

The Rigor of Case-Based Causal Analysis

Busting Myths through a
Demonstration

Estelle Raimondo



WORLD BANK GROUP
World Bank • IFC • MIGA

IEG Methods and Evaluation Capacity Development Working Paper Series

© 2023 International Bank for Reconstruction and Development / The World Bank
1818 H Street NW
Washington, DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

ATTRIBUTION

Please cite the report as: Raimondo, Estelle. 2023. *The Rigor of Case-Based Causal Analysis: Busting Myths through a Demonstration*. IEG Methods and Evaluation Capacity Development Working Paper Series. Independent Evaluation Group. Washington, DC: World Bank.

MANAGING EDITORS

Jos Vaessen
Ariya Hagh

EDITING AND PRODUCTION

Amanda O'Brien

GRAPHIC DESIGN

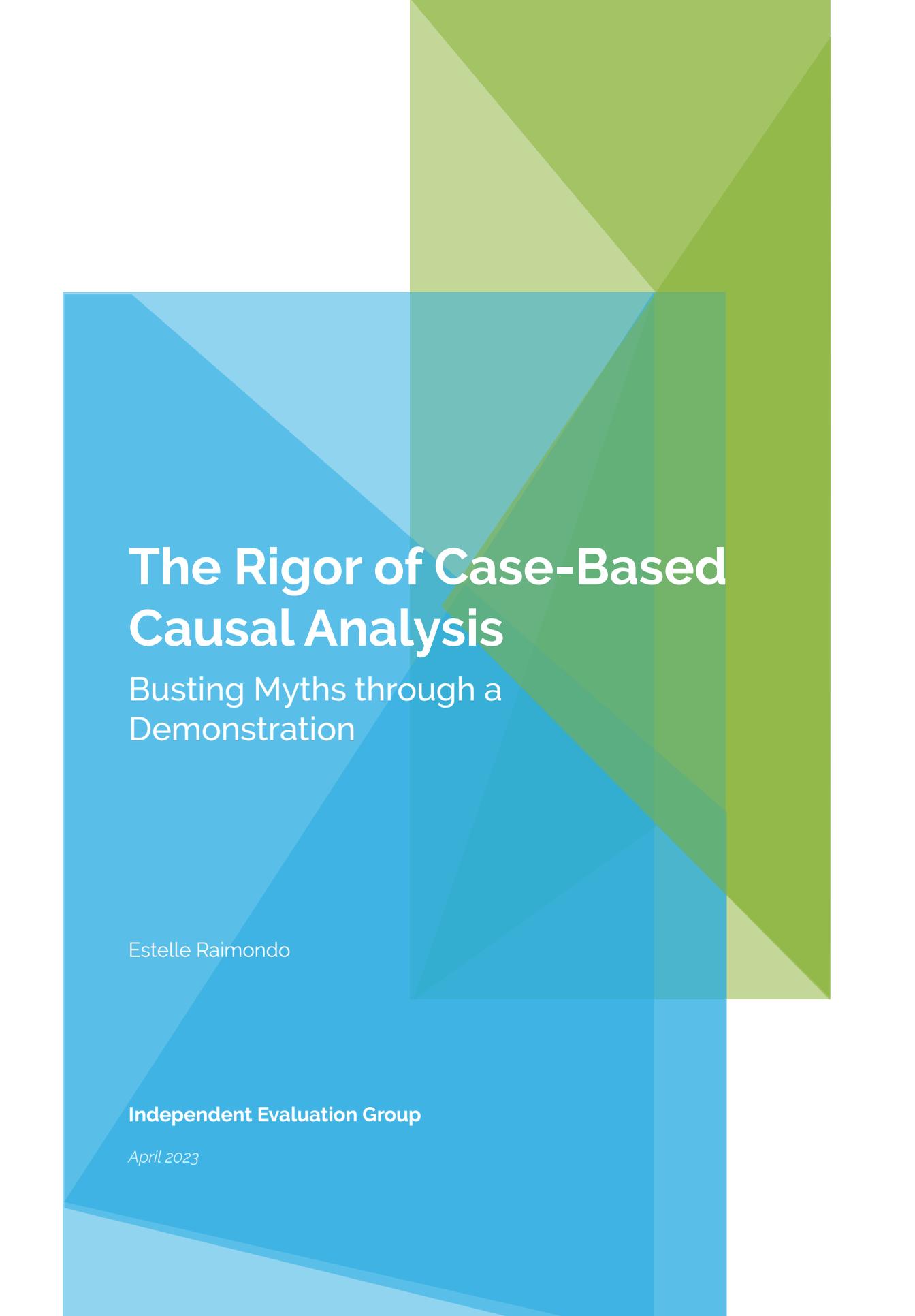
Lúisa Ulhoa

This work is a product of the staff of The World Bank with external contributions. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

RIGHTS AND PERMISSIONS

The material in this work is subject to copyright. Because The World Bank encourages dissemination of its knowledge, this work may be reproduced, in whole or in part, for noncommercial purposes as long as full attribution to this work is given.

Any queries on rights and licenses, including subsidiary rights, should be addressed to World Bank Publications, The World Bank Group, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2625; e-mail: pubrights@worldbank.org.



The Rigor of Case-Based Causal Analysis

Busting Myths through a
Demonstration

Estelle Raimondo

Independent Evaluation Group

April 2023

CONTENTS

Abbreviations	iv
Author	vi
Abstract	viii
Introduction	x
1. Designing for Causal Inference and Generalizability	2
How to Infer Causality	4
How to Generalize from Case-Based Evidence	5
Overarching Design	7
Structure of the Causal Theory	9
Defensible Case Selection	10
Contributory Factors Selection	12
Systematic Qualitative Data Collection	13
2. Analysis	16
3. Illustrating Findings	22
Conclusion	28
Bibliography	32

ABBREVIATIONS

CDM Clean Development Mechanism
ERPA Emission Reduction Purchase Agreement
QCA qualitative comparative analysis

AUTHOR

Estelle Raimondo, Senior Evaluation Officer

Author Affiliation

Independent Evaluation Group, World Bank Group

ABSTRACT

Several myths persist within research and evaluation circles about the power and limitations of evaluation designs that use cases (or case studies) as their primary empirical material (case-based evaluation designs). Using a real-world application, this paper busts two myths regarding the use of case-based designs in evaluations that aim to answer effectiveness questions and unpack the relationships between interventions and observed changes in outcomes (broadly known as causal analysis): that case studies cannot be used for causal analysis and that it is impossible to generalize from case studies. Through a detailed demonstration of how the evaluation of the World Bank's support to carbon finance has been designed and implemented, the paper undoes these preconceived ideas about the inferential, explanatory, and generalizability power of case-based evaluation designs.

INTRODUCTION

Over the past decade, the debate on the rigor of various approaches to impact evaluation and causal analysis has made significant strides. There is increasing consensus that questions must drive the methodology for this type of evaluation and not the reverse (Ravallion 2020), that some types of causal questions about interventions that are being evaluated simply do not lend themselves to (quasi-)experimental designs (Stern et al. 2012), and that various methodologies have different comparative advantages for answering various types of causal questions (AFD 2022; Befani 2012; Quadrant Conseil 2017). Epistemologically, some common ground is emerging regarding the redefinition of standards of rigor (for example, Jimenez et al. 2018; Johnson and Rasulova 2017), moving away from the idea that any one methodology is the gold standard for all quality criteria for research.

In another institutional movement, clearinghouses created to promote (quasi-)experimental impact evaluations are now exploring avenues for incorporating other approaches (Dixon and Bamberger 2022). The number of articles on how to integrate qualitative methods into impact assessments to complement quantitative methods has increased dramatically over the past few years. However, the level of rigor with which quantitatively driven impact evaluations incorporate qualitative approaches remains weak, as a recently conducted review of impact evaluations shows (Jimenez et al. 2018).

Recognition of the need to expand the range of evaluation methods used for causal analysis is also increasing (Stern et al. 2012; Jimenez et al. 2018). As a community of practice, evaluators have started to experiment with various methods of causal analysis, borrowing from other disciplines of the social sciences and adapting to real-world evaluation constraints (Schmitt 2020). The literature on applying alternative approaches to impact assessments has thus flourished, predominantly regarding combining theory-based approaches that use case studies as their primary empirical material, such as contribution analysis (for example, Delahais and Toulemonde 2012; Kane et al. 2021; Ton et al. 2019); realist evaluation approaches (Kazi 2003); process tracing (Befani 2021; Raimondo 2020; Rothgang and Lageman 2021); and qualitative comparative analysis (Befani 2016; Hanckel et al. 2021). Yet this literature remains emergent, and the applicability, quality, and usefulness of these various approaches require more testing.

