Tell me more.

Findings from a Scoping Review on the Reporting of Field Experiments in Sociology

Gianluca Argentin, University of Milan Bicocca

Giovanni Abbiati, University of Brescia

Davide Azzolini, Fondazione Bruno Kessler (FBK-IRVAPP)

Gabriele Ballarino, University of Milan

Davide Cartagini, University of Milan

Elisa Manzella, University of Milan Bicocca

Marta Pellegrini, University of Cagliari

Loris Vergolini, University of Bologna

Keywords: Randomized Controlled Trials (RCTs); Field Experiments; Causal Inference; Research Transparency; Sociological Methods

Corresponding author:

Gianluca Argentin

Department of Sociology and Social Research, Via Bicocca degli Arcimboldi, 8 – 20126 Milan (Italy)

phone +390264487574

gianluca.argentin@unimib.it

Abstract

Randomized controlled trials (RCTs) are increasingly used in sociology to generate causal evidence. While the uptake of RCTs in sociology has lagged behind other disciplines, recent growth signals a shift toward what some now call “experimental sociology.” However, the methodological promise of RCTs depends not only on random assignment but also on transparent reporting practices that allow readers to assess internal validity, implementation quality, and the generalizability of findings. In this article, we conduct a systematic scoping review of 91 field experiments published since 2000 in high-ranking sociological journals. Using an adapted version of the CONSORT-SPI checklist, we evaluate reporting accuracy across a range of design and implementation indicators. We find substantial variability and frequent omissions in how sociological RCTs are reported, with implications for replicability and the cumulative advancement of causal knowledge. Our findings suggest that sociologists have yet to fully adopt reporting standards developed in other disciplines, and that greater attention to documentation could strengthen the credibility and utility of experimental evidence. We offer concrete recommendations for researchers, reviewers, and journal editors to improve transparency and comparability in sociological field experiments.

1. Introduction

Empirical studies based on Randomized Controlled Trials (RCTs) have gained increasing traction within sociological research. While the discipline was initially slower than economics or political science to incorporate experimental designs (Green and Gerber 2003; Jackson and Cox 2013), the past decade has witnessed a marked expansion in their application across substantive areas. This growth is documented in recent reviews (Jackson and Cox 2013; Baldassarri and Abascal 2017; Mitze and Manago 2022) and has led some to refer to this emerging domain as "experimental sociology" (Barrera et al. 2025), signaling both the institutionalization of experimental methods and their integration into core sociological concerns.

Although RCTs are widely regarded as a rigorous method for establishing internal validity—defined as the ability to identify causal relationships through controlled assignment of treatment—they have long been critiqued for their limitations in external validity, particularly with respect to generalizability and relevance across sociological domains. Yet this dichotomy risks oversimplifying both the strengths and vulnerabilities of experimental designs. As Deaton and Cartwright (2018) argue, the credibility of causal claims derived from RCTs hinges not only on random assignment but also on a host of design, implementation, and interpretive choices that can introduce fragility into the estimation process. Even minor departures from ideal conditions—such as noncompliance, attrition, or poorly specified interventions—can substantially compromise inference.

Accurate and transparent reporting is essential for identifying sources of fragility in experimental designs. Without clear documentation of key design and implementation choices, it becomes difficult to evaluate the credibility of causal claims, assess risks to internal validity, or replicate findings. Fields with a longer history of experimental research, particularly in clinical and psychological sciences, have recognized that inadequate reporting can obscure methodological flaws and inflate the apparent robustness of causal claims (Andrew et al. 1994; Grant et al. 2018). Sociological applications of RCTs would similarly benefit from standardized reporting practices to mitigate threats to validity and support critical appraisal of findings.

This paper interrogates the extent to which the emerging body of sociological RCTs fulfills the methodological promise most often associated with the experimental approach: internally valid causal inference. We examine recent experimental studies to evaluate whether, and how consistently, the core design elements necessary for unbiased estimation are reported in published accounts. Specifically, we investigate the gap between the theoretical strength of RCTs as a tool for causal identification and the practical realities of their implementation, as reflected in the reporting practices of sociological research.

Our focus is not on outcomes per se, but on the integrity of the design and reporting processes that underpin claims of causal inference. Drawing on CONSORT-SPI standards (CONsolidated Standards of Reporting Trials for Social and Psychological Interventions) (Grant et al. 2018), we assess whether sociological RCTs attend to the full range of methodological considerations required for credible inference—including randomization procedures, allocation concealment, protocol deviations, attrition, and sample size calculations. Through this lens, we ask: To what extent is the internal (and, where possible, external) validity of sociological RCTs supported by the transparency and completeness of their reporting?

We address this question through a systematic scoping review (Tricco et al. 2018), employing a structured and transparent approach to identify, code, and analyze published accounts of RCTs appearing in leading sociology journals. Our analysis focuses specifically on field experiments, which we define as randomized trials conducted in natural settings with participants drawn from populations targeted by the intervention, interacting in their ordinary social contexts. This design feature makes field experiments particularly valuable for assessing causal effects under ecologically valid conditions—an attribute especially relevant in sociological inquiry.

We further restrict our sample to studies published in high-ranking journals, for two reasons. First, field experiments embedded in real-world contexts offer particular theoretical and methodological value in sociology. Second, journals with higher disciplinary visibility are more likely

to enforce standards of design and reporting, making them especially relevant to our research question.

We use this sample of RCTs to describe how the method is applied in sociology, providing what is, to our knowledge, the first systematic overview of its implementation in the field. We directly assess reporting accuracy and, in doing so, identify common pitfalls in the design and execution of sociological field experiments. When key decisions around design and implementation are not reported, the resulting evidence becomes difficult to evaluate or replicate—and this may also reflect limited awareness of the risk of bias inherent in experimental research. Our results show that the reporting of RCTs in sociology remains inconsistent, even when limited to high-ranking journals. Inadequate reporting is problematic, as it can obscure weak designs or poor implementation—both of which compromise internal validity. When added to the known limitations of experimental methods, this issue further reduces the potential of RCTs to generate credible causal evidence in sociology.

We conclude that the early adoption of RCTs in the discipline has not fully capitalized on methodological standards established elsewhere. Overall, attention to reporting clarity and completeness appears limited: key design elements are often omitted, making published experiments difficult to assess and replicate. This is particularly concerning given that individual RCTs are vulnerable to bias, tend to have limited internal validity, and typically lack external validity due to the importance of context in sociological processes. Only through replication across settings—and the comparison of results across studies—can experimental findings be consolidated and generalizability improved, thereby addressing longstanding critiques of experimental evidence (Pawson and Tilley 1997). In this light, poor reporting represents a major obstacle to the accumulation of reliable and replicable causal knowledge. In short, even well-designed RCTs cannot fulfill their evidentiary promise if they are poorly reported.

The paper is structured as follows. Section 2 presents the background and motivation, highlighting the value of RCTs for causal inference in sociology and the importance of reporting accuracy for assessing internal validity. Section 3 describes our review methodology, from journal selection to coding procedures, including our use of the CONSORT-SPI checklist (Grant et al. 2018) to evaluate reporting quality. Section 4 provides a descriptive overview of the field experiments in our sample. Section 5 offers a detailed assessment of reporting accuracy, while Section 6 examines its evolution over time and relationship to key study features. Section 7 concludes with a summary of findings and recommendations for improving the design and reporting of sociological RCTs.

2. Background and motivation

2.1 The potential of field experiments for causal investigation in sociology

Among quantitative social scientists, randomized controlled trials (RCTs) are widely recognized as the "gold standard" for establishing causal relationships between variables (Green and Gerber 2003; Berk 2005; Gangl 2010). Their appeal lies in their ability to address the core challenge of causal inference: distinguishing the effect of a given independent variable X on an outcome Y from the influence of confounding factors. By randomly assigning units to treatment and control groups, RCTs eliminate systematic differences in observed and unobserved covariates across groups, thereby mitigating endogeneity and isolating the causal impact of X on Y (Morgan and Winship 2007). Getting rid of all the other confounders is especially important in the study of social phenomena, where causal structures are often complex and units (e.g., individuals, groups) are not interchangeable due to heterogeneity in background characteristics, preferences, and social positioning.

External validity remains a persistent challenge in experimental research, as RCTs often rely on convenience samples, making it difficult—if not impossible—to generalize findings beyond the specific populations studied, whether from a single trial or across multiple ones (Deaton and Cartwright 2018). This challenge is amplified in laboratory and vignette-based experiments, which

account for a substantial share of experimental causal evidence in the social sciences (Barrera et al. 2025). In these settings, researchers construct fully controlled—physical or simulated—environments in an attempt to elicit natural responses from participants (Harrison and List 2004; Jackson and Cox 2013). Yet despite these efforts, the artificial nature of the context often shapes participants' behavior in ways that are difficult to interpret outside the experimental setting, raising additional concerns about the generalizability of results.

By contrast, field experiments offer a powerful tool for studying causal processes in real-world conditions. They assess the effects of real interventions on real outcomes among individuals drawn from target populations, acting in their natural environments—in some cases without awareness of their participation in a study (Gerber and Green 2012; Barrera et al. 2025). As Gerber and Green (2012) note, "field experiments take many forms," with varying degrees of "fieldness," but what unites them is that the context is not constructed but taken as given. At one end of the spectrum, field experiments involve minimal researcher control beyond maintaining random assignment in complex, real-world settings (Torgerson and Torgerson 2008). At the other end, lab, vignette, and survey experiments create fully designed contexts, often detached from the social dynamics under investigation, and pay a price in terms of ecological and behavioral realism.

Another key strength of field experiments is their potential to inform policy and practice. It is no coincidence that RCTs are most widely used in applied fields—especially among policy evaluators—where the relevance of results for real-world decision-making is predominant (Luca and Bazerman 2020). This close link between evidence production and policy application helps explain the growing interest in field experimentation, which has expanded well beyond its origins in clinical medicine to psychology, economics, and political science. Although sociology has adopted this trend more gradually, the number of sociological field experiments has grown substantially in recent years (Baldassarri and Abascal 2017; Mitze and Manago 2022).

We identify at least three factors behind this shift. First, there is increasing recognition among sociologists of the need for credible causal inference (Goldthorpe 2000, 2004, 2015), particularly in light of the counterfactual turn in the discipline (Morgan and Winship 2007), which has highlighted the limitations of associational analyses based on regression models. Second, there is a growing consensus that the macro-level patterns traditionally studied in sociology must be explained through micro-level causal processes (Boudon 1979; Coleman 1990; Goldthorpe 2000). Third, the growing visibility of *cameral sociology*—that is, sociology oriented toward informing policy—has raised the importance of producing actionable, policy-relevant evidence (Boudon 2002; Goldthorpe 2004). This shift has likely been reinforced by broader institutional trends promoting evidence-based policymaking and by targeted funding initiatives—particularly in education—that support experimental research designs.

2.2 Criticisms towards RCTs

The recent rise in the use of RCTs in sociology—whether to generate theoretical insight or inform policy—has been accompanied by a growing set of critiques. Some scholars urge caution in interpreting experimental findings, pointing to the intrinsic limitations of the method (Deaton and Cartwright 2018), while others offer more fundamental objections to the logic and application of RCTs. We focus here on the latter.

A key analytical criticism targets what RCT proponents view as the method's principal strength: the reduction of complex social phenomena to a well-specified, unidirectional relationship between a limited set of variables. In this view, causal inference is achieved through the random manipulation of X to estimate its isolated effect on Y. However, random assignment is not feasible for many core sociological variables—such as gender, race, social origin, or class—making experimental methods (and in some cases, the counterfactual framework itself) inapplicable (Holland 1986; Goldthorpe 2015). Moreover, sociological inquiry often involves multidimensional and interdependent causal processes, which resist the reduction of causal structure to a single manipulable

treatment. In contrast, the statistical modelling of robust associations—based on random samples and multivariate regressions—has long served as a core empirical strategy in sociology. While this approach cannot, on its own, establish causality, it allows researchers to build and test theoretically informed models of complex relationships among variables (Goldthorpe 2015; Lundberg, Johnson, and Stewart 2021).

Other critics argue that RCTs insufficiently account for contextual variation. Reduced-form causal estimates may hold in some settings but fail to replicate in others—not merely due to problems of external validity, but because the meaning and operation of causal mechanisms are themselves context-dependent. From this perspective, evidence from RCTs, when abstracted from context, may appear inconsistent or even contradictory, undermining its utility for policy design. Rather than focusing solely on average treatment effects, causal inference should incorporate the interaction between context, mechanism, and outcome (Pawson and Tilley 1997).

Further objections concern the ethical and political dimensions of RCTs. A central ethical critique targets the legitimacy of random assignment, which by design withholds potentially beneficial interventions from part of the sample. While such concerns have long been debated in the medical field (Hellman and Hellman 1991; Torbjörn 2022), they also raise important ethical questions in sociological settings, particularly when withholding treatment may affect access to services or opportunities for disadvantaged populations (Baldassarri and Abascal 2017).

From a political standpoint, RCTs have been criticized for their apparent neutrality in the evaluation of social policies. Critics argue that, beneath their technical rigor, experiments often pre-define the intervention (X), the outcome (Y), and the target population, thereby sidelining meaningful input from those most affected. This may lead to a form of "social engineering" in which key normative and political choices are made without deliberation or participation (Fatterman 2000). In this light, participatory evaluation approaches—which involve stakeholders in defining problems, goals, and metrics—are sometimes presented as alternatives to experimental methods (Garaway 1995).

2.3 A hidden source of weakness in evidence from RCTs: reporting accuracy

Despite these criticisms, interest in RCTs among sociologists has steadily grown—albeit at a slower pace than in other social sciences (Jackson and Cox 2013; Baldassarri and Abascal 2017; Mitze and Manago 2022). In what follows, we document this growing engagement while also underscoring the need for greater methodological awareness in the design, implementation, and reporting of RCTs within the discipline.

The central point, as noted earlier, is that RCTs are powerful but fragile instruments for causal inference. Their main appeal lies in providing a credible estimate of the causal effect of manipulating x on Y. However, this goal can be compromised by a range of methodological and practical factors. Randomization alone does not guarantee valid results. While it ensures balance between treatment and control groups on average, in any single trial—particularly with smaller samples—randomization may fail to balance key confounders purely by chance. Beyond this issue, internal validity can also be threatened by the many assumptions and design choices embedded in RCTs, which often go unreported but can significantly shape results.

