Discussion Paper N° 2020|02

How to Broker the Evaluation of Public Policies? A Strategy to Reveal Inherent Uncertainties in Policy-Decision Making

Ernesto San Martín & Trinidad González-Larrondo

Facultad de Matemáticas
Pontificia Universidad Católica de Chile
Av. Vicuña Mackenna 4860, Macul
lies.mat.uc.cl/

How to Broker the Evaluation of Public Policies? A Strategy to Reveal Inherent Uncertainties in Policy-Decision Making

Ernesto San Martín[1, 2, 4] and Trinidad González-Larrondo[1, 2, 3]
[1]*Interdisciplinary Laboratory of Social Statistics, Pontificia Universidad Católica de Chile, Chile*
[2]*Faculty of Mathematics, Pontificia Universidad Católica de Chile, Chile*
[3]*School of Psychology, Faculty of Social Sciences, Pontificia Universidad Católica de Chile, Chile*
[4]*The Economics School of Louvain, Université Catholique de Louvain, Belgium*

January 10, 2020

Abstract

Evidence-based practices (EBP) seem to be the right way to relate science and policy. Although this relationship has been criticized by Policy Analysis and Psychology, critics have not analyzed the way in which scientific evidence becomes scientific knowledge, assuming the latter is relevant for the design and implementation of public policies. This paper offers a strategy that shows to what extent scientific knowledge is mediated by subjectivities that can be identified at the methodological level; and how this is an achievement at the methodological level of the scientist's honest broker role. The paper develops the strategy at the conceptual and theoretical level, which jeopardize the EBP paradigm. Two case studies are developed in the Appendix.

Key-words: Causal Inference; Evidence Based Practices; Honest Broker of Policy Alternatives; Policy Decisions; Political Values; Prescriptive Decision Theory; Scientific Knowledge.

1 Introduction

Most of the time, three agents come together in a public policy decision process: policymakers, politicians and scientists. Policymakers perform *decisions* because policy "is a decision, a commitment to particular course of action" (Pielke Jr, 2007), "it is the sum total of government action, from signals of intent to the final outcomes" (Cairney, 2016). Politicians do the necessary "bargaining, negotiation, and compromise in pursuit of desired ends related to *political values*" (Pielke Jr, 2007, the italics are ours). Thus, as it is widely known, the political values shape the meaning and scope of the action. Whatever the political values are, public policies intend to improve the welfare of society through interventions. The meaning of society's well-being depends on those values, (implicitly) assuming a causal relationship between interventions and welfare.

1.1 Public policy implementation and uncertainty

Such a casual relation, which can be considered as a belief (Müller, Garcia-Retamero, Cokely, & Maldonado, 2011), is subject to uncertainty due at least to some of the following issues: a specific intervention could imply outcomes that are inconsistent with the expectations of the policymaker (Pielke Jr, 2007, Chapter 5), leading even to unintended and perverse impacts (Van Woensel, 2020, Chapter 3); a policymaker can understand and try to solve a policy problem in many ways because a single event can take place in a multitude of possible states (Cairney, 2016); and a policymaker takes a decision with a partially known welfare function and/or probabilities of the outcomes (Manski, 2007a, Chapter 11).

In a context of uncertainty, science seems to be useful because it also looks for welfare of society and, consequently, their findings are considered as progress of and for humanity. Just some leading examples: the American Psychological Association is called "to make a positive impact on critical societal issues" and "to promote the advancement, communication, and application of psychological science and knowledge to benefit society and improve lives"[1]. The mission of the American Educational Research Association is "to promote the use of research to improve education and serve the public good"[2]. Science then plays the role of reducing the uncertainty of the causal intervention-welfare relationship. To do this, scientific research not only produces evidence, but through an inter-subjective agreement defines criteria that allow to graduate the quality of scientific evidence, in addition to developing statistical methodologies relevant to such graduation (for instance, systematic revision of randomized clinical trials or meta-analysis; causal inference for observational data; and so on). This is the so-called *Evidence-Based Practice* (EBP) framework, from which Evidence Based Medicine and Evidence Based Policy are key instances. For references and applications of Evidence Based Medicine, see Cochrane Initiative (www.cochrane.org); and for Evidence Based Policy, see The Campbell Collaboration (campbellcollaboration.org).

1.2 The features of the EBP paradigm

The EBP paradigm is highly appreciated by both academicians and politicians because "evidence *matters* for public policymaking" (Parkhurst, 2017, p.4) given that a rigorous use of it could avoid unnecessary harms, and evidence tells us "what works". These advantages are due to the fact that the EBP framework intends to solve the problem of the gap between policy practices and research, improving thus the quality of policies and interventions (Cochrane, 1972; OECD, 2007; European Commission, 2017). Politicians and policymakers perceive quantitative evidence-based research as the more legitimate framework because of its eventual accuracy and trustworthiness (Wiseman, 2010), which leads to consider it as the way to ensure "objectivity" (King, 2010). This perspective stresses that, when relating science and public policy, the emphasis is placed on the quality of the scientific evidence used. The challenge to reduce uncertainty in decision-making is accordingly to recognize and rely on the best available evidence defined according to the inter-subjectivity international agreement mentioned above (Sackett, Straus, Richardson, Rosenberg, & Haynes, 2000; Clearinghouse, 2007; Head, 2010; Burns, Rohrich, & Chung,

[1]From https:www.apa.orgmonitor201905ceo; retrieved on September 22, 2019.
[2]From https:www.aera.net/About-AERA; retrieved on September 22, 2019.

2011; Schünemann, 2016; Rousseau & Gunia, 2016; Ryan & Hill, 2016). The motto of Evidence Based Medicine, of Evidence Based Policy and, in general, of Evidence Based Practice is accordingly that *policy decisions should be based on the better available evidence* (Davies & Boruch, 2001).

1.3 The consequences of the EBP paradigm

According to the EBP paradigm, the welfare of society is ensured by science (knowledge, research, applications). This fits with the *linear model of science* (V. Bush, 1945; Godin, 2006, 2011; Pielke Jr, 2007, 2012), a paradigm that brings politics into science rather than science into policy. In fact, it argues that science can and should compel political outcomes given that *right science bases better decision-making*. The purpose of the policy action is, consequently, shaped by "right science" in the sense that the right scientific results deserve to become public policies. This attitude corresponds to the politicization of science by scientist, being its final stage to substitute politics by science.

Following Pielke Jr (2007)'s taxonomy of the role of scientist in policy and politics, the previous attitude could lead to an advocacy scientist role: under the linear model paradigm, science becomes a convenient and necessary means for removing certain options from a debate without explicitly dealing with disputes over values (politics). For examples of this attitude, see Manski (1992)'s analysis of Friedman's perspective on voucher and social mobility; or Pielke Jr (2007)'s analysis of scientists' attitude regarding Lomborg's *The Skeptical Environmentalist*. The conclusions are determined according to a standardized quality of evidence, which is related to what is considered correct science through an uncritical revision of an inter-subjective agreement –in this case, Evidence Based Medicine and Evidence Based Policy standards.

1.4 Limitations of the EBP paradigm due to additional uncertainties

Although the decision of public policies seeks to decrease uncertainty through the EBP paradigm, there exists additional uncertainties that limit the power of EBP. These uncertainties emerge when the decision making reasoning is analyzed under a descriptive decision theory rather than a normative theory: instead of assessing the decision making with respect to normative theories as those developed by Wald (1950) and Von Neumann, Morgenstern, and Kuhn (2007), the challenge is to describe how and to what extend the context affects the decision making; for details, see Kahneman and Tversky (1979), Parmigiani and Inoue (2009) and Peterson (2017).

A first uncertainty is due to biases of scientific policy advisors. As a matter of fact, following Van Woensel (2020), scientific policy advisers collect information and on its basis formulate alternative policy options. In this process, biases can cause systematically distorted perceptions of facts and opinions, which lead to incorrectly weight the evidence. Van Woensel (2020) proposes a tool (the *bias-wheel*) for bias-awareness in scientific advising. Six types of biases are necessary to take into account in order to acquire a critical view on the own biases: research biases, cultural and value biases, attention biases, interest biases, availability biases and associative biases; for details, see Van Woensel (2020, Chapter 2). By using the bias-awareness tool, the one's biases toward new information will be reduced, leading to less uncertain evidence-based policy decisions.

A second uncertainty has been made explicit by the Policy Analysis perspective. The distinction between normative and descriptive decision theories allows to realize that policy decisions are not ultimately based on scientific evidence. Several reasons explain this: sometimes scientific evidence does not interest; in other occasions, it is used to manipulate, pushing thus a specific political value; sometimes it is necessary to consider institutional and organizational mechanisms that influence decisions; and so on. The point is that a policy decision making is complex and subject to political ambiguities; see, among many others, Rycroft-Malone and Bucknall (2011), Cairney (2016), Cairney, Oliver, and Wellstead (2016), Cairney (2017), Parkhurst (2017), Newman (2017) and Strassheim (2017).

A third uncertainty comes from deviations of both policymakers and politicians from the normative theory. As a matter of fact, prospect theory (a descriptive decision theory; see Kahneman & Tversky, 1979) has been applied to policy decision making (see, among many others, Dacey, 2003; Levy, 2003). Linde and Vis (2017) has applied two concepts of the prospect theory, namely *probability weighting* and *reflection effect*, to experimentally analyze the decisions made by a sample of Dutch parliamentarians; for details on these concepts, see Levy (1992) and Kahneman and Tversky (2013). Linde and Vis (2017) builds prospects such as those presented by Allais (1953) and Kahneman and Tversky (1979). In particular, they frame prospects in terms of profits and their reflections in terms of losses; this type of prospects is inspired in the classical Asian disease problem (Kahneman & Tversky, 1984), where a choice between voting in favor of implementing program A or voting against it in which case program B is implemented. The conclusions are that decisions depend on the *framing* of the decision situation (Tversky & Kahneman, 1981), but when participants are "familiar with the type of decision" then decisions are taken according to the normative theory.

1.5 Are policies based on scientific evidence or scientific knowledge?

The previous critical frameworks of the EBP paradigm share a common feature, namely to critically describe how policymaker *relate to* the scientific evidence. Such a relation can be affected either by the scientific advisors biases, or by the complexity and ambiguity of the political context, or by deviations from a normative decision theory. It could be concluded that if such uncertainties were non-existent, the EBP paradigm would work without limitations. This conclusion is reinforced because the three critical frameworks already described do not develop (up to the best of our knowledge) a critique of the way in which *scientific evidence* becomes *scientific knowledge*. It is important to make a distinction between scientific knowledge and scientific evidence: "Knowledge is the set of conclusions that one draws by combining evidence with assumptions about unobserved quantities" (Manski, 2013a). What it is really useful for political decision making is scientific knowledge, and not just evidence. The evidence is not a subject whose light and majesty we bow to (paraphrasing Toland, 1702, one of the first to describe the properties of evidence in the context of reason and its powers), but an object that a subject (the scientist) uses to share (some) knowledge.

What we mean by *assumptions about unobserved quantities*? On the one hand, let us recall a terminological distinction: *assumption* here does not correspond to the concept of *statistical hypothesis* in the sense that such hypothesis can be tested through the hypothesis testing techniques (Oaks, 1986; Lehmann, 2011; Gigerenzer, 2004) and eventually falsifiable (for the potential relationships between hypothesis testing and the popperian concept of falsification, see Oaks, 1986; Lehmann, 1995; Mayo & Spanos, 2006). On the other hand, the concept of *assumptions* is a component of the empirical research, which can schematically be represented as follows (Manski, 2013b):

$$\text{Assumptions} \quad + \quad \text{Data / Empirical evidence} \quad \Longrightarrow \quad \text{Conclusions / Recommendations.}$$

Here *assumptions* correspond to the explanation or formation of the studied phenomenon, which is the spring of the substantive theory (Mouchart & Russo, 2011; Wolpin, 2013; Wunsch, Mouchart, & Russo, 2014; Felline, 2018). Paraphrasing Koopmans (1949), statistical inference unsupported by a substantive theory applies to whatever statistical regularities and, consequently, stable relationships can be discerned in the data. Such purely empirical relationships are likely to be due to the presence and persistence of underlying structural relationships. However, the direction of this deduction cannot be reversed (from the empirical to the structural relationships) expect possibly with the help of a substantive theory which specifies the form of the structural relationships. The possibility of this deduction necessarily implies to solve an *identification problem* at the model construction level. As Haavelmo (1997) puts it,

> The one type of problems, seemingly paradoxical, grew out of a rather intricate consequence of successful economic theory. Strangely enough, the fact is that if an economic theory, an economic relation, is a good theory, true to reality, it may not be possible to quantify it by using data from the economic of which that relation is a part. This is the so-called "problem of identification" (p.14).

These considerations are applicable to all social and health sciences, among many others. This is a requirement of a scientific inquiry that focuses on the explanation of the phenomena, and not (only) on the mere prediction with implicit assumptions of structural stability (as is the case with most of the recent data science techniques).

Technically speaking, an observed phenomenon is captured through a set of probability distributions, defining thus the so-called *statistical model* (Fisher, 1922). This modelling approach is due to the fact that observed phenomena are heterogeneous and therefore subject to variability. Now, such probability distributions include parameters, which represent characteristic of the observed population under study. However, most of the time, scientists look deducing facts based on the observed phenomenon: when these facts can not be represented as a functional of the statistical model (that is, when such facts are not represented through a parameter), we face an identification problem. An identification problem can only be resolved through assumptions. There is no single way to solve it: when an assumption is chosen and combined with the evidence, we gain *a* specific scientific knowledge. If we change the assumptions, a *different* scientific knowledge could arise. For details and discussion, see San Martín (2016) and San Martín (2018).

We argue that the evaluation of public policies requires this type of assumptions leading to solve inherent identification problems. In this paper, we not only show that there is not a unique way to establish such assumptions, but we argue that a scientist should explicitly explore more than one assumption in order to expand policy choices. By so doing, it is possible to make explicit how the scientific evidence becomes scientific knowledge, which in turn is mediated by subjective choices. In this way, we do not intend to look for additional elements leading to improve the EBP paradigm, but rather to consider the relationship science-policy-politics as a problem to be studied, looking for a *prescriptive solution* (Manski, 2019) of the relationship between science, policy and politics, respecting the difference there exists between them.

1.6 Organization of the paper

This paper is organized as follows: the strategy to transform scientific evidence in scientific knowledge is developed in Section 2, where the attention is focused on the evaluation of a public policy based on an evidence generated in the context of an observational study. In Section 3 the findings established in the theoretical part are complemented by using the decision theory developed by Wald (1950) and Savage (1951): we use the minimax-regret criterion to base a decision leading to define a fractional treatment rules. The ethical problems underlying it are also discussed. Taking into account the previous results, Section 4 develops a discussion about the role of science in policy and politics. The proof of the main results are gathered in Appendix A. Two applications are developed in Appendices B and C.