In addition, a few myths and misunderstandings linger, despite several attempts at busting them (Flyvbjerg 2006; Widner, Woolcock, and Ortega Nieto 2022), and hinder wider adoption of case-based approaches to causal analysis in evaluation practice. The first of these relates to the inferential power of case-based approaches to such analysis (Cartwright 2022). There is a common misunderstanding that causal claims can be built only on approaches involving analysis of large numbers of observations using counterfactual thinking. The second has to do with the generalizability (external validity) of causal claims built through case-based methods. The misconception here is that evidence generated through analysis of cases is necessarily anecdotal and case specific and cannot be transferred to other cases. Taken together, these two myths fuel the argument that case-based approaches to causal analysis fail to generate findings and evidence that are useful for informing policies and thus lack value in practice.

This paper busts these myths by demonstrating that (i) the application of case-based and theory-based methods can lead to robust causal inferences and explanations and hence fill important knowledge gaps on the impact of development interventions, and (ii) it is possible to generalize from case studies, and in so doing, to generate practical and useful information on the inner workings of complex interventions and the conditions under which interventions are more or less successful.

The demonstration presented here is based on a causal analysis of the World Bank's support to carbon finance through the development of Emission Reduction Purchase Agreements between 1999 and 2012. Carbon finance mechanisms enable high-polluting countries to offset their carbon emissions; the World Bank has supported increasing use of these mechanisms through its support for projects that introduce new technologies in developing countries, such as renewable energy projects, reforestation projects, or waste capture projects, generating reductions in greenhouse gas emissions that high-polluting countries can in turn buy as part of their carbon-offsetting efforts. An Emission Reduction Purchase Agreement is a contractual vehicle through which such a system can work. The World Bank plays the role of trustee of a carbon fund, agreeing to pay for the purchase of a certain quantity of an asset (emission reduction); payment takes place on delivery. The evaluation question that motivated the causal analysis undertaken in this paper is, How effective have the main World Bank Group carbon finance interventions been in (i) catalyzing and developing carbon markets and leveraging private investments, (ii) reducing greenhouse gas emissions, and (iii) generating demonstration effects for technologies and carbon finance?

The demonstration is organized in three sections. The first section lays out the methodology used in the causal analysis, emphasizing principles in the design of

the analysis that address causal inference and generalizability. The second section shows how within-case causal analysis was conducted; exposition of the application of cross-case causal analysis follows. The third section shows how the findings can be illustrated and integrated into other analyses. The paper concludes with a discussion on the applicability of case-based approaches, their relative strengths, and their limitations.

1

DESIGNING FOR CAUSAL INFERENCE AND GENERALIZABILITY



How to Infer Causality



Structure of
Causal Theory



Overarching Design



Systematic Qualitative
Data Collection

How to Infer Causality

To a certain extent, the myths discussed in the introduction persist because of a lack of awareness of the various causal theories that underlie different techniques for and approaches to impact assessment (Cartwright 2022). Answering critical evaluation questions regarding what works in interventions, for whom, under what circumstances, how, and why (which is the crux of the impact evaluation enterprise) requires a combination of these techniques and approaches to causal inference. A brief recap is in order.

Traditional accounts of causality are based on the idea that cause and effect are regularly found together, and that causality can be identified by observing patterns of regularity (Befani 2012). In practice, methods of statistical causal modeling are used to check the presence of both intervention and outcome in a large number of cases, under the assumption that causal association grows stronger as the number of cases in which both are present increases. However, because it is impossible to consider all possible cases that would share only cause and effect (referred to as Mill's method of agreement), researchers resort to an alternative solution: comparing cases that are by all accounts similar except for cause and effect (Mill's method of difference, also known as counterfactual thinking), in which the causal inference concentrates on establishing a singular cause-effect relationship between an independent variable (an intervention) and an outcome (the dependent variable). The main causal question of interest focuses on the attribution of a marginal (net) effect to the intervention. Although the preferred methodology for establishing these types of causal relationships is (quasi-)experiments that seek to create a counterfactual situation (Befani 2012), thought experiments such as those in rapid impact assessments or the emerging application of virtual reality also involve counterfactual thinking (Rowe 2019; Gürerk et al. 2014). These techniques are best suited to measuring the average effect of an intervention on an outcome of interest. However, they often black-box the steps that logically link the intervention to the observed outcome.

An alternative causal theory, equally relevant to evaluation, focuses on causation as being, by its nature, attributable to multiple different but plausible combinations of factors. According to this theory, multiple causal pathways can lead to the same outcome, and a combination of factors are at play in each pathway. According to this conception of causality, most often an effect has no single cause; rather, a combination of causally relevant factors generates a particular outcome. Moreover, multiple combinations of factors may lead to the same outcome, and a given outcome may result from either the presence or absence of a particular factor or set of factors, depending on the context (Rihoux and Ragin 2009, 8). This theory is best leveraged

to answer causal questions that are particularly salient in evaluation, such as, for whom and under what circumstances did a particular intervention work or not work? and What role did the intervention play, among other factors, in producing a particular outcome? Cross-case research and evaluation methods are best suited to answer these types of questions.

Finally, generative or mechanisms-based causal theories, inspired by scientific realism (Bhaskar 1975; Glennan 1996; Pawson 2013; Schmitt 2020), seek to identify the intervening causal process (or causal chain), made up of interlocking causal mechanisms, between an independent variable (the intervention) and the outcome of interest. The key causal question of interest in approaches based on these theories is, why and how has the intervention made a difference in the outcome of interest? Approaches such as process tracing, realist evaluation, contribution analysis, impact pathway analysis, and causal mediation have a comparative advantage in answering causal questions of this type (for more details, see Beach and Pedersen 2019; Befani 2012, 2021; and Raimondo and Beach, forthcoming).

In the last two of these approaches, the strength of the causal inference obtained depends critically on the specificity of the causal theory that is posited as underlying the intervention and explicit and detailed description of the theorized causal packages and mechanisms that explain why a particular intervention would contribute to a given outcome (Raimondo and Beach, forthcoming). For approaches of these two types, typical causal theories consisting of a few boxes and arrows do not suffice as support for strong causal inferences. Instead, explicitly laying out the underlying causal assumptions and mechanisms are fundamental parts of the design of such approaches, as the next section shows.

How to Generalize from Case-Based Evidence

Another myth that continues to inhibit the adoption of case-based methods involves the issue of generalizability of findings from case studies. A common argument against using case-based approaches in evaluation is that one cannot validly generalize from a single case. Yet as Flyvbjerg (2006, 219) argued early on, this argument “if not directly wrong, is so oversimplified as to be grossly misleading.” Woolcock (2013, 2022) and others argue that case studies have a comparative advantage in providing key facts necessary to determine a causal claim’s level of generalizability, especially in instances of complex interventions. Case studies help elucidate (i) the contextual factors that help explain whether an intervention that works in one instance will also work in another; (ii) the process mechanisms that help establish what parts of an intervention worked or proved to be broken; (iii) the ingredients

needed for a particular intervention to work, including resources, skills, capacities, and laws; and (iv) an intervention’s trajectory of change, including how long it takes for change to materialize, progress, or deteriorate. In turn, Bennett’s (2022) answer to whether case studies generalize is that it depends on prior knowledge of causal mechanisms, understanding of populations of cases, and patterns in contextual factors that enable or disable causal mechanisms. Generalization here relies on the soundness and power of the causal explanation as opposed to laws of large numbers.

In addition, three principles can be usefully leveraged for determining whether findings emerging from case-based methods of causal analysis are generalizable. The first of these should inform the design of the methodology. Specifically, at the outset of an evaluation, what Rihoux and Ragin (2009) call an area of homogeneity must be defined to establish the boundaries and scope within which cases will be selected for analysis. Then, when selecting cases for inclusion in the analysis, the evaluator should strive to maximize variation within this particular homogeneity space. Cases included in the analysis must be sufficiently similar to one another to be comparable along certain dimensions (that is, they must have enough in common to be productively compared); as the saying goes, comparing apples and oranges is not useful. However, within this well-defined area of homogeneity, it is important that evaluators maximize the diversity of cases across the minimum established number of cases that can feasibly be studied given the time and resource constraints bounding the evaluation. This will enhance the potential for (modest) generalizability to cases that belong to the same population as the cases examined and that share sufficient contextual elements to enable their shared elements to be used to explain variations in outcome (for more details, see Rihoux and Ragin 2009, 21–25, and Bennett 2022).