Rather than focusing on the "standard" threats to the robustness of causal estimates—such as sample size limitations or contextual factors specific to a study—we emphasize a broader set of implementation conditions that can themselves introduce substantial bias into RCT results (Torgerson and Torgerson 2008; Gerber and Green 2012; Barrera et al. 2025). To assess the credibility of an RCT's findings, other researchers must be able to evaluate the many design and implementation choices that shaped the study, along with the characteristics of the context in which it was conducted. This requires not only methodological awareness on the part of experimenters but also meticulous attention to detail in the planning and execution of the trial. Most importantly, these decisions and potential sources of bias must be documented transparently and reported in full to enable critical appraisal and replication.

As is well established in disciplines with a longer tradition of using RCTs, results from a single experiment should be replicated and synthesized with findings from similar studies to obtain a

more reliable estimate of a given causal mechanism. This process requires both replicability and comparability—two conditions that depend critically on the accuracy and completeness of reporting. Specifically, researchers must document all relevant design decisions and implementation features in sufficient detail to allow others to evaluate, reproduce, or build upon the original study. The recent replication crisis in both the natural and social sciences has underscored the dangers of opaque research practices and insufficient documentation (Begley and Ellis 2012; Open Science Collaboration 2015; Gould et al. 2025). From this perspective, inadequate reporting in sociological RCTs represents a hidden source of fragility in experimental evidence—one that has been recognized and addressed in other disciplines for decades (Andrew et al. 1994).

3. Methods

We conducted a systematic scoping review to identify and code the eligible studies, following the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) guidelines for scoping reviews (Tricco et al., 2018).

3.1 Search Procedures

Our search strategy to locate RCTs published in sociological journals proceeded in two steps. First, we identified a set of high-ranking sociological journals and selected potentially eligible studies all papers published in those journals. Our selection of high-quality sociological journals was based on three international rankings: ScimagoJR (category: sociology and political science), Web of Science (WOS, category: sociology), SCOPUS (subject area: social sciences).[1] The top-ranking 150 journals in WOS and SCOPUS as well as the top-ranking 200 journals in ScimagoJR were selected, based on two inclusion criteria: (i) the journal publishes papers in English; (ii) the journal addresses sociological topics or is specific to sociological streams of research (e.g., *Journal of Marriage and Family*). We then excluded journals specialized in psychology, political science and economics. The selection was conducted independently by three authors of this article who assessed whether each

journal was meeting the two criteria stated above. We retained all the journals positively evaluated by at least one of the evaluators. After excluding duplicates, this resulted in a list of 195 high-ranking sociological journals (Table A in Supplementary materials).

We then searched and selected RCTs published in these journals. We searched for studies through Web of Science-WOS (category: sociology) and SCOPUS (subject area: social sciences) databases as of November 2022. A combination of keywords related to study design (e.g. RCT, control group) was used to search WOS, restricting the search to titles and abstracts. In SCOPUS we added the term "sociolog*" to the search string used in WOS. We limited our search in both databases to records published in journals, in English, and after 2000. We then conducted a second database search in SCOPUS and WOS in December 2022, with other combinations of terms, to specifically identify field experiments in sociology. Complete search strings are reported in the Supplementary materials. Second, we used the web tool *Paperfetcher* to search the high-ranking journals of our list that were not indexed in SCOPUS or WOS. For this subset of journals, we restricted the search to records that included the word "random" or "experiment". Four journals were not indexed in *Paperfetcher – Current Perspectives in Social Theory, Archives Europeennes de Sociologie, Journal of Elections, Public Opinion and Parties*, we therefore hand-searched their table of contents.

3.2 Eligibility Criteria

We defined four eligibility criteria in order to determine whether to include or exclude the records in this review. (i) Studies were required to have used an experimental design, in which participants or groups of participants were randomly assigned to a control group and (at least) one treatment group compared. (ii) As stated above, we restricted our sample to field experiments, and we excluded clinical, survey/vignette and lab trials. Because "fieldness" exists on a continuum (Gerber and Green 2012), an additional screening step was necessary. We treated the natural setting as the primary dimension as the defining criterion and excluded any experiment conducted in artificial

or simulated environments. (iii) Studies were required include an impact evaluation of an intervention on participants' outcomes. We excluded studies that focused only on mediation analysis. (iv) Follow-up studies of previous field RCTs were excluded, as these often lack detailed reporting of study design and implementation, relying instead on references to earlier publications. Independent replications of previous RCTs, instead, were included if they met the other criteria. No restrictions were applied based on the topic of substantive area the study.

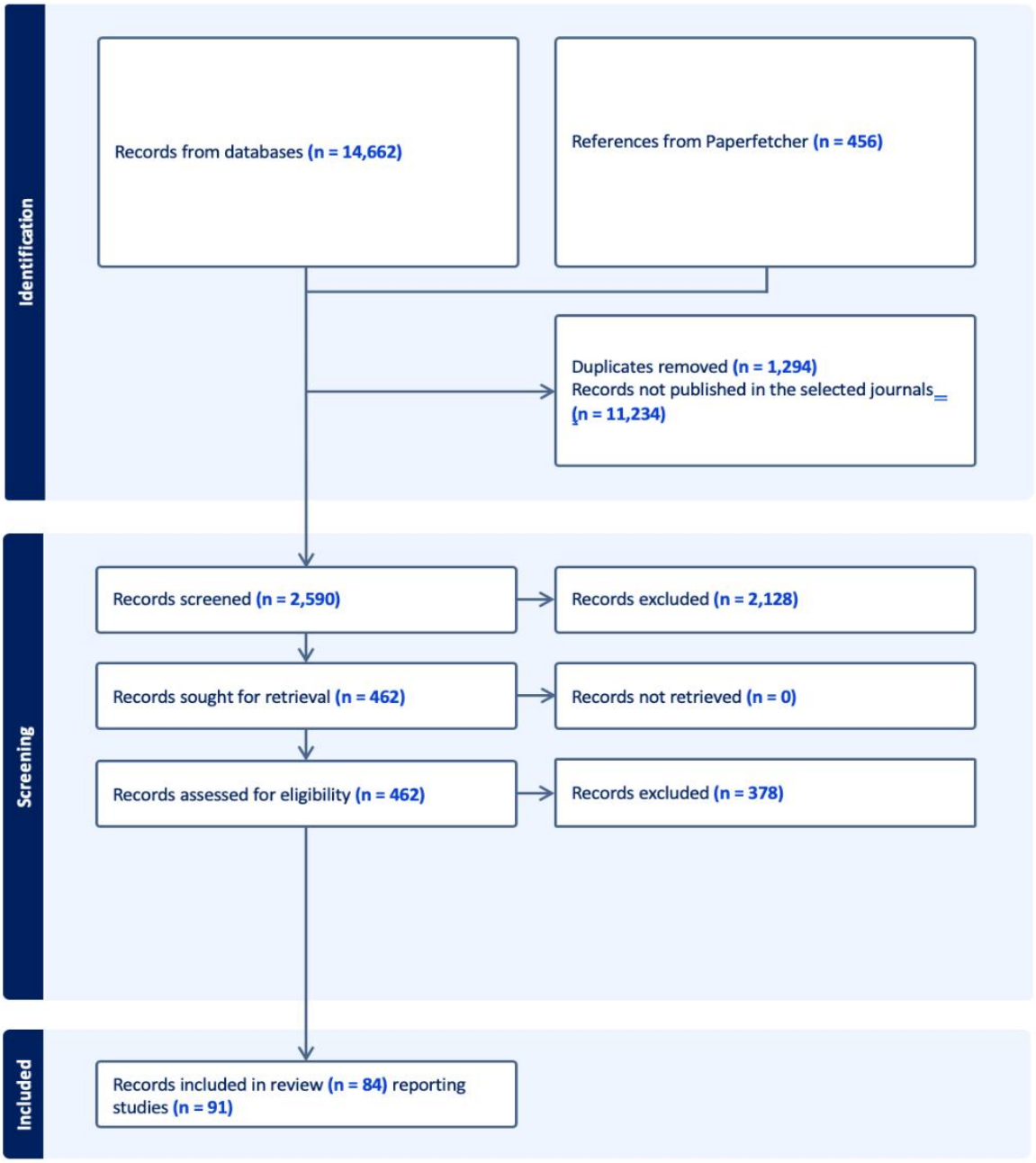
3.3 Selection process

The search procedure in SCOPUS, WOS, and *Paperfetcher* yielded a total of 15,118 papers, as shown in Figure 1.

After removing 1,294 duplicates, 13,939 records remained for further screening. Of these, 2,590 studies were published in the high-ranking sociological journals identified in the previous selection stage, and were thus retained. The title and abstract of the identified records were screened in duplicate by two independent reviewers using Covidence, a software platform designed to support systematic reviews through screening, full-text review, and coding. We conducted an initial training session, in which all authors of this paper piloted the screening tool developed for the selection process by reviewing the same 30 abstracts. Subsequent training sessions allowed coders to refine the screening tool, reaching a 90% agreement rate across reviewers. After this initial stage, the abstracts were independently screened in duplicate by two authors. When disagreements occurred, they were resolved in regular weekly meetings involving at least three authors, until consensus was reached.

We then retrieved the full texts of the 462 studies retained from the abstract screening and double-reviewed them, applying the eligibility criteria defined above. This stage was conducted by four authors—all sociologists with expertise in RCTs—following a process similar to that used during abstract screening. The initial agreement rate for this phase was 70%, rising to 100% after bi-weekly discussions to resolve discrepancies. A total of 84 records reporting 91 studies met our eligibility criteria and were included in the review (Figure 1).

Figure 1. Flowchart of the selection process



3.4 Data extraction

We used the CONSORT-SPI checklist (Grant et al., 2018) as the foundation for developing a codebook with a classificatory aim. CONSORT provides a standardized, comprehensive, and authoritative framework for trial reporting. In the early days of clinical research, biomedical studies—

where RCTs are the dominant source of evidence—were frequently criticized for poor reporting, making it difficult to evaluate the validity and reproducibility of findings. In response, the CONSORT checklist was developed within the medical community as a practical tool to promote complete and transparent trial documentation. First published in 1994 (Andrew et al., 1994), it was updated in 2010 (Schulz, Altman, and Moher 2010) and again in 2025 (Hopewell et al., 2025). The CONSORT-SPI extension was introduced in 2018 to address the specific reporting needs of social and psychological interventions, adding nine items to the original checklist for a total of 26 items and associated sub-items.

The original CONSORT-SPI checklist organizes its 26 indicators into six areas, reflecting key components of an RCT report:

(i) Introduction, which describes the study rationale and research questions;

(ii) Methods, detailing trial design, population, intervention, outcomes, and sample;

(iii) Randomization, with eight items dedicated specifically to the random assignment process;

(iv) Results, including data on both trial implementation and outcome estimates;

(v) Discussion, covering interpretation of findings, limitations, and external validity; and

(vi) Other information, such as trial registration, protocol availability, and stakeholder involvement—factors that may indirectly influence internal validity.

To suit the goals of our review, two main adaptations were made to the original checklist. First, we modified the binary coding scheme (Yes/No + Not Applicable), which we considered insufficient to capture situations in which reporting partially fulfilled a criterion. Including partial compliance categories was necessary to avoid unduly penalizing authors in a discipline where formal reporting standards are not yet well established. For example, item 1b (structured abstract) cannot be fairly assessed using binary coding in a literature where this format is still uncommon. Our modified scheme allows more nuanced assessments.

Second, we extended the checklist by adding 30 new items to better capture both reporting quality (7 items, such as ethical approval, flow diagrams, or trial data availability) and the substantive/methodological characteristics of the literature (25 items, such as intervention type, country of implementation, presence of declared effects, or whether a sociologist was among the first two authors). Some of these additional items allow open-ended responses—for instance, items a0e–a0g include a brief description of the scientific field of the study.

In total, our coding framework comprised 73 items. The full codebook, including the new items and modifications to existing CONSORT-SPI items (along with justifications), is available in the Supplementary Materials.

We piloted the codebook on five studies and revised it after discussion among all authors. Bi-weekly training sessions followed, during which ten studies were jointly coded to align interpretations. Once a 90% agreement rate was reached, the studies were assigned to pairs of independent coders. Discrepancies were resolved through discussion with a third experienced coder. As noted above, when a paper reported multiple experiments, each experiment was coded separately.

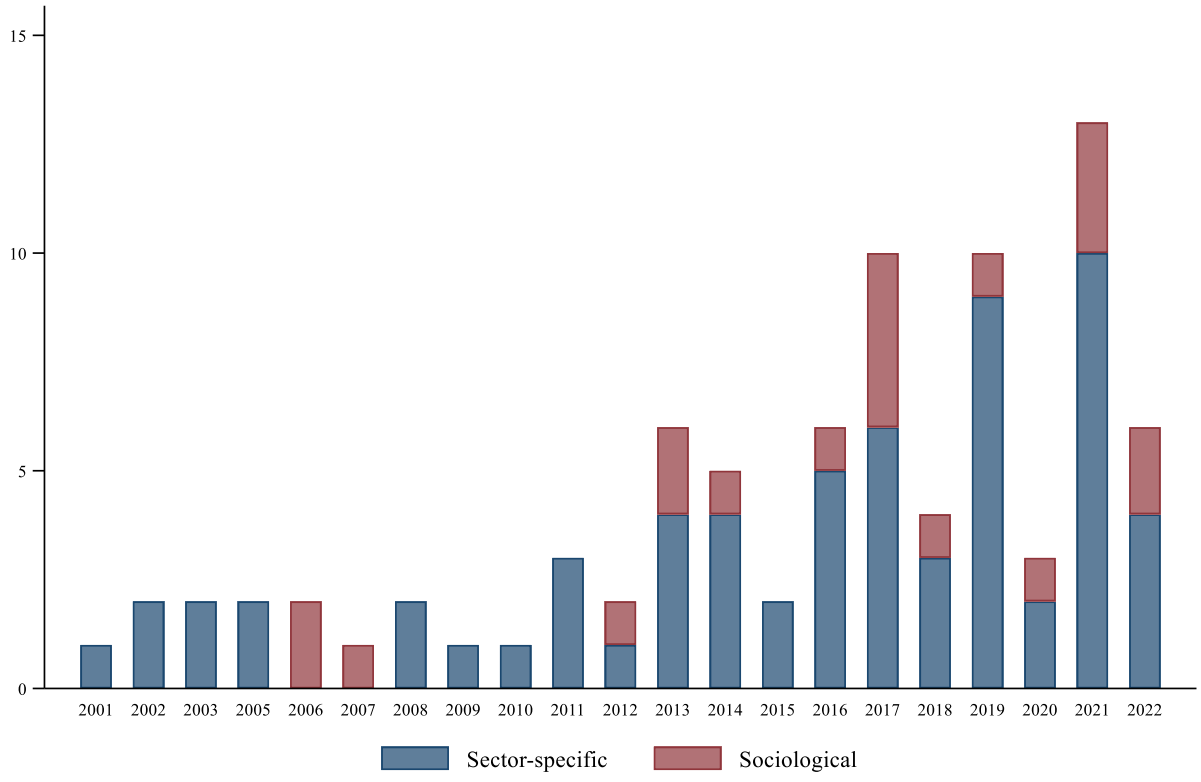
4. Results I: Descriptive overview of the studies

Figure 2 shows the evolution over time in the frequency of the 91 field experiments identified in our review, by year of publication and journal type (generalist sociology journals vs. specialized subfield journals). The full list of journals is provided in the Supplementary Materials.