2 How to construct a prescriptive role for science in policy decisions?

Following Pielke Jr (2007), science can play a constructive role in policy if organizations and individuals are willing to expand the range of options available to policymakers "by serving as Honest Broker of Policy Alternative" (p.141). In the context of policy debate the following questions may be useful to ask suppliers of policy analyses that are based on science: *(1) If your recommendation is indeed based on scientific results, what scientific information would be necessary to change your policy position? (2) A range of policies is consistent with particular scientific results. What is the full range of options that you see as consistent with the state of science in order to achieve particular desired ends?* (Pielke Jr, 2007, pp.141-142).

We propose to broker the evaluation of public policies in the context of observational studies using partial identification. The *basic elements* of the context are the following: (1) The public policy decision problem consists in implementing an innovation or not. (2) The evidence available to take a decision is an observational study, which is summarized by two conditional probabilities, $P(Y = 1 \mid Z = 1, X)$ and $P(Y = 1 \mid Z = 0, X)$, where Y is a policy relevant outcome. Without loosing generality, we assume that it is a binary random variable: when $Y = 1$ we will say that the statistical unit (or, more briefly, individual) experiences the expected beneficial outcome, and if $Y = 0$, that the beneficial outcome is not experienced by the individual. $Z = 1$ means that the individual is under an *innovation*, whereas $Z = 0$ that is under the *status quo*; and X is a set of covariates characterizing subpopulations. (3) The decision should be based on a comparison between $P(Y_1 = 1 \mid X)$ and $P(Y_0 = 1 \mid X)$, where Y_1 is the outcome as if all the individuals were under the innovation, and similarly for Y_0. The *problem* is that those conditional probabilities are unidentified. This can be viewed by considering the following fundamental decompositions (which follows from the Law of Total Probability):

$$P(Y_1 = 1 \mid X) = P(Y_1 = 1 \mid X, Z = 1)P(Z = 1 \mid X) + P(Y_1 = 1 \mid X, Z = 0)P(Z = 0 \mid X); \quad (2.1)$$
$$P(Y_0 = 1 \mid X) = P(Y_0 = 1 \mid X, Z = 1)P(Z = 1 \mid X) + P(Y_0 = 1 \mid X, Z = 0)P(Z = 0 \mid X). \quad (2.2)$$

The evidence alone is enough to identify $P(Z = 1)$ –the proportion of individuals under innovation, and therefore those under the status quo. It is also enough to identify $P(Y_1 = 1 \mid Z = 1, X)$ because it corresponds to the distribution of the outcome that an individual would experience under the innovation when it is actually under the innovation ($Z = 1$); that is, $P(Y_1 = 1 \mid X, Z = 1) = P(Y = 1 \mid X, Z = 1)$. By a similar argument, $P(Y_0 \mid Z = 0, X) = P(Y = 1 \mid X, Z = 0)$ and, therefore, it is identified. However, the evidence is not enough to identify $P(Y_1 = 1 \mid Z = 0, X)$, that is, the distribution of the outcome that an individual would experience under the innovation when it is actually under the status quo. By a similar argument, $P(Y_0 = 1 \mid Z = 1, X)$ is unidentified. This is the so-called Fundamental Problem of Causal Inference (Holland, 1988).

According to the Honest Broker perspective (Pielke Jr, 2007), a public policy evaluation should consider different policy alternatives. These ones intend to make explicit the complexity of the decision making process: once we have empirical evidence in the form of an observational study, it is reasonable to assume that there is more than one alternative to drive the implementation of innovation or the maintenance of the status quo.

2.1 The standard solution

Before displaying the complexity of the decision making process by making explicit different alternatives, let us revisit the standard solution to the identification problem which is based on the *exogenous switching condition* (Maddala, 1983) or *strong ignorability condition* (Rosenbaum & Rubin, 1983). It asserts statistical independence of the realized treatments and the outcomes, conditional on a set of covariates. This condition mimics the definition of a randomized experiment (Fisher, 1935; Cochran & Chambers, 1965; Angrist & Pischke, 2008), but it is not empirically refutable (Manski, 2007a): it can be justified in specific applications only, the justification essentially depending on the choice of specific covariates X. More precisely, it is assumed that

$$(Y_0, Y_1) \perp\!\!\!\perp Z \mid X, \quad (2.3)$$

which combined with (2.1) and (2.2) implies that

$$P(Y_1 = 1 \mid X, Z = 1) = P(Y_1 = 1 \mid X, Z = 0) = P(Y_1 = 1 \mid X);$$
$$P(Y_0 = 1 \mid X, Z = 1) = P(Y_0 = 1 \mid X, Z = 0) = P(Y_0 = 1 \mid X).$$

Condition (2.3) deserves two remarks and one question:

1. It is an identification restriction: it restricts the parameter space of $P(Y_1 = 1 \mid X)$ and $P(Y_0 = 1 \mid X)$ to two points, which precisely are the information provided by the observational study.

2. It does not solve the Fundamental Problem of Causal Inference which states that "we do not observe a same statistical unit under both the innovation and the status quo" (Holland, 1988). Condition (2.3) states, for instance, that the distribution of the outcome that an individual would experience under the innovation when it is actually under the status quo is exactly equivalent to the distribution of the outcome that an individual should experience under the innovation when it is actually under the the innovation.

3. Is it possible to provide a policy interpretation of such a condition? Following the logic of the empirical research (see Section 1.5, condition (2.3)) is combined with evidence in order to provide policy recommendations. However, it is important to stress what kind of policy behavior and/or political value is being affected by such a condition. We will come back to this later.

2.2 Optimistic policymaker

Using Manski (2007a)'s approach on partial identification, we propose six solutions leading to solve of the Fundamental Problem of Causal Inference. These solutions are obtained after comparing $P(Y_1 = 1 \mid X, Z = 1)$ with $P(Y_0 = 1 \mid X, Z = 1)$, and $P(Y_1 = 1 \mid X, Z = 0)$ with $P(Y_0 = 1 \mid X, Z = 0)$. A first solution is based on two *policymaker's optimistic perspectives*: the first one is characterized by a *convincing policymaker's behavior*, whereas the second one by an *ambiguous policymaker's perspective*.

2.2.1 Convincing policymaker's optimistic perspective

A convincing policymaker's optimistic (CPO) perspective is defined by the following inequalities:

$$\begin{array}{ll} \text{(i)} & P(Y_1 = 1 \mid X, Z = 1) > P(Y_0 = 1 \mid X, Z = 1); \\ \text{(ii)} & P(Y_1 = 1 \mid X, Z = 0) \geq P(Y_0 = 1 \mid X, Z = 0). \end{array} \tag{2.4}$$

This perspective is characterized by (i) promoting the idea that if individuals who are actually under the innovation, were under the status quo, their performance would be worse; that is, the status quo harms individuals under the innovation –this corresponds to condition (2.4.i). And (ii) if individuals who are actually under the status quo were under the innovation, their performance would be at least as the performance under the status quo; that is, the innovation does not harm individuals under the status quo -this corresponds to condition (2.4.ii).

What the impact of the CPO perspective on the identifiability of both $P(Y_1 = 1 \mid X)$ and $P(Y_0 = 1 \mid X)$ is? Or, equivalently, what are all the plausible values of $P(Y_1 = 1 \mid X)$ and $P(Y_0 = 1 \mid X)$ that are compatible with the CPO perspective? These questions can be answered after combining (2.4) with the basic decompositions (2.1) and (2.2). By so doing, we obtain the following partial identification intervals:

$$\begin{array}{ll} \text{(i)} & P(Y = 1 \mid X) \leq P(Y_1 = 1 \mid X) \leq P(Y = 1 \mid X) + P(Y = 0, Z = 0 \mid X); \\ \text{(ii)} & P(Y = 1, Z = 0 \mid X) \leq P(Y_0 = 1 \mid X) < P(Y = 1 \mid X). \end{array} \tag{2.5}$$

For a proof, see Appendix A.1. The intervals (2.5) make explicit the consequences of the CPO perspective at both behavioral and expectation levels:

1. The interval (2.5.i) shows that a CPO perspective will improve the current situation represented by $P(Y = 1 \mid X)$, that is, the proportion of individuals experiencing the beneficial expected outcome. By rewriting the upper bound as

$$P(Y = 1, Z = 1 \mid X) + P(Y = 1, Z = 0 \mid X) + P(Y = 0, Z = 0 \mid X),$$

   it can be described how the policymaker's behavior and action are made explicit, and on which groups:

   (a) The positive promotion of the innovation (that is, condition (2.4.i)) is a message for individuals under the innovation: the policymaker trusts that such a message will reinforce actions leading to experience the beneficial expected outcome. This corresponds to $P(Y = 1, Z = 1 \mid X)$. This proportion is a granted result in the sense that it does not require the actual intervention of the policymaker.

   (b) The non-negative promotion of the innovation (that is, condition (2.4.ii)) is a message for individuals under the status quo experiencing the beneficial expected outcome. This corresponds to $P(Y = 1, Z = 0 \mid X)$. This is also a granted result that does not require the intervention of the policymaker.

   (c) The question is, on which group the implementation of the innovation intends to change the expected outcome? On the proportion of individuals who are under the status quo and does not experience the beneficial expected outcome; that is, $P(Y = 0, Z = 0 \mid X)$.

Summarizing, the implementation of the innovation requires, first, making a distinction between those who do require the implementation and those who don't. Secondly, it requires to trust that those who don't require the implementation will continue to maintain their attitude, which is ensured by the positive message underlying the CPO perspective.

Now, if the policymaker acts against her perspective (that is, she advocates for the status quo rather than the innovation, which can be done by concrete political actions), then the actual proportion of individuals experiencing the beneficial expected outcome (namely, $P(Y = 1 \mid X)$) will decrease, being the worse scenario represented by those individuals under the status quo experiencing the beneficial expected outcome; that is, $P(Y = 1, Z = 0 \mid X)$. That is, a contradictory policymaker's behavior negatively affects those individuals that are under the innovation (that is, $P(Y = 1, Z = 1)$), leaving intact the behavior of the individuals under the status quo (that is, $P(Z = 0 \mid X) = P(Y = 1, Z = 0 \mid X) + P(Y = 0, Z = 0 \mid X)$). This explains why the worse scenario is represented by $P(Y = 1, Z = 0 \mid X)$.

In the standard literature of causal inference, the decision of implementing the innovation or not is based on the Average Treatment Effect (ATE), which is defined by $P(Y_1 = 1 \mid X) - P(Y_0 = 1 \mid X)$: if it is statistically positive, the decision is to implement the innovation; for details, see Angrist and Pischke (2008). The intervals (2.5) allow us to obtain a partial identification interval for the ATE, that is,

$$0 < P(Y_1 = 1 \mid X) - P(Y_0 = 1 \mid X) \leq P(Y = 1, Z = 1 \mid X) + P(Y = 0, Z = 0 \mid X).$$

Under the CPO perspective, the ATE is strictly positive with an upper bound depending on both the proportion of the individuals under the innovation who experience the beneficial expected outcome and the proportion of the individuals under the status quo who don't experience the beneficial outcome: this is precisely the group on which the policymaker should focus the intervention.

2.2.2 Ambiguous policymaker's optimistic perspective

Although optimistic, the attitude of the policymaker is not necessarily convincing, but it could be *ambiguous*. An ambiguous policymaker's optimistic (APO) perspective is characterized by the following inequalities:

$$\begin{array}{ll}\text{(i)} & P(Y_1 = 1 \mid X, Z = 1) \geq P(Y_0 = 1 \mid X, Z = 1); \\ \text{(ii)} & P(Y_1 = 1 \mid X, Z = 0) \geq P(Y_0 = 1 \mid X, Z = 0).\end{array} \tag{2.6}$$

In other word, the APO perspective consists is believing that the implementation of the innovation is at least as better as the status quo. As a matter of fact, (2.6.i) means that the actual proportion of individuals experiencing the expected beneficial outcome under the innovation will be at least equal to the corresponding proportion if those individuals were under the status quo . Similarly, (2.6.ii) means that among the individuals that are actually under the status quo, the proportion of them experiencing the expected beneficial outcome as if they were under the innovation will be at least equal to the actual proportion under the status quo.

In order to obtain all the plausible values of $P(Y_1 = 1 \mid X)$ and $P(Y_0 = 1 \mid X)$ which are compatible with the APO perspective, we combine (2.6) with the fundamental decompositions (2.1) and (2.2). The following partial identification intervals are obtained:

$$\begin{array}{ll}\text{(i)} & P(Y = 1 \mid X) \leq P(Y_1 = 1 \mid X) \leq P(Y = 1 \mid X) + P(Y = 0, Z = 0 \mid X); \\ \text{(ii)} & P(Y = 1, Z = 0 \mid X) \leq P(Y_0 = 1 \mid X) \leq P(Y = 1 \mid X).\end{array} \tag{2.7}$$

The proof is similar to the proof implying intervals (2.5). The current situation, characterized by $P(Y = 1 \mid X)$, cannot be considered as a *sure result* because it is the point common to both the decision of implementing the innovation and to maintaining the status quo. In other words, the groups characterized by $P(Y = 1, Z = 0)$ and $P(Y = 1, Z = 1)$ could change their behavior due to the APO perspective.

2.3 Conservative policymaker

When we evaluate the impact of a public policy, it is also necessary to consider a conservative policymaker. At the political level, a conservative policymaker's perspective may represent the opposition to a Government which in turn is promoting an innovation. Similarly to the optimistic policymaker perspective, we can distinguish two conservative perspectives: convincing policymaker's conservative (CPC) perspective and an ambiguous policymaker's conservative (APC) perspective.

2.3.1 Convincing policymaker's conservative perspective

A CPC perspective is defined by the following inequalities:

$$\begin{array}{ll} \text{(i)} & P(Y_1 = 1 \mid X, Z = 1) < P(Y_0 = 1 \mid X, Z = 1); \\ \text{(ii)} & P(Y_1 = 1 \mid X, Z = 0) \leq P(Y_0 = 1 \mid X, Z = 0). \end{array} \tag{2.8}$$

In other words, the CPC perspective is characterized by (i) promoting the idea that if individuals who are actually under the innovation, were under the status quo, their performance would be better; that is, the innovation harms individuals under the status quo –this corresponds to condition (2.8.i). And (ii) if individuals who are actually under the status quo were under the innovation, their performance would be at most as the performance under the status quo; that is, the status quo does not harm individuals under the innovation -this corresponds to condition (2.8.ii).

