

Natural Experiments in Management Research

Simone Santoni^{•◦} and Jost Sieweke^{*}

[•]Bayes Business School (formerly Cass)

[◦]Soundcloud

^{*}Vrije Universiteit Amsterdam

June 24, 2021

(Structured draft — do not circulate)

INTRODUCTION

The quest for empirical identification in management research has created substantial attention around ‘natural experiments,’ a form of causal inquiry that has been traditionally popular in economics (Meyer1995; Rosenzweig2000) and political science (Dunning2008). The premise to conduct a natural experiment is the presence of a ‘naturally’ occurring event — such as new regulations and laws, natural disasters, or economic and political crises — that heterogeneously influences the units of a population (dunning2012; Robinson2009). Insofar as such event generates random or as-if random variations in the environment, scholars can mimic the experimental ideal in which units are split into a treatment and a control group or receive different levels of the treatment. Ultimately, this opens up the possibility of inferring causal effects when the substantive relationship at hand is difficult to investigate in a laboratory setting and/or require operating costly, impractical, or unethical field experiments.

Although naturally occurring events can turn into opportunities to conduct causal research, there are limited guidelines that help management scholars prepare and review papers that implement the natural experiment research design. To fill this gap, we highlight the strengths and weaknesses of natural experiments as operated in the field of management studies and propose actionable suggestions to assess and communicate the validity of natural experiments.

To do so, we critically review the population of 147 natural experiments published across seventeen top-tier management journals. Our review aims to address the following research questions: *R1 — How do management scholars claim the random or as-if random nature of environmental variation at the core of a natural experiment?* *R2 — How do they claim the empirical and substantive relevance of a natural experiment?* *R3 — How do they claim the credibility of the statistical model encapsulated in a natural experiment design?*

This work is organized as follows. The next two sections briefly introduce the key features of the ‘standard natural experiment,’¹ along with the evaluative framework we use to analyze the individual natural experiments. The following section describes the selection of the reviewed studies. Then, we present the key insights that emerge from our analysis and conclude with a suggested checklist that helps management scholars to exploit the opportunities of causal inference offered by naturally occurring events. The online appendix presents an integrated set of Python scripts to implement standard natural experiments and to assess their validity.

NATURAL EXPERIMENTS AND CAUSAL INFERENCE

The standard natural experiment resembles the design of a randomized experiment. Naturally occurring events (such as earthquakes (Belloc2016)) are supposed to determine a statistical unit’s treatment status. As show in Figure 1, the availability of longitudinal data makes possible to estimate the causal effect of the treatment by

¹ In his comprehensive, cross-disciplinary analysis of the literature, Dunning dunning2012 identifies three forms of natural experiments: Standard natural experiments; instrumental variables (Angrist, 1990); regression discontinuity designs (Thistlethwaite & Campbell, 1960). In the interest of clarity and integrity, our review concentrates on standard natural experiments, whose origin goes back to the highly acclaimed and impactful research Dr John Snow (Snow, 1855) conducted on the diffusion of cholera in the mid 19th century London. In this paper we use the term ‘natural experiment’ to exclusively refer to standard natural experiments.

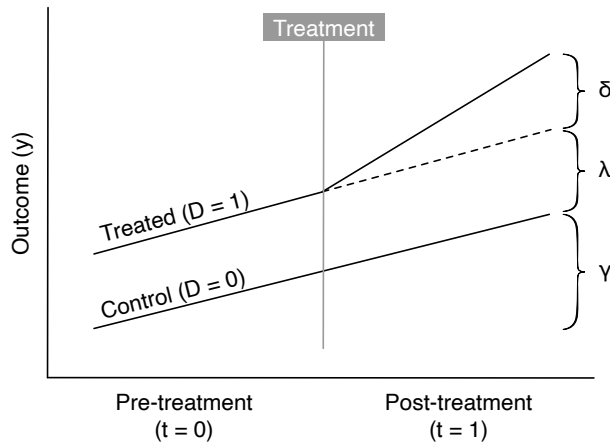


Figure 1: Visual representation of the standard natural experiment. The underlying population regression function is $y = \gamma t + \lambda D + \delta tD$, where γ , λ , and δ represent the systematic difference in the outcome across the treated and control cases, the trend effect, and the difference in the outcome that is due to the treatment. For the sake of clarity, we represent the case in which $\delta > 0$.

contrasting the pre-post change in the outcome variable y across the control group (γ) and the treatment group ($\gamma + \delta$).