Second, the strength of the generalizability of findings that emerge from case-based methods of causal analysis will depend on the patterns of convergence of evidence identified in the cases studied across different contexts (different countries, projects, and so on). If patterns are found to recur across cases within a sample, despite variability in the cases selected, the likelihood of finding similar patterns elsewhere increases.

Third, as explained by Rihoux and Ragin (2009, 12), a “good index of the quality of causal explanations could be precisely their ability to withstand refutation when confronted with new cases.” If patterns of findings in a particular causal analysis converge with existing evidence present in the related literature, for instance, or if they converge with one or more out-of-sample cases, then the credibility of claims regarding the generalizability of those findings increases.

These three considerations were central to the design of the Emission Reduction Purchase Agreement (ERPA) causal analysis conducted here, as the next section explains.

Overarching Design

The overarching design of the case-based causal analysis used in the current study followed the logic of theory testing that Trochim (1985, 1989) popularized as pattern matching. As figure 1.1 illustrates, pattern matching involves an attempt to connect two patterns: a theoretical pattern and an observed (or empirical) pattern. The bottom part of the figure shows the theoretical realm. In this study, the causal theory (or theory of change) originated from a review of the existing literature (consisting of both a structured literature review and a lighter review of specific themes related to carbon finance).¹ These were supplemented with consultations with carbon finance experts and validation from the World Bank's Carbon Finance Unit. The conceptualization task involved transforming the ideas that emerged from the literature review and consultations into a graphical representation, ultimately generating a set of propositions for each part of the theorized causal process.

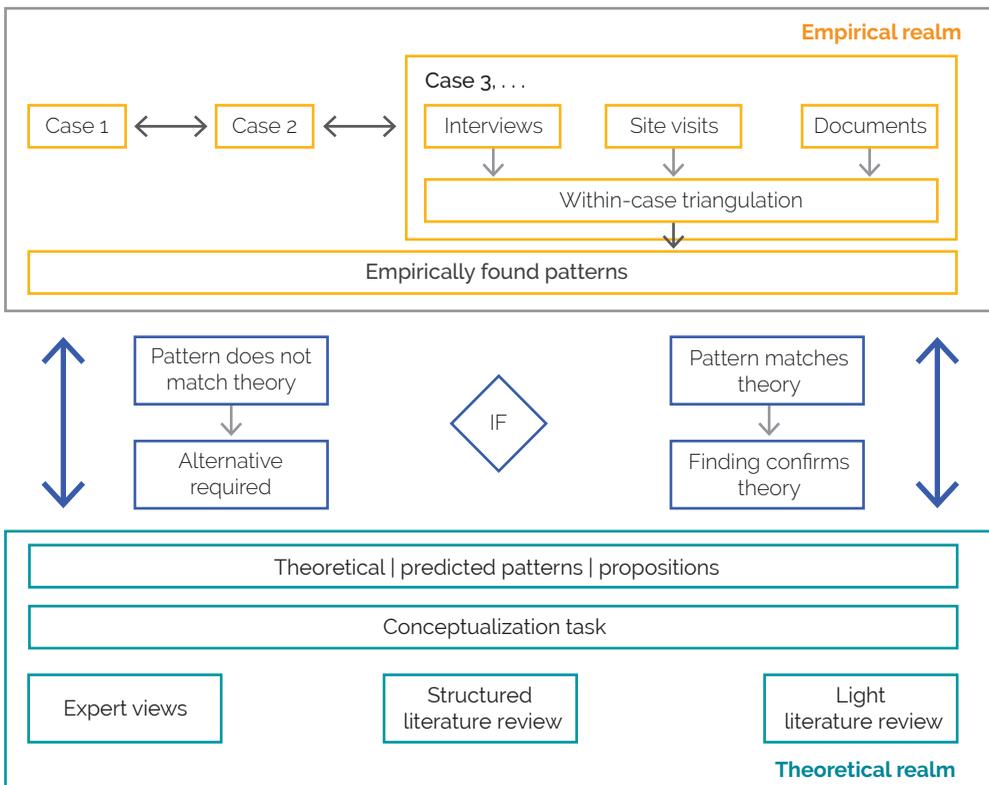
The top part of the figure depicts the empirical realm. The empirical strategy used in the study applied two different case-based methods (with a comparative advantage in providing evidence for causal analysis) to the 16 ERPA cases. These consisted of a within-case causal analysis, following the logic of process tracing, and a cross-case causal analysis, following the logic of qualitative comparative analysis (QCA). For each ERPA case, the study team traced the contribution of the World Bank and other critical actors and variables throughout the process of intervention development, implementation, and follow-through. Data collection broadly included review of documents related to the intervention, field visits, and a series of interviews with key stakeholders engaged throughout the ERPA cycle and beyond. Patterns of convergence and divergence that emerged across cases were systematically analyzed using the logic of QCA, ultimately generating a robust empirical base.

The middle part of the figure represents the inferential task, which attempts to link or match the theoretical and empirical patterns. To the extent to which the theoretical and empirical patterns match one another, the posited causal theory is validated, and the same empirically observed pattern can be predicted to exist in cases similar to those studied. A causal analysis of this type uses techniques imbued with the idea that causality is complex (Cartwright 2004, 2007). These techniques accept the premise that different causal pathways, made up of different combinations of variables (or causal packages), can lead to the same outcome. This notion, known as multiple conjunctural causation in the literature (for example, De Meur and Rihoux 2002), considers causality as context- and conjuncture-specific and refutes a number of assumptions at the core of common statistical approaches to causal inference. Notably, the following are not assumed:

- Linearity and permanent causality
- Homogeneity of the unit of analysis
- Additivity
- Causal symmetry (Rihoux and Ragin 2009, 9)

In the current study, this inferential task was performed for each part of the causal process through QCA, with formalization using Boolean logic for a few links of the process in which causality was theorized to be particularly complex. The overall design of the analysis had four key elements: a detailed causal theory, a defensible case selection, identification of a select number of variables to be systematically scrutinized and investigated in the empirical inquiry, and a detailed plan for systematic collection of qualitative information within and across cases. The next section details these four elements.

Figure 1.1. Overall Evaluation Design: Pattern Matching

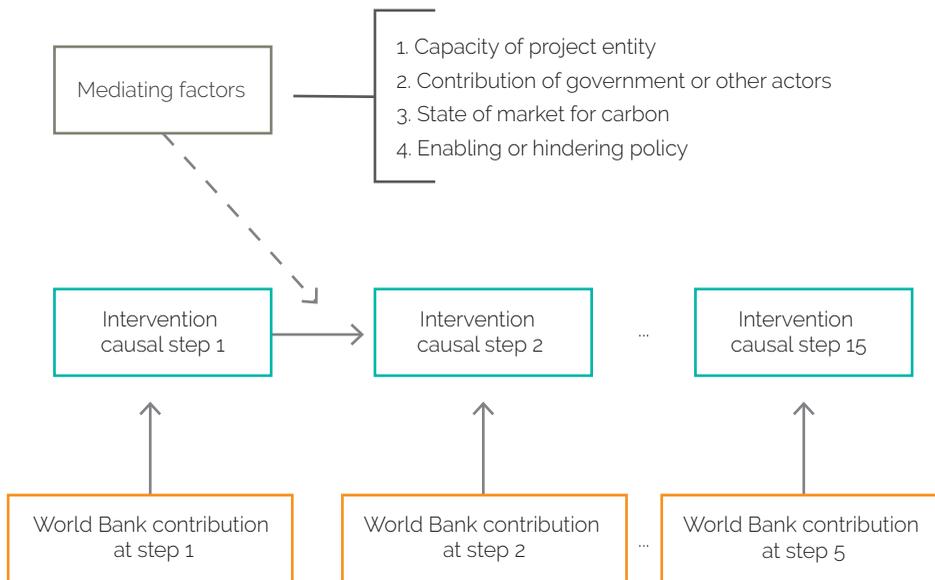


Source: Independent Evaluation Group.

