophy of science. Revue de l'Institut International de Statistique, 25, 7–22. DOI: 10.2307/1401671
[21] Manski, C.F., 2013. Diagnostic testing and treatment under ambiguity: Using decision analysis to inform clinical practice. Proceedings of the National Academy of Sciences, 110, 2064–2069. DOI: 10.1073/pnas.1221405110
[22] Diemer, E., Zuccolo, L., and Swanson, S.A., 2023. Partial identification of the average causal effect in multiple study populations: The challenge of combining Mendelian randomization studies. Epidemiology, 34, 20–28. DOI: 10.1097/EDE.0000000000001526
[23] Frisch, R., 1934. Statistical confluence analysis by means of complete regression systems. Technical Report, University Institute for Economics, Oslo, Norway.
[24] Reiersøl, O., 1941. Confluence analysis by means of lag moments and other methods of confluence analysis. Econometrica, 9, 1–24. DOI: 10.2307/1907171
[25] Horowitz, J.L. and Manski, C.F., 1995. Identification and robustness with contaminated and corrupted data. Econometrica, 63, 281–302. DOI: 10.2307/2951627
[26] Molinari, F., 2008. Partial identification of probability distributions with misclassified data. Journal of Econometrics, 144, 81–117. DOI: 10.1016/j.jeconom.2007.12.003
[27] San Martín, E. and Alarcón-Bustamante, E., 2022. Dissecting Chilean surveys: The case of missing outcomes. Chilean Journal of Statistics, 13, 17–46. DOI: 10.32372/ChJS.13-01-02
[28] Duncan, O.D. and Davis, B., 1953. An alternative to ecological correlation. American Sociological Review, 18, 665–666. DOI: 10.2307/2088122
[29] Fréchet, M., 1960. Les tableaux dont les marges sont données. TEST, 11, 3–18. DOI: 10.1007/BF03009203
[30] Cross, P.J. and Manski, C.F., 2002. Regressions, short and long. Econometrica, 70, 357–368. DOI: 10.1111/1468-0262.00279
[31] Mebane, W.R. and Poast, P., 2003. Causal inference without ignorability: Identification with nonrandom assignment and missing treatment data. Political Analysis, 21, 233–251. DOI: 10.1093/pan/mps043
[32] Keele, L. and Minozzi, W., 2013. How much is Minnesota like Wisconsin? Assumptions and counterfactuals in causal inference with observational data. Political Analysis, 21, 193–216. DOI: 10.1093/pan/mps041
[33] Keele, L., 2015. The statistics of causal inference: A view from political methodology. Political Analysis, 23, 313–335. DOI: 10.1093/pan/mpv007
[34] Breskin, A., Westreich, D., Cole, S., and Edwards, J., 2019. Using bounds to compare the strength of exchangeability assumptions for internal and external validity. American Journal of Epidemiology, 188, 1355–1360. DOI: 10.1093/aje/kwz060
[35] Holland, P.W., 1986. Statistics and causal inference. Journal of the American Statistical Association, 81, 945–960. DOI: 10.1080/01621459.1986.10478354
[36] Vis, B., 2023. Heuristics and political elites' judgment and decision-making. Oxford University Press, Oxford, UK.