As in the previous cases, after combining (2.8) with the fundamental decompositions (2.1) and (2.2), we obtain all the plausible values of $P(Y_1 = 1 \mid X)$ and $P(Y_0 = 1 \mid X)$ that are compatible with the CPC perspective. These ones are given by the following partial identification intervals:

$$\begin{array}{ll} \text{(i)} & P(Y = 1, Z = 1 \mid X) \leq P(Y_1 = 1 \mid X) \leq P(Y = 1 \mid X); \\ \text{(ii)} & P(Y = 1 \mid X) < P(Y_0 = 1 \mid X) \leq P(Y = 1 \mid X) + P(Y = 0, Z = 1). \end{array} \tag{2.9}$$

The proof is similar to the proof implying interval (2.5). Let us comment the consequences of the intervals (2.9):

1. The interval (2.9.ii) shows that a CPC perspective will improve the current situation represented by $P(Y = 1 \mid X)$, that is, the proportion of individuals experiencing the beneficial expected outcome. By rewriting the upper bound as

$$P(Y = 1, Z = 1 \mid X) + P(Y = 1, Z = 0 \mid X) + P(Y = 0, Z = 1 \mid X),$$

   it can be described how the policymaker's behavior is made explicit and on which groups:

   (a) The positive promotion of the status quo (that is, condition (2.8.ii)) is a message for individuals under the status quo: the policymaker trusts that such a message will reinforce actions leading to experience the beneficial expected outcome. This corresponds to $P(Y = 1, Z = 0 \mid X)$. This is a granted result in the sense that the group characterized by $P(Y = 1, Z = 0 \mid X)$ does not require the intervention of the policymaker.

   (b) The non-negative promotion of the status quo (that is, condition (2.8.i)) is a message for individuals under the innovation who experience the beneficial expected outcome. This corresponds to $P(Y = 1, Z = 1 \mid X)$. This is also a granted result in the sense that the group characterized by $P(Y = 1, Z = 1 \mid X)$ that does not require the intervention of the policymaker.

   (c) The question is, on which group the implementation of the status quo intends to change the expected outcome? On the group characterized by the proportion of individuals who are under the innovation and does not experience the beneficial expected outcome; that is, $P(Y = 0, Z = 1 \mid X)$.

Summarizing, the implementation of the status quo requires, on the one hand, to make a distinction between those who do require the implementation and those who don't. On the other hand, it requires to trust that those who don't require the implementation will continue to maintain their attitude, which is ensured by the message underlying the CPC perspective.

Now, if the policymaker acts against her perspective (that is, she advocates for the innovation rather than the status quo, which can be done by actual political actions), then the actual proportion of individuals experiencing the beneficial expected outcome (namely, $P(Y = 1 \mid X)$) will decrease, being the worse scenario represented by those individuals under the innovation experiencing the beneficial expected outcome, namely $P(Y = 1, Z = 1 \mid X)$. This means that the contradictory policymaker's behavior affects negatively those individuals that are under the innovation (that is, $P(Y = 1, Z = 0)$), leaving intact the behavior of the individuals under the innovation experiencing the beneficial outcome (that is, $P(Y = 1, Z = 1 \mid X)$). This explains why the worse scenario is represented by $P(Y = 1, Z = 1 \mid X)$.

As in the previous case, the intervals (2.9) allow us to obtain a partial identification interval for the ATE, which is given by

$$-P(Y = 1, Z = 0 \mid X) - P(Y = 0, Z = 1 \mid X) \leq P(Y_1 = 1 \mid X) - P(Y_0 = 1 \mid X) < 0.$$

The ATE is not only strictly negative (that is, supporting the status quo), but also its minimum bound is given by both the proportion of the individuals under the status quo who experience the beneficial expected outcome, and the subpopulation of the individuals under the innovation who don't experience the beneficial outcome: this is precisely the group on which the policymaker should focus the intervention.

2.3.2 Ambiguous policymaker's conservative perspective

It is also possible to define an ambiguous policymaker's conservative perspective as follows:

$$\begin{array}{ll} \text{(i)} & P(Y = 1, Z = 1 \mid X) \leq P(Y_1 = 1 \mid X) \leq P(Y = 1 \mid X); \\ \text{(ii)} & P(Y = 1 \mid X) \leq P(Y_0 = 1 \mid X) \leq P(Y = 1 \mid X) + P(Y = 0, Z = 1). \end{array} \tag{2.10}$$

Comments and discussions are similar to those developed in Section 2.2.2 and therefore are left to the reader.

2.4 What if the policy decision harms all the individuals?

As it was discussed above, the scientific evidence is mainly characterized by two conditional probabilities, namely $P(Y = 1 \mid X, Z = 1)$ –the proportion of individuals experiencing the beneficial outcome when they are actually under the innovation; and $P(Y = 1 \mid X, Z = 0)$ –the proportion of individuals experiencing the beneficial outcome when they are actually under the status quo. We argue that a public policy evaluation should consider the possibility that a policy decision could harm both groups.

In order to quantify how severe the damage is, we need to write in probabilistic terms this possibility. Let us call both $P(Y_1 = 1 \mid X, Z = 1)$ and $P(Y_0 = 1 \mid X, = 0)$ *factual probabilities*, and $P(Y_1 = 1 \mid X, Z = 0)$ and $P(Y_0 = 1 \mid X, Z = 1)$ *counterfactual probabilities*. Factual probabilities represent what actually we observe, whereas counterfactual probabilities represent what would occur if an individual under innovation (resp., status quo) were under the status quo (resp., innovation). Let suppose that we are ready to accept that *factual is better than counterfactual*, which in probabilistic terms is written as follows:

$$\begin{array}{ll}\text{(i)} & P(Y_1 = 1 \mid X, Z = 1) > P(Y_0 = 1 \mid X, Z = 1); \\ \text{(ii)} & P(Y_1 = 1 \mid X, Z = 0) < P(Y_0 = 1 \mid X, Z = 0).\end{array} \tag{2.11}$$

That is, (i) if individuals who are actually under the innovation, were under the status quo, their performance would be worse and, consequently, the innovation is the ideal scenario for those individuals –this corresponds to condition (2.11.i). And (ii) if individuals who are actually under the status quo were under the innovation, their performance would be worse and, consequently, the status quo is the ideal scenario for those individuals -this corresponds to condition (2.11.ii).

It can be noticed that if policy decision is to implement the innovation, the individuals under the status quo will be negatively affected; and if the policy decision is to maintain the status quo, the individuals under the innovation will be negatively affected. Thus, whatever the policy decision is, one or the other group is harmed. It is possible to quantify the damage. As a matter of fact, by combining (2.11) with the fundamental decompositions (2.1) and (2.2), we get the following intervals of partial identification:

$$\begin{array}{ll}\text{(i)} & P(Y = 1, Z = 1 \mid X) \leq P(Y_1 = 1 \mid X) < P(Y = 1 \mid X); \\ \text{(ii)} & P(Y = 1, Z = 0 \mid X) \leq P(Y_0 = 1 \mid X) < P(Y = 1 \mid X).\end{array} \tag{2.12}$$

For a proof, see Appendix A.2. It can be concluded that, whatever the policy decision is, the current situation represented by $P(Y = 1 \mid X)$ will be worse. How bad the situation could be? If the decision is to implement the innovation, the proportion of individuals experiencing the beneficial expected outcome could decrease up to $P(Y = 1, Z = 1 \mid X)$, that is, the proportion of individuals under the innovation and experiencing the beneficial outcome. If the decision is to maintain the status quo, the proportion of individuals experiencing the beneficial expected outcome could decrease up to $P(Y = 1, Z = 0 \mid X)$, that is, the proportion of individuals under the status quo and experiencing the beneficial outcome. In other words, the worst scenarios are actually represented by the proportion of individuals experiencing the beneficial outcome in each group. Thus, in a first approach, taking into account that the greatest welfare of society is pursued, the policy decision will depend on a comparison between the lower bounds of both intervals. Consequently, a treatment assignment rule will run as follows:

1. If $P(Y = 1, Z = 0 \mid X) < P(Y = 1, Z = 1 \mid X)$, then the decision is to implement the innovation.

2. If $P(Y = 1, Z = 1 \mid X) < P(Y = 1, Z = 0 \mid X)$, then the decision is to maintain the status quo.

Let us remark that the decision is unclear if the ATE is used as a criterion because the corresponding interval contains 0:

$$-P(Y = 1, Z = 0 \mid X) < P(Y_1 = 1 \mid X) - P(Y_1 = 0 \mid X) < P(Y = 1, Z = 1 \mid X).$$

It is, therefore, desirable to look for a criterion in order to assess the previous treatment assignment rule and eventually to improve it. We postpone the discussion to Section 3.

Remark 2.1 It is possible to characterize the scenario *factual is at least as better as counterfactual*. It is enough to rewrite (2.11) as

$$\begin{array}{rl} \text{(i)} & P(Y_1 = 1 \mid X, Z = 1) \geq P(Y_0 = 1 \mid X, Z = 1); \\ \text{(ii)} & P(Y_1 = 1 \mid X, Z = 0) \leq P(Y_0 = 1 \mid X, Z = 0). \end{array} \tag{2.13}$$

We leave to the reader to complete details and discussions about this scenario.

2.5 What if the policy decision benefits all individuals?

Finally, let us consider the possibility that whatever the policy decision is, it benefits both groups – the group under the innovation and the group under the status quo. This means to assume that *the counterfactual is better than the factual*, which in probabilistic terms is written as follows:

$$\begin{array}{rl} \text{(i)} & P(Y_1 = 1 \mid X, Z = 1) < P(Y_0 = 1 \mid X, Z = 1); \\ \text{(ii)} & P(Y_1 = 1 \mid X, Z = 0) > P(Y_0 = 1 \mid X, Z = 0). \end{array} \tag{2.14}$$

Thus, (i) if individuals who are actually under the innovation, were under the status quo, their performance would be better and, consequently, the status quo is the ideal scenario for those individuals –this corresponds to condition (2.14.i). And (ii) if individuals who are actually under the status quo were under the innovation, their performance would be better and, consequently, the innovation is the ideal scenario for those individuals -this corresponds to condition (2.14.ii).

In this case, whatever the policy decision, both groups benefit positively. The corresponding social benefit can be quantified. It is enough to combine (2.14) with the fundamental decompositions (2.1) and (2.2). We obtain the following intervals of partial identification:

$$\begin{array}{rl} \text{(i)} & P(Y = 1 \mid X) < P(Y_1 = 1 \mid X) \leq P(Y = 1 \mid X) + P(Y = 0, Z = 0 \mid X); \\ \text{(ii)} & P(Y = 1 \mid X) < P(Y_0 = 1 \mid X) < P(Y = 1 \mid X) + P(Y = 0, Z = 1 \mid X). \end{array} \tag{2.15}$$

The proof is similar to that underlying intervals (2.12). It can be concluded that, whatever the policy decision is, the current situation represented by $P(Y = 1 \mid X)$ will improve. How good the situation could be? If the decision is to implement the innovation, the proportion of individuals experiencing the beneficial expected outcome could increase up to $P(Y = 1 \mid X) + P(Y = 0, Z = 0 \mid X)$ and, consequently, the focus of intervention is on the individuals under the status quo who don't experience the beneficial outcome. If the decision is to maintain the status quo, the proportion of individuals experiencing the beneficial expected outcome could increase up to $P(Y = 1 \mid X) + P(Y = 0, Z = 1 \mid X)$ and, consequently, the focus of intervention is on the individuals under the innovation who don't experience the beneficial outcome. Thus, in a first approach, taking into account that the greatest welfare of society is pursued, the policy decision will depend on a comparison between the upper bounds of both intervals. Consequently, a treatment assignment rule will run as follows:

1. If $P(Y = 0, Z = 0 \mid X) > P(Y = 0, Z = 1 \mid X)$, then the decision is to implement the innovation.

2. If $P(Y=0, Z=0 \mid X) < P(Y=0, Z=1 \mid X)$, then the decision is to maintain the status quo.

As in the previous case, the decision is unclear if the ATE is used as a criterion: its partial identification interval contains 0, namely

$$-P(Y=0, Z=1 \mid X) < P(Y_1=1 \mid X) - P(Y_1=0 \mid X) < P(Y=0, Z=0 \mid X).$$

Similarly to the previous case, it is desirable to look for an additional criterion allowing to eventually improve this treatment assignment rule.

2.6 Overview of the solutions

Let us summarize the four solutions CPO, CPC, factual better than counterfacual (FBC) and counterfactual better than factual (CBF), discussed in Sections 2.2.1, 2.3.1, 2.4 and 2.5, respectively. These results are summarized as follows:

	$P(Y_1=1 \mid X)$		$P(Y_0=1 \mid X)$	
Scenario	**Lower bound**	**Upper bound**	**Lower bound**	Upper bound
CPO	$P(Y=1 \mid X)$	$P(Y=1 \mid X) + P(Y=0, Z=0 \mid X)$	$P(Y=1, Z=0 \mid X)$	$P(Y=1 \mid X)$
CPC	$P(Y=1, Z=1 \mid X)$	$P(Y=1 \mid X)$	$P(Y=1 \mid X)$	$P(Y=1 \mid X) + P(Y=0, Z=1 \mid X)$
FBC	$P(Y=1, Z=1 \mid X)$	$P(Y=1 \mid X)$	$P(Y=1, Z=0 \mid X)$	$P(Y=1 \mid X)$
CBF	$P(Y=1 \mid X)$	$P(Y=1 \mid X) + P(Y=0, Z=0 \mid X)$	$P(Y=1 \mid X)$	$P(Y=1 \mid X) + P(Y=0, Z=1 \mid X)$

Regarding CPO and CPC, it should be confessed that the respective policy recommendations are *verbatim* a narrative repetition of the corresponding perspective. Thus, the policy recommendation under the CPO perspective is to implement the innovation, whereas under the CPC perspective, is to maintain the status quo. Then, what the information provided by the partial identification interval is? It provides the uncertainty inherent to each perspective: such an uncertainty is estimated using the scientific evidence summarized by both $P(Y=1 \mid X, Z=1)$ and $P(Y=1 \mid X, Z=0)$.

Similarly, the FBC perspective always harms both individuals under the innovation and individuals under the status quo. The policy recommendation is still a *verbatim* narrative of this perspective. However, these intervals show how severe is the damage in each group. Similar comments can be done for the CBF perspective.

An illustration of these considerations is developed in Appendix B.

2.7 A plausible policy interpretation of strong ignorability?

The four solutions, namely CPO CPC, FBC and CBF perspectives, can be compared between them. Such a comparison leads us to answer one of the questions proposed in Section 2.1, namely the eventual policy interpretation of the strong ignorability condition (2.3).

Different patterns can be obtained depending on the order we impose between $P(Y=1 \mid X)$, $P(Y=0 \mid X)$, $P(Z=1 \mid X)$ and $P(Z=0 \mid X)$. It is possible to distinguish 16 different patterns. We focus our attention on two of them: $P(Z = 0 \mid X) < P(Y = 0 \mid X) < P(Y = 1 \mid X) < P(Z = 1 \mid X)$ in Figure 1, and $P(Y = 1 \mid X) < P(Z = 0 \mid X) < P(Z = 1 \mid X) < P(Y = 1 \mid X)$ in Figure 2 (for simplicity, in the figures we omit to condition on X).