First, we observe a sharp increase in the number of experimental papers published from 2011 onwards, in line with the findings of previous reviews (Jackson and Cox 2013; Baldassarri and Abascal 2017; Mitze and Manago 2022). Second, sector-specific journals host most of the selected papers, accounting for 64 out of 84, while only 20 were published in generalist sociological journals. Sector-specific journals are also primarily responsible for the observed uptick in the data. We

speculate that this might reflect a greater receptivity in sector-specific journals to contributions from researchers across disciplines within the social sciences (e.g., education, economics, psychology), where the experimental method is more commonly used than in sociology. This interpretation is supported by an additional analysis we conducted on author affiliations: in only 26 of the 84 papers was a sociologist listed among the first two authors. Nonetheless, this occurs more frequently in sociological journals (13 out of 20 articles). These findings underscore the relatively recent emergence of experimental research in sociology—at least in the domain of field experiments, which postdate laboratory studies (Barrera et al. 2025). In addition, these results highlight the substantial contribution of researchers from outside the field and the fact that field experiments frequently lead to interdisciplinary research. This may reflect the diverse skill sets required to design and implement real-world interventions while evaluating their impact through an RCT.

Figure 2 – Evolution over time of papers reporting field RCTs appeared in top sociological journals



Note: journals included are listed in the Supplementary materials, table A*. See text for the inclusion criteria.

Table 1. Key features of the experiments - I

	%	N
Research topic		
Public administration	4.4	4
Education	36.3	33
Family	7.7	7
Health	8.8	8
Labour market	5.5	5
Law	8.8	8
Politics	11.0	10
Social psychology	12.1	11
Welfare	5.5	5
Area		
English-speaking countries	64.8	59
Europe (UK excluded)	22.0	20
Africa	5.5	5
Asia	6.6	6
Latin America	1.1	1
Has the mechanism been explored		
Yes	47.3	43
Is the experiment a replication?		
Yes	4.4	4
N=91		

From this point forward, we take the experiment (N = 91), rather than the paper, as our unit of analysis. Table 1 reports the key characteristics of these experiments. The range of research topics is broad, but education accounts for more than one-third of the total. Many of these studies (25 out of 33) focus on educational inequalities, a topic with a long-standing tradition in sociology and one that aligns with the political momentum behind evidence-based policy. Regarding the geographical distribution of the studies, it is unsurprising that the English-speaking world dominates, with 59 of the 91 experiments conducted in those countries (most in the U.S.), followed by the European Union (20), Asia (6), Africa (5), and Latin America (1). The focus on education and the predominance of

Anglo-Saxon countries in our sample suggest that increased investment in evidence-based education in the U.S. and U.K. played a significant role in driving experimental research in sociology (Torgerson and Torgerson 2012).

Finally, we find that fewer than half of the studies (43 out of 91) attempt to investigate the mechanisms behind intervention effects, either by using design features (e.g., multi-arm experiments) or through secondary qualitative or quantitative analysis. In the remaining cases, only treatment effects are reported. This pattern partially confirms the critique that RCT-based evaluations often treat interventions as "black boxes" (Pawson and Tilley 1997). Despite growing awareness of the replication crisis, only 4.4% of the studies are replications of prior experiments—even though it is widely acknowledged that the evidence from a single RCT is insufficient and should be corroborated by replications. This low replication rate may reflect both the relatively recent diffusion of RCTs in sociology and an artifact of our sample, which focuses on high-ranking journals that often prioritize novelty and innovation.

Table 2 provides additional descriptive information on the RCTs published in highly ranked sociological journals. In terms of outcome measurement—i.e., what the intervention sought to change—a large proportion of studies relied on questionnaires: 35 experiments measured self-reported attitudes and 30 measured self-reported behavior. These results underscore once again the relevance of field experiments, where participants' presence in natural settings may mitigate artificial context effects on attitudes and opinions. Real-world behavioral data—typically obtained from administrative or platform-based sources—served as the primary outcome in 34 experiments. Educational experiments, in particular, often relied on standardized assessments (e.g., test scores), which were used as outcomes in 21 cases. This further supports the claim that investments in evidence-based education played a significant role in driving sociological RCTs. Simulated attitudes—more typical of vignette studies—were rare, serving as outcomes in only two cases, consistent with our focus on field experiments.

Table 2. Key features of the experiments - II

Type of outcome (possible multiple elements) - %		N
Real world behaviour	37.3	34
Reported attitude	38.5	35
Reported behaviour	33.0	30
Simulated attitude	2.2	2
Standardized assessment	19.8	18
Nr of outcomes (%)		
1	27.5	25
2-5	36.3	33
6 or more	36.3	33
Elements involved in the intervention (possible multiple elements) - %		
Incentives	17.6	16
Education	28.6	26
Information - communication modes	13.2	12
Information – content	31.9	29
Mentoring	20.9	19
Reminder	5.5	5
Training	20.9	19
Psychological interventions	7.7	7
Other	9.9	9
Number of randomized units (N)		
5th percentile	15	
Median	211	
95th percentile	66,442	
Number of beneficiaries (N)		
5th percentile	45	
Median	741	
95th percentile	78,872	
N=91		

Regarding the number of outcomes assessed, fewer than one in four experiments (22.6%) examined only a single outcome, while most assessed between 2 and 5 (38.1%) or 6 or more (39.3%) outcomes. Randomization was typically conducted on a median of 211 units, though this varied

widely. The number of final experimental units used to estimate treatment effects was substantially higher, with a median of 741, due to the frequent use of clustered designs. In many educational studies, for instance, randomization occurs at the class or school level, while outcomes are measured at the student level. The wide range in the number of experimental units highlights the heterogeneity in field experiment scale—from small pilot trials to large-scale interventions.

Lastly, we offer a descriptive overview of the intervention types featured in these studies, shedding light on the current trajectory of experimental research in sociology. Information-based treatments are the most common, often combined with educational, training, or mentoring components. Educational interventions follow closely, along with training, mentoring, and incentive-based treatments, which appear in roughly equal proportions. Taken together, these findings suggest that at least three factors are shaping the growth of field experimentation in sociology. First, field experiments tend to focus on causes that are easily manipulable, with information-based interventions offering a practical entry point. Second, the availability of clusters (e.g., schools, classrooms) facilitates the logistical demands of randomization. Third, as noted above, funding directed toward evidence-based research—especially in education—has likely played a critical role in supporting the expansion of RCTs in the field. Overall, a combination of disciplinary openness, methodological opportunity, and institutional support appears to be driving this experimental turn in sociology—one that will likely shape future replications and new trials.

5. Results II: The accuracy of reporting

In this section we assess the accuracy of the reporting using the adapted CONSORT-SPI checklist. Figures 3 through 6 concern the original items of CONSORT-SPI, while figure 7 includes a selection of the additional items. In the figures, the bar length represents the percentage of experiments that are compliant with each item. The colors of the bars distinguish levels of compliance: blue indicates full

compliance, maroon indicates partial compliance (i.e., when insufficient detail is provided relative to the benchmark), and khaki denotes items that are not applicable.[2] The items are organized according to the six sections of the CONSORT-SPI checklist.

A quick glance at the three figures reveals three main findings. First, only 8 out of 41 items exceed an 80% rate of full compliance, and only 11 out of 41 meet the threshold for either full or at least partial compliance. Although we do not have a benchmark for direct comparison, this clearly indicates a generally low level of reporting accuracy. Second, we observe substantial variation in compliance across items and domains, suggesting that reporting accuracy is far from uniform. The magnitude of these differences allows us to identify specific areas in need of improvement, as discussed below. Third, we note widespread incompleteness in many items, visually reflected in the size of the maroon bars.

As noted earlier, we aim to go beyond a descriptive assessment of reporting accuracy by interpreting these results as indicators of authors' awareness of the procedures necessary to properly design and implement RCTs. What emerges is a concerning lack of understanding about the safeguards required to avoid biased results or to enable the scientific community to assess or replicate the reported evidence.

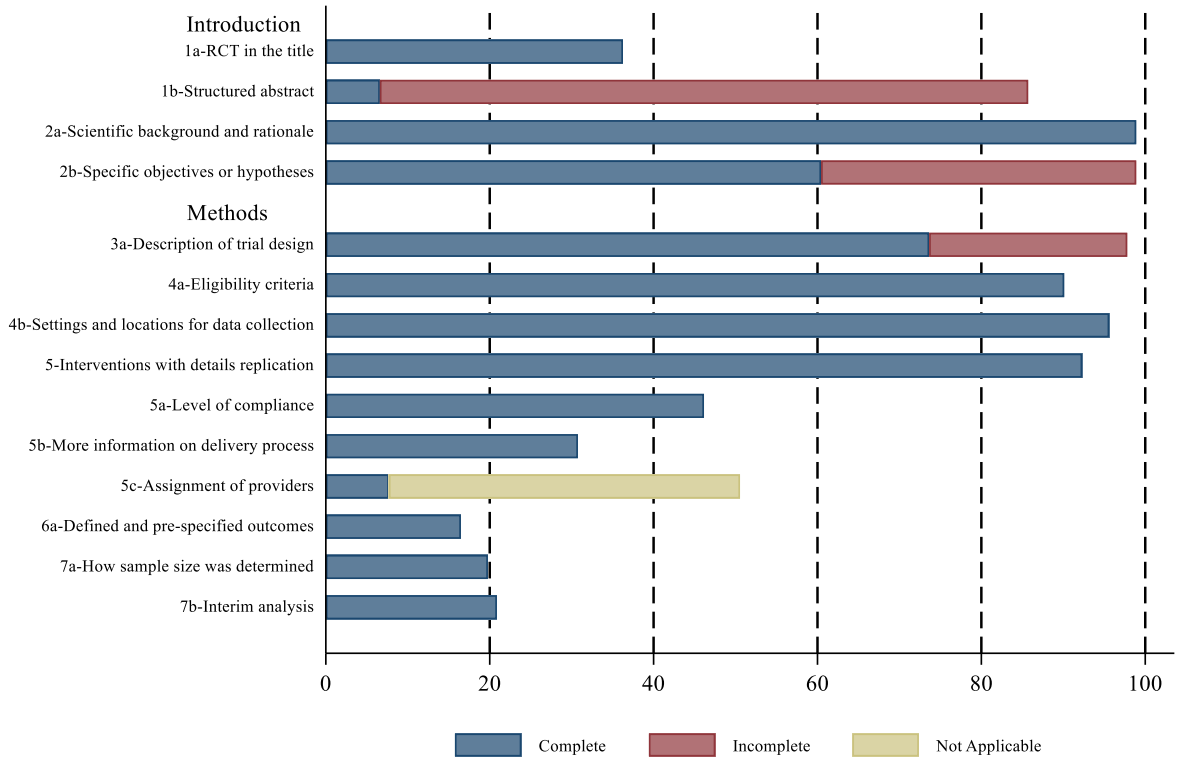
A closer inspection allows us to distinguish between shortcomings related to a lack of awareness and those related to a lack of replicability. In Figure 3, items in the "Introduction" section appear relatively well reported: over 80% of abstracts present, at least partially, the relevant background and state the study's objectives and hypotheses (items 1b, 2a, 2b). It is worth noting that the incompleteness observed for item 1b (Structured Abstract) likely reflects journal submission guidelines rather than deliberate omissions by authors.

By contrast, another item provides more telling information—namely, the low share of papers that declare in the title that the study is an RCT. This suggests that many sociologists remain unaware of the importance of making experimental studies easily identifiable in order to support systematic

reviews and meta-analyses aimed at synthesizing rigorous causal evidence. Of course, this visibility depends on the prior existence of a substantial body of comparable studies—a condition still largely unmet in sociology.

With a few exceptions, most items in the CONSORT-SPI "Methods" section show relatively low levels of reporting accuracy. Some areas stand out positively: a majority of studies provide clear descriptions of the experimental design (item 3a), eligibility criteria (4a), data collection procedures (4b), and the intervention itself (5). However, reporting quality declines significantly for several other key indicators.

Figure 3. Reporting accuracy of the observed experiments – I (%)



Two items are especially problematic: compliance with the intervention protocol (5a) and the determination of sample size (7a). In the first case, low reporting is particularly concerning, since the degree of compliance may influence the choice of modeling strategies and, therefore, the interpretation of the results. The lack of information on item 5b (delivery details of the intervention)

further undermines interpretability. In the second case, most studies do not report conducting power calculations to determine sample size ex ante, often because sample size is shaped by external constraints, such as available resources. Still, this practice risks confusing null results with a lack of statistical power, especially in trials with small samples of randomized units.