Assessing the Validity of Natural Experiment Designs

How do (management) scholars evaluate a standard natural experiment research design? According to Dunning (**dunning2012**), the validity of a natural experiment should be assessed against three criteria (see Figure 2). First, scholars should prove the random nature of the treatment or, at least, defend the plausibility of as-if random. In the *randomized standard natural experiment*, it is important that the assignment process is truly random. Although this may seem obvious, this condition is sometimes violated, even in the context of lotteries (**Starr1997**). In the case of *as-if randomization*, it is vital the assignment process is: i) independent of factors that are related to the outcome, and ii) not affected by unit's self-selection into treatment or control conditions. As Dunning points out, the researcher has to make a compelling case for this assertion (or refrain from claiming he/she conducts a natural experiment). In-depth knowledge of the context (e.g., industry regulatory frameworks), qualitative evidence about the naturally occurring event (e.g., a new law), and quantitative evidence at the event- and unit-level are essential ingredients to defend the plausibility of as-if random assignment, and, ultimately, to sustain the natural experiment.

Second, the naturally occurring event should reveal the wider “*theoretical, substantive, and/or policy issues*” (**Dunning2012**) that motivate the study. For example, the sudden, premature death of a star scientist (**Azoulay2010**) create the premises for a natural experiment that quantifies the spill-over effect of collaborating with academics who are prominent in their fields of research.

Finally, the statistical model should fit with the characteristics of the naturally occurring event. In the case of a randomized standard natural experiment, simplicity and transparency should take precedence in the data analysis stage. Particularly, Neyman's potential outcomes framework (**Splawa1990**), namely, a treated Vs. control mean comparison test, should be used *prima facie*. At the same time, some

statistical adjustments may be required even in presence of a random (or as-if random) treatment. For example, the Stable-Unit-Treatment-Value-Assumption (SUTVA) may be violated insofar as the treatment status of a unit i interferes with the potential outcome of unit j . Such a concern is central in the Belloc and colleagues's (Belloc2016) study about the impact of earthquakes on institutional change at the city-level in the Middle-Ages northern and central Italy. Both the distribution and timing of earthquakes are random. However, the probability a control city will move from autocratic regimes to self-government is also a function of the information key actors are exposed to, such as the transition choices treated neighboring cities make. In this case, scholars may want to adopt some statistical adjustments to model the correlation of residuals induced by the geographical proximity of any pair of units.

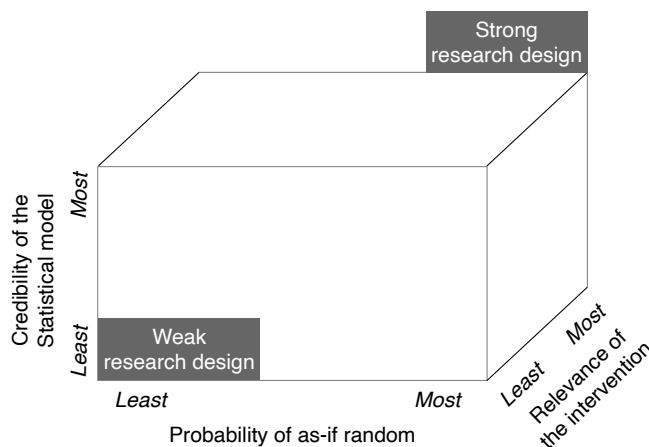


Figure 2: Visual representation of Dunning's validity framework (dunning2012)