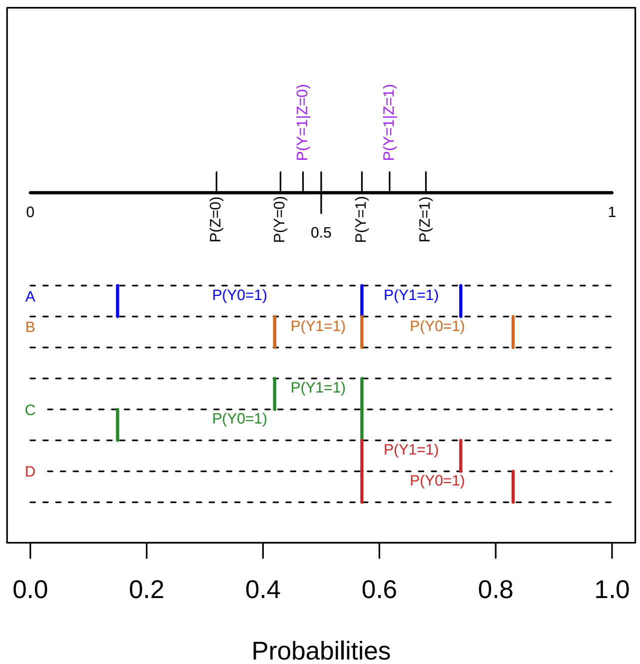

Figure 1: Overview of 4 solutions assuming that $P(Z = 0 \mid X) < P(Y = 0 \mid X) < P(Y = 1 \mid X) < P(Z = 1 \mid X)$

In general terms, the decision of implementing the innovation or not depends on the pursuit of the social welfare. This means to improve the current situation represented by $P(Y = 1 \mid X)$. Under the ignorability condition we have that

$$P(Y_1 = 1 \mid X) = P(Y_1 = 1 \mid X, Z = 1) = P(Y = 1 \mid X, Z = 1),$$
$$P(Y_0 = 1 \mid X) = P(Y_0 = 1 \mid X, Z = 0) = P(Y = 1 \mid X, Z = 0).$$

Now, for the case depicted in Figure 1, it follows that

$$P(Y = 1 \mid X, Z = 0) < P(Y = 1 \mid X) < P(Y = 1 \mid X, Z = 1). \tag{2.16}$$

Therefore, the solution obtained with the strong ignorabilty condition is compatible with the CPO perspective. Moreoever, it can be proved that (2.16) is equivalent to

$$cov\left(\mathbb{1}_{\{Y=1\}}, \mathbb{1}_{\{Z=1\}} \mid X\right) > 0.$$

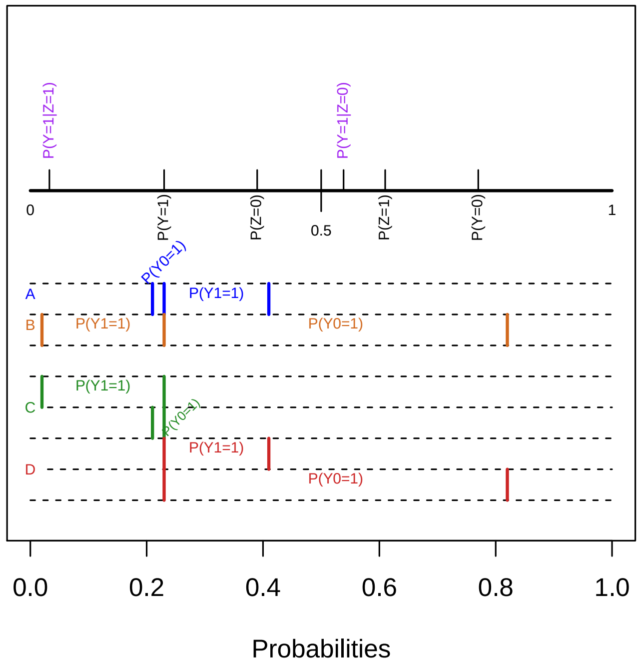

Figure 2: Overview of 4 solutions assuming that $P(Y = 1 \mid X) < P(Z = 0 \mid X) < P(Z = 1 \mid X) < P(Y = 1 \mid X)$

That is, the innovation is positively correlated with the expected outcome: this explains why the strong ignorability condition is compatible with the CPO perspective.

However, the situation is different if we consider the case depicted in Figure 2. This case implies that

$$P(Y = 1 \mid X, Z = 1) < P(Y = 1 \mid X) < P(Y = 1 \mid X, Z = 0). \tag{2.17}$$

Therefore, the solution obtained with the strong ignorabilty condition is compatible with the CPC perspective. Condition (2.17) is equivalent to

$$cov\left(\mathbb{1}_{\{Y=1\}}, \mathbb{1}_{\{Z=0\}} \mid X\right) > 0.$$

That is, the status quo is positively correlated with the expected outcome; this explains why the strong ignorability condition is compatible with the CPC perspective.

Summarizing, both $P(Y = 1 \mid X, Z = 1)$ and $P(Y = 1 \mid X, Z = 0)$ must belong to some partial identification interval and, therefore, the strong ignorability have not an “intrinsic” policy interpretation. The previous cases actually show that the policy interpretation of the solution obtained under strong ignorability depends on the association between outcome and treatment, as well as on the order between $P(Y = 1 \mid X), P(Y = 0 \mid X), P(Z = 1 \mid X)$ and $P(Z = 0 \mid X)$.

3 A Decision Problem

Let us come back to the FBC perspective. The policy decision will always harm those under the innovation and those under the status quo. A first solution was based on the comparison of the lower bounds of each partial identification intervals, namely

1. If $P(Y = 1, Z = 0 \mid X) < P(Y = 1, Z = 1 \mid X)$, then the decision is to implement the innovation.
2. If $P(Y = 1, Z = 1 \mid X) < P(Y = 1, Z = 0 \mid X)$, then the decision is to maintain the status quo.

This solution is extreme in the sense that all individuals, whether under innovation or the status quo, will be harmed by the policy decision. Consequently, we can look for a policy decision based on fractional treatment assignment; that is, deciding to which proportion of individuals to apply the innovation and to which one, the status quo. This problem can be solved by applying the decision theory due to Wald (1950) and Savage (1951).

3.1 Basic framework of decision theory

We follow Manski (2005b), Manski (2007a) and Manski (2007b) for a presentation of the decision theory. We consider two treatment, $t = 0$ (status quo) and $t = 1$ (innovation), so $T = \{0, 1\}$. The policymaker's problem is to choose one of the treatment. Each member j of the population J has a response function $Y_j : T \longrightarrow \mathcal{Y}$ mapping treatments $t \in T$ into outcomes $Y_j(t) \in \mathcal{Y}$. The policymaker, who must assign a treatment to each member of J, observe covariates $X_j \in \mathcal{X}$ for each individual j: the policymaker can differentiate individuals with different values of X, but she cannot distinguish between individuals with the same observed covariates. This is due to the fact that covariates induce a partition (or equivalent class) in the set of individuals. Thus, $X : J \longrightarrow \mathcal{X}$ is the random variable mapping individuals into their covariates. Given that the policymaker observes the covariates of all the individuals, she knows the covariate distribution $P(X = \xi) > 0$ for $\xi \in \mathcal{X}$. The basic steps of the decision process are the following:

A policymaker must choose among a set of feasible actions: A feasible treatment rule assigns either (i) all individuals with the same observed covariates to one treatment, or (ii) randomly allocates such individuals across the treatments. Let Δ denote the space of functions that map $T \times \mathcal{X}$ into the unit interval and that satisfy the adding-up conditions

$$\delta(\cdot, \cdot) \in \Delta \Longrightarrow \delta(0, \xi) + \delta(1, \xi) = 1 \quad \forall \xi \in \mathcal{X}.$$

The feasible treatment rules are the elements of Δ: these elements satisfy conditions (i) and (ii) mentioned above.

A policymaker wants to choose a feasible treatment rule that maximizes population mean welfare: The welfare from assignment treatment t to individual j is $u_j(t) \doteq u[Y_j(t), t, X_j]$ (here $\doteq$ denotes "defined as") , where $u : \mathcal{Y} \times T \times \mathcal{X} \longrightarrow \mathbb{R}$. For each treatment rule δ, the mean welfare that would be realized with choice of rule δ is

$$U(\delta, P) \doteq \sum_{\xi \in \mathcal{X}} P(X = \xi) \; \{ \, \delta(0, \xi) \, E[u(0) \mid X = \xi] + \delta(1, \xi) \, E[u(1) \mid X = \xi] \, \} \, .$$

This follows from $E[u] = E[E(u \mid X)]$ after interpreting $\delta(t, \xi)$ as a conditional probability. Here $U(\delta, P)$ is the social welfare when treatment rule δ is applied to a population with distribution of treatment response P. The policymaker wants to solve the problem

$$\max_{\delta \in \Delta} U(\delta, P). \tag{3.1}$$

The maximum is achieved by a rule that allocates all individuals with covariates ξ to a treatment solving the problem

$$\max_{t \in \{0,1\}} E[u(t) \mid X = \xi].$$

Consequently, the population welfare achieved by an optimal rules is

$$U^*(P) \doteq \sum_{\xi \in \mathcal{X}} P(X = \xi) \; \left\{ \max_{t \in \{0,1\}} E[u(t) \mid X = \xi] \right\} . \tag{3.2}$$

Criteria for planning under uncertainty: A policymaker who knows the treatment-response distribution $P(Y(t) \mid X)$, with $t \in \{0, 1\}$, can choose an optimal rule. However, in the case addressed in this paper, the policymaker does not know the response distributions, but observes a study population in which treatments have been selected and outcomes realized. Whatever the empirical evidence and maintained assumptions may be, let Γ index the set of feasible *state of nature*. Thus, $\{P_\gamma : \gamma \in \Gamma\}$ is the set of values that the policymaker deems feasible for the distribution of treatment response. Note that this set is related to the partial identification intervals. It was mentioned above that the policymaker wants to choose a feasible treatment rule that maximize the population mean welfare, which essentially reduces to solve problem (3.1). However, when the state of nature is unknown, the policymaker faces the ambiguity of ranking treatment rules and therefore the policymaker does not know which rule is optimal. As a matter of fact, let $\delta, \delta' \in \Delta$ be any two feasible treatment rules. Rule δ dominates rule δ' if $U(\delta, P_\gamma) \geq U(\delta', P_\gamma)$ for all $\gamma \in \Gamma$ and $U(\delta, P_\gamma) > U(\delta', P_\gamma)$ for some $\gamma \in \Gamma$. It is clear that the policymaker should not choose dominated actions. It is accordingly necessary to reduce the choice to undominated treatment rules, that is, such that $U(\delta, P_\gamma) > U(\delta', P_\gamma)$ for some $\gamma \in \Gamma$ and $U(\delta, P_\gamma) < U(\delta', P_\gamma)$ for other $\gamma \in \Gamma$. It is difficult to determine which treatment rules are undominated. Hence the usual practice is to perform the maximization over the full set Δ of feasible treatment rules. In the case where the state of nature is unknown, this maximization problem requires a criterion to choose an admissible tretement rule:

1. Wald (1950) focused on the *maxmin* criterion, which means that the policymaker would solve

$$\max_{\delta \in \Delta} \min_{\gamma \in \Gamma} U(\delta, P_\gamma).$$

This criterion has a normative foundation in competitive games. In a competitive game, the policymaker choose a treatment rule from Δ. Then a probability treatment from Γ is chosen by an opponent whose objective is to minimize the realize outcome (Manski, 2007a).

2. Savage (1951) proposes the *minimax regret* criterion, which means that the policymaker would solve

$$\min_{\delta \in \Delta} \max_{\gamma \in \Gamma} \left\{ U^*(P_\gamma) - U(\delta, P_\gamma) \right\}, \tag{3.3}$$

where $U^*(P_\gamma)$ is the optimal population mean welfare that would be achievable if it were known that $P = P_\gamma$ in (3.2). This criterion suggests that the policymaker chooses a treatment rule that minimizes the maximum loss of welfare that results from not knowing the probability of the treatment response. As a matter of fact, suppose that the policymaker choose some treatment δ and that the probability of treatment response is some P_γ. The loss in potential welfare resulting from choice of this treatment rule is

$$\max_{\delta' \in \Delta} U(\delta', P_\gamma) - U(\delta, P_\gamma) = U^*(P_\gamma) - U(\delta, P_\gamma).$$

This loss, called *regret*, is the nonnegative difference between the maximum welfare attainable with P_γ and the welfare attained by choice of δ. The true probability of treatment response is unknown, but we can compute the maximum regret of action δ over Γ. This is

$$\max_{\gamma \in \Gamma} \left\{ U^*(P_\gamma) - U(\delta, P_\gamma) \right\}.$$

A minimax-regret rule choose a treatment rule that minimize the previous function, which means to solve (3.3). For details, see Manski (2007a).

3.2 Minimax-regret for the FBC perspective

Using the notation of Section 2, outcomes Y_0 and Y_1 correspond to $Y(0)$ and $Y(1)$, respectively. These outcomes are viewed as the beneficial expected outcomes, so we define the welfare function $u(t)$ with $t \in \{0, 1\}$ as $u(t) \doteq Y(t)$. Let $\delta \doteq \delta(1, \xi)$. Following Manski (2007b)'s arguments, it can be verified that the minimax-regret rule solves the problem

$$\min_{\delta \in [0,1]} \max \left\{ (1 - \delta) \left[(e_{1\xi} p_{1\xi} + u_{1\xi} p_{0\xi}) - (e_{0\xi} p_{0\xi} + l_{0\xi} p_{1\xi}) \right], \delta \left[(e_{0\xi} p_{0\xi} + u_{0\xi} p_{1\xi}) - (e_{1\xi} p_{1\xi} + l_{1\xi} p_{0\xi}) \right] \right\}$$

where $e_{0\xi} = P(Y = 1 \mid X = \xi, Z = 0)$, $e_{1\xi} = P(Y = 1 \mid X = \xi, Z = 1)$, $p_{0\xi} = P(Z = 0 \mid X = \xi)$ and $p_{1\xi} = P(Z = 1 \mid X = \xi)$. Moreover, $l_{1\xi}$ and $u_{1\xi}$ are the lower and upper bound of interval (2.12.i); and $l_{0\xi}$ and $u_{0\xi}$ are the lower and upper bound of interval (2.12.ii). The solution of this problem is given by (see Manski, 2007b, Proposition 1):

1. The innovation is applied to all the population characterized by ξ (that is, $\delta = 1$) if

$$P(Y=1, Z=0 \mid \xi) < \\ P(Y=1,Z=1 \mid \xi) + [P(Y=1,Z=1 \mid \xi)P(Z=0 \mid \xi) - P(Y=1 \mid \xi)P(Z=1 \mid \xi)]\,.$$

   In this case, the innovation dominates the status quo.