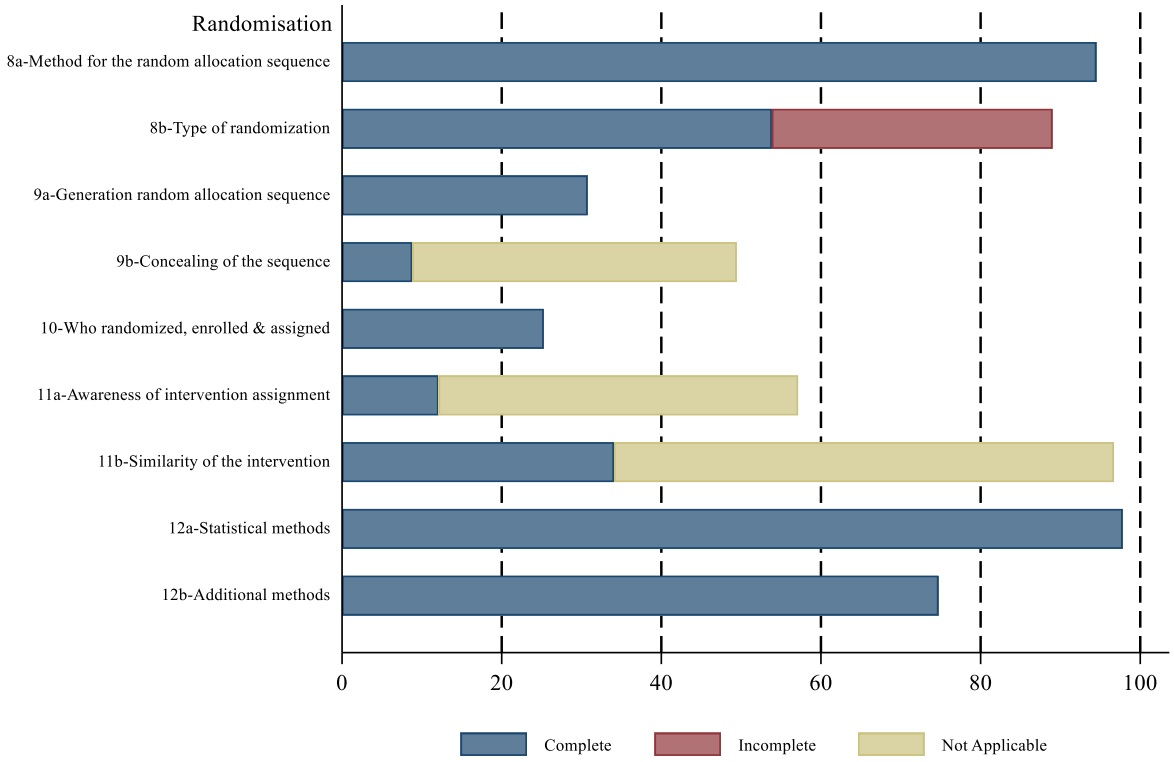
Both compliance monitoring and power analysis are standard in RCT methodology but less common in the survey-based research tradition more typical of quantitative sociology. The limited reporting on these elements suggests insufficient familiarity with experimental design. Much of the methodological information appears intended to explain what was tested, without acknowledging the need for replicability or documentation that enables replication.

Similarly low levels of reporting accuracy are found for interim analysis (7b), when applicable. We interpret this as a sign of limited awareness among sociologists about the role of interim results in experimental integrity. Ignoring interim data prevents researchers from adapting the study design if problems arise during implementation.

Only a few studies report how intervention providers were selected or assigned to their role (item 5c). This is troubling, as many interventions rely on forms of expert-delivered content—such as information, training, or mentoring—where provider expertise is central to implementation. Insufficient detail about providers introduces several risks: (i) treatment heterogeneity across experimental units; (ii) underestimation of implementation barriers or enabling factors, particularly when provider engagement influences effectiveness; and (iii) limited replicability, due to the unreported tacit knowledge embedded in how interventions are delivered.

Another key concern is the pre-specification of outcomes. While most studies describe outcomes clearly, very few pre-specify them in advance (6a). This combination increases the risk that researchers examine multiple outcomes post hoc, potentially inflating the likelihood of false positives—or worse, cherry-picking statistically significant results.

Figure 4. Reporting accuracy of the observed experiments – II (%)



The "Randomization" section of the CONSORT-SPI checklist includes items that address both the technical implementation and the substantive implications of assigning subjects to treatment and control groups. Reporting accuracy across these items is highly variable.

The first six items in this section detail four core steps of the randomization process: how the random sequence was generated and implemented (items 8a and 8b); whether and how allocation was concealed (9a and 9b); who performed the randomization (10); and who was aware of group assignments (11). Reported compliance ranges from excellent to poor: 94.5% and 88.9% for sequence generation and implementation, but only 30.8%, 8.8%, 25.3%, and 12.1% for the remaining items, respectively. However, items 9b and 11a apply only to a subset of studies.

We find high accuracy in describing the type of randomization, but much lower accuracy regarding its implementation (item 10), and especially regarding allocation concealment (item 9b). The former may reflect the fact that details about implementation are sometimes reported in separate technical or policy documents rather than in journal articles. In contrast, poor reporting on

concealment likely reflects more structural challenges: concealment procedures can be difficult to implement in sociological field experiments, where researchers and participants often operate in open settings. For example, delayed-treatment designs are frequently used in education and training interventions to ensure that control groups are eventually treated—for both ethical and practical reasons.

Nevertheless, researchers could be more transparent about any steps taken to mitigate bias introduced by knowledge of treatment status. The widespread opacity surrounding these technical aspects of randomization—the methodological core of an RCT—suggests that many authors adopt a simplistic view of randomization (e.g., "flipping a coin"), without attending to the complexities of real-world assignment procedures. This reflects both a lack of preparation on the part of some researchers and a parallel lack of editorial scrutiny regarding these details.

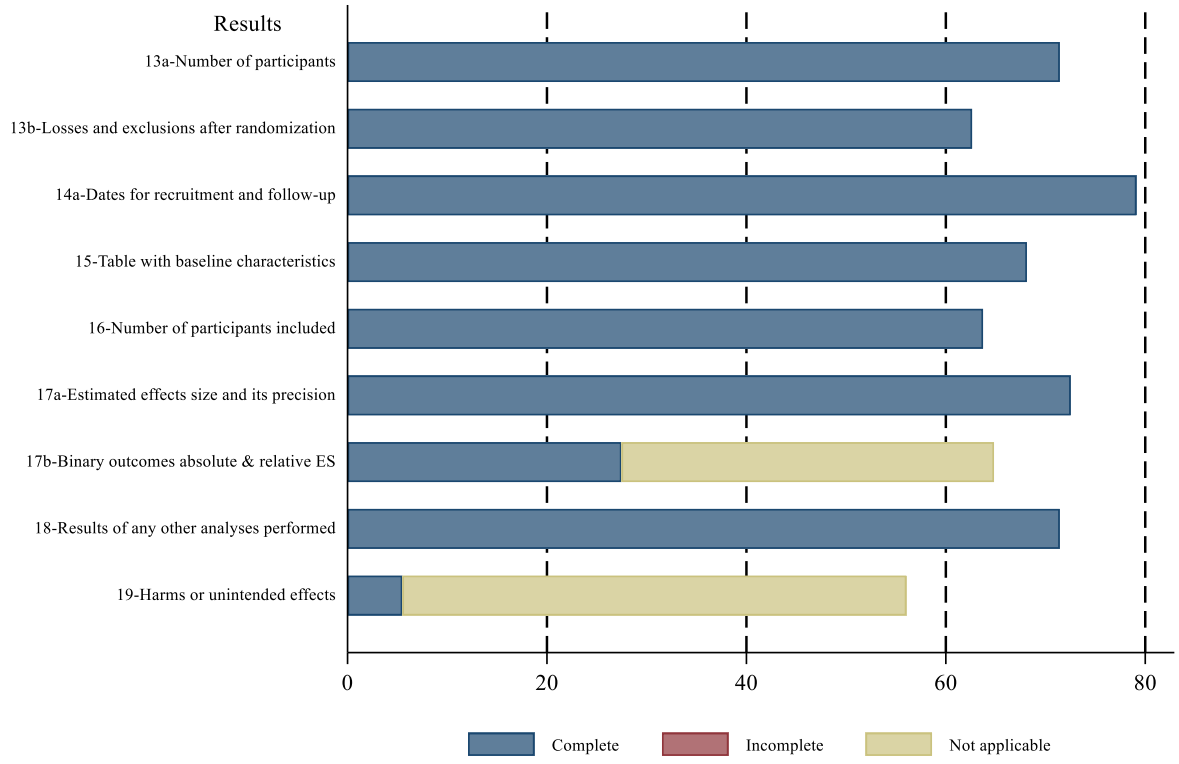
As for the four remaining items in the randomization section, reporting is generally stronger. Items 12a and 12b, concerning methods for comparing group outcomes and conducting additional analyses, are well reported. Item 11b applies only to studies that feature multiple interventions. The overall high accuracy in reporting analytical strategies is not surprising, given the long-standing familiarity with multivariate analysis among quantitative sociologists.

Figures 5 and 6 summarize reporting accuracy for the final three CONSORT-SPI domains: Results, Discussion, and Other Important Information. The "Results" section addresses the clarity and completeness of the findings, including participant flow, data collection periods, attrition, effect sizes, and uncertainty around estimates. Compared to earlier sections, this area shows less variability, with most items exhibiting moderate to good levels of compliance.

These patterns likely reflect the fact that sociologists are accustomed to presenting and interpreting quantitative findings. However, there is still room for improvement, particularly because much of the relevant information—such as data collection details (item 14)—can be easily incorporated into manuscripts. In contrast, low reporting accuracy on item 13b (losses and exclusions

after randomization) and item 16 (sample size used in the final analysis) makes it difficult for readers and reviewers to evaluate the robustness of estimated effects.

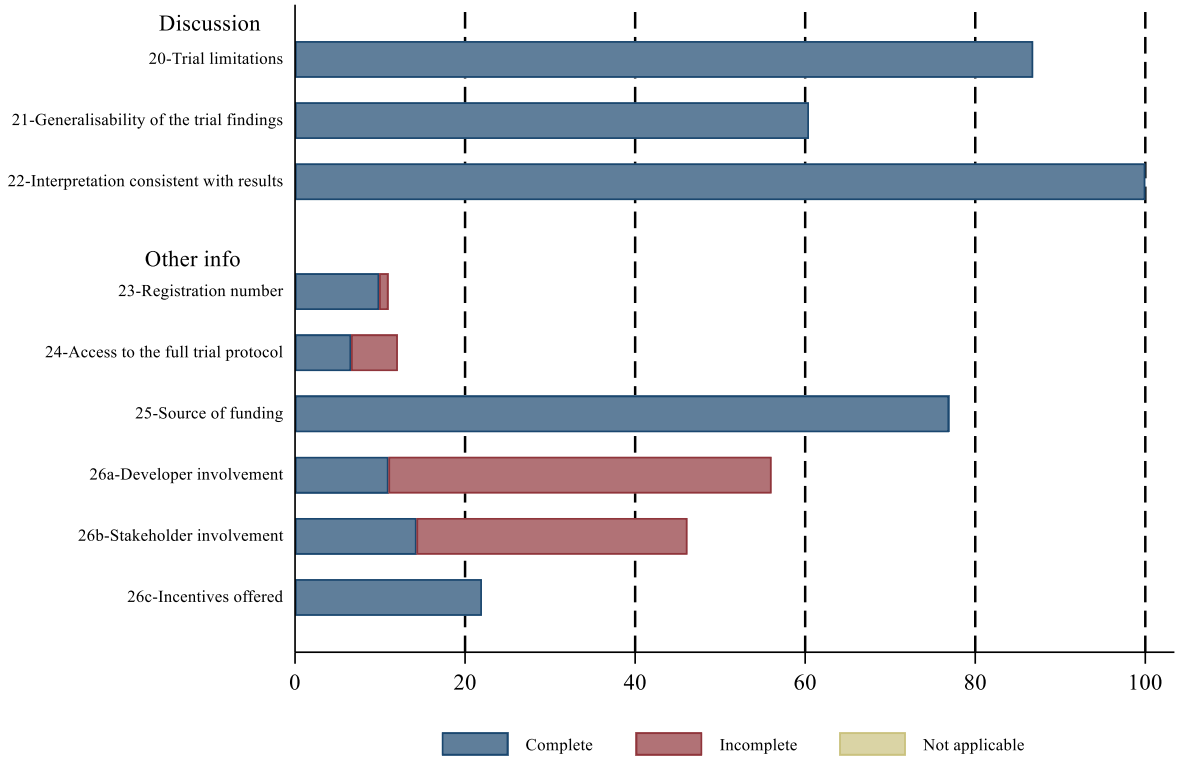
Figure 5. Reporting accuracy of the observed experiments – III (%)



Items 17a and 17b, which address effect size reporting and the presentation of binary outcomes, show relatively high compliance. Accurate reporting in these areas is essential for interpreting the magnitude of results and enabling meta-analysis, suggesting an encouraging degree of rigor in this part of the literature.

The weakest-performing item in this section is item 19, which concerns the reporting of harms or unintended effects. While this topic is understandably less central in sociological than in clinical research, authors should nonetheless attend to the possibility that social interventions may have unintended or adverse consequences (e.g., Petrosino, Turpin-Petrosino, and Finckenauer 2000; Barbetta, Canino, and Cima 2023). Greater attention to this aspect would improve transparency and ethical reflection.

Figure 6. Reporting accuracy of the observed experiments – IV (%)



The "Discussion" section includes three items. Item 20 (trial limitations) is generally well addressed, with an average compliance rate of about 87%. Item 22 (interpretation of findings) is consistently reported, achieving 100% accuracy. However, reporting on item 21—discussion of external validity—is considerably weaker, with fewer than two-thirds (60.4%) of studies explicitly addressing generalizability.

We offer two possible explanations for this gap. First, authors may feel discouraged from emphasizing limited generalizability due to publication incentives. Second, given that experimental effects depend jointly on the randomized sample, context, and implementation, it may be difficult to adequately address external validity in the limited space typically allocated to limitations sections. Regardless, the underreporting of external validity concerns remains a notable weakness.

Items in the final CONSORT-SPI section focus on issues of transparency in the research process, including researcher roles, stakeholder involvement, and protocol availability. These items consistently show low to very low compliance. This not only reinforces concerns about a naïve or underdeveloped approach to experimental design in sociology, but also raises potential concerns about undisclosed conflicts of interest or strategic reporting aligned with stakeholder preferences.