DATA & METHODS

Sample of Studies

Consistently with review articles recently published in the Journal of Management (gonzalez2018; rindova2018), we restrict our literature review to a selection of prominent journals such as Academy of Management Journal, Administrative Science Quarterly, Entrepreneurship Theory and Practice, Journal of Business Ethics, Journal of Business Venturing, Journal of Management, Journal of Management Studies, The Leadership Quarterly, Management Science, Organization Science, Organization Studies, Research Policy, Strategic Entrepreneurship Journal, Strategic Management Journal, Strategic Organization. Using the search engine embedded in each journal's web page (see Table A1) reported in the Appendix), we searched for articles published after January 2000 that show the token "natural experiment*" in the full text.

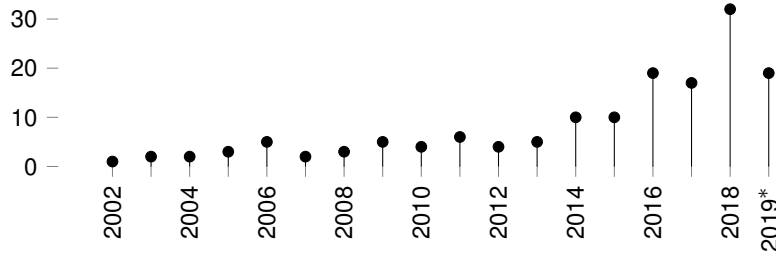


Figure 3: Inter-temporal distribution of natural experiment studies. The 2019 bucket, marked with an asterisk, contains both published and accepted, online articles.

We retrieved 499 publications, 201 of which were eventually included in the review. Figure A5, reported in the Appendix, illustrates the counts of excluded studies across the various categories. For example, We excluded 33 non-empirical publications, such as theoretical articles ([makadok2011](#)) or review articles ([shaver2020](#)), and articles that: i) recall the empirical evidence produced by previous natural experiments, ii) indicate natural experiments as a possible way to overcome the limitations or expand on the study at hand ([hsu2006](#)), iii) use the instrumental variable design ([zoloty2018](#)) or the regression discontinuity design ([flammer2015](#)). Finally, we excluded twelve qualitative studies adopting the logic of natural experiment ([powell2017](#)).

Furthermore, we filtered out articles whose authors claim to adopt a natural experiment research design, whereas, in fact, they operate: (i) correlational designs on observational data ($N = 53$), data produced in the context of business simulations ($N = 1$), or game shows ($N = 2$); (ii) field experiments ($N = 1$); (iii) quasi-experiment/matching designs ($N = 3$); (iv) twin studies ($N = 4$). Figure 3 reports the inter-temporal distribution of the retained studies.

Coding Schema

RESULTS

Assessing the As-If Random Nature of the Treatment

This section focuses on the diagnostics that can be used to assess and argument the (as-if) random nature of the environmental variation at the center of the natural experiment. Specifically, this section surveys and articulates the following:

- diffusion/role of qualitative diagnostics to appreciate:
 - units' information about the treatment
 - units' incentives to self-select into the treatment (control) group
 - unit's capacity to self-select into the treatment (control) group
- diffusion/role of quantitative diagnostics (e.g., balance test) to compare and contrast treated and control units along relevant dimensions

Domain	Variable	Values of the Variable
Study features	Demographics	Authors; journal; year
	Level of theorizing	Between- or within-units design
Research design features	Level of the empirical model	Between- or within-units design
	Nature of the variation	Random; as-if random; not random
	Analytical strategy	DiD; mean-comparison; IV
	Natural experiment source	Open coding (e.g., CEO sudden death)
	Conceptual development level	Single-, multi-, or cross-level
Plausibility of as-if random	Role of information	False/True
	Role of incentives	False/True
	Individual capacity to self-select	False/True
Relevance of the intervention	Justification for the natural experiment	Methodological role; substantive role; both
	LATE interpretation	False/True
Credibility of statistical model	Statistical adjustment via covariates	False/True
	Statistical adjustment via matching	False/True
	Derivation of standard errors	Standard s.e.; clustered s.e.
	SUTVA considerations	False/True
Use of qualitative evidence	To sustain relevance of the intervention	False/True
	To sustain plausibility of as-if-random	False/True

Table 1: Coding schema.