2. The status quo is applied to all the population characterized by ξ (that is, $\delta = 0$) if

$$P(Y=1, Z=1 \mid \xi) < \\ P(Y=1,Z=0 \mid \xi) + [P(Y=1,Z=0 \mid \xi)P(Z=1 \mid \xi) - P(Y=1 \mid \xi)P(Z=0 \mid \xi)]\,.$$

   In this case, the status quo dominates the innovation.

3. The innovation is applied to a δ fraction of the population characterized by ξ and the status quo to a $(1-\delta)$ fraction of the same population, where δ is given by

$$\delta = \frac{[P(Y=1,Z=1 \mid \xi) - P(Y=1,Z=0 \mid \xi)] + [P(Y=1 \mid \xi)P(Z=0 \mid \xi) - P(Y=1,Z=0 \mid \xi)P(Z=1 \mid \xi)]}{P(Y=1 \mid \xi) - P(Y=1,Z=1 \mid \xi)P(Z=0 \mid \xi) - P(Y=1,Z=0 \mid \xi)P(Z=1 \mid \xi)}$$

   In this case, neither treatment (innovation or status quo) dominates.

According to the first treatment assignment rule, the innovation is applied to all the population characterized by $X = \xi$ if

$$P(Y=1,Z=0 \mid \xi) < P(Y=1,Z=1 \mid \xi). \tag{3.4}$$

However, the minimax-regret solution offer a different criterion: the previous inequality needs to be complemented by a second term which in turn is lower than the dependence between Y and Z, namely

$$P(Y=1,Z=1 \mid \xi)P(Z=0 \mid \xi) - P(Y=1 \mid \xi)P(Z=1 \mid \xi) < \\ P(Y=1,Z=1 \mid \xi) - P(Y=1 \mid \xi)P(Z=1 \mid \xi) = cov\left(\mathbb{1}_{\{Y=1\}}, \mathbb{1}_{\{Z=1\}} \mid \xi\right).$$

This means that such a second term could be negative and therefore the criterion to choose $\delta = 1$ is more severe than criterion (3.4). Similar comments can be done for the implementation of the status quo to all the population characterized by $X = \xi$.

3.3 Ethical issues of fractional treatment rules

It seems relevant to point out the ethical issues raised by fractional treatment rules. Typically, the standard practice in social planning is to consider singleton treatment rules, that is, to assign a unique treatment to all the population. However, as Manski (2007a) argues, "implementation of a fractional minimax-regret rule is appealing because it enables society to diversify a risk that is privately indivisible. An individual cannot diversify; a person receives either [innovation] or [status quo]. Yet society can diversify by having positive fractions of the population receive each treatment" (p. 233); see also Manski (2005a).

Fractional treatment rules ensure, therefore, a diversification of the risk at the society level. This aspect is pertinent when the policy decision will always harm all the population. The singleton treatment rule (based on criterion (3.4)) does not diversify the risk in the sense that all individuals under status quo are more harmed than individuals under innovation. One might ask why not "equally" diversify the damage between both groups: the fractional treatment rule is the way to diversify the risk. Let us mention that this type of rules are consistent with the principle *equal treatment of equals* in the *ex ante* sense "that observationally identical people have the same probability of receiving a particular treatment"; but fractional rules "violate the principle in the *ex post* sense that observationally identical persons ultimately receive different treatments" (Manski, 2007a, p. 233-234). This type of considerations is widely discussed in the prospect theory literature, particularly in the context of the *Asian disease-type problems*; for details, see Li and Xie (2006), Jullien (2013), Manski (2013b, Chapter 4) and Dreisbach and Guevara (2017).

4 The role of science in policy

Pielke Jr (2007) describes the Honest Broker of Policy Alternative in the following terms:

> The defining characteristic of the honest broker of policy alternatives is an effort to expand (or at least clarify) the scope of choice for decision-making in a way that allows for the decision-maker to reduce choice based on his or her own preferences and values (pp. 2-3).

The Honest Broker perspective is considered as the most relevant and appropriate scientific role in a renewed relationship between science, policy and politics (Turnhout, Stuiver, Klostermann, Harms, & Leeuwis, 2013; Pedersen, 2014; Stone, 2017; Van Woensel, 2020). Its difficulties has been also analyzed, proposing strategies to overcome them (e.g., Andereggen, Vischer, & Boutellier, 2012).

4.1 The impact of the Honest Broker role at the methodological level

In a different vein, we develop a methodological strategy leading to make explicit how to be an Honest Broker in the context of public policy evaluations. We realize that this is possible when the identification problems and their eventual solution are made explicit. It is important to note that there is a temptation to build models in such a way that the identification problem is hidden or solved without justifying implicit assumptions:

> One might regard problems of identifiability as a necessary part of the specification problem. We would consider such a classification acceptable, *provided the temptation to specify models in such a way as to produce identifiability of relevant characteristics is resisted. Scientific honesty* demands that the specification of a model be based on prior knowledge of the phenomenon studied and possibly on criteria of simplicity, but *not on the desire for identifiability of characteristics in which the researcher happens to be interested* (Koopmans & Reiersøl, 1950, pp.169-170; the italics are own).

In passing, it should be said that most of the methodological statistical contributions seem to be motivated by solving identification problems through strong assumptions (Tamer, 2010). Moreover, some Bayesian statisticians claim that “unidentifiability causes no real difficulty in the Bayesian approach” (Lindley, 1972), a statement which is conceptually incorrect; for a discussion, see San Martín (2018) and the references therein.

4.2 Implications of an Honest Broker strategy

The strategy developed in this paper shows that an identification problem in the evaluation of public policies can be solved in different ways. Each way is associated with a specific belief of a policymaker which in turn concerns benefits and harms of a policy innovation as well as on the statu quo: when these beliefs are combined with scientific evidence, the uncertainty inherent in that belief arises. We emphasize that it is a matter of *beliefs* because these ones cannot be empirically tested. When the different ways of solving the identification problem were analyzed, it became clear that a political intervention is not reduced to *actions* but also to *speak*. Thus, for example, we saw that in the CPO and CPC perspectives, the policy maker is confident that a *convincing speech* will reinforce the actions of a group of individuals who already experience the expected benefit outcome; and additionally we identify the group in which the *actions* (interventions) should be carried out. In this regard, it is pertinent to refer to Hannah Arendt and her thoughts on political speaking:

> Although I am of the opinion that one must not appear and act in public self-consciously, still I know that in every action the person is expressed as in no other human activity. *Speaking is also a form of action*. That is one venture (pp. 37-38 Arendt, 2013);

for a more detailed digression, see Arendt (1958). This type of speaking could be related to the concept of *performative sentences* introduced by Austin (1962). The name is derived from ’perform’, the usual verb with the noun ’action’: it indicates that the issuing of the utterance is the performing of an action –it is not normally thought of as just saying something. In this respect, see Hasselbladh and Bejerot (2017).

The analysis of the different ways in which the identification problem can be solved also allowed us to make explicit that a policy decision can harm a population or benefit it. We argue that a public policy evaluation should consider these possibilities. In doing so, fractional assignment treatment rules appear as a viable alternative, although they face ethical problems, particularly those related to the problem of Asian disease. At the level of policy evaluation, these considerations are relevant because fractional rules are consistent with the principle that *all observationally identical people have the same probability of receiving a particular treatment* in an *ex-ante* sense. As a matter of fact, the design and implementation of policies is performed in the *ex-ante* sense –for instance, progressive implementations of public policies or random choice of groups where an innovation is implemented.

4.3 The challenge of the proposed strategy

The strategy of public policy evaluation proposed in this paper should be considered not as a normative solution, but as a prescriptive solution: to know the implications of decisions after choosing specific beliefs. It may actually be considered as a qualitative analysis tool, which could be useful to a team of policy advisors. In this sense, Lichtenberg's words are a good aphoristic summary of the proposed strategy:

> When we teach men *how* to think, instead of *what* to think, then we also reduce misunderstanding. It is a sort of baptism into the mysteries of humanity (quoted by Lorenzen, 1987, p.3).

One of the main methodological challenges of this strategy is to relate different ways of solving an identification problem with political behaviors and political values. This requires not only transdisciplinary teams, but also (fluid) communication with politicians and policy makers. For an additional example, see the Case Study 2 in Appendix C.

This strategy intends to be *a* methodological realization of the Honest Broker of policy alternatives role. It makes explicit that scientific knowledge is mediated by subjective choices, but also shows what (could) mean(s) a rigorous scientific knowledge. This perspective collaborates to the necessity of considering all the sides of a public policy problem, even if those sides are contradictory. This seems to be the way to serve politics and policy:

> The statesman is surrounding, encircled by an army of experts. So that now the question of action lies between the statesman and experts. The statesman has to make the final decision. He can hardly do that realistically, since he can't know everything himself. He must take the advice of experts, indeed of experts who in principle always have to contradict each other. Isn't that so? Every reasonable statesman summons experts with opposing points of view. Because he has to see the matter from all sides. That's true, isn't it? He has to judge between them. And this judgement is a highly mysterious process –in which, then, common sense is made manifest (pp.35-36 Arendt, 2013).

Appendix

A Proofs

A.1 Proof of inequalities (2.5)

Using (2.4.ii), (2.1) implies that

$$P(Y_1 = 1 \mid X, Z = 1)P(Z = 1 \mid X) + P(Y_0 = 1 \mid X, Z = 0)P(Z = 0 \mid X) \leq P(Y_1 = 1 \mid X).$$

However, $P(Y_1 = 1 \mid X, Z = 1) = P(Y = 1 \mid X, Z = 1)$ and $P(Y_0 = 1 \mid X, Z = 0) = P(Y = 1 \mid X, Z = 0)$. It follows that $P(Y = 1 \mid X) \leq P(Y_1 = 1 \mid X)$, namely the lower bound of (2.5.i). On the other hand, taking into account that $P(Y_1 = 1 \mid X, Z = 0) \leq 1$, (2.1) implies that

$$\begin{aligned} P(Y_1 = 1 \mid X) &\leq P(Y_1 = 1 \mid X, Z = 1)P(Z = 1 \mid X) + P(Z = 0 \mid X) \\ &= P(Y = 1 \mid X, Z = 1)P(Z = 1 \mid X) + P(Y = 1, Z = 0 \mid X) + P(Y = 0, Z = 0 \mid X) \\ &= P(Y = 1 \mid X) + P(Y = 0, Z = 0 \mid X). \end{aligned}$$

This proves the upper bound of (2.5.i).

Similarly, using (2.4.i), (2.2) implies

$$
\begin{aligned}
P(Y_0 = 1 \mid X) &< P(Y_1 = 1 \mid X, Z = 1)P(Z = 1 \mid X) + P(Y_0 = 1 \mid X, Z = 0)P(Z = 0 \mid X) \\
&= P(Y = 1 \mid X, Z = 1)P(Z = 1 \mid X) + P(Y = 1 \mid X, Z = 0)P(Z = 0 \mid X) \\
&= P(Y = 1 \mid X).
\end{aligned}
$$

This proves the upper bound of (2.5.ii). On the other hand, taking into account that $P(Y_1 = 1 \mid X, Z = 1) \geq 0$, (2.2) implies that

$$
\begin{aligned}
P(Y = 1, Z = 0 \mid X) &= P(Y = 1 \mid Z = 0, X)P(Z = 0 \mid X) \\
&= P(Y_0 = 1 \mid Z = 0, X)P(Z = 0 \mid X) \\
&\leq P(Y_0 = 1 \mid X);
\end{aligned}
$$

that is, the lower bound of (2.5.ii).

A.2 Proof of inequalities (2.12)

Using (2.11.ii), (2.1) implies that

$$
\begin{aligned}
P(Y_1 = 1 \mid X) &< P(Y_1 = 1 \mid X, Z = 1)P(Z = 1 \mid X) + P(Y_0 = 1 \mid X, Z = 0)P(Z = 0 \mid X) \\
&= P(Y = 1 \mid X).
\end{aligned}
$$

On the other hand, $P(Y_1 = 1 \mid X, Z = 0) \in [0, 1]$, and therefore (2.1) implies that

$$
P(Y = 1, Z = 1 \mid X) = P(Y_1 = 1 \mid X, Z = 1)P(Z = 1 \mid X) \leq P(Y_1 = 1 \mid X).
$$

Thus, (2.12.i) is obtained. Using a similar argument, we obtain (2.12.ii).

A.3 Proof of (2.16)

First, note that $P(Y = 1 \mid X, Z = 1) > P(Y = 1 \mid X)$ is equivalent to $P(Y = 1 \mid X, Z = 1) > P(Y = 1 \mid X, Z = 0)$. Secondly,

$$
\begin{aligned}
cov\left(\mathbb{1}_{\{Y=1\}}, \mathbb{1}_{\{Z=1\}} \mid X\right) &= E\left(\mathbb{1}_{\{Y=1\}}\mathbb{1}_{\{Z=1\}} \mid X\right) - E\left(\mathbb{1}_{\{Y=1\}} \mid X\right) E\left(\mathbb{1}_{\{Z=1\}} \mid X\right) \\
&= P(Y = 1, Z = 1 \mid X) - P(Y = 1 \mid X)P(Z = 1 \mid X).
\end{aligned}
$$

In passing, let us mention that this type of equality provides the correct way to define correlations between categorial variables.

B Case study 1

In this Appendix, we report a case study where the approach developed in the main text was applied to a Chilean public policy problem.

B.1 The Chilean National System of Quality Assurance of Education

After five years of political discussion, the Chilean government approved in 2011 the Law 20,529 on the *National System of Quality Assurance of Education*, positioning it as the most significant reform of the Chilean educational system. Its main objective is to make a systemic reform to improve and achieve a continuous improvement of students learning. According to the Law, the primary requirement to accomplish this is to increase the educational capabilities of principals and teachers (Law 20.529, Art. 2). To meet its objectives, one of the pillars of the reform of the educational system consisted of the generation of new educational institutions. One of them is the National Agency of Quality of Education; it is in charge of implementing a centralized monitoring system of the performance of schools.