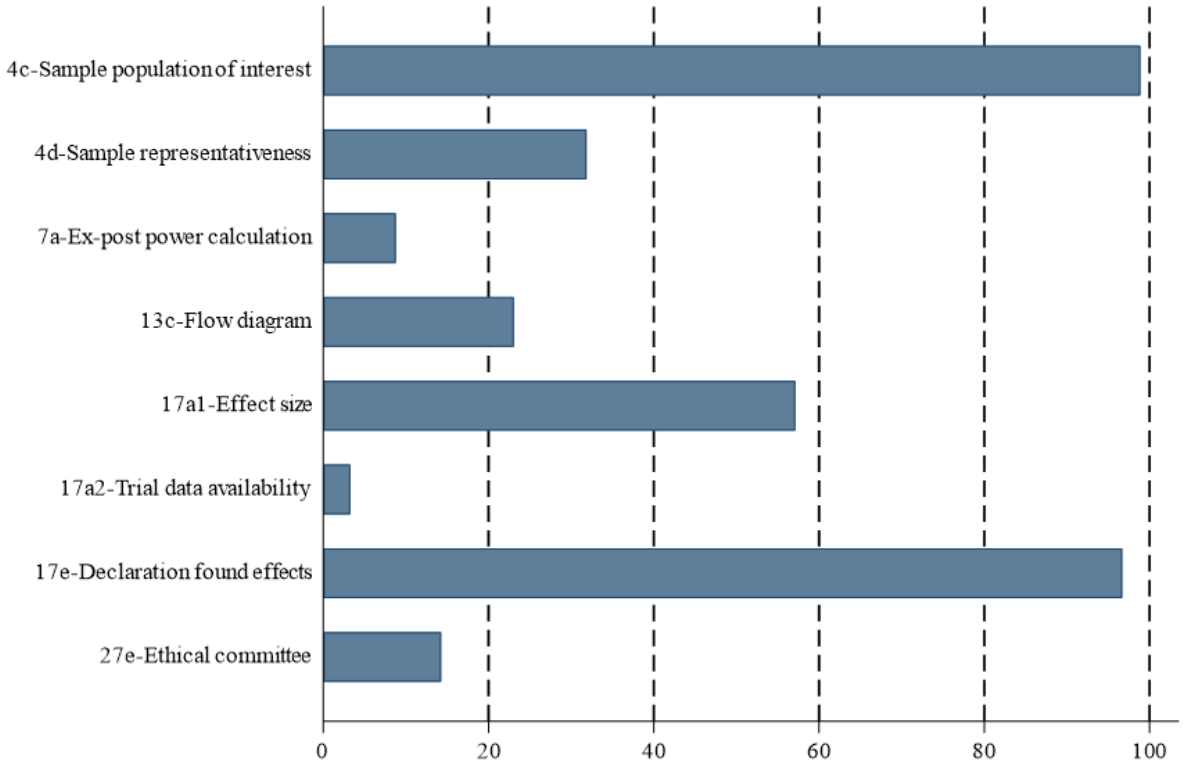
The two most important indicators in this domain—preregistration (item 23) and protocol availability (item 24)—score among the lowest across the entire checklist. These omissions are troubling, as they point to structural sources of opacity in RCT planning and reporting. Without preregistration or a documented protocol, researchers are free to adjust plans post hoc, raising the risk of "fishing for effects," cherry-picking results, or even unconscious bias in selective reporting (Bohannon 2016; Shrout and Rodgers 2018). These practices are widely seen as contributors to the replication crisis in psychology and behavioral economics. Preregistration, by contrast, is increasingly regarded as a crucial safeguard, and making it mandatory has proven effective effective against bias in medical field (Phillips et al 2017).

This issue is particularly acute in field experiments, where intervention designers and evaluators often overlap or collaborate. Consistent with this, items 26a and 26b, which concern stakeholder and developer involvement, are rarely reported in detail. The lack of transparency in these roles limits replicability and makes it difficult to evaluate how the intervention was shaped.

Access to trial data (item a17a2) is another critical weakness: only a small minority of studies offer open data, which further limits the potential for replication and review.

Figure 7 also presents results for additional items we introduced to complement CONSORT-SPI. These items, coded as binary indicators (compliant vs. non-compliant), pertain to methods (e.g., target population, sample representativeness, ex post power calculations), results (e.g., flow diagrams, effect size reporting), and transparency (e.g., ethical approval, data availability, declaration of effects).

Figure 7. Reporting accuracy of the observed experiments – III (additional, non-CONSORT items; % of complete)



Our data show that randomized samples often include units resembling the intervention's target population (item a4c), but are rarely representative in a statistical sense (item a4d). This is frequently due to the reliance on opt-in programs, where researchers control allocation but not initial recruitment. While common in RCTs, this reliance on convenience sampling raises serious concerns about external validity. Contextual features interacting with treatment effects are also often unreported, compounding this limitation.

This limitation should be acknowledged explicitly, as it affects both the policy implications and theoretical generalizability of findings. However, such recognition remains rare.

Finally, nearly all reviewed studies report finding statistically significant effects (item a17e). While not surprising—given the well-documented problem of publication bias (Brodeur et al. 2016; Bernardi, Chakhaia, and Lutz 2017)—this pattern is especially troubling in the context of policy-

relevant field experiments, where null findings are often harder to publish. Limited data access further compounds this problem, as most journals in sociology do not require aut7hors to share data or code.

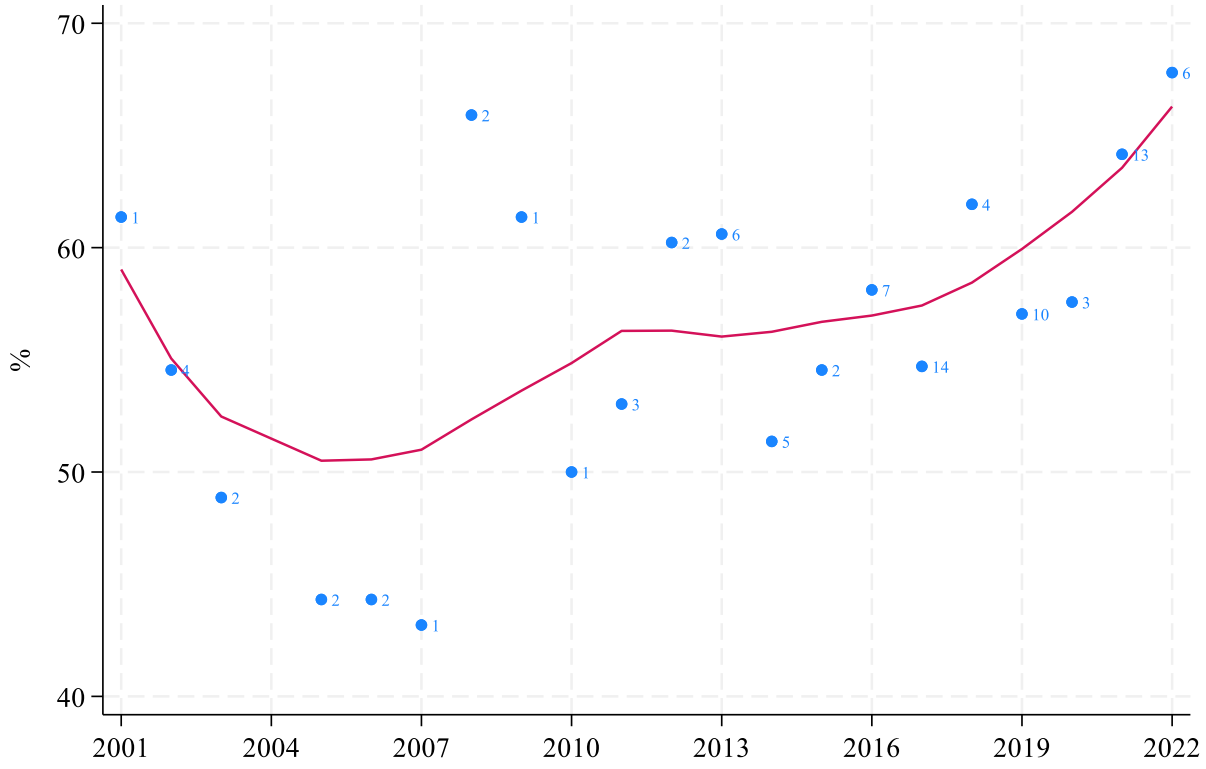
Ethical oversight (item a27e) is also underreported: fewer than one in five studies mention approval by an ethics committee. This suggests that norms and procedures around ethical review remain underdeveloped among sociologists conducting RCTs, even when disadvantaged populations are involved. We expect that growing awareness—and increased scrutiny from funders and journals—will gradually lead to improvements in this area.

6. Results III: the reporting accuracy over time, and its relation to the experiments' features

We now examine whether reporting accuracy has improved over time. To assess this, we computed an additive reporting accuracy index for each experiment, summing the number of items fully compliant with the CONSORT-SPI checklist (excluding those where the "not applicable" option was selected).[3] We then averaged this index by year of publication (2000–2022). The index includes 44 items, yielding a theoretical range from 0 to 44. The overall mean score is 25.4 (SD = 4.4). For ease of interpretation, we rescaled the index to range from 0 to 100, resulting in an average reporting accuracy of 57.7%.

Figure 8 shows a scatterplot of the reporting accuracy index by year of publication, along with a locally weighted (loess) regression line interpolating the data. The number of experiments per year is indicated next to each data point. Excluding the early years—when the number of studies was very low—the overall pattern suggests a slow but steady improvement in reporting accuracy over time, reaching approximately 70% compliance with the extended CONSORT-SPI guidelines by 2022.

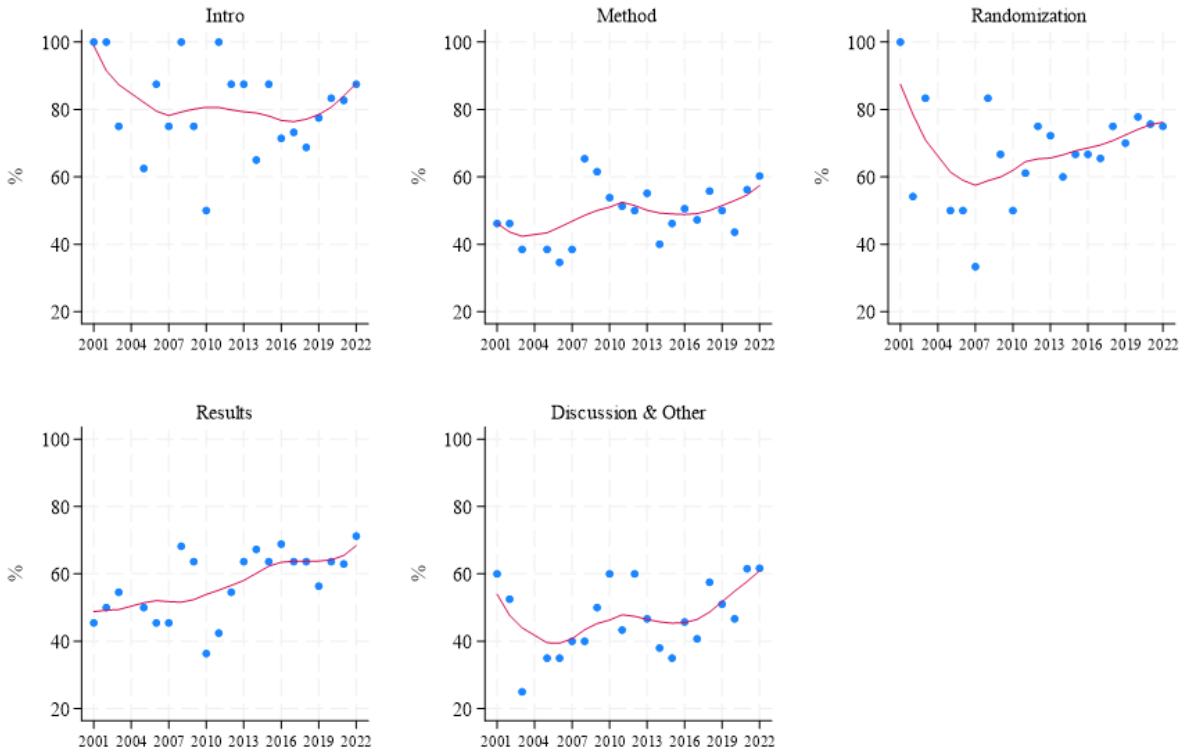
Figure 8. Accuracy of reporting by publication year (percentages)



Note: scatter plot and locally weighted regression estimates with a 0.6 bandwidth.

As noted above, the CONSORT-SPI checklist is divided into six sections. For each section, we calculated a reporting accuracy index by counting the number of items in that section that complied with the guidelines, using the same approach as for the overall index. We merged the "Discussion" and "Other" sections, since the "Discussion" domain contains only three items. Figure 9, similarly to Figure 8, shows the yearly variation in reporting accuracy for each section. Notably, the "Randomization" section exhibits a trend that roughly mirrors the overall pattern—but it consistently shows the lowest reporting accuracy across the entire period. The "Introduction" section remains largely unchanged over time, although it began with relatively high accuracy levels. The remaining three sections show more pronounced change: after some volatility in the early years—likely due to small sample sizes—reporting accuracy remains stable until around 2017, followed by a clear upward trend. Taken together, these results suggest that most sections have seen an improvement in reporting accuracy in recent years.

Figure 9. Accuracy of reporting by publication year and CONSORT checklist section (percentages).



Note: scatter plot and locally weighted regression estimates with a 0.6 bandwidth.

It is noteworthy that the share of pre-registered studies in our sample has increased significantly in recent years, rising from 0% before 2016 to 50% in 2022, and doubling relative to 2021—a trend that signals growing methodological awareness and alignment with best practices.

Finally, we estimated a set of statistical models to examine how reporting accuracy varies according to the characteristics of the experiments. Table 3 presents the results of three ordinary least squares (OLS) regressions: the first column models the overall reporting accuracy index, while the second and third columns focus on two aggregated sub-indices based on the CONSORT-SPI checklist—"Intro, Method & Randomization" and "Results, Discussion & Other"—to avoid constructing indices based on too few items. All models include covariates capturing key features of the experiments, as described in Tables 1 and 2.

Table 3. OLS models of the experiments' reporting accuracy index, in general and according to the CONSORT-SPI checklist section.