Assessing the Relevance of the Treatment

This section focuses on the empirical and substantive relevance of the environmental variation at the center of a natural experiment. Particularly, this section reviews and discusses the following:

- diffusion/role of qualitative diagnostics to show:
 - the empirical, substantive, and policy relevance of the natural experiment
 - the external validity (non-idiosyncrasy) of the natural experiment
 - exclusion of ‘bundling of treatments’ (i.e., environmental variations affecting the outcome through multiple causal pathways)
- diffusion/role of placebo tests supporting the magnitude of the average treatment estimation on the treated (ATT)
- diffusion/role of local average treatment estimation (LATE) considerations

Assessing the Credibility of the Statistical Model

This section focuses on the credibility of the statistical model encapsulated in the natural experiment design. Specifically, this section surveys and articulates the following:

- diffusion/rationale of model based adjustments (instead of simple mean-comparison tests):
 - adjustment via control covariates

- adjustment via matching
- diffusion of SUTVA considerations and associated model adjustments (see the derivation of standard errors)

GUIDELINES

This section wraps up around the literature review results and provides actionable guidelines to better leverage the natural experiment design. The preliminary analysis of the coded data ² seem to indicate future natural experiments could:

- better integrate qualitative evidence and institutional knowledge in order to establish the as-if random nature of the treatment
- provide a more systematic discussion of the conditions under which a treatment can plausibly be considered as-if random (see the point on units' information, incentives, and capacity to self-select into the treatment group)
- pay equal attention to the empirical and substantive relevance of the treatment (that is, the possibility to reveal and/or detail important theoretical mechanisms by exploiting naturally-occurring events)
- provide a thorough assessment of the strengths and weaknesses of relying on a certain naturally-occurring event — i.e., explaining what the pros and cons are in terms of empirical identification (see LATE aspects) and theorizing opportunities
- use model-based adjustments (such as matching and control covariates) when there is no ground to establish the (as-if) random nature of the treatment. Indeed, the comparative advantage of natural experiments over alternative designs (e.g., quasi-experiments) also comes from the possibility to conduct causal inference by means of simple, transparent statical models. In other words, there should be good reasons to move from a design-based causal inference strategy to a model-based one (e.g., piggybacking on models that jointly use matching, DiD, and a long list of control covariates)
- consider the interactions among as-if random, relevance, and credibility elements. For example, the credibility of a model should be assessed against the nature of the treatment (random, as-if random, not random) and the process through which it is administered (see SUTVA)

² **TODO:** We have already coded the 147 studies in the sample. Although further analyses are needed to reveal clear patterns, some interesting elements seem to emerge.

CODA

.

REFERENCES

APPENDIX A — LITERATURE SEARCH

Reproducibility of the literature search

On June 19, 2021, we retrieved the candidate studies for the review using the search tool available in each journal's website. This allowed us to look for the instances in which the 'natural experiment*' token Table A1 reports the set of urls we visited. For the journals published by 'John Wiley & Sons' we were not allowed to compose our search query using the wild-car symbol '*.' Therefore, to ensure the commensurability of results across journals, we run two separate queries, namely, 'natural experiment' and 'natural experiments.'