This monitoring system implies to evaluate the achievement of the learning standards as well as to promote standards on the performance of both teachers and principals. The essential part of this process is the School Performance Categorization that leads to classify schools in four performance levels (high, medium, medium-low and insufficient); each category is derived from an indicator that incorporates four dimensions of quality (Law 20.529, Art. 18): the first of these comprises learning standards (67%), which indicate the level of learning achieved by students. The second includes progress measures (3.3%), which indicate an improvement in results over time. The third area is SIMCE scores (3.3%), where the SIMCE is a national standardized test in Mathematics and Language, plus other disciplines when correspond ; for details on the SIMCE test, see Meckes and Carrasco (2010) and Manzi and Preiss (2013). The final dimension is related to Indicators of Personal and Social Development (26.4%). It should be mentioned that if a school classified as insufficient does not improve its performance level after four years, it risks to be closed (Law 20.529 Art. 11).

B.1.1 Integrated leadership model in the Chilean policy context

According to the Law 20,529, the improvement of students learning requires to increase the educational capabilities of principals and teachers. The Law states that the main function of the principal is to lead the Institutional Educational Project; this implies to manage pedagogical, administrative and financial aspects (Art. 5). From an educational theoretical perspective, this requires to disseminate and apply a series of initiatives to improve the school leadership. The Chilean educational system has been implementing this on different actions across the time as the definition of Indicative Performance Standards for Educational Establishments and their Sustainers (in what follows, EID) (MINEDUC, 2005). Based on these initial advances, in 2015 the Ministry of Education proposed a final version of the standard s, establishing thus the Framework for Good School Leadership (MBD) (MINEDUC, 2014b; MINEDUC, 2015). These standards were defined using the first version of the framework (MINEDUC, 2005), the EID and international literature on school leadership, international empirical research, discussion with civil society, among others type of sources of information.

Currently, the EID and the MBD constitute the general framework considered by the National Agency of Quality of Education to improve the capabilities of the members of a school community. The specific practices determined in the EID are directly related to the school leadership literature as well as to the school leadership dimensions as established by the MBD:

1. Develop and implement a shared vision on the mission and goals of the school.
2. To improve and empowered teachers' and other staff members' abilities and leadership.
3. Coordinating, monitoring, and evaluating curriculum, instruction and assessment.
4. Promote inclusive and collaborative organizational culture, stimulating community relationships among school members.
5. Manage an effective organizational performance ensuring school resources, disseminating results and defining roles of the school members.

Dimensions 1, 3 and 5 are mostly related to collaborative instructional leadership, whereas dimensions 2 and 4 are more related to the transformational leadership model. As a matter of fact, in the collaborative instructional model the principal is oriented to educational development looking for conditions that directly impact on the quality of curriculum and instruction, considering the opinion, perspective and work of teachers. This requires to develop and implement practices leading to coordinate the members of an organization (the school). By so doing, it is expected to mobilize them to the search of a set of common objectives (Robinson, Lloyd, & Rowe, 2008; Hallinger, 2003).

The transformational leadership model provides intellectual direction and aims at innovating within the organization, while empowering and supporting teachers as partners in decision making (Marks & Printy, 2003). In the school context, the effective functions of transformational leadership that has been identified are the following: (a) mission centered (developing a widely shared vision for the school, building consensus about school goals and priorities), (b) performance centered (holding high performance expectations, providing individualized support, supplying intellectual stimulation), and (c) culture centered (modeling organizational values, strengthening productive school culture, building collaborative cultures, and creating structures for participation in school decisions). For details, see Leithwood, Day, Sammons, Harris, and Hopkins (2006); Leithwood and Jantzi (2006); T. Bush and Glover (2003); Waters, Marzano, and McNulty (2003); Barber and Mourshed (2007); Pont, Nusche, and Moorman (2008); Seashore, Leithwood, Wahlstrom, and Anderson (2010); Horn and Marfán (2010); Weinstein and Hernández (2014).

Summarizing, in the Chilean policy context, both collaborative instructional leadership and transformational leadership operate "in tandem" which, according to Marks and Printy (2003)'s terminology, is called *integrated leadership*.

B.1.2 Information and guidance: the inputs provided to schools by the Agency of Quality

The Agency of Quality of Education promotes the school leadership standards by means of a report of results which communicates to each school its performance level obtained according to the School

Performance Categorization Methodology (ACE, 2016). The report is organized in such a way that is useful to facilitate and orientate principals' and teachers' actions regarding school improvement. As a matter of fact, the performance category level obtained by a school is explained in terms of the expected achievements attained by the students in each of the 10 indicators used to construct such a category level (for the construction, see Section B.1). Associated to this information, the Agency of Quality offers three types of supports: (a) online resources that orientate the analysis of the results gathered in the report in such a way that the resulting reflections help to take intra-school decisions related to school improvement, particularly curricular and pedagogical changes; (b) progressive evaluation, a new voluntary and self-applied evaluation that intends to provide information on students progress in Language at second grade of primary school (this is the only grade at which the test is available); (c) learning visits that, thanks to the active participation of the school community, allow to identify practices significant for the development of students as well as to the institutional improvement (Decree 17, 2014).

These supports are aligned with the integrated leadership model. It should be mentioned that this alignment is emphasized by reporting the positive marginal effects of leadership and teacher feedback on the learning results in Language and Mathematics as measured by the SIMCE test. Among the orientations proposed by the Agency of Quality, three are relevant in this respect: (a) teachers are invited to promote a school climate of respect, and provide feedback to students; (b) principals are invited to base curricular and pedagogical decisions on the results obtained by the school; (c) to involve parents/guardians in the design of the mission of the school (Decree 17, 2014).

B.2 The impact of school leadership on intra-school practices

Due to the high stakes consequences of the National System of Quality Assurance of Education, the Chilean Government decided to start in 2015 the implementation of the school monitoring system as a trial period. During this period, the National Agency of Quality of Education applies a national survey to collect information on five dimensions: (a) general conceptions of school communities on issues such as quality of education, the National System of Quality Assurance of Education, and the National Agency of Quality of Education; (b) general knowledge of school communities on the four school performance levels and the associated educational interpretation; (c) perceptions and general evaluation of school communities about the School Classification Methodology; (d) perceptions of school communities about the relationships between the school performance level and support actions provided by the Agency of Quality, which may include the eventual close of a low-performing school; (e) intra-school changes performed by school communities as a reaction to the obtained school performance level (for details, see DESUC-CEPPE, 2015).

The sampling framework of the National Survey corresponds to a stratified-multistage sampling procedure. This procedure was applied to each region of the country (Chile is politically divided in 15 regions). For each region, the official percentages of type of schools is available: public schools (fully financed by the respective county), subsidized schools (financed by both the county and parents/guardians) and private schools (financed by parents/guardians only); for details, see Bellei, Vanni, Valenzuela, and Contreras (2016). At the first stage of the sampling, schools were chosen with a probability proportional to the percentage of type of schools. When a school was chosen, its principal as well as the head of the technical-pedagogical unit were chosen. At the second sampling stage, from each school at most four teachers were randomly chosen. To the final sample, specific weights were applied in such a way that the sampling distribution of schools were similar to the national distribution of schools. The weights were constructed using the number of schools by region and, conditionally to a region, the number of schools by type of school (for details, see Kalton & Flores-Cervantes, 2003). The final sample includes 385 schools, 385 principals, 385 heads of the technical-pedagogical units, and 1218 teachers; at a confidence level of .95, the error of the sample is +/-.2 (for details, see DESUC-CEPPE, 2015).

B.2.1 Relating school leadership and intra-school practices

We focus our attention on two questions from this national survey, question Q34 and Q39:

Q34. Which of the following activities was or will be implemented in your school as a consequence of the school performance level which was informed during the pilot implementation of the school monitoring system?

- **A.** Teachers receive in a writing format the information regarding the obtained school performance level.
- **B.** Meetings with the teachers to explain the obtained performance level.
- **C.** Principals and teacher discuss, analyze or evaluate the results obtained by the school.

Q39. According to the information you manage, which of the following actions was or will be implemented in your school as a consequence of the school performance level which was informed during the pilot implementation of the school monitoring system?

- **A.** Perform pedagogical and curricular changes.
- **B.** Introduce adjustments to the Plan of School Improvement (this plan is agreed between each school and the Ministry of Education).
- **C.** Propose changes to the institutional educative project.

These questions belongs to the fifth dimension mentioned above and deals with the relationship between school leadership and two specific intra-school practices, namely pedagogical and curricular changes. This relationship is an explicit intention of the Quality Agency that promotes that decisions related to school improvement are based on information related to the school performance level (see Section B.1). More specifically, question Q34 inquires how principals share with teachers the information provided the Agency of Quality. Respondent should choose among three levels at which the information is shared: mere reception of a writing communication, meetings in which principals explain the obtained results, a joint discussion and analysis of the information. The last level represents actions undertaken by principals in order to share with teachers information relevant to take the better joint decisions on school improvement. This is precisely related to both collaborative instructional leadership and transformational leadership –that is, integrated leadership. Question Q39 describes different intra-school changes. We focus our attention to the pedagogical and curricular changes because they are more immediate for students as well as directly related to the school improvement[3]. Question P39 inquires therefore on aspects related to the effects of integrated leadership. Following the substantive framework proposed by Marks and Printy (2003), Q39 can be considered as an effect of Q34 which in turn represents integrated school leadership. As a matter of fact, "when the principal elicits high levels of commitment

[3] The changes described in alternatives **B** and **C** of question Q39 have a potential impact at the school level and at the official relationship between the school and the educational authorities.

and professionalism from teachers and works interactively with teachers in a shared instructional leadership capacity, schools have the benefit of integrated leadership; they are organizations that learn and perform at high levels" (Marks & Printy, 2003, p.393). Moreover, question Q39 is precisely what the Agency of Quality of Education expects as an effect of the joint analysis of the information regarding the obtained school performance level. Summarizing, question Q34 is substantively interpreted as the innovation policy (or innovation, to be denoted by Z) that the Agency of Quality of Education searches to implement in the Chilean schools. The question Q39 corresponds to the outcome (to be denoted by Y) which is generated by the treatment policy Z: this causal relationship, illustrated in Figure 3, is justified by the theoretical integrated school leadership framework that underlines the action theory of the National System of Quality Assurance of Education.

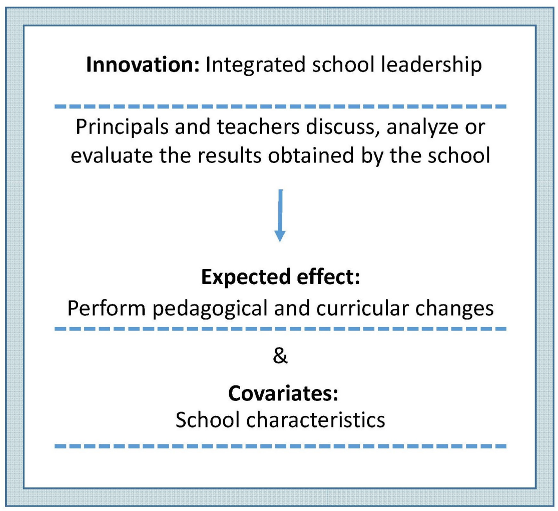

Figure 3: Causal relationship between integrated school leadership (innovation) and pedagogical and curricular changes (outcome)

B.2.2 Summary of the observational evidence

As discussed in the main text, the observational evidence should be summarized in probabilistic terms. The substantive considerations already exposed lead to relate the question Q34 to the innovation Z, and the question Q39 to the outcome Y. In order to operationalize both Y and X, two considerations should be taken into account:

1. The Agency of Quality of Education applied the survey as a means to inquire if intra-school changes at both the pedagogical and the curricular levels are decided on the basis of the information regarding the school performance level.

2. According to the theoretical framework already exposed, this type of changes is in practice performed done by teachers as a result of integrated leadership.

Therefore, both Y and Z will be operationalized taken into account the information provided by the teachers only. Given that the survey was applied to at most four teachers by school, three different binary codifications for the outcome can be considered:

- $Y_a = 1$ if at least one teacher of a school reports pedagogical and curricular changes, $Y_a = 0$ otherwise.
- $Y_b = 1$ if 50% of teachers of a school reports pedagogical and curricular changes, $Y_b = 0$ otherwise.
- $Y_c = 0$ if 100% of teachers of a school reports pedagogical and curricular changes, $Y_c = 0$ otherwise.

Similarly, three binary codifications for the innovation assignment variable can be defined:

- $Z_a = 1$ if at least one teacher of a school reports discussion and analysis of the information, $Z_a = 0$ otherwise.
- $Z_b = 1$ if 50% of teachers r of a school reports discussion and analysis of the information, $Z_b = 0$ otherwise.
- $Z_c = 1$ if 100% of teachers r of a school reports discussion and analysis of the information, $Z_c = 0$ otherwise.

Using these variables, Tables 1 and 2 summarize the information collected in the survey for public schools and subsidized schools. Different combinations of outcomes and innovation assignments can be used to analyze the empirical evidence. In order to show the stability of the observational evidence, we focus our attention on three of them: (Y_b, Z_b), (Y_c, Z_c) and (Y_b, Z_c). Tables 3 and 4 summarize the estimation of the corresponding conditional probabilities. In the three cases, it is observed that the proportion of schools (public or subsidized) reporting pedagogical and curricula changes under the presence of school integrated is greater than the proportion of schools reporting such changes under the status quo. The only difference between these cases deals with the precision of the estimates of the conditional probabilities: for the case (Y_b, Z_b), the precision is better because the number of observations in each cell is greater than in the other cases. Consequently, the observational evidence suggests to promote that principals and teachers adequate their intra-school practices to the innovation (integrated school leadership). However, from a public policy perspective, the relevant question is the following:

> What would be the percentage of public and subsidized schools performing curricular and pedagogical changes if both principals and teachers were to modify their intra-school practices in line with school integrated leadership?

This question is answered using the strategy developed in the main text.

Table 1: Results for Public Schools

	$Z_a = 0$	$Z_a = 1$	$Z_b = 0$	$Z_b = 1$	$Z_c = 0$	$Z_c = 1$
$Y_a = 0$	25	13	30	8	35	3
$Y_a = 1$	5	104	34	74	88	20
$Y_b = 0$	28	42	47	24	65	6
$Y_b = 1$	2	75	18	59	59	17
$Y_c = 0$	30	97	62	65	111	16
$Y_c = 1$	0	20	3	17	12	8

Table 2: Results for Subsidized Schools

	$Z_a = 0$	$Z_a = 1$	$Z_b = 0$	$Z_b = 1$	$Z_c = 0$	$Z_c = 1$
$Y_a = 0$	32	17	35	14	48	1
$Y_a = 1$	23	129	48	104	123	29
$Y_b = 0$	44	54	57	41	94	4
$Y_b = 1$	11	92	26	77	77	26
$Y_c = 0$	52	123	77	98	154	22
$Y_c = 1$	5	24	6	20	17	9

Table 3: Estimation of the conditional probabilities for Public Schools

Case	$P(Y = 1 \mid Z = 1, \text{Public})$	$P(Y = 1 \mid Z = 0, \text{Public})$	$P(Z = 1 \mid \text{Public})$
(Y_b, Z_b)	0.70	0.28	0.56
(Y_c, Z_c)	0.33	0.10	0.16
(Y_b, Z_c)	0.74	0.48	0.16

Table 4: Estimation of the conditional probabilities for Subsidized Schools

Case	$P(Y = 1 \mid Z = 1, \text{Subsidized})$	$P(Y = 1 \mid Z = 0, \text{Subsidized})$	$P(Z = 1 \mid \text{Subsidized})$
(Y_b, Z_b)	0.65	0.31	0.59
(Y_c, Z_c)	0.29	0.10	0.15
(Y_b, Z_c)	0.87	0.45	0.15

B.3 Policy Evaluation

Using the probabilities summarized in the first row of Tables 3 and 4, we obtain the lower and upper bounds of the partial identification intervals for the CPO, CPC, FBC and CBF perspectives; the results are summarized in Tables 5 and 6, respectively.