	General	Intro, Method & Randomization	Result, Discussion & Other
Period (ref: 2000-2008)			
2009-2015	1.012	0.451	2.629
	(3.192)	(4.049)	(3.207)
2016-2022	3.644	1.644	6.299*
	(3.007)	(3.215)	(3.364)
Area of the study (ref: US)			
Europe	-1.577	-2.063	-0.944
	(2.869)	(2.857)	(3.463)
Developing countries	6.281**	1.683	7.495**
	(2.553)	(2.653)	(3.363)
Sector of the journal (ref: Sectorial)			
Sociological	1.747	2.589	3.104
	(2.415)	(2.992)	(2.891)
Scimago quartiles (ref: Q1)			
Q2-Q4	-9.774***	-10.016***	-12.237***
	(2.883)	(3.029)	(3.682)
Type of intervention (ref: incentives)			
Education	-5.523*	-6.401*	-7.091
	(3.195)	(3.368)	(4.383)
Information	-2.241	0.715	-4.928
	(3.556)	(3.560)	(4.432)
Training	-10.057**	-11.183**	-11.510**
	(4.090)	(4.365)	(5.140)
Other	-3.401	-1.843	-8.147**
	(3.030)	(3.157)	(3.940)
Constant	59.813***	64.707***	59.386***
	(2.795)	(2.870)	(3.299)
Observations	91	91	91
R^2	0.318	0.253	0.323

Note: the reference categories are respectively: 2001-2008; US, Canada, Australia; Sectorial; Q1; Incentives. Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The indexes are expressed in percentage of compliance.

Two findings from Table 3 stand out. The first concerns journal quality, measured using Scimago quartiles.[4] Papers published in higher-ranking journals tend to report with greater accuracy, both overall and across the two aggregated CONSORT-SPI sections. The second key finding concerns the type of intervention: nearly all intervention categories—with only one marginal exception—show negative coefficients relative to incentive-based interventions, which serve as the reference category.[5] This indicates that RCTs evaluating incentive-based interventions are associated with higher reporting accuracy. We suggest three possible explanations for this pattern: (i) incentive-based interventions are typically monetary or financial in nature, making them relatively easy to define, report, and randomly assign, compared to more complex programmatic treatments; (ii) these interventions often target a specific and measurable outcome, and are frequently designed to assess cost-effectiveness, which tends to facilitate more structured reporting; (iii) because incentive-based programs often involve multiple stakeholders, they require greater coordination prior to implementation, which may, in turn, promote more transparent and accurate documentation downstream.

The positive time trend observed in previous analyses is confirmed in the regression models, although the effect is not particularly strong: the estimate reaches statistical significance only for the "Results, Discussion & Other" section ($p < 0.10$). The models also indicate that, on average, studies conducted in developing countries are associated with higher reporting accuracy. One possible explanation is that field experiments have a longer-standing tradition in development research, where they are often implemented by multidisciplinary teams with greater familiarity with experimental protocols and reporting standards.

7. Conclusions

In this paper, we presented a systematic scoping review of field experiments in sociology, focusing on 91 studies published in highly ranked journals between 2000 and 2022. We first described the key

features of these RCTs, then turned to their reporting accuracy, using an expanded version of the CONSORT-SPI checklist (Schulz, Altman, and Moher 2010). Our aim was not only to assess reporting quality, but also to learn from how scholars are designing and implementing field experiments in sociology—a field where the use of RCTs remains relatively recent and less widespread than in neighboring disciplines.

We found that the number of published RCTs has increased over time, but this growth has occurred mainly in sector-specific journals rather than generalist sociological outlets. Many of these studies were authored by multidisciplinary teams, with non-sociologists often listed among the lead authors. Topics and outcomes varied widely, reflecting the broad scope of sociology and the dispersed nature of the community using experiments—making it harder to establish shared methodological standards. Nonetheless, we observed a concentration in education and training studies, many of which involved informational or incentive-based interventions. This first wave of sociological field experiments appears shaped by contexts where RCTs are more easily implemented—that is, where treatments are highly manipulable, randomization logistics are feasible (e.g., via clustering), and funding is available due to political support for evidence-based interventions.

Turning to our central concern—reporting accuracy—we find considerable gaps across the board. Many studies offer insufficient detail to allow meaningful replication or even robust assessment of validity. This opacity poses a serious problem, especially in light of the replication crises facing other fields where RCTs are widely used (e.g., psychology, behavioral economics). As late adopters, sociologists could potentially learn from past mistakes, but our review suggests the field may be repeating them, albeit on a slower timeline.

Editorial constraints may partly explain low reporting accuracy, especially in high-ranking journals where space is limited. But in the case of RCTs, omitting technical details undermines the very claims these studies are designed to support. Editors bear particular responsibility here, as they

can enforce reporting standards and provide space (e.g., online supplements) for technical documentation.

Beyond these structural factors, we identify three critical areas requiring greater attention from sociologists conducting RCTs.

First, there is a general lack of pre-registration. While this practice has increased in recent years, only about 50% of sociological RCTs in our sample were pre-registered. Combined with the common practice of estimating treatment effects on multiple outcomes, this absence of pre-registration leaves experimental evidence particularly vulnerable to questionable causal claims. Pre-registration is one of the most effective safeguards against undesirable opportunistic behaviors, such as "fishing for effects" or, in the absence of adequate statistical awareness, mistaking random significant differences for real effects. This issue is further exacerbated by the structural pressures of academic publishing. Running a field experiment is a costly endeavor—in terms of time, effort, and resources—and the "publish or perish" environment creates strong incentives to report positive results (Grimes, Bauch, and Ioannidis 2018). Pre-registration can mitigate this pressure by promoting mutual accountability among researchers, making it more difficult to engage in intentional or unconscious data manipulation. Although sometimes perceived as a bureaucratic hurdle, the cost–benefit balance strongly favors broader adoption, as already seen in disciplines such as medicine. Beyond promoting transparency, pre-registration offers substantive advantages: it encourages researchers to (i) specify the behavioral theory motivating the intervention; (ii) identify the social mechanisms expected to be triggered; and (iii) clarify the assumptions underlying the intervention's design and expected effects. By requiring early articulation of these components, pre-registration and publicly available protocols increase the likelihood of successful implementation and reduce the risk of misleading conclusions (Ioannidis 2005).

Second, reporting on the roles of researchers, funders, and implementers is often minimal or opaque. This lack of transparency limits the community's ability to detect potential conflicts of

interest—especially when the same parties fund, design, and evaluate the intervention. Such arrangements create incentives to produce positive results, and the problem is exacerbated when there is no preregistration or published protocol. This pattern is frequent in contexts where experiments are relatively new (Abbiati et al. 2022), but it has been documented even in more established RCT domains, such as education (Wolf et al. 2020).

Third, the design choices that shape the internal and external validity of field experiments are frequently underreported. Sociologists often face unique challenges: they lack standardized outcome measures, often rely on self-reports, and must judge the appropriate follow-up period without strong empirical guidance. Masking is rarely feasible, and placebos are nearly nonexistent. These choices—about outcomes, timing, and measurement—should be made explicit in both protocols and published reports, yet they are often absent or only briefly mentioned. Notably, over one-third of reviewed studies failed to discuss external validity at all.

These findings reinforce the conclusion that many sociological field experiments reflect a limited understanding of the methodological fragilities inherent to RCTs. For instance, power calculations—a basic feature of experimental design—are reported in only one out of five studies. Greater methodological awareness is needed: randomization alone does not ensure credible causal inference, and accurate reporting is essential to evaluate the robustness of results (Andrew et al. 1994; Torgerson and Torgerson 2008; Gerber and Green 2012; Deaton and Cartwright 2018).

Summing up, there is substantial room for improvement in how RCTs are reported in sociology—but the trend is encouraging. Reporting accuracy has increased in recent years, suggesting growing awareness and adoption of better practices. We hope this paper contributes to a broader discussion about the need for tailored reporting guidelines in sociology, building on CONSORT-SPI but adapted to the realities of social research. Establishing shared benchmarks—through collective efforts involving both researchers and journal editors—could elevate the quality, credibility, and cumulative value of experimental research in the discipline.

8. Limitations

Our study has three main limitations, which we hope future research will address as part of a broader effort to assess the quality of experimental evidence in sociology through direct evaluations of reporting practices.

First, we limited our review to field experiments, meaning that our conclusions may not extend to other forms of experimental research. For example, we do not know whether survey-based or vignette experiments—more established within sociological methodology—would perform better on the CONSORT-SPI checklist, owing to their longer-standing integration into the discipline.

Second, we focused exclusively on articles published in highly ranked journals, which may bias our findings. While some experimental reports—possibly associated with the RCTs we reviewed—might offer richer methodological detail, they are often not referenced in published papers. Still, we deliberately examined the most visible outlets in the field, where reporting practices are likely to influence how future sociological experiments are designed, conducted, and published.

Finally, our adaptation of the CONSORT-SPI 2018 checklist highlights the absence of discipline-specific standards in sociology. To our knowledge, no reporting guidelines currently exist that are tailored to sociological experiments. Our development of additional items and refinements to existing criteria reflects an extensive deliberation within our author team, but it may still carry biases stemming from the team's composition and expertise. We fully acknowledge this limitation and view our contribution as a first step toward a collective conversation within the sociological community about the role of RCTs and the need for transparent, standardized reporting.

NOTES

1. The selection was conducted in November 2022.

2. Items allowing the “Incomplete” option include: 1b; 2b; 3a; 8b; 17b; 23; 24; 26a; 26b, while those allowing “Not applicable” include: 5c; 9b; 11a; 11b; 17b; 19.

3. Categories complete and incomplete are jointly coded as 1, while non-compliance as 0. To some extent, the compliance is then overestimated. The accuracy index does not consider those six items (cs5c; cs9b; cs11a; cs11b; cs17b; cs19) including the “Not applicable” answer option.

4. The Scimago quartiles refer to the publication year.

5. It could be the case that an intervention includes more than one instrument (see table 2).

REFERENCES

Abbiati, Giovanni, Gianluca Argentin, Davide Azzolini, Gabriele Ballarino, and Loris Vergolini. 2022. "Experimental research in education: An appraisal of the Italian experience." Schweizerische Zeitschrift Fuer Soziologie , 48(1): 21-46.

Andrew, Erik, Anis Aslam, Tom Chalmers, Mildred Cho, Mike Clarke, David Felson, Peter Gøtzsche, Richard Greene, Alejandro Jadad, Wayne Jonas, Terry Klassen, Paul Knipschild, Andreas Laupacis, Curtis L. Meinert, David Moher, Graham Nichol, Andy Oxman, Marie-France Penman, Stuart Pocock, Joan Reisch, David Sackett, Kenneth Schulz, Judy Snider, Peter Tugwell, Jon Tyson, France Varin, Wikke Walop, Sharon Walsh, and George Wells. 1994. "A Proposal for Structured Reporting of Randomized Controlled Trials." *JAMA* 272(24):1926–1931.

Baldassarri, Delia, and Michael Abascal. 2017. "Field Experiments Across the Social Sciences." *Annual Review of Sociology* 43:41–73.

Barbetta, Gian Paolo, Paolo Canino, and Sara Cima. 2023. "Let's Tweet Again? Social Networks and Literature Achievement in High School Students." *Education Finance and Policy* 18(4):676–707.

Barrera, Davide, Klarita Gërxhani, Bernhard Kittel, Luis Miller and Tobias Wolbring. 2025. *Experimental Sociology. Outline of a Scientific Field*. Cambridge: Cambridge University Press

Begley, Charles G., and Lee M. Ellis. 2012. "Raise Standards for Preclinical Cancer Research." *Nature* 483:531–533.

Berk, Richard A. 2005. "Randomized Experiments as the Bronze Standard." *Journal of Experimental Criminology* 1(4):417–433.

Bernardi, Fabrizio, Lila Chakhaia, and Lutz Leopold. 2017. "Sing Me a Song with Social Significance': The (Mis)Use of Statistical Significance Testing in European Sociological Research." *European Sociological Review* 33(1):1–15.

Bohannon, John. 2016. "About 40% of Economics Experiments Fail Replication Survey." *Science*. (https://www.science.org/content/article/about-40-economics-experiments-fail-replication-survey-rev2)

Boudon, Raymond. 1979. *The Logic of Social Action*. London: Routledge.

Boudon, Raymond. 2002. "Sociology That Really Matters: European Academy of Sociology, First Annual Lecture." *European Sociological Review* 18(3):371–378.

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg. 2016. "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics* 8(1):1–32.

Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, MA: Belknap Press of Harvard University Press.

Deaton, Angus, and Nancy Cartwright. 2018. "Understanding and misunderstanding randomized controlled trials" *Social Science & Medicine* 210: 2-21.

Fatterman, David. 2000. *Foundations of Empowerment Evaluation*. Thousand Oaks, CA: Sage.

Gangl, Markus. 2010. "Causal Inference in Sociological Research." *Annual Review of Sociology* 36:21–47.

Garaway, Gail B. 1995. “Participatory Evaluation.” *Studies in Educational Evaluation* 21(1):85–102.

Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: Norton.

Goldthorpe, John E. 2000. *On Sociology: Numbers, Narratives, and the Integration of Research and Theory*. Oxford: Oxford University Press.

Goldthorpe, John E. 2004. “Sociology as Social Science and Cameral Sociology: Some Further Thoughts: European Academy of Sociology, Third Annual Lecture, Paris, 25 October 2003.” *European Sociological Review* 20(2):97–105.

Goldthorpe, John E. 2015. *Sociology as a Population Science*. Cambridge: Cambridge University Press.