Structure of the Causal Theory

The causal theory developed for the current study, composed of 15 causal steps, was constructed iteratively, drawing from a structured review of the literature on the design and effectiveness of carbon finance interventions and consultations with key experts within and outside the Bank Group. It diverges from more traditional representations of causal theories, which follow an input-activity-output-outcome logic. Instead, the causal theory’s structure encompasses three different elements, which the figure traces: the theory’s 15 causal steps, the World Bank’s expected contribution at each step, and assumptions about other contributing factors that mediate the causal relationship among the parts of the theorized causal process (figure 1.2).

Figure 1.2. Structure of the Causal Theory



Source: Independent Evaluation Group.

The causal theory developed for the current study is rather complex and seeks to generate unique theoretical patterns. More complex theoretical patterns usually make it more difficult to construe sensible alternative patterns that would predict the same result (Befani 2021; Beach and Pedersen 2019). For each link in the theorized causal process, the causal theory makes explicit the Bank Group’s expected contribution and the specific causal assumptions that define the circumstances under which it is more or less likely that a particular causal process will occur.

Defensible Case Selection

Case selection is a critical part of designing case-based causal analysis and must be executed carefully to maximize the chances of both valid causal inference and (modest) generalizability of the analysis's findings (Bennett 2022; Rihoux and Ragin 2009). Several considerations must be weighed carefully in choosing a case selection strategy. As explained earlier, cases must be selected within a specific area of homogeneity, so that they have enough in common to be compared and so that characteristics shared across cases can be used to explain the variabilities in outcomes. For that reason, in the current analysis, a most-similar-different-outcome selection strategy was applied first. In the current example, two major elements were homogeneous across ERPA cases selected for the study. First, the process of asset creation was almost identical across cases because the United Nations had codified it in the Clean Development Mechanism (CDM). To be eligible for World Bank support, projects in all of the ERPAs studied had to abide by a few rules and complete a number of steps to generate emission reduction credits. Second, the type of support, that is, the intervention itself, was rather homogeneous across cases. World Bank support consisted of advocating for carbon finance with governments and specific entities involved in ERPA-related projects, in providing technical assistance to those entities at various steps in the process of creating assets (in the form of greenhouse gas reductions that could be bought by high-polluting countries), and in promoting due diligence to ensure compliance with the CDM process. Additionally, cases representing different degrees of success in carbon finance were selected, relying on external databases that captured whether specific ERPAs had achieved their emission reduction targets and on a preliminary screening of the entire portfolio of World Bank-supported ERPAs. Moreover, to increase the likelihood that the findings from the analysis would have internal validity, multiple cases in the same country and involving the same category of technology were included.

But, potential generalizability of the findings relies in part on how well the cases included in the study represent the broader universe of World Bank ERPA interventions. Therefore, case selection in this study sought to reach a maximum degree of heterogeneity over a minimum number of cases. It was informed by a preliminary review of the World Bank's entire portfolio of ERPAs. An additional consideration was the need to accommodate other components of the evaluation, notably the inclusion of country-level case studies for which the countries had already been selected (based on other relevant selection criteria). In cases that involved the constraints of preselected countries, the following additional selection criteria for ERPA cases were used:

- Ensuring representation of the four primary categories of technologies used in ERPAs, with the objective that a case study selected for inclusion in the

current study involved at least one case from each of four categories: afforestation or reforestation, hydropower, other (nonhydro) renewable energy, and waste management

- Ensuring representation of various levels of country capacity for carbon finance, with the objective that a case study selected for inclusion in the current study involved countries with at least four different levels of country capacity
- Ensuring representation of various levels of maturity of the CDM process and carbon market, with the objective that a case study selected for inclusion in the current study involved cases that spanned at least a 20-year horizon
- Considering the need to keep the number of case studies selected manageable for in-depth analysis
- Considering practical challenges for organizing data collections (for example, selecting among cases in China that were in geographic proximity)

As table 1.1 illustrates, the final case selection ensured that the unfolding of the causal process within the cases selected could be compared (i) within countries and across technologies; (ii) within technologies and across countries; and (iii) within technologies and within countries, across both positive and negative outcomes.

Table 1.1. Case Selection

Technology	Chile	China	Colombia	Ethiopia	Uganda	Total
Afforestation or reforestation	•	•	• •	•	•	6
Hydropower	•	• •			•	4
Other (nonhydro) renewable energy	• •		•			3
Waste management	•		•		•	3
Total	5	3	4	1	3	16

Source: Independent Evaluation Group.

Note: Dots represent presence of specific technologies in different country cases.

Contributory Factors Selection

In keeping with the set-theoretic research tradition and QCA in particular, the choice of variables (also known as conditions) for inclusion in the study was both theoretically and inductively informed, with insight gained from knowledge generated during three pilot cases, for identifying the key elements in the cases studied that needed to be considered. The imperative of avoiding the “many variables, few cases” dilemma, common in approaches of the type used here, also guided the selection of variables (Befani 2016; Rihoux and Ragin 2009).

Selection of variables for study proceeded in three stages: first, all possible explanatory variables and assumptions that could have influenced the likelihood a particular ERPA project would move from one step to the next along the theorized causal process were listed and embedded within the causal theory developed for the study. These variables and assumptions were then categorized into a broad typology of variables and assumptions recurrent in several parts of the causal process. Next, this typology was pilot tested for three cases in Chile. After completion of the pilot, the selection process for variables was revised and systematized, resulting in the ultimate choice of five variables for study, grouped into two main categories:

Contribution of key players in the process:

1. Efficacy of Bank Group contribution (main intervention of interest)
2. Capacity of project entities (implementing agency or project owners)
3. Support of external players (government entities, third parties, trader associations, other donors)

Enabling environment:

4. Conduciveness of policy environment (for example, regulations, other carbon-related policies, government subsidies)
5. Conduciveness of market environment (for example, carbon market conditions)

Selection of this subset of variables increased the likelihood the study would identify the core elements of the causal mechanisms at work in the cases studied while preserving the parsimony required by the approach. As described in more detail later in the paper, case data collection was thus primarily deductive and involved trying to identify the presence or absence of the five variables selected for study. However, inductive inquiry was also incorporated, to ensure that additional explanatory variables that explained the outcomes of interest were not missed. All case authors

were instructed to tease out additional explanations not included in the causal theory selected for the study. For instance, semistructured interviews were conducted, involving a sequential purchase of information approach, starting with broad open-ended questions on the factors that facilitated or hindered the process of asset creation and the outcomes of interest, followed by structured questioning on the subset of variables of interest.

Systematic Qualitative Data Collection

The robustness of the findings in studies like the current one depends greatly on the quality, consistency, and reliability of the data collected. The granularity and context specificity of the findings they generate make case-based approaches particularly useful. Evaluative case studies thus generally involve thick descriptions of the interventions studied, with rich examples. But, in the current study, the comparative nature of the approach and the relatively large sample size of 16 cases demanded a consistent approach to data collection across cases. To ensure the right balance between granularity of any resulting causal explanations and the consistency and reliability of the data collected across cases, the study team developed a structured case study template made up of a number of questions to be answered in a detailed case narrative and a matrix for synthesizing and structuring the qualitative data collected, as presented in figure 1.3. The template was used in the pilot in Chile and then refined based on the pilot experience.

For each case included in the study, investigators gathered evidence (that is, data) through reviewing project documents and conducting interviews with key stakeholders and site visits during a field mission lasting one to two weeks. Data collection involved eight local investigators and four case leaders. The case leaders identified and selected local investigators for the cases in the four remaining countries included in the study based on their knowledge of the CDM process and specific technologies used within each country. Both a methods expert and a study coordinator trained all investigators. The training sought to ensure investigators had a good understanding of the study objective and case study template and to advise them on how to conduct the tracing work involved in the investigative process, including how to identify the right stakeholders and documents to consult, how to look for “fingerprints” of the process, and how to judge the probative value of any evidence obtained (for example, seeking access to the full evidentiary record and gauging the trustworthiness of the sources; Beach and Pedersen 2019; Raimondo and Beach, forthcoming). The training was recorded and shared with the investigators for future reference during data collection.

Two levels of quality assurance were put in place to ensure the completeness, consistency, and accuracy of data collected. First, each case leader reviewed the work of investigators working on the same country. Second, the study coordinator checked all the cases for quality, comparing the level of evidence gathered across cases to ensure comparability.

Figure 1.3. Template for Qualitative Data Collection for Each Case

Step in the causal chain	Outcome	World Bank Group contribution	Project entity capacity	Government contribution	Market environment	Policy environment
Step 1						
Step 2						
Step 3, ...						
Step 15						

Triangulation of

1. Multiple interviews
2. Document review
3. Data review
4. Site visit

Source: Independent Evaluation Group.