Journal	Address of the search page
Academy Of Management Journal .	https://journals.aom.org/search/advanced
Administrative Science Quarterly...	https://journals.sagepub.com/search/advanced?SeriesKey=asqa
Entrepreneurship Theory & Practice	https://journals.sagepub.com/search/advanced?SeriesKey=etpb
Journal of Business Ethics.....	https://link.springer.com/search?query=&search-within=Journal&facet-journal-id=10551
Journal of Business Venturing.....	https://www.sciencedirect.com/journal/journal-of-business-venturing
Journal of Management.....	https://journals.sagepub.com/search/advanced?SeriesKey=joma
Journal of Management Studies....	https://onlinelibrary.wiley.com/search/advanced?publication=14676486&text1=
Management Science.....	https://pubsonline.informs.org/action/doSearch?SeriesKey=mns
Organization Science.....	https://pubsonline.informs.org/action/doSearch?SeriesKey=ors
Organization Studies.....	https://journals.sagepub.com/search/advanced?SeriesKey=oss
Research Policy	https://www.sciencedirect.com/journal/research-policy
Strategic Entrepreneurship Journal.	https://onlinelibrary.wiley.com/search/advanced?publication=1932443x&text1=
Strategic Management Journal....	https://onlinelibrary.wiley.com/search/advanced?publication=10970266&text1=
Strategic Organization	https://journals.sagepub.com/search/advanced?SeriesKey=soq
The Leadership Quarterly.....	https://www.sciencedirect.com/journal/the-leadership-quarterly

Table A1: Sample of target journals along with search page addresses.

Sampling

To consistently sample the studies for the review, two co-authors independently went through a random sample of twenty items and created a tentative set of exclusion categories. Then, the whole team of co-authors collectively evaluated the codes included in each set of exclusion categories, normalized the codes, and created the sampling workflow portrayed in Figure A4.

First, we removed non-empirical works, such as conceptual studies (eden2021), literature reviews (shaver2020), meta-analysis (geyskens2006), calls for papers (jacquart2020), or editorial notes (breschiz2020). With few exceptions (sieweke2020), the NE design plays a peripheral role in this category's studies. Second, we removed ten qualitative empirical works that denote the empirical setting as a 'natural experiment' () or that mirror the logic of a natural experiments (powell2017), a case study design that Eisenhardt recently labeled "racing design" since "*cases (often ventures) begin at the same time with similar initial conditions like founders, location, and funding, and then 'race' to some natural endpoint like IPO, unicorn valuation, or other temporal*