Gould, Elliot, Hannah S. Fraser, Timothy H. Parker, Shinichi Nakagawa, Simon C. Griffith, Peter A. Vesk, Fiona Fidler, Daniel G. Hamilton, Robin N. Abbey-Lee, Jessica K. Abbott, Luis A. Aguirre, Carles Alcaraz, Irith Aloni, Drew Altschul, Kunal Arekar, Jeff W. Atkins, Joe Atkinson, Christopher M. Baker, Meghan Barrett, Kristian Bell, Suleiman Kehinde Bello, Iván Beltrán, Bernd J. Berauer, Michael Grant Bertram, Peter D. Billman, Charlie K. Blake, Shannon Blake, Louis Bliard, Andrea Bonisoli-Alquati, Timothée Bonnet, Camille Nina Marion Bordes, Aneesh P. H. Bose, Thomas Botterill-James, Melissa Anna Boyd, Sarah A. Boyle, Tom Bradfer-Lawrence, Jennifer Bradham, Jack A. Brand, Martin I. Brengdahl, Martin Bulla, Luc Bussière, Ettore Camerlenghi, Sara E. Campbell, Leonardo L. F. Campos, Anthony Caravaggi, Pedro Cardoso, Charles J. W. Carroll, Therese A. Catanach, Xuan Chen, Heung Ying Janet Chik, Emily Sarah Choy, Alec Philip Christie, Angela Chuang, Amanda J. Chunco, Bethany L. Clark, Andrea Contina, Garth A. Covernton, Murray P. Cox, Kimberly A. Cressman, Marco Crotti, Connor Davidson Crouch, Pietro B. D’Amelio, Alexandra Allison de Sousa, Timm Fabian Döbert, Ralph Dobler, Adam J. Dobson, Tim S. Doherty, Szymon Marian Drobniak, Alexandra Grace Duffy, Alison B. Duncan, Robert P. Dunn, Jamie Dunning, Trishna Dutta, Luke Eberhart-Hertel, Jared Alan Elmore, Mahmoud Medhat Elsherif, Holly M. English, David C. Ensminger, Ulrich Rainer Ernst, Stephen M. Ferguson, Esteban Fernandez-Juricic, Thalita Ferreira-Arruda, John Fieberg, Elizabeth A. Finch, Evan A. Fiorenza, David N. Fisher, Amélie Fontaine, Wolfgang Forstmeier, Yoan Fourcade, Graham S. Frank, Cathryn A. Freund, Eduardo Fuentes-Lillo, Sara L. Gandy, Dustin G. Gannon, Ana I. García-Cervigón, Alexis C. Garretson, Xuezhen Ge, William L. Geary, Charly Géron, Marc Gilles, Antje Girndt, Daniel Gliksman, Harrison B. Goldspiel, Dylan G. E. Gomes, Megan Kate Good, Sarah C. Goslee, J. Stephen Gosnell, Eliza M. Grames, Paolo Gratton, Nicholas M. Grebe, Skye M. Greenler, Maaike Griffioen, Daniel M. Griffith, Frances J. Griffith, Jake J. Grossman, Ali Güncan, Stef Haesen, James G. Hagan, Heather A. Hager, Jonathan Philo Harris, Natasha Dean Harrison, Sarah Syedia Hasnain, Justin Chase Havird, Andrew J. Heaton, María Laura Herrera-Chaustre, Tanner J. Howard, Bin-Yan Hsu, Fabiola Iannarilli, Esperanza C. Iranzo, Erik N. K. Iverson, Saheed Olaide Jimoh, Douglas H. Johnson, Martin Johnsson, Jesse Jorna, Tommaso Jucker, Martin Jung, Ineta Kačergytė, Oliver Kaltz, Alison Ke, Clint D. Kelly, Katharine Keogan, Friedrich Wolfgang Keppeler, Alexander K. Killion, Dongmin Kim, David P. Kochan, Peter Korsten, Shan Kothari, Jonas Kuppler, Jillian M. Kusch, Malgorzata Lagisz, Kristen Marianne Lalla, Daniel J. Larkin, Courtney L. Larson, Katherine S. Lauck, M. Elise Lauterbur, Alan Law, Don-Jean Léandri-Breton, Jonas J. Lembrechts, Kiara L’Herpiniere, Eva J. P. Lievens, Daniela Oliveira de Lima, Shane Lindsay, Martin Luquet, Ross MacLeod, Kirsty H. Macphie, Kit Magellan, Magdalena M. Mair, Lisa E. Malm, Stefano Mammola, Caitlin P. Mandeville, Michael Manhart, Laura Milena Manrique-Garzon, Elina Mäntylä, Philippe Marchand, Benjamin Michael Marshall, Charles A. Martin, Dominic Andreas Martin, Jake Mitchell Martin, April Robin Martinig, Erin S. McCallum, Mark McCauley, Sabrina M. McNew, Scott J. Meiners, Thomas Merkling, Marcus Michelangeli, Maria Moiron, Bruno Moreira, Jennifer Mortensen, Benjamin Mos, Taofeek Olatunbosun Muraina, Penelope Wrenn Murphy, Luca Nelli, Petri Niemelä, Josh Nightingale, Gustav Nilsonne, Sergio Nolazco, Sabine S. Nooten, Jessie Lanterman Novotny, Agnes Birgitta Olin, Chris L. Organ, Kate L. Ostevik, Facundo Xavier Palacio, Matthieu Paquet, Darren James Parker, David J. Pascall, Valerie J. Pasquarella, John Harold Paterson, Ana Payo-Payo, Karen Marie Pedersen, Grégoire Perez, Kayla I. Perry,

Patrice Pottier, Michael J. Proulx, Raphaël Proulx, Jessica L Pruett, Veronarindra Ramananjato, Finaritra Tolotra Randimbiarison, Onja H. Razafindratsima, Diana J. Rennison, Federico Riva, Sepand Riyahi, Michael James Roast, Felipe Pereira Rocha, Dominique G. Roche, Cristian Román-Palacios, Michael S. Rosenberg, Jessica Ross, Freya E. Rowland, Deusdedith Rugemalila, Avery L. Russell, Suvi Ruuskanen, Patrick Saccone, Asaf Sadeh, Stephen M. Salazar, Kris Sales, Pablo Salmón, Alfredo Sánchez-Tójar, Leticia Pereira Santos, Francesca Santostefano, Hayden T. Schilling, Marcus Schmidt, Tim Schmoll, Adam C. Schneider, Allie E. Schrock, Julia Schroeder, Nicolas Schtickzelle, Nick L. Schultz, Drew A. Scott, Michael Peter Scroggie, Julie Teresa Shapiro, Nitika Sharma, Caroline L. Shearer, Diego Simón, Michael I. Sitvarin, Fabrício Luiz Skupien, Heather Lea Slinn, Grania Polly Smith, Jeremy A. Smith, Rahel Sollmann, Kaitlin Stack Whitney, Shannon Michael Still, Erica F. Stuber, Guy F. Sutton, Ben Swallow, Conor Claverie Taff, Elina Takola, Andrew J. Tanentzap, Rocío Tarjuelo, Richard J. Telford, Christopher J. Thawley, Hugo Thierry, Jacqueline Thomson, Svenja Tidau, Emily M. Tompkins, Claire Marie Tortorelli, Andrew Trlica, Biz R. Turnell, Lara Urban, Stijn Van de Vondel, Jessica Eva Megan van der Wal, Jens Van Eeckhoven, Francis van Oordt, K. Michelle Vanderwel, Mark C. Vanderwel, Karen J. Vanderwolf, Juliana Vélez, Diana Carolina Vergara-Florez, Brian C. Verrelli, Marcus Vinícius Vieira, Nora Villamil, Valerio Vitali, Julien Vollering, Jeffrey Walker, Xanthe J. Walker, Jonathan A. Walter, Pawel Waryszak, Ryan J. Weaver, Ronja E. M. Wedegärtner, Daniel L. Weller, Shannon Whelan, Rachel Louise White, David William Wolfson, Andrew Wood, Scott W. Yanco, Jian D. L. Yen, Casey Youngflesh, Giacomo Zilio, Cédric Zimmer, Gregory Mark Zimmerman, and Rachel A. Zitomer. 2025. “Same Data, Different Analysts: Variation in Effect Sizes Due to Analytical Decisions in Ecology and Evolutionary Biology.” *BMC Biology* 23(35).

Grant, Sean, Evan Mayo-Wilson, Paul Montgomery, Geraldine Macdonald, Susan Michie, Sally Hopewell, and David Moher. 2018. “CONSORT-SPI 2018 Explanation and Elaboration: Guidance for Reporting Social and Psychological Intervention Trials.” *Trials* 19:1–18.

Green, Donald P., and Alan S. Gerber. 2003. “The Underprovision of Experiments in Political Science.” *Annals of the American Academy of Political and Social Science* 589:94–112.

Grimes, David R., Chris T. Bauch, and John P. A Ioannidis. 2018. “Modelling Science Trustworthiness under Publish or Perish Pressure.” *Royal Society Open Science* 5:171511.

Harrison, Glenn W., and John A. List. 2004. “Field Experiments.” *Journal of Economic Literature* 42:1009–1055.

Hellman, Samuel, and Deborah S. Hellman. 1991. “Of Mice but Not Men: Problems of the Randomized Clinical Trial.” *New England Journal of Medicine* 324:1585–1589.

Holland, Paul W. 1986. “Statistics and Causal Inference” *Journal of the American Statistical Association* 81(396): 945–960.

Hopewell, Sally, An-Wen Chan, Gary S. Collins, Asbjørn Hróbjartsson, David Moher, Kenneth F. Schulz, and Isabelle Boutron. 2025. “CONSORT 2025 Statement: Updated Guideline for Reporting Randomised Trials.” *The Lancet*.

Ioannidis, John P. A. 2005. “Why Most Published Research Findings Are False.” *PLoS Medicine* 2(8):e124.

Jackson, Michelle, and David R. Cox. 2013. “The Principles of Experimental Design and Their Application in Sociology.” *Annual Review of Sociology* 39:27–49.

Luca, Michael, and Max H. Bazerman. 2020. *The Power of Experiments: Decision Making in a Data-Driven World*. Cambridge, MA: MIT Press.

Lundberg, Ian, Rebecca Johnson, and Stewart M. Brandon. 2021. “What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory.” *American Sociological Review* 86(3):532–565.

Morgan, Stephen L., and Christopher Winship. 2007. *Counterfactuals and Causal Inference. Methods and Principles for Social Research.* Cambridge: Cambridge University Press.

Mitze, Timo D., and Brita Manago. 2022. “The Past, Present, and Future of Experimental Methods in the Social Sciences.” *Social Science Research* 108.

Open Science Collaboration. 2015. “Estimating the Reproducibility of Psychological Science.” *Science* 349(6251).

Pawson, Ray, and Nick Tilley. 1997. *Realistic Evaluation*. London: Sage.

Petrosino, Anthony, Carolyn Turpin-Petrosino, and James O. Finckenauer. 2000. “Well-Meaning Programs Can Have Harmful Effects! Lessons from Experiments of Programs Such as Scared Straight.” *Crime & Delinquency* 46(3):354–379.

Phillips Adam T., Desai Nihar R., Krumholz Harlan M., Zou Constance X., Miller Jennifer E. and Ross Joseph S. 2017. “Association of the FDA Amendment Act with trial registration, publication, and outcome reporting.” *Trials* 18(1):333.

Schulz, Kenneth F., Douglas G. Altman, and David Moher. 2010. “CONSORT Group. CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials.” *BMC Medicine* 8:18.

Shrout, Patrick E., and Joseph L. Rodgers. 2018. “Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis.” *Annual Review of Psychology* 69:487–510.

Torbjörn, Callréus. 2022. “The Randomised Controlled Trial at the Intersection of Research Ethics and Innovation.” *Pharmaceutical Medicine* 36:287–293.

Torgerson, Carole, and David Torgerson. 2008. *Designing Randomised Trials in Health, Education and the Social Sciences*. New York: Springer.

Torgerson, Carole, and David Torgerson. 2012. “The Educational Importance of Ethos” pp. 203-214 in Education Matters 60 years of the British Journal of Educational Studies, edited by Arthur, James, Jon Davison, and Richard Pring. London: Routledge.

Tricco, Andrea C., Erin Lillie, Wasifa Zarin, Kelly K. O'Brien, Heather Colquhoun, Danielle Levac, David Moher, Micah D. J Peters, Tanya Horsley, Laura Weeks, Susanne Hempel, Elie A. Akl, Christine Chang, Jessie McGowan, Lesley Stewart, Lisa Hartling, Adrian Aldcroft, Michael G. Wilson, Chantelle Garritty, Simon Lewin, Christina M. Godfrey, Marilyn T. Macdonald, Etienne V. Langlois, Karla Soares-Weiser, Jo Moriarty, Tammy Clifford, Özge Tunçalp, and Sharon E. Straus. 2018. “PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation.” *Annals of Internal Medicine* 169(7):467–473.

Wolf, Rebecca, Jessica Morrison, Alexandra Inns, Robert Slavin, and Karen Risman. 2020. “Average Effect Sizes in Developer-Commissioned and Independent Evaluations.” *Journal of Research on Educational Effectiveness* 13(2):428–447.

Supplementary materials

A. List of selected journals

B. Complete strings for database search

C. Codebook. New and revised items description

D. Studies references

A. List of selected journals

Journal	Type	ISSN	eISSN
Accounting, Organizations And Society	Sectorial	3613682	
Acta Sociologica	Sociological	00016993	15023869
Administration And Society	Sectorial	953997	15523039
Agriculture And Human Values	Sectorial	0889048X	15728366
American Behavioral Scientist	Sectorial	15523381	27642
American Journal Of Cultural Sociology	Sociological	20497113	20497121
American Journal Of Economics And Sociology	Sociological	00029246	15367150
American Journal Of Evaluation	Sectorial	10982140	
American Journal Of Sociology	Sociological	00029602	15375390
American Sociological Review	Sociological	00031224	19398271
Annals Of The American Academy Of Political And Social Science	Sectorial	15523349	27162
Annual Review Of Law And Social Science	Sectorial	15503585	15503631
Archives Europeennes De Sociologie	Sociological		
Armed Forces & Society	Sectorial	0095327X	15560848
Berliner Journal Fur Soziologie	Sociological	08631808	18622593
Biodemography And Social Biology	Sectorial	19485565	19485573
British Journal Of Sociology	Sociological	00071315	14684446
British Journal Of Sociology Of Education	Sociological	01425692	14653346
Cambridge Journal Of Regions, Economy And Society	Sociological	17521386	17521378
Canadian Journal Of Sociology-Cahiers Canadiens De Sociologie	Sociological	03186431	03186431
Canadian Review Of Sociology-Revue Canadienne De Sociologie	Sociological	17556171	1755618X
Chinese Sociological Review	Sociological	21620555	21620563