Endnotes

¹ For a more detailed explanation of the differences between various forms of literature review, see the Independent Evaluation Group's Methods Paper Series publication on conducting structured literature reviews in evaluation (Fenton Villar 2022).

2

ANALYSIS



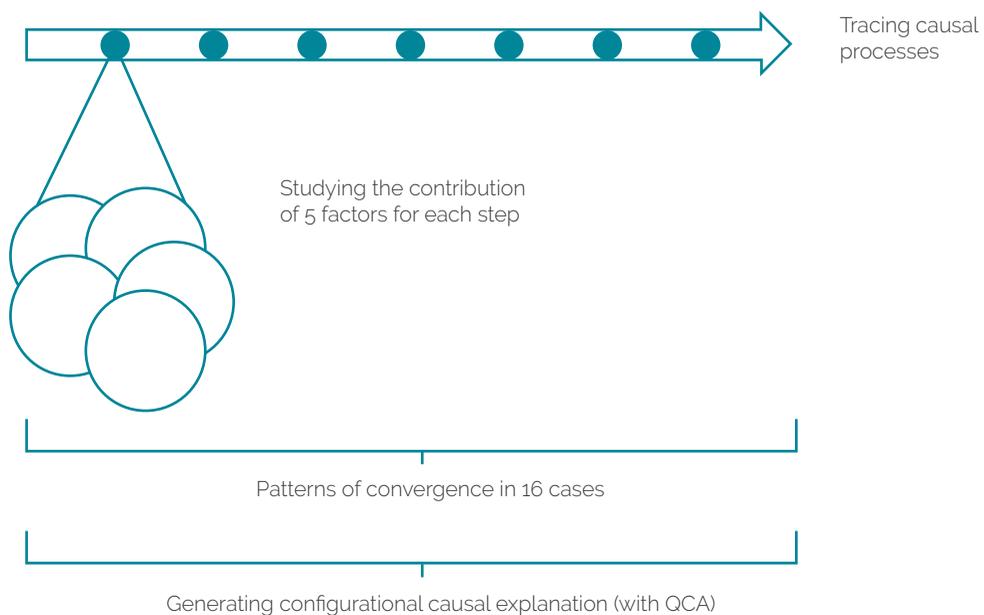
Structure of the
Causal Analysis



Qualitative Data
Processing in
Configurational Table

The causal analysis in the current study proceeded in four stages, as figure 2.1 shows. The first stage was a within-case causal analysis conducted for each of the 16 cases selected for review. For the purposes of this analysis, within-case evidence was defined as “evidence from within the temporal, spatial, or topical domain defined as a case” (Bennett and Checkel 2015, 8). The causal theory developed for the study was traced throughout each case, and the causal contributions of the five identified variables of interest at each step of the theorized causal process were systematically categorized as *present* or *absent*, based on rich descriptions developed for each case (as discussed earlier) using triangulated evidence from multiple interviews, document review, data review, and a site visit. Assessments of this evidence resulted in 16 independent case narratives about the causal contributions of the Bank Group and other contributory actors and variables, each including rich and deep description. These narratives and their accompanying case matrixes were collated in a database of qualitative data, which formed the basis for subsequent steps in the analysis.

Figure 2.1. Structure of the Causal Analysis



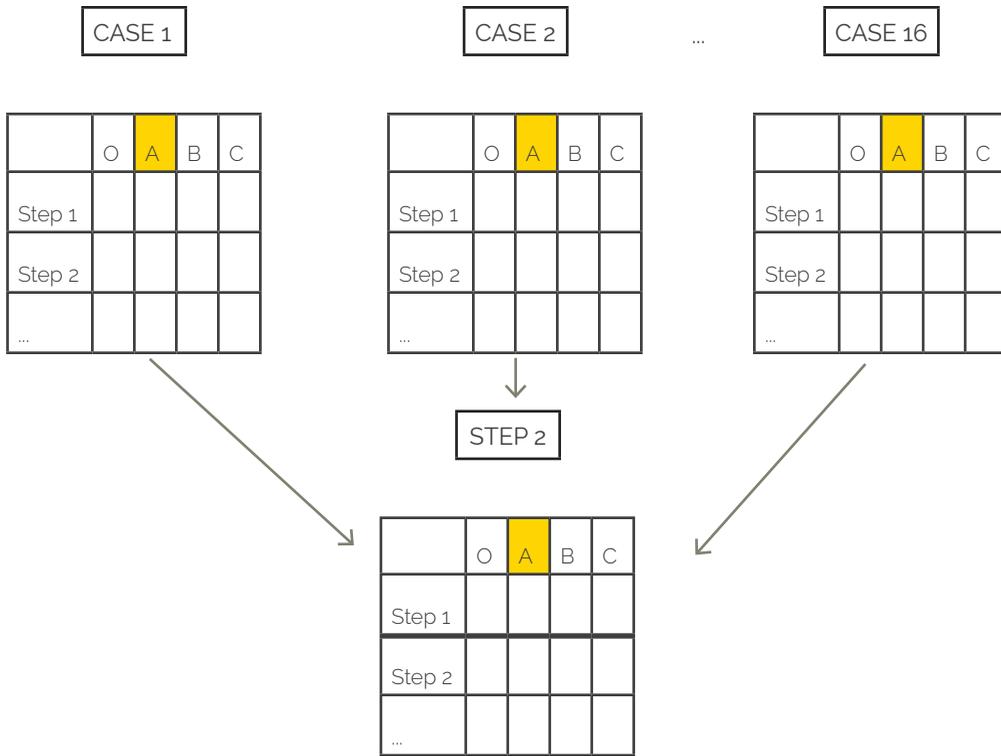
Source: Independent Evaluation Group.

Note: QCA = qualitative comparative analysis.

In the second stage, the study team performed a cross-case causal analysis, consisting of a systematic analysis of patterns of convergence and divergence across cases for each step in the theorized causal process. First, the team flipped the case matrixes to

create configurational tables (with the cases in the rows and the variables in the columns) for each of the 15 steps of the theorized causal process, as figure 2.2 illustrates.

Figure 2.2. Qualitative Data Processing in Configurational Table



Source: Independent Evaluation Group.

Cases were then grouped by outcome level and by level of Bank Group contribution. These groupings of cases ultimately elicited various patterns of regularity in the contribution of various variables to the outcomes at each step in the theorized causal process. For relevant causal steps, patterns of convergence by technology or by country capacity were also identified.

In the third stage of the analysis, the study team checked the empirical patterns that emerged from the cross-case comparison against the posited causal theory to determine those patterns that fit the theory (matches) and those that did not (mismatches). The causal theory in contributory hypothesis was formalized at this stage. In addition, in identified outlier cases (that is, cases in which explanatory variables other than those included in the causal theory were found that explained the outcomes of interest), the study team systematically checked whether the outliers constituted a refutation of the causal theory or instead illustrated a broken

causal process or unmet assumptions. During this stage, the study team put the case evidence in constant dialogue with the theory.

In the fourth and final stage of the analysis, given the causal complexity underlying the explanations of the five main outcomes of interest, the team formally tested the causal theory using crisp-set QCA, a well-established technique that uses Boolean minimization to “simplify complex data structures in a logical and holistic manner”¹ (Ragin 2014, viii). Application of the technique requires transforming the main outcome variable and other variables of interest into binary variables (that is, variables that can take the values 0 or 1). For the current study, the study team reformulated all of the variables that captured the contribution of key actors as either *high* or *low* and all variables that captured another explanatory variable as either *present* or *absent*. At this stage of the analysis, the team had gained adequate substantive knowledge about each case and adequate theoretical knowledge about the most relevant variables included in the analysis to be able to adjudicate consistently across all variables and cases.

In addition, during the transformation of the variables of interest into binary variables, there was also continuous dialogue between the cases and the data set to ensure proper calibration of the variables. For each outcome of interest, the study’s methods expert generated truth tables using the software fsQCA and Venn diagrams using the software Tosmana. (Both the truth tables and the Venn diagrams are illustrated in the next section.) In transforming the variables of interest into binary variables, the study team applied good-practice principles of QCA, notably by ensuring that once the variables were binary, the study included a mix of cases with positive and negative outcomes; that there was sufficient variation for each variable; and that there were no counterintuitive configurations (for example, configurations in which all 1 variables led to a 0 outcome). For three of the theorized causal steps, the Boolean minimizations yielded contradictory configurations (that is, similar configurations that yielded different outcomes); for those steps, the study team sought resolution of the contradictions by examining the coherence of the data included in the cases. The team found that all contradictions could be resolved by reconsidering the ratings or reexamining the way one of the variables was operationalized.