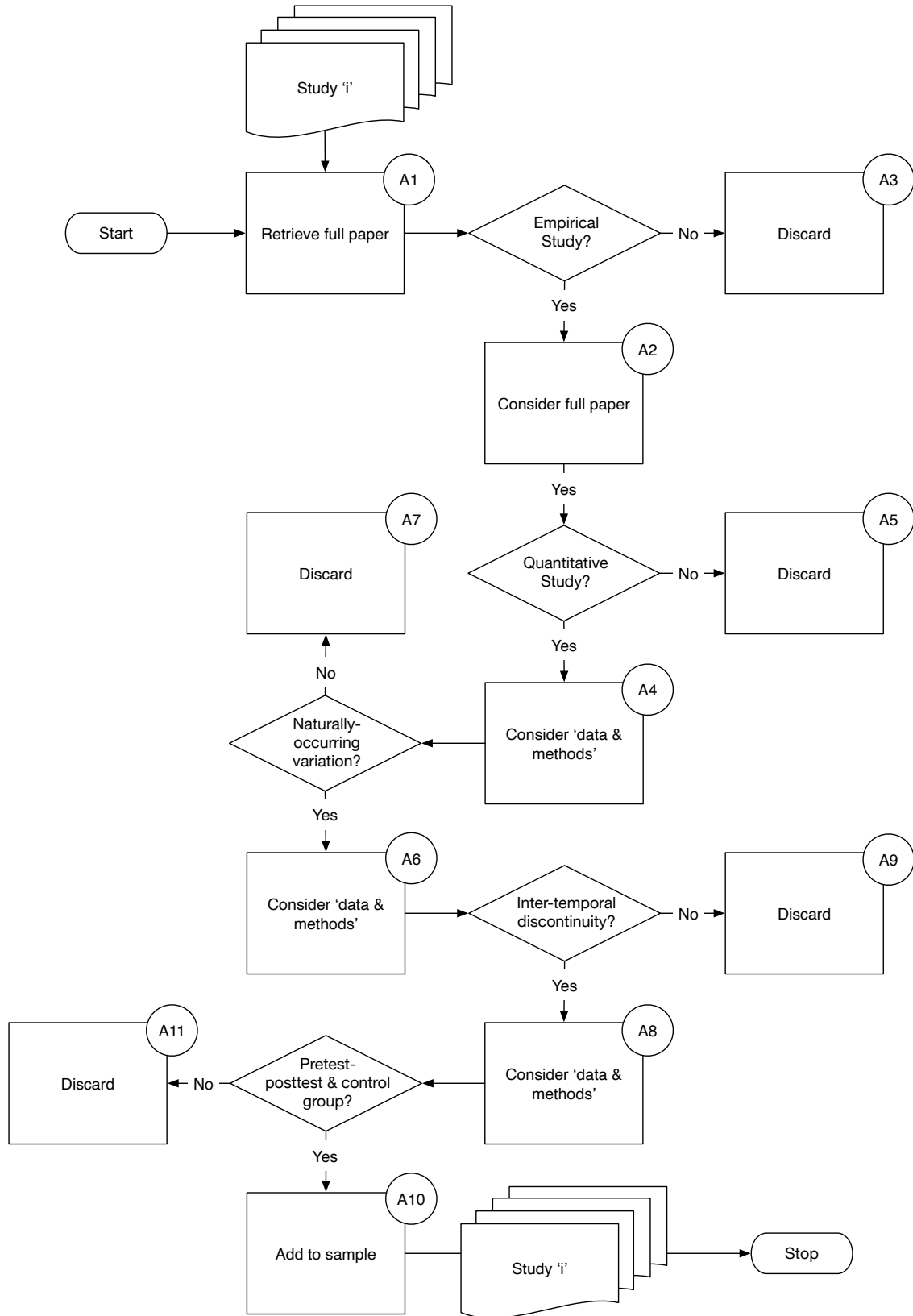


Figure A4: Sampling workflow. Notes.
— Circles denote activities.

marker.” (eisenhardt2021). Third, we removed quantitative empirical works that limit to refer to previous natural experiment studies (stevens2021), discuss a project’s limitations in regard to the lack of a natural experiment (chen2020), or indicate natural experiments as a future research avenues (xie0000). Fourth, we removed empirical works that claim to use a natural experiment whereas they do not exploit any sudden naturally occurring, sudden variations in the data. Such a category includes studies that — in fact — use field experiments (lee2017), quasi-experiments (azoulay2014), laboratory/online experiments (laurieromartinez2014), twin research design (nicolaou2008), or correlational research designs (boyle2011). Fifth, we removed empirical works that exploit a naturally occurring variation to operate the instrumental variable (zolotoy2018) or the regression discontinuity research design (flammer2015). Finally, we removed empirical works that exploit a sudden, naturally-occurring variation but lack either a comparison group (corbo2016) or pre-treatment data (desjardine2019).