Cities	Sectorial	2642751	
Comparative Migration Studies	Sectorial	2214594X	
Comparative Studies In Society And History	Sectorial	00104175	14752999
Contemporary Sociology-A Journal Of Reviews	Sociological	00943061	19398638
Contributions To Indian Sociology	Sociological	00699667	09730648
Convergencia-Revista De Ciencias Sociales	Sociological	24485799	24485799
Crime And Justice	Sectorial	21530416	1923234
Cultural Sociology	Sociological	17499755	17499763
Current Perspectives In Social Theory	Sociological		
Current Sociology	Sociological	00113921	14617064
Deviance Et Societe	Sectorial	03787931	03785807
Drustvena Istrazivanja	Sectorial	13300288	13300288
Du Bois Review-Social Science Research On Race	Sectorial		
Early Childhood Research Quarterly	Sectorial	8852006	
Eastern European Countryside	Sectorial	12328855	23008717
Economic And Social Review	Sectorial	00129984	00129984
Economy And Society	Sectorial	03085147	14695766
Educational Technology And Society	Sectorial	11763647	14364522
Environmental Sociology	Sociological	23251042	
Ethnic And Racial Studies	Sectorial	01419870	14664356
Ethnography	Sociological	14661381	17412714
European Journal Of Social Theory	Sociological	13684310	14617137
European Societies	Sociological	14616696	14698307
European Sociological Review	Sociological	02667215	14682672
Evaluation	Sectorial	13563890	14617153

Filosofija-Sociologija	Sociological	02357186	02357186
Food Policy	Sectorial	3069192	
Futures	Sectorial	163287	
Gender & Society	Sociological	08912432	15523977
Health And Place	Sectorial	13538292	
Health And Social Care In The Community	Sectorial	9660410	13652524
Health Sociology Review	Sociological	14461242	18393551
Housing Studies	Sectorial	14661810	2673037
Human Ecology Review	Sectorial	10744827	22040919
Human Studies	Sectorial	01638548	1572851X
Innovation-The European Journal Of Social Science Research	Sectorial	13511610	14698412
International Journal Of Comparative Sociology	Sociological	00207152	17452554
International Journal Of Public Opinion Research	Sectorial	14716909	9542892
International Journal Of Urban And Regional Research	Sectorial	3091317	14682427
International Political Sociology	Sectorial	17495679	17495687
International Review For The Sociology Of Sport	Sociological	10126902	14617218
International Sociology	Sociological	02685809	14617242
Internet Research	Sectorial	10662243	
Journal For The Scientific Study Of Religion	Sectorial	00218294	14685906
Journal Of Aging And Health	Sectorial	15526887	8982643
Journal Of Applied Behavior Analysis	Sectorial	19383703	218855
Journal Of Benefit-Cost Analysis	Sectorial	21522812	21945888
Journal Of Contemporary Ethnography	Sectorial	08912416	15525414
Journal Of Criminal Justice	Sectorial	472352	
Journal Of Cultural Economy	Sectorial	17530350	17530369

Journal Of Democracy	Sectorial	10455736	10863214
Journal Of Early Adolescence	Sectorial	15525449	2724316
Journal Of Economic Inequality	Sectorial	15691721	15738701
Journal Of Elections, Public Opinion And Parties	Sectorial	17457289	
Journal Of European Public Policy	Sectorial	14664429	13501763
Journal Of Family Violence	Sectorial	8857482	15732851
Journal Of Gambling Studies	Sectorial	10505350	15733602
Journal Of Health And Social Behavior	Sectorial	00221465	21506000
Journal Of Historical Sociology	Sociological	09521909	14676443
Journal Of Immigrant & Refugee Studies	Sectorial	15562948	15562956
Journal Of Information Technology And Politics	Sectorial	19331681	1933169X
Journal Of Law And Society	Sectorial	0263323X	14676478
Journal Of Leadership And Organizational Studies	Sectorial	19397089	15480518
Journal Of Marriage And Family	Sectorial	00222445	17413737
Journal Of Mathematical Sociology	Sectorial	0022250X	15455874
Journal Of Organizational Behavior	Sectorial	10991379	8943796
Journal Of Policy Analysis And Management	Sectorial	2768739	15206688
Journal Of Public Administration Research And Theory	Sectorial	14779803	10531858
Journal Of Race, Ethnicity And Politics	Sectorial	20566085	
Journal Of Racial And Ethnic Health Disparities	Sectorial	21968837	
Journal Of Rural Studies	Sectorial	7430167	
Journal Of Service Research	Sectorial	10946705	15527379
Journal Of Sex Research	Sectorial	224499	15598519
Journal Of Social And Personal Relationships	Sectorial	2654075	
Journal Of Sociolinguistics	Sectorial	13606441	14679841

Journal Of Sociology	Sociological	14407833	17412978
Journal Of Sport & Social Issues	Sectorial	01937235	15527638
Journal Of The History Of Sexuality	Sectorial	10434070	15353605
Journal Of Urban Affairs	Sectorial	7352166	14679906
Journals Of Gerontology - Series B Psychological Sciences And Social Sciences	Sectorial	10795014	17585368
Kolner Zeitschrift Fur Soziologie Und Sozialpsychologie	Sociological	00232653	1861891X
Law And Society Review	Sectorial	239216	15405893
Media Culture & Society	Sectorial	01634437	14603675
Mobilities	Sectorial	17450101	1745011X
Mobilization	Sectorial	1086671X	19381514
New Media And Society	Sectorial	14614448	14617315
Policing And Society	Sectorial	10439463	
Policy And Politics	Sectorial	3055736	14708442
Policy And Society	Sectorial	14494035	
Policy Studies Journal	Sectorial	0190292X	15410072
Polish Sociological Review	Sectorial	12311413	12311413
Politics & Society	Sectorial	00323292	15527514
Public Administration	Sectorial	14679299	333298
Public Administration Review	Sectorial	333352	15406210
Public Choice	Sectorial	15737101	485829
Public Opinion Quarterly	Sociological	0033362X	15375331
Qualitative Research	Sociological	14687941	17413109
Qualitative Sociology	Sociological	01620436	15737837
Race & Class	Sectorial	03063968	17413125
Race And Justice	Sectorial	21533687	

Race And Social Problems	Sectorial	18671748	18671756
Rationality And Society	Sociological	10434631	14617358
Research In Social Stratification And Mobility	Sociological	02765624	18785654
Review Of Religious Research	Sectorial	0034673X	22114866
Revista Espanola De Investigaciones Sociologicas	Sociological	02105233	19885903
Revista Internacional De Sociologia	Sociological	00349712	1988429X
Revue Francaise De Sociologie	Sociological	00352969	19585691
Rural Sociology	Sociological		
Science Technology And Human Values	Sectorial	1622439	15528251
Sexuality Research And Social Policy	Sectorial	15536610	18689884
Social Compass	Sociological	00377686	14617404
Social Development	Sectorial	14679507	0961205X
Social Forces	Sociological	00377732	15347605
Social Indicators Research	Sociological	03038300	15730921
Social Issues And Policy Review	Sociological	17512409	17512395
Social Justice Research	Sociological	08857466	15736725
Social Movement Studies	Sectorial	14742837	14742829
Social Networks	Sectorial	03788733	18792111
Social Policy And Administration	Sectorial	1445596	14679515
Social Problems	Sociological	00377791	15338533
Social Science Quarterly	Sociological	00384941	15406237
Social Science Research	Sociological	0049089X	10960317
Society	Sectorial	01472011	19364725
Society & Animals	Sectorial	10631119	15685306
Society & Natural Resources	Sectorial	08941920	15210723

Society And Mental Health	Sectorial	21568693	21568731
Socio-Economic Review	Sociological	14751461	1475147X
Sociológia - slovak sociological review	Sociological	00491225	13368613
Sociologia Ruralis	Sociological	00380199	14679523
Sociological Forum	Sociological	08848971	15737861
Sociological Inquiry	Sociological	00380245	1475682X
Sociological Methodology	Sociological	00811750	14679531
Sociological Methods & Research	Sociological	00491241	15528294
Sociological Perspectives	Sociological	07311214	15338673
Sociological Quarterly	Sociological	00380253	15338525
Sociological Research Online	Sociological	13607804	13607804
Sociological Review	Sociological	00380261	1467954X
Sociological Science	Sociological		23306696
Sociological Spectrum	Sociological	02732173	15210707
Sociological Theory	Sociological	07352751	14679558
Sociologicky Casopis-Czech Sociological Review	Sociological	00380288	2336128X
Sociologie Du Travail	Sociological	00380296	17775701
Sociologisk Forskning	Sociological	00380342	00380342
Sociologus	Sociological	00380377	18655106
Sociology Compass	Sociological	17519020	17519020
Sociology Of Education	Sectorial	00380407	19398573
Sociology Of Health & Illness	Sectorial	01419889	14679566
Sociology Of Race And Ethnicity	Sociological	23326492	23326505
Sociology Of Religion	Sectorial	10694404	17598818
Sociology Of Sport Journal	Sectorial	07411235	15432785

Sociology-The Journal Of The British Sociological Association	Sociological	00380385	14698684
Sotsiologicheskie Issledovaniya	Sociological	01321625	01321625
South European Society And Politics	Sectorial	13608746	17439612
Soziale Welt-Zeitschrift Fur Sozialwissenschaftliche Forschung Und Praxis	Sociological	00386073	00386073
Sport In Society	Sectorial	17430437	17430445
Teaching Sociology	Sociological	0092055X	1939862X
Technology In Society	Sectorial	0160791X	
Telos	Sectorial	00906514	1940459X
Tempo Social	Sectorial	01032070	18094554
Theoretical Criminology	Sectorial	14617439	13624806
Theory And Research In Social Education	Sectorial	933104	21631654
Theory And Society	Sociological	03042421	15737853
Theory, Culture And Society	Sociological	2632764	14603616
Urban Affairs Review	Sectorial	10780874	
Urban Review	Sectorial	15731960	420972
Voluntas	Sectorial	9578765	15737888
Work And Occupations	Sectorial	07308884	15528464
Work Employment And Society	Sectorial	09500170	14698722
Work, Aging And Retirement	Sectorial	20544650	
World Development	Sectorial	0305750X	18735991
Young	Sectorial	11033088	17413222
Youth & Society	Sectorial	0044118X	15528499
Zeitschrift Fur Soziologie	Sociological	03401804	23660325

B. Complete strings for database search

Database	Strategy	Results
SCOPUS	TITLE-ABS-KEY (rct OR random* OR "randomized controlled trial" OR "randomised controlled trial") OR TITLE-ABS-KEY ("control group" AND experiment*) AND PUBYEAR > 1999 AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SUBJAREA , "SOCI")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SRCTYPE , "j"))	Search run in November 2022 12,214 results
Web of Science	AB=(RCT or random* or "randomized controlled trial" or "randomised controlled trial") OR ("control group" and experiment*) Limiters: Date Published: after 2000 English language Academic journals	Search run in November 2022 2,046 results
SCOPUS	TITLE-ABS-KEY ("field experiment*" OR "field trial*") AND (sociolog*) AND PUBYEAR > 1999 AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SUBJAREA , "SOCI")) AND (LIMIT-TO (LANGUAGE , "English"))	Search run in December 2022 800 results
Web of Science	TITLE =("field experiment*" OR "field trial*") or AB=("field experiment*" OR "field trial*") Limiters: Date Published: after 2000 English language Academic journals	Search run in December 2022 128 results

C. Codebook. New and revised items' description

Item code	Description and differences with the official CONSORT-SPI	Coding	Item developed for the study
a0a, a0b, a0c	**Brief description:** A brief description on the experiment: type of intervention; the dependent variables; targeted population.	• OPEN	YES
a0e, a0f, a0g	**Scientific field:** General and specific disciplinary field of the research, as well as the country or countries where the experiments habe been conducted.	• OPEN	YES
1b	**Structured summary:** The original item requires a complete and structured description of the article. Consort's guidelines recommend to structure it according to a predesigned template which includes *trial design, methods, results, and conclusions*. We included a third category in the coding which allows us to relax the Cconsort standards, since structured abstracts are not required in most sociological journals.	• YES_COMP = All fields are described and delimited with headlines; • YES_INCOMP = There are no headlines, but the abstract is sufficiently detailed, i.e. it describes all aspects required; • NO = There are no headlines and it is not detailed.	NO
2b	**Objectives or hypotheses:** See Grant et al. (2018) for the item's description. The original item requires to have explicitly stated objectives and/or hypotheses. We added a third category among the possible answers to identify studies that fall between the two original categories.	• YES_COMP = RQs or hypotheses are explicitly stated; • YES_INCOMP = RQs or hypotheses are not explicitly stated but can be inferred; • NO = Neither implicitly nor explicitly.	NO
a2c1, a2c2, a2c3, a2c_other	**Type of intervention:** Typology developed by the research team.	• Incentives = All kind of incentives, e.g. economic and non; • Education = Focus on education; • Training, professional development = Focus occupation; • Mentoring = 1 to 1 training, individualisation; • Reminder; • Information - content = e.g. Interventions on information barriers; • Information - communication modes; • Other = None of the above.	YES

3a	**Trial design (allocation ratio):** See Grant et al. (2018). We allowed for a third category to identify studies non fully compliant.	• YES_COMP = Both the type of design and the allocation ratio are included in the article; • YES_INCOMP = Only the allocation ratio is presented (e.g. in tables, can be inferred) OR only the design; • NO = Authors did not include neither the trial design nor the allocation ratio.	NO
a3c	**Type of trial:** The experiment's design can significantly influence the quality and quantity of information presented by the authors.	• Explicitly pilot = Clearly described as a pilots; • Standard = Standard field experiments, developed to answer a specific RQ or policy issue; • Extension = Experiments designed by drawing on past researches/RCTs; • Other = Does not fit in the classification above.	YES
a3d	**N of experimental arms:** Each experiment may consist of two (treatment/control) or more experimental arms.	• NUMBER	YES
a3e	**Experiment replica:** Unlike a follow-up study, where researchers collect data at multiple time points, a replication typically differs in its authors and objectives from the original work. The goal of such replication is usually to validate the findings of the original experiment and assess the reliability and generalizability of its results.	• YES • NO	YES
a4c	**Population:** This item assesses whether the authors have reported whether they sampled from the population of interest.	• YES • NO	YES
a4d	**Representativeness:** Whether the experimental sample is representative of the population of interest.	• YES • NO	YES
a5a	**Fidelity:** Degree to which the intervention has been implemented as originally intended. It might happen that the intervention is not carried out by the researchers, e.g. the analysis on the impact of a policy OR the intervention does not need to evaluate the fidelity, e.g. reminder-based interventions. For	• YES • NO • NA	YES

	these cases we added a third category -*NA*-.		
5c	**How intervention providers were assigned:** See Grant et al. (2018) for the item's description. It might happen that the intervention does not require providers OR their assignment was considered not to be essential for the experimental design. For these cases