Endnotes

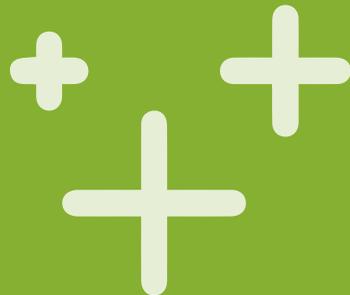
¹ In Boolean minimization, a long, complex expression is reduced to a more parsimonious expression. As it relates to the type of analysis used in this study, the process can be summarized as follows: “If two Boolean expressions (combining multiple factors) differ in only one causal variable yet produce the same outcome, then the causal variable that distinguishes the two expressions can be considered irrelevant and can be removed to create a simple, combined expression” (Ragin 2014, 83).

3

ILLUSTRATING FINDINGS



Qualitative Comparative
Analysis Venn Diagram
for Co-benefits



Pathways to
Co-benefits

This section illustrates the nature of the findings that were generated, using the approach described, by demonstrating how the causal analysis materialized for one of the outcomes of interest: generating co-benefits. (For a detailed report of the full results, see World Bank 2018.)

The literature has discussed the degree to which carbon finance projects foster local community co-benefits as a useful secondary benefit from their outputs (Hultman, Lou, and Hutton 2020). The methodological approach used in the current study sheds significant light on the causal pathways that led to the different outcomes among these projects in regard to achieving this second objective.

First, the patterns of causality emerging from the cross-case evidence in the current study echo the literature in finding that afforestation or reforestation projects have the most potential among the types of projects studied to generate significant local co-benefits. Indeed, across all countries included in the current study, afforestation or reforestation projects examined generated direct co-benefits to local communities. The projects in these cases were developed under the BioCarbon Fund, one of three World Bank carbon funds with an explicit objective of generating co-benefits for the communities involved. However, the afforestation or reforestation cases examined in this study have inherent characteristics that require providing incentives to local communities, more so than other technologies such as renewable energy. In some cases, this is because the entity carrying out the project needed to enter into lease agreements with landholders; in other cases, the entity was a rural development agency, and the project was part of a larger rural development program with the dual goals of improving or diversifying community livelihoods and enhancing environmental conditions via better land management practices and sustainable forest management.

Second, beyond the nature of the technology that characterized the carbon finance intervention, cross-case analysis identified several other contributory factors. For one, all projects carried out in Colombia across all four technologies studied and all hydraulic projects carried out in China generated co-benefits. In most of the other cases studied, however, the projects provided limited community co-benefits or none at all.

For another, the within-case causal analysis traced the theorized pathways to co-benefits and identified the unique contribution of the five actors and variables for each project with a high degree of specificity and many explanatory details. Once the cross-case analysis within each technology was conducted, patterns of regularity emerged, as did patterns of exception. The within-case causal analysis was necessary to explain outlier patterns. For example, almost none of the hydro projects examined generated direct co-benefits to the communities they served, the exception be-

ing the hydro project in China. That project was carried out in an area of China with significant populations of ethnic minorities, triggering World Bank environmental and social safeguards, which in turn triggered an additional causal mechanism. Specifically, revenues generated from the certified emission reduction achieved in the project were used to finance the Ethnic Minority Plan prepared to comply with the World Bank's Environmental and Social Framework safeguard policies. The plan included a range of beneficiaries, including a local health clinic, a temple, a road maintenance unit, and village education officials, thus generating co-benefits.

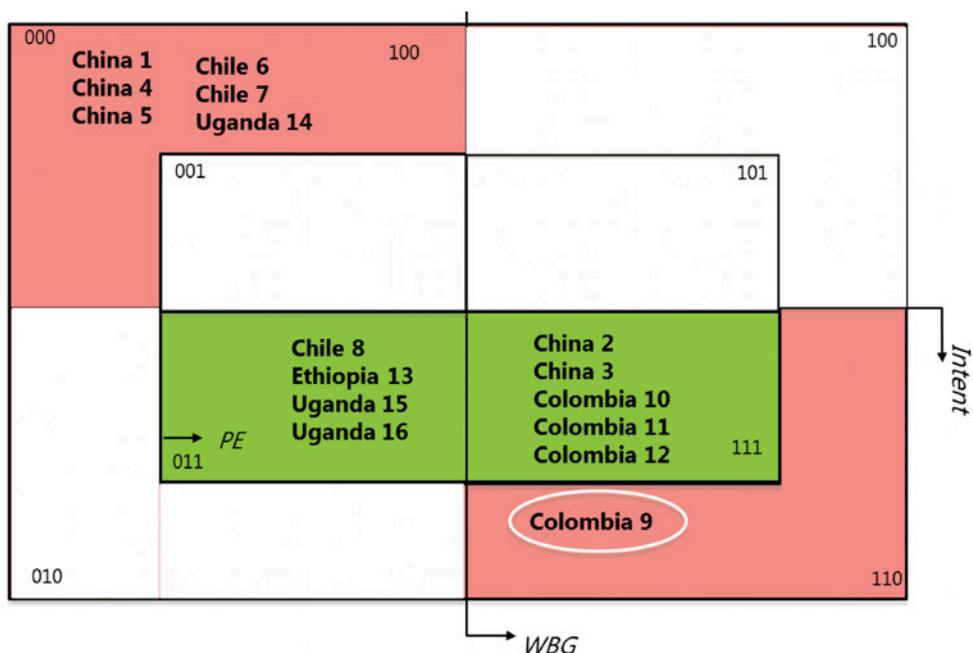
Analyses across cases within countries were also useful in identifying and testing the causal contribution of specific factors. For example, in all four cases in Colombia that were reviewed, the projects provided direct co-benefits to the communities the projects served, regardless of the technology involved. In all four cases, the World Bank's Environmental and Social Framework safeguards policies enhanced the projects' direct socioeconomic benefits to the communities they served. In the Jepiroche wind project, which was implemented in Indigenous peoples' territory, the World Bank played an extremely prominent role and provided incentives for the implementation of its Environmental and Social Framework safeguards policies by offering a premium price in the ERPA, contingent on high-quality implementation of those policies.

Finally, in addition to the patterns of regularity emerging from analysis by technology and country, application of the crisp-set QCA algorithm revealed two main causal pathways, one leading to positive outcomes and one leading to negative outcomes. As illustrated in figure 3.1 and table 3.1, the pathway to positive outcomes (the green boxes in the figure) combines a strong intent to achieve co-benefits at the project design stage with a demonstrated commitment, throughout the project, to achieving those co-benefits on the part of the entity carrying out the project. Within this causal pathway, local co-benefits were more likely to be achieved. In some cases, the World Bank was instrumental in ensuring that there was an explicit and deliberate intent to generate co-benefits at the project design stage, including through its safeguards policies, as noted earlier, specifically those regarding Indigenous peoples; however, the QCA results highlight that this was neither a necessary nor a sufficient condition for achieving co-benefits.

Conversely, when there was limited intent to provide project co-benefits, the entity conducting the project felt neither compelled nor committed to serving the community, and the World Bank had limited say in the project beyond ensuring compliance with safeguards, co-benefits were unlikely to be generated (the red boxes in the figure). The results from the case analysis are summarized in table 3.1. Note that table 3.1 summarizes the causal pathways to change as a function of two necessary

conditions (intent and commitment), highlighting adherence to those conditions across the cases explored. Furthermore, potential outliers in the analysis are also highlighted, as was the case with Colombia, which did not generate co-benefits despite the presence of both the intent and Bank Group support to do so.

Figure 3.1. Qualitative Comparative Analysis Venn Diagram for Co-benefits



Source: Independent Evaluation Group.

Note: Colombia 9 is an outlier case in which the causal process broke early on (explaining the lack of co-benefits, despite both intent to provide such benefits and World Bank Group support). PE = project entity.

Table 3.1. Pathways to Co-benefits

	Positive Outcome	Negative Outcome
Pathways to change	"Intent and commitment"	"No intent, no commitment, no pressure"
Boolean expression	INTENT × PE	intent × pe × wbg
Groups of cases	(China 2, China 3, Colombia 10, Colombia 11, Colombia 12) and (Chile 8, Ethiopia 13, Uganda 15, Uganda 16)	(China 1, China 4, China 5, Chile 6, Chile 7, Uganda 14)
Robustness measures	Consistency = 1.0 Coverage = 1.0	Consistency = 1.0 Coverage = 0.85

Source: Independent Evaluation Group.