Table 5: Impact evaluation for Public Schools

	$P(Y_1 = 1 \mid \text{Public})$		$P(Y_0 = 1 \mid \text{Public})$	
Scenario	**Lower bound**	**Upper bound**	**Lower bound**	**Upper bound**
CPO	0.52	0.84	0.12	0.52
CPC	0.39	0.52	0.52	0.73
FBC	0.39	0.52	0.12	0.52
CBF	0.52	0.84	0.52	0.73

Table 6: Impact evaluation for Subsidized Schools

	$P(Y_1 = 1 \mid \text{Subsidized})$		$P(Y_0 = 1 \mid \text{Subsidized})$	
Scenario	**Lower bound**	**Upper bound**	**Lower bound**	**Upper bound**
CPO	0.51	0.79	0.13	0.51
CPC	0.38	0.51	0.51	0.72
FBC	0.38	0.51	0.13	0.51
CBF	0.51	0.79	0.51	0.72

The discussion is focused on public schools; for subsidized schools the considerations below are similar because the estimates are quite similar.

Discussion with respect to the CPO perspective: This perspective is characterized by saying that *the innovation is better than status quo*. Therefore, it is expected that the implementation of the innovation will improve the current situation, namely 52% of the schools experiencing pedagogical and curricular changes. However, the identification region allows to provide information on the following question: How good is the innovation compared with the status quo? According to the results, if the policymaker acts according to her belief, the implementation of the innovation improves the current situation with a maximum of 84%. Nevertheless, the policymaker needs to focus the innovation on $P(Y = 0, Z = 0 \mid \text{Public}) = 31\%$ of the schools, trusting that her message and actions will reinforce the behavior of the schools characterized by $P(Y = 1, Z = 1 \mid \text{Public}) = 39\%$ and $P(Y = 1, Z = 0 \mid \text{Public}) = 12\%$.

Discussion with respect to the CPC perspective: This perspective is characterized by saying that *the status quo is better than innovation*. Therefore, it is expected that the implementation of the innovation will worse the current situation, namely 52% of the schools experience pedagogical and curricular changes. How bad is the innovation compared to the status quo? According to the results, if the policymaker acts contrary to her belief, the implementation of the innovation will worsen the current situation with a minimum of 39%. If she acts according to her belief, the maintenance of the status quo will improve the current situation with a maximum of 73%. Nevertheless, the policymaker needs to focus the status quo on $P(Y = 0, Z = 1 \mid \text{Public}) = 21\%$ of the schools, trusting that her message and actions will reinforce the behavior of the schools characterized by $P(Y = 1, Z = 1 \mid \text{Public}) = 39\%$ and $P(Y = 1, Z = 0 \mid \text{Public}) = 12\%$.

Discussion with respect to the FBC perspective: This perspective is characterized by saying that *the factual situation of schools is better than the counterfactual*. Therefore, if the policymaker decides to implement the innovation or not, in both situations she will take a bad decision. How bad is it? According to the results, if the policymaker takes the wrong decision of implementing the innovation, then the current situation will worsen with a minimum of 39%. However, if the policymaker takes the wrong decision of maintaining the status quo, then the current situation will worsen with a minimum of 15%. Consequently, the maintenance of the status quo is less bad than the implementation of the innovation.

Discussion with respect to the CBF perspective: This perspective is is characterized by saying that *the counterfactual situation of schools is better than the factual*. Therefore, if the policymaker decides to implement the innovation or not, in both situations she will take a good decision. How good is it? According to the results, if the policymaker takes the good decision of implementing the innovation, then the current situation will improve with a maximum of 84%: this means that the policymaker needs to focus the innovation on $P(Y = 0, Z = 0 \mid \text{Public}) = 32\%$ of the schools. If the policymaker takes the good decision of maintenance of the status quo, then the current situation will improve with a maximum of 73%: this means that the policymaker needs to focus the status quo on $P(Y = 0, Z = 1 \mid \text{Public}) = 21\%$ of the schools. Consequently, the implementation of the innovation is much better than the maintenance of the status quo.

C Case study 2

The current situation of the Chilean society is a serious political crisis, mixed with high levels of violence from both State and minority groups. The crisis began with an increase in the price of the subway: student protests over this fact, which in a couple of days became a massive and systematic protest. This transition seems to be due to a perception of abuse: the increase in the price of the subway was perceived as an abuse by the State, which led to the demonstration against several abuses committed in the past, such as, for example, drug collusion. Prices or toilet paper. Quickly, this "protest against abuse" included protests against the pension system (Chile has an individual capitalization system, symptoms of low pensions), as well as against low health coverage of both the private and public system.

The government's reaction focused on restoring order and social security, as well as speeding up some laws that tried to improve the low pensions of a proportion of citizens of the lower socioeconomic levels, as well as laws regulating the price of medicines. In a sense, the government's reaction was carried out through specific policies, which were in line with the political values underlying the government coalition. However, citizens realized that these policies would not solve the crisis because they were based on political values enshrined in a Constitution of the Republic that ensures the existence of the current pension and health system, systems that are perceived as "abusive systems". Therefore, the discussion shifted from policy to politics: citizens began calling for a new constitution since the Constitution of the Republic of 1980 was drafted under the military dictatorship of Pinochet.

Politicians realized the change of discussion and signed an agreement that set a "road map" that would eventually end up in the drafting of a new constitution. This agreement was signed on November 15, 2020. Three stages were agreed: the realization of a plebiscite in which the citizenship can pronounce itself in favor or not of a new constitution, as well as of the body that will write it. If the option in favor of a newconstitution wins, there will be an electoral period so that in October 2020 the members of that body are chosen. Once elected, they will have one year to write the new constitution. There will be a ratifying referendum in which citizens will pronounce themselves in favor or against the new constitutional text.

After the signing of the agreement, the political parties nominated a technical commission in charge of specifying the details of both the referendum and the editorial bodies of the new constitution. After just over a month of work, this proposal was discussed and approved in the Chamber of Deputies on December 24, 2019. The approved text corresponds to the Law 21.200.

C.1 Structure of the national referendum

The Article 130 of the Law 21.200 states that the national referendum will be held on April 26, 2020. Citizens will receive two electoral cards, which will contain the following question:

First electoral card: Do you want a new constitution?

- I approve.
- I reject.

Second electoral card: What kind of body should write the new constitution?

- Mixed constitutional convention (Integrated equally by popularly elected members and parliamentarians in office).
- Constitutional convention (Integrated exclusively by popularly elected members).

It is not at all clear why citizens who reject a new constitution should answer the second electoral card: if a citizen rejects a new constitution, he or she is not interested in discussing what kind of body should write the eventual new constitution. However, an opposite argument could be the following: if the option *I approve a new constitution* wins, citizens who rejected the new constitution has the opportunity to safeguard the current situation (the status quo) by ruling on the eventual drafting body. This last interpretation makes explicit that the national referendum is unfair with citizens who will approve a new constitution.

An alternative argument is the following: a constitutional convention corresponds to a constituent assembly. For a part of Chilean citizens and politicians, the expression *constituent assembly* is synonymous with *institutional chaos* because, using a heuristic of representativeness (Kahneman & Tversky, 1972;

Vis, 2019), they mention that the weakening of democracy in Venezuela is due to the new institutionality that was born from a constituent assembly. To avoid this interpretation, the second Bachelet government (2014-2018) introduced the figure of *mixed constitutional convention* as a way of ensuring that a political-institutional renewal is ensured by the current institutionality; that is, innovation is protected by the status quo. Therefore, citizens may want a new constitution, but that the new constitutional text is to some extent safeguarded by the political status quo represented by the representatives in office

C.2 Identification problem

Once the referendum has been held, it will be interesting to explore which of the previous interpretations is backed by empirical evidence: it is of interest to know the proportion of citizens rejecting the new constitution and choosing either a mixed constitutional convention or a constitutional convention. It is relevant to assess whether the structure of the national referendum is really unfair or not with respect to the citizens rejecting a new constitution. It is also of interest to know if the citizens who accept a new constitution consider the constitutional convention as a dangerous mechanism of political innovation. However, since there is no way to match the two electoral cards, the empirical evidence is not enough to choose one of these interpretations. In other words, we face an identification problem. As a matter of fact, let us define the following binary variables:

$$X = \begin{cases} 1, & \text{if a citizen approves a new constitution;} \\ 0, & \text{if a citizen rejets a new constitution.} \end{cases}$$

$$Y = \begin{cases} 1, & \text{if a citizen chooses a constitutional convention;} \\ 0, & \text{if a citizen chooses a mixed constitutional convention.} \end{cases}$$

What we will observe is the proportion of citizens rejecting a new constitution, namely $P(X = 0)$; the proportion of citizens approving a new constitution, namely $P(X = 1)$; the proportion of citizens choosing a mixed constitutional convention, namely $P(Y = 0)$; and the proportion of citizens choosing a constitutional convention, namely $P(Y = 1)$. The political discussion implicitly assumes that X and Y are dependent variables (*e.g.*, correlated). The data reveals the marginal distributions of X and of Y, but *not* the joint distribution of (X, Y); see Table 7 for a representation of the problem. There is not a *unique* way to derive the entries of a contingency table with known margins. This problem was considered by Fréchet (1960a) and Fréchet (1960b) (see also Rüschendorf, 1981), establishing the partial identifiability of the joint distribution of (X, Y), namely

$$\max\{0, P(X = i) + P(Y = j) - 1\} \leq P(X = i, Y = j) \leq \min\{P(X = i), P(Y = j)\} \tag{C.1}$$

for $(i, j) \in \{0, 1\}^2$. Note that if one entry of the table is chosen in the corresponding interval (C.1), the other three ones can automatically be obtained. For instance, if we fix $P(X = 0, Y = 0)$ then

$$\begin{aligned}
&P(X = 1, Y = 0) = P(Y = 0) - P(X = 0, Y = 0); \\
&P(X = 0, Y = 1) = P(X = 0) - P(X = 0, Y = 0); \\
&P(X = 1, Y = 1) = P(X = 1) - P(X = 1, Y = 0) = P(Y = 1) - P(X = 0, Y = 1).
\end{aligned}$$

Table 7: Identification problem underlying the Chilean national referendum

	$Y=0$	$Y=1$	Total
$X=0$	$P(X=0,Y=0)$	$P(X=0,Y=1)$	$P(X=0)$
$X=1$	$P(X=1,Y=0)$	$P(X=1,Y=1)$	$P(X=1)$
Total	$P(Y=0)$	$P(Y=1)$	1

Just an example: suppose that $P(X = 0) = 0.4$, $P(X = 1) = 0.6$, $P(Y = 0) = 0.3$ and $P(Y = 1) = 0.7$. According to (C.1), $P(X00, Y = 0)$ is between 0 and 0.3. In this case, it can not be assumed that all citizens who rejected the new constitution, voted by a mixed constitutional convention because $P(X = 0, Y = 0) \leq 0.3$. However, it can be assume that all citizens that voted by a mixed constitutional convention, rejected a new constitution; that is, $P(X = 0, Y = 0) = 0.3$. It follows that 10% of citizens who rejected a new constitution, voted for a constitutional convention (that is, $P(X = 0, Y = 01) = 0.1$); that no citizen who approved a new constitution, voted for a mixed constitutional convention (that is, $P(X = 1, Y = 0) = 0$)); and that all the citizens who approved a new constitution, voted for a constitutional convention.

What we learn from this example can be summarized as follows:

1. If $P(Y = 0) < P(X = 0)$, then it could be that some citizens rejecting a new constitution, voted for a mixed constitutional convention.
2. If $P(Y = 1) > P(X = 1)$, then it could be that all citizens approving a new constitution, voted for a constitutional convention.

However, these conjectured highly depend on a value that is fixed in one entry of the joint probability distribution of (X, Y).

Acknowledgements: The first author was partially supported by the FONDECYT Project No. 1181261. The second author was partially supported by a Ph. D. grant from the Corporación Nacional de Ciencia y Tecnología CONICYT from the Chilean government. Part of this work was presented by the second author at the VIII European Congress of Methodology, Jena, Germany (July, 2018); by the second author in a Symposium on *Methodological contributions to improve policy evaluations in highly uncertain contexts* at the International Meeting of the Psychometric Society IMPS 2019, Santiago, Chile (July, 2019); and by the first author as an invited conference at the International Meeting of the Psychometric Society iMPS 2019, Santiago, Chile (July, 2019).

References

ACE. (2016). *Sistema de Aseguramiento de la Calidad. Resultados Categoría de Desempeño 2016*. Santiago de Chile: Author.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica*, *21*, 503–546.

Andereggen, S., Vischer, M., & Boutellier, R. (2012). Honest but broke: The dilemma of universities acting as honest brokers. *Technology in Society*, *34*, 118–126.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Arendt, H. (1958). The Modern Concept of History. *The Review of Politics*, *20*, 570–590.

Arendt, H. (2013). *Hannah Arendt: The Last Interview and Other Conversations*. Melville House.

Austin, J. L. (1962). *How to Do Things with Words*. Clarendon Press.

Barber, M., & Mourshed, M. (2007). *How the world's best-performing schools systems come out on top*. McKinsey & Company.

Bellei, C., Vanni, X., Valenzuela, J., & Contreras, D. (2016). School improvement trajectories: an empirical typology. *School Effectiveness and School Improvement*, *27*(3), 275–292.

Burns, P. B., Rohrich, R. J., & Chung, K. C. (2011). The levels of evidence and their role in evidence-based medicine. *Plastic and Reconstructive Surgery*, *128*, 305.

Bush, T., & Glover, D. (2003). *School Leadership: Concepts and Evidence*. National College for School Leadership.

Bush, V. (1945). *Science, the endless frontier. Report to the President on a Program for Postwar Scientific Research* (Tech. Rep.). Office of Scientific Research and Development, Washington DC.

Cairney, P. (2016). *The Politics of Evidence-Based Policy Making*. Palgrave Macmillan.