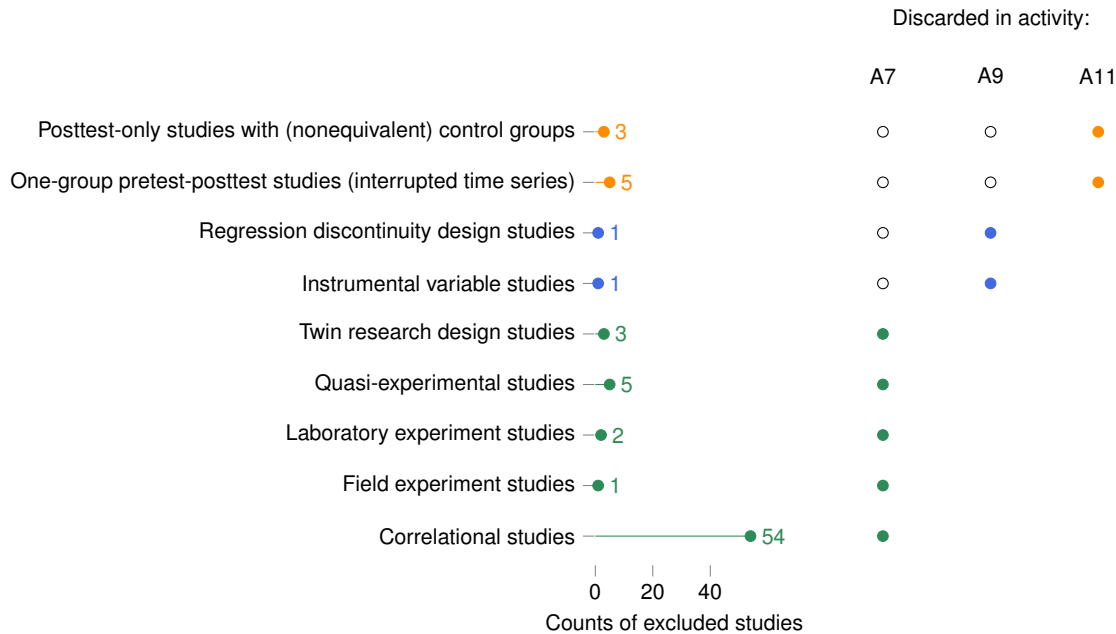


Figure A5: Distribution of the retrieved studies across exclusion categories. Notes. — As per Figure A4, in activity ‘A7,’ we discard the studies that does not present a naturally occurring variation; in ‘A9,’ we discard the studies whose naturally occurring variation is not associated with any inter-temporal discontinuities; in ‘A11,’ we discard the studies that lack either a pretest observation or a control case.

References

APPENDIX B — SAMPLE OF STUDIES

References

APPENDIX C — NATURAL LANGUAGE PROCESSING

References














Counts of articles by publication year																								
Journal	Trend	'00	'01	'02	'03	'04	'05	'06	'07	'08	'09	'10	'11	'12	'13	'14	'15	'16	'17	'18	'19	'20	'21	
Academy of Management Journal		0	10	12	102	0	1	0	0	0	0	1	0	12	102	0	1	0	0	0	0	1	0	
Academy of Management Journal		0	10	12	102	0	1	0	0	0	0	1	0	12	102	0	1	0	0	0	0	1	0	
Academy of Management Journal		0	10	12	102	0	1	0	0	0	0	1	0	12	102	0	1	0	0	0	0	1	0	
Academy of Management Journal		0	10	12	102	0	1	0	0	0	0	1	0	12	102	0	1	0	0	0	0	1	0	
Academy of Management Journal		0	10	12	102	0	1	0	0	0	0	1	0	12	102	0	1	0	0	0	0	1	0	
Academy of Management Journal		0	10	12	102	0	1	0	0	0	0	1	0	12	102	0	1	0	0	0	0	1	0	
Academy of Management Journal		0	10	12	102	0	1	0	0	0	0	1	0	12	102	0	1	0	0	0	0	1	0	
Academy of Management Journal		0	10	12	102	0	1	0	0	0	0	1	0	12	102	0	1	0	0	0	0	1	0	
Academy of Management Journal		0	10	12	102	0	1	0	0	0	0	1	0	12	102	0	1	0	0	0	0	1	0	
Academy of Management Journal		0	10	12	102	0	1	0	0	0	0	1	0	12	102	0	1	0	0	0	0	1	0	
Academy of Management Journal		0	10	12	102	0	1	0	0	0	0	1	0	12	102	0	1	0	0	0	0	1	0	
Academy of Management Journal		0	10	12	102	0	1	0	0	0	0	1	0	12	102	0	1	0	0	0	0	1	0	
Academy of Management Journal		0	10	12	102	0	1	0	0	0	0	1	0	12	102	0	1	0	0	0	0	1	0	

Table B1: Distribution of the sampled studies over time and across journals.