Note: In Boolean language, expressions in capital letters mean "present," those in lowercase signify "absent." *Consistency* refers to the ratio of the number of cases in which a particular variable (or combination of variables) is present that are successful to the total number of cases in which that particular variable (or combination of variables) is present. It equals 1 when all the cases in which the particular variable (or combination of variables) is present are successful. *Coverage* refers to the ratio of the number of cases in which a particular variable (or combination of variables) is present to the total number of successful cases. When it equals 1, then the particular variable is not only sufficient but also necessary, for success: it is present in all successful cases.

CONCLUSION

This paper set out to bust two myths related to the selection of methods for meaningful inference in evaluation, moving away from the idea that any specific approach represents a gold standard for the quality of research and evidence generation. As noted earlier, evaluators have started to experiment with mixed methods approaches to causal analysis, combining theory-based methods that use case studies and other qualitative inputs as their primary empirical material. The use of contribution analysis, process tracing, and QCA, among other approaches, is increasingly supported in the literature, helping to refute the first myth—the misconception that causal claims can only be derived from quantitative or “large-n” methodologies in a counterfactual framework. As shown in this paper, case-based and theory-based methods can generate robust causal inferences and fill important knowledge gaps in evaluations. To refute the second myth, the analysis in this paper showed that findings from case-based work can be generalized to other contexts, thereby generating practical insights relevant to complex interventions and the conditions that influence their relative success.

As evaluation practitioners continue to experiment with a range of approaches and methods for answering a variety of causal questions regarding increasingly complex interventions, the need for guidance has increased for both project commissioners and project evaluators on what approach and method to select for a particular intervention and how to ensure that the approach and method are carried out thoroughly. Widner, Woolcock, and Ortega Nieto (2022) offer some useful principles for deciding whether and when case-based approaches can offer robust insights for drawing causal inferences and for determining how far those insights can be extended from the cases that provided them. These principles resonate with the experience of the team involved in carrying out the current study, which involved combining within- and cross-case analysis in the evaluation of a selection of projects in the World Bank’s carbon finance portfolio.

First, as Widner and colleagues highlight, it remains that “quantitative analysis of large numbers of discrete cases is more effective at estimating the strength of the relationship between causes and outcomes that can both be measured quantitatively” (Widner et al. 2022, 4). In that sense, if the primary causal question to be answered is, How much of an effect (on average) has a particular intervention had on a specific measurable outcome?, then case-based approaches of the kind described earlier in this paper will not provide a valid, useful answer. But, case-based approaches have other comparative advantages when it comes to

providing causal explanations and identifying the role contextual or implementation conditions play in successful or unsuccessful outcomes of interventions. Notably, case-based approaches help (i) identify causal mechanisms to open the black box processes connecting causes and outcomes; (ii) elicit how processes of change unfold; (iii) explain the circumstances under which causal mechanisms are or not triggered; and (iv) provide what Woolcock (2013, 2022) calls key facts for determining whether a particular intervention could work in other cases. As Woolcock (2013, 95) asserts, “the higher the complexity, the more salient (even necessary) inputs from analytic case studies become as contributors to the decision-making process” regarding whether particular interventions could be effectively scaled or replicated in other contexts.

That said, the promise of case-based approaches can be fulfilled only if certain conditions are met (Johnson and Rasulova 2017). First, evaluators should ensure the defensibility of the causal inferences they draw from the cases they study, with precisely specified causal theories, diligent consideration of alternative explanations, and assessment of the trustworthiness and probative value of the evidence brought to bear to support causal inference in the cases examined (Beach and Pedersen 2019; Cartwright 2022; Mahoney 2000).

Second, evaluators should carefully delimit the boundaries within which their generalization applies. In most instances, the degree of generalization will be modest (Rihoux and Ragin 2009), and it will be delimited to the class of cases that share the variables determined to be necessary or sufficient to trigger the causal mechanisms identified. In conducting case-based evaluations, the following five principles, inspired by and adapted from Widner, Woolcock, and Ortega Nieto (2022), are of particular importance:

1. Articulating a plausible causal theory that is informed by a thorough review of the literature and practitioners’ experience, is specific enough, and proposes plausible explanations for outcomes for interest and relevant alternatives to those explanations.
2. Selecting cases for study according to clear and transparent criteria that are pragmatic but do not yield too much to convenience. Researchers should try to include in their studies both cases with positive outcomes and those with negative outcomes.
3. Articulating clear hypotheses about a handful of contributory factors that will be the object of close scrutiny across the cases reviewed while leaving space for inductive inquiry and the possibility of stumbling on important additional factors to consider.

4. Providing evidence that has been carefully weighed, often triangulated across sources, and is considered trustworthy. Researchers should ensure that the evidence they provide is as unique as possible to the causal explanation they propose and should be transparent about alternative explanations that they cannot rule out.
5. Being open about the caveats and limitations and as transparent in the process as possible so that others can check or debate the conclusions reached

Now, even the most thorough case-based design has limitations. Testing a theory against a small number of chosen cases is inevitably a perilous exercise, especially when the number of cases available for study is limited and the number of causal factors that might explain the outcomes is large. Scenarios can quickly arise in which the complexity of phenomena overwhelms the number of observations. In such scenarios, careful reviews of the existing literature can often help narrow the causal field, but not always. Sometimes, exploratory process tracing should be undertaken first. Careful tracing in single cases can help reveal links among activities, actors, the ways they behave and influence others, and ultimately the outcomes of interest. The information these links provide on implementation challenges can also contribute to generating hypotheses about the variables that must hold for change to take place.

There are also undeniable practical challenges to carrying out case-based work that should not be underestimated. Time is often the scarcest resource and may preclude evaluation teams from going deep enough in their analysis to yield useful conclusions. Organizational politics can also be hard to navigate, especially regarding case selection, access to key informants, and what information can or cannot be used as evidence (Aston 2022).

BIBLIOGRAPHY

- AFD (Agence Française de Développement). 2022. “Evaluation d’impact—cartographie des usages.” AFD, Paris. <https://www.afd.fr/fr/ressources/evaluation-impact-cartographie-usages>.
- Aston, Thomas. 2022. “The Comeback of the Case Study?” *Medium*, January 11, 2022. <https://thomasmtaston.medium.com/the-comeback-of-the-case-study-461c441fe89d>.
- Beach, Derek, and Rasmus Brun Pedersen. 2019. *Process-Tracing Methods*. Ann Arbor: University of Michigan Press.
- Befani, Barbara. 2012. “Modes of Causality and Causal Inference.” Background paper in Elliot Stern, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies, and Barbara Befani, “Broadening the Range of Designs and Methods for Impact Evaluation: Report of a Study Commissioned by the Department for International Development,” Working Paper 38, Department for International Development, London, A1–A12.
- Befani, Barbara. 2016. *Pathways to Change: Evaluating Development Interventions with Qualitative Comparative Analysis (QCA)*. Report 2016:05. Stockholm: Expert Group for Aid Studies. <https://eba.se/en/reports/pathways-to-change-evaluating-development-interventions-with-qualitative-comparative-analysis-qca/4157/>.
- Befani, Barbara. 2021. *Credible Explanations of Development Outcomes: Improving Quality and Rigour with Bayesian Theory-Based Evaluation*. Report 2021:03. Stockholm: Expert Group for Aid Studies. <https://eba.se/en/reports/cridible-explanations-of-development-outcomes-with-bayesian-theory-based-evaluation/17287/>.
- Bennett, Andrew. 2022. “Drawing Contingent Generalizations from Case Studies.” In *The Case for Case Studies: Methods and Applications in International Development*, edited by Jennifer Widner, Michael Woolcock, and Daniel Ortega Nieto, 195–218. Cambridge, UK: Cambridge University Press. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/31D76BE9C-37D459E2B153D43C4B3B647/9781108427272AR.pdf/The_Case_for_Case_Studies.pdf?event-type=FTLA.
- Bennett, A., and Checkel, J. T., eds. 2015. *Process Tracing*. Cambridge University Press.
- Bhaskar, Roy. 1975. *A Realist Theory of Science*. Abingdon, UK: Routledge.
- Cartwright, Nancy. 2004. “Causation: One Word, Many Things.” *Philosophy of Science* 71 (5): 805–819. doi:10.1086/426771.

- Cartwright, Nancy. 2007. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge, UK: Cambridge University Press.
- Cartwright, Nancy. 2022. "How to Learn about Causes in the Single Case." In *The Case for Case Studies: Methods and Applications in International Development*, edited by Jennifer Widner, Michael Woolcock, and Daniel Ortega Nieto, 29–51. Cambridge, UK: Cambridge University Press. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/31D76BE9C37D459E2B153D43C4B-3B647/9781108427272AR.pdf/The_Case_for_Case_Studies.pdf?event-type=FTLA.
- Delahais, Thomas, and Jacques Toulemonde. 2012. "Applying Contribution Analysis: Lessons from Five Years of Practice." *Evaluation* 18 (3): 281–293. doi:10.1177/1356389012450810.
- De Meur, Gisèle, and Benoit Rihoux. 2002. *L'Analyse quali-quantitative comparée (AQQC QCA): approche, techniques et applications en sciences humaines*. Louvain-la-Neuve, Belgium: Academia-Bruylant.
- Dixon, Vibecke, and Michael Bamberger. 2022. "Incorporating Process Evaluation into Impact Evaluation: What, Why and How." 3ie Working Paper 50, International Initiative for Impact Education, New Delhi.
- Fenton Villar, Paul. 2022. "Structured Literature Reviews: Building Transparency and Trust in Standards of Reporting Evidence." IEG Methods and Evaluation Capacity Development Working Paper Series. Independent Evaluation Group, World Bank, Washington, DC.
- Flyvbjerg, Bent. 2006. "Five Misunderstandings about Case-Study Research." *Qualitative Inquiry* 12 (2): 219–245. doi:10.1177/1077800405284363.
- Glennan, Stuart S. 1996. "Mechanisms and the Nature of Causation." *Erkenntnis* 44: 49–71. doi:10.1007/BF00172853.
- Gürerk, Özgür, Andrea Bönsch, Lucas Braun, Christian Grund, Christine Harbring, Thomas Kittsteiner, and Andreas Staffeldt. 2014. "Experimental Economics in Virtual Reality." MPRA Paper 66617, Munich Personal RePEc Archive, Munich. <https://mpa.ub.uni-muenchen.de/66617/>.
- Hanckel, Benjamin, Mark Petticrew, James Thomas, and Judith Green. 2021. "The Use of Qualitative Comparative Analysis (QCA) to Address Causality in Complex Systems: A Systematic Review of Research on Public Health Interventions." *BMC Public Health* 21: 877. doi:10.1186/s12889-021-10926-2.
- Hultman, Nate, Jiehong Lou, and Stephen Hutton. 2020. "A Review of Community Co-benefits of the Clean Development Mechanism (CDM)." *Environmental Research Letters* 15 (5): 053002. doi:10.1088/1748-9326/
- Jimenez, Emmanuel, Hugh Waddington, Neeta Goel, Audrey Prost, Andrew Pullin, Howard White, Shaon Lahiri, and Anmol Narain. 2018. "Mixing and Matching: Using Qualitative Methods to Improve Quantitative Impact Evaluations (IEs)

- and Systematic Reviews (SRs) of Development Outcomes.” *Journal of Development Effectiveness* 10 (4): 400–421. doi:10.1080/19439342.2018.1534875.
- Johnson, Susan, and Saltanat Rasulova. 2017. “Qualitative Research and the Evaluation of Development Impact: Incorporating Authenticity into the Assessment of Rigour.” *Journal of Development Effectiveness* 9 (2): 263–276. doi:10.1080/19439342.2017.1306577.
- Kane, Robin, Carlisle Levine, Carlyn Orians, and Claire Reinelt. 2021. “Contribution Analysis: A Promising Method for Assessing Advocacy’s Impact.” *New Directions for Evaluation* 2021 (171): 45–57. doi:10.1002/ev.20471.
- Kazi, M. A. 2003. *Realist Evaluation in Practice*. London: SAGE. <https://dx.doi.org/10.4135/9781849209762>.
- Mahoney, James. 2000. “Strategies of Causal Inference in Small-N Analysis.” *Sociological Methods & Research* 28 (4): 387–424. doi:10.1177/0049124100028004001.
- Marx, A., B. Rihoux, and C. Ragin. 2014. “The Origins, Development, and Application of Qualitative Comparative Analysis: The First 25 Years.” *European Political Science Review* 6 (1): 115–142.
- Pawson, Ray. 2013. *The Science of Evaluation: A Realist Manifesto*. Los Angeles: Sage.
- Quadrant Conseil. 2017. “How Can Impact Be Evaluated?” <https://www.quadrant-conseil.fr/ressources/impacttree.html>.
- Ragin, C. C. 2014. *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies*. Berkeley and Los Angeles: University of California Press.
- Raimondo, Estelle. 2020. “Getting Practical with Causal Mechanisms: The Application of Process-Tracing under Real-World Evaluation Constraints.” *New Directions for Evaluation* 2020 (167): 45–58. doi:10.1002/ev.20430.
- Raimondo, Estelle, and Derek Beach. Forthcoming. “Process Tracing Methods in Evaluation.” In *Research Handbook on Program Evaluation*, edited by Kathryn E. Newcomer and S. Mumford. Cheltenham, UK: Elgar.
- Ravallion, Martin. 2020. “Should the Randomistas (Continue to) Rule?” Working Paper 27554, National Bureau of Economic Research, Cambridge, MA. <https://www.nber.org/papers/w27554>.
- Rihoux, Benoit, and Chares C. Ragin. 2009. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Applied Social Research Methods. Los Angeles: Sage. doi:10.4135/9781452226569.
- Rothgang, Michael, and Bernhard Lageman. 2021. “The Unused Potential of Process Tracing as Evaluation Approach: The Case of Cluster Policy Evaluation.” *Evaluation* 27 (4): 527–543. doi:10.1177/13563890211041676.
- Rowe, Andy. 2019. “Rapid Impact Evaluation.” *Evaluation* 25 (4): 496–513. <https://doi.org/10.1177/1356389019870213>.

- Schmitt, J. 2020. "The Causal Mechanism Claim in Evaluation: Does the Prophecy Fulfill?" *New Directions for Evaluation* 2020 (167): 11–26.
- Stern, Elliot, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies, and Barbara Befani. 2012. "Broadening the Range of Designs and Methods for Impact Evaluation: Report of a Study Commissioned by the Department for International Development." Working Paper 38, Department for International Development, London.
- Ton, Giel, John Mayne, Thomas Delahais, Jonny Morell, Barbara Befani, Marina Apgar, and Peter O'Flynn. 2019. "Contribution Analysis and Estimating the Size of Effects: Can We Reconcile the Possible with the Impossible?" Practice Paper 20, Centre for Development Impact, Institute of Development Studies, Brighton, UK. <https://www.ids.ac.uk/publications/contribution-analysis-and-estimating-the-size-of-effects-can-we-reconcile-the-possible-with-the-impossible/>.
- Trochim, W. M. 1985. "Pattern Matching, Validity, and Conceptualization in Program Evaluation." *Evaluation Review* 9 (5): 575–604.
- Trochim, W. M. 1989. "Outcome Pattern Matching and Program Theory." *Evaluation and Program Planning* 12(4): 355–366.
- Widner, Jennifer, Michael Woolcock, and Daniel Ortega Nieto, eds. 2022. *The Case for Case Studies: Methods and Applications in International Development*. Cambridge, UK: Cambridge University Press. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/31D76BE9C37D459E2B153D43C4B-3B647/9781108427272AR.pdf/The_Case_for_Case_Studies.pdf?event-type=FTLA.
- Woolcock, M. 2013. "Using Case Studies to Explore the External Validity of 'Complex' Development Interventions." *Evaluation* 19 (3): 229–248.
- Woolcock, M. 2022. "Will It Work Here? Using Case Studies to Generate 'Key Facts' About Complex Development Programs." In *The Case for Case Studies: Methods and Applications in International Development*, edited by Jennifer Widner, Michael Woolcock, and Daniel Ortega Nieto, 87–116. Cambridge, UK: Cambridge University Press.
- World Bank. 2018. *Carbon Markets for Greenhouse Gas Emission Reduction in a Warming World*. Independent Evaluation Group. Washington, DC: World Bank.



IEG
INDEPENDENT
EVALUATION GROUP

WORLD BANK GROUP
World Bank • IFC • MIGA

The World Bank
1818 H Street NW
Washington, DC 20433