Cairney, P. (2017). Evidence-based best practice is more political than it looks: a case study of the "Scottish Approach". *Evidence & Policy: A Journal of Research, Debate and Practice*, *13*, 499–515.

Cairney, P., Oliver, K., & Wellstead, A. (2016). To bridge the divide between evidence and policy: reduce ambiguity as much as uncertainty. *Public Administration Review*, *76*, 399–402.

Clearinghouse, W. W. (2007). Welcome to WWC. *Retrieved April*, *5*, 2019.

Cochran, W. G., & Chambers, S. (1965). The Planning of Observational Studies of Human Populations. *Journal of the Royal Statistical Society. Series A*, *128*, 234–266.

Cochrane, A. L. (1972). *Effectiveness and Efficiency: Random Reflections on Health Services*. Nuffield Provincial Hospitals Trust London.

Dacey, R. (2003). The S-Shaped Utility Function. *Synthese*, *135*(2), 243–272.

Davies, P., & Boruch, R. (2001). The Campbell Collaboration: Does for Public Policy what Cochrane Does for Health. *British Medical Journal Publishing Group*, *323*, 294-295.

Decree 17. (2014). *Aprueba la metodología de ordenación de todos los establecimientos educacionales reconocidos por el estado conforme a lo dispuesto en el inciso cuarto del artículo 17.* (`https://www.leychile.cl/Navegar?idNorma=1060182&idVersion=2014-12-01`)

DESUC-CEPPE. (2015). *Encuesta Nacional de Percepciones, Opiniones y Actitudes de Sostenedores, Equipos Directivos y Docentes sobre la Macha Blanca del Sistema de Aseguramiento de Calidad de la Educación: Fase Implementación Ordenación* (Tech. Rep.). Santigo de Chile: Instituto de Sociología, Pontificia Universidad Católica de Chile & Centro de Estudios de Políticas y Prácticas en Educación.

Dreisbach, S., & Guevara, D. (2017). The Asian Disease Problem and the Ethical Implications of Prospect Theory. *Noûs*, *53*, 613-638.

European Commission. (2017). *Mechanism for Evidence Based Policy-Making in Education. Eurydice Report*. Luxembourg: Publications Office if the European Union.

Felline, L. (2018). Mechanisms Meet Structural Explanation. *Synthese*, *195*, 99–114.

Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London. Series A*, *222*, 309–368.

Fisher, R. A. (1935). *The Design of Experiments*. Oliver and Boyd.

Fréchet, M. (1960a). Les tableaux dont les marges sont données. *Trabajos de Estadística*, *11*, 1–18.

Fréchet, M. (1960b). Sur les tableaux dont les marges et des bornes sont données. *Revue de l'Institut International de Statistique*, *28*, 10–32.

Gigerenzer, G. (2004). Mindless Statistics. *The Journal of Socio-Economics*, *33*, 587–606.

Godin, B. (2006). The Linear Model of Innovation: The Historical Construction of an Analytical Framework. *Science, Technology, & Human Values*, *31*(6), 639–667.

Godin, B. (2011). The Linear Model of Innovation: Maurice Holland and the Research Cycle. *Social Science Information*, *50*(3-4), 569–581.

Haavelmo, T. (1997). Econometrics and the Welfare State. *The American Economic Review*, *87*, 13–15.

Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education*, *33*(3), 329–352.

Hasselbladh, H., & Bejerot, E. (2017). Performative Policy: The Case of Swedish Healthcare Reforms. *Critical Policy Studies*, *11*, 291–310.

Head, B. W. (2010). Reconsidering evidence-based policy: Key issues and challenges. *Policy and Society*, *29*, 77-94.

Holland, P. (1988). Causal Inference, Path Analysis, and Recursive Structural Equation Models. *Sociological Methodology*, *18*, 449–84.

Horn, A., & Marfán, J. (2010). Relación entre liderazgo educativo y desempeño escolar: Revisión de la investigación en chile. *Psicoperspectivas*, *9*(2), 82–104.

Jullien, D. (2013). *Asian Disease-type of Framing of Outcomes as an Historical Curiosity* (Tech. Rep.). Groupe de Recherche en Droit, Economie, Gestion (GREDEG CNRS), Université Nice Sophie Antipolis.

Kahneman, D., & Tversky, A. (1972). Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology*, *3*, 430–454.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, *47*, 263–291.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*, 341–350.

Kahneman, D., & Tversky, A. (2013). Prospect Theory: An Analysis of Decision Under Risk. In *Handbook of the fundamentals of financial decision making: Part i* (pp. 99–127). World Scientific.

Kalton, G., & Flores-Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics*, *19*(2), 81-97.

King, T. (2010). *Statistics in Society: three case studies in the uk* (Tech. Rep.). Southampton Statistical Sciences Research Institute.

Koopmans, T. C. (1949). Identification Problems in Economic Model Construction. *Econometrica*, *17*, 125-144.

Koopmans, T. C., & Reiersøl, O. (1950). Identification of Structural Characteristics. *The Annals of Mathematical Statistics*, *21*, 165–181.

Lehmann, E. L. (1995). Neyman's Statistical Philosophy. *Probability and Mathematical Statistics*, *15*, 29–36.

Lehmann, E. L. (2011). *Fisher, Neyman, and the Creation of Classical Statistics*. Springer Science & Business Media.

Leithwood, K., Day, C., Sammons, P., Harris, A., & Hopkins, D. (2006). *Successful school leadership: What it is and how it influences pupil learning*. Nottingham: National College for School Leadership.

Leithwood, K., & Jantzi, D. (2006). Transformational school leadership for large-scale reform: Effects on students, teachers, and their classroom practices. *School Effectiveness and School Improvement*, *17*(2), 201–227.

Levy, J. S. (1992). An Introduction to Prospect Theory. *Political Psychology*, *13*, 171–186.

Levy, J. S. (2003). Applications of Prospect Theory to Political Science. *Synthese*, *135*, 215–241.

Li, S., & Xie, X. (2006). A new look at the "asian disease" problem: A choice between the best possible outcomes or between the worst possible outcomes? *Thinking & reasoning*, *12*, 129–143.

Linde, J., & Vis, B. (2017). Do Politicians Take Risks Like the Rest of Us? An Experimental Test of Prospect Theory under MPs. *Political Psychology*, *38*, 101–117.

Lindley, D. V. (1972). *Bayesian Statistics: A Review* (Vol. 2). SIAM.

Lorenzen, P. (1987). *Constructive Philosophy*. Univ of Massachusetts Press.

Maddala, G. (1983). *Qualitative and Limited Dependent Variable Models in Econometrics*. Cambridge: Cambridge University Press.

Manski, C. F. (1992). Educational Choice (Vouchers) and Social Mobility. *Economics of Education Review*, *11*, 351–369.

Manski, C. F. (2005a). *Fractional Treatment Rules for Social Diversification of Indivisible Private Risks* (Tech. Rep.). National Bureau of Economic Research.

Manski, C. F. (2005b). *Social Choice with Partial Knowledge of Treatment Response*. Princeton University Press.

Manski, C. F. (2007a). *Identification for Prediction and Decision*. Harvard University Press.

Manski, C. F. (2007b). Minimax-regret treatment choice with missing outcome data. *Journal of Econometrics*, *139*, 105–115.

Manski, C. F. (2013a). Diagnostic testing and treatment under ambiguity: using decision analysis to inform clinical practice. *Proceedings of the National Academy of Sciences*, *110*, 2064–2069.

Manski, C. F. (2013b). *Public Policy in an Uncertain World: Analysis and Decisions*. Harvard University Press.

Manski, C. F. (2019). *Patient Care Under Uncertainty*. Princeton University Press.

Manzi, J., & Preiss, D. (2013). Educational assessment and educational achievement in South America. *International guide to student achievement*, 472–474.

Marks, H. M., & Printy, S. M. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, *39*(3),

370–397.
Mayo, D. G., & Spanos, A. (2006). Severe testing as a basic concept in a Neyman–Pearson philosophy of induction. *The British Journal for the Philosophy of Science*, *57*, 323–357.
Meckes, L., & Carrasco, R. (2010). Two decades of SIMCE: an overview of the National Assessment System in Chile. *Assessment in Education: Principles, Policy & Practice*, *17*(2), 233–248.
MINEDUC. (2005). *Marco para la Buena Dirección. Criteris para el Desarrollo Profesional y Evaluación de Desempeño*. Santiago de Chile: Unidad de Gestión y Mejoramiento Educativo. División de Educación General. Ministerio de Educación. República de Chile.
MINEDUC. (2014b). *Estándares indicativos de desempeño para los establecimientos educacionales y sus sostenedores* (Unidad de Currículum y Evaluacíon ed.). Santiago de Chile.
MINEDUC. (2015). *Marco para la Buena Dirección y el Liderazgo Escolar*. Santiago de Chile: Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas, CPEIP. Ministerio de Educación. República de Chile.
Mouchart, M., & Russo, F. (2011). Causal explanation: Recursive decompositions and mechanisms. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the Sciences* (pp. 317–337). Oxford University Press.
Müller, S. M., Garcia-Retamero, R., Cokely, E., & Maldonado, A. (2011). Causal beliefs and empirical evidence: Decision-making processes in two-alternative forced-choice tasks. *Experimental Psychology*, *58*, 324–332.
Newman, J. (2017). Deconstructing the debate over evidence-based policy. *Critical Policy Studies*, *11*, 211–226.
Oaks, M. (1986). *Statistical Inference: A Commentary for the Social and Behavioral Sciences.* Wiley.
OECD. (2007). *Evidence in Education: Linking Research and Policy*. Paris: OECD.
Parkhurst, J. (2017). *The Politics of Evidence: From Evidence-Based Policy to the Good Governance of Evidence*. Routledge.
Parmigiani, G., & Inoue, L. (2009). *Decision Theory: Principles and Approaches* (Vol. 812). John Wiley & Sons.
Pedersen, D. B. (2014). The political epistemology of science-based policy-making. *Society*, *51*, 547–551.
Peterson, M. (2017). *An Introduction to Decision Theory. Second Edition*. Cambridge University Press.
Pielke Jr, R. A. (2007). *The Honest Broker: Making Sense of Science in Policy and Politics*. Cambridge University Press.
Pielke Jr, R. A. (2012). “Basic Research” as a Political Symbol. *Minerva*, *50*(3), 339–361.
Pont, B., Nusche, D., & Moorman, H. (2008). *Improving School Leadership, Volume 1 Policy and Practice: Policy and Practice* (Vol. 1). OECD Publishing.
Robinson, V., Lloyd, C., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, *44*(5), 635-674.
Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, *70*, 41–55.
Rousseau, D. M., & Gunia, B. C. (2016). Evidence-based Practice: The Psychology of EBP Implementation. *Annual Review of Psychology*, *67*, 667–692.

Rüschendorf, L. (1981). Sharpness of Fréchet-bounds. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *57*, 293–302.

Ryan, R., & Hill, S. (2016). *How to GRADE the Quality of the Evidence* (Tech. Rep.). Cochrane Consumers and Communication Group.

Rycroft-Malone, J., & Bucknall, T. (2011). *Models and Frameworks for Implementing Evidence-Based Practice: Linking Evidence to Action (Vol. 2).* John Wiley & Sons.

Sackett, D. L., Straus, S. E., Richardson, W., Rosenberg, W., & Haynes, R. B. (2000). *Evidence Based Medicine: How to Practice and Teach EBM (2nd ed.).* Churchill Livingstone.

San Martín, E. (2016). Identification of Item Response Theory Models. In W. van der Linden (Ed.), *Handbook of Item Response Theory: Models, Statistical Tools, and Applications. Volumen 2* (pp. 127–150).

San Martín, E. (2018). Identifiability of Structural Characteristics: How Relevant is it for the Bayesian Approach? *Brazilian Journal of Probability and Statistics*, *32*, 346–373.

Savage, L. J. (1951). The Theory of Statistical Decision. *Journal of the American Statistical Association*, *46*, 55–67.

Schünemann, H. J. (2016). Interpreting GRADE's levels of certainty or quality of the evidence: GRADE for statisticians, considering review information size or less emphasis on imprecision? *Journal of Clinical Epidemiology*, *75*, 6–15.

Seashore, K., Leithwood, K., Wahlstrom, K., & Anderson, S. (2010). *Investigating the links to improved student learning: Final report of research findings*. The Wallace Foundation.

Stone, G. D. (2017). Dreading CRISPR: GMOs, honest brokers, and mertonian transgressions. *Geographical Review*, *107*, 584.

Strassheim, H. (2017). Bringing the political back in: reconstructing the debate over evidence-based policy. a response to Newman. *Critical Policy Studies*, *11*, 235–245.

Tamer, E. (2010). Partial Identification in Econometrics. *Annual Review of Economics*, *2*, 167–195.

Toland, J. (1702). *Christianity not Mysterious, or, A Treatise Shewing, That there is nothing in the Gospel Contrary to Reason, nor Above it: And that no Christian Doctrine can be properly call'd a Mystery*. Printed for Sam. Buckley at the Dolphin over against St. Dunstans Church in Fleetstreet.

Turnhout, E., Stuiver, M., Klostermann, J., Harms, B., & Leeuwis, C. (2013). New Roles of Science in Society: Different Repertoires of Knowledge Brokering. *Science and Public Policy*, *40*, 354–365.

Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, *211*(4481), 453–458.

Van Woensel, L. (2020). *A Bias Radar for Responsible Policy-Making: Foresight-Based Scientific Advice*. Palgrave Macmillan, Springer Nature.

Vis, B. (2019). Heristic and Pplitical elites' Judgement and Decision-Making. *Political Studies Review*, *17*, 41–52.

Von Neumann, J., Morgenstern, O., & Kuhn, H. W. (2007). *Theory of Games and Economic Behavior (Commemorative Edition).* Princeton University Press.

Wald, A. (1950). *Statistical Decision Functions*. John Wiley and Sons, Inc.

Waters, T., Marzano, R., & McNulty, B. (2003). *Balanced Leadership: What 30 Years of Research Tells Us about the Effect of Leadership on Student Achievement. A Working Paper*. ERIC.

Weinstein, J., & Hernández, M. (2014). Políticas hacia el liderazgo directivo escolar en chile: Una mirada comparada con otros sistemas escolares de américa latina. *Psicoperspectivas*, *13*(3), 52–68.
Wiseman, A. W. (2010). The Uses of Evidence for Educational Policymaking: Global Contexts and International Trends. *Review of Research in Education*, *34*, 1–24.
Wolpin, K. I. (2013). *The Limits of Inference without Theory*. The MIT Press.
Wunsch, G., Mouchart, M., & Russo, F. (2014). Functions and Mechanisms in Structural-Modelling Explanations. *Journal for General Philosophy of Science*, *45*, 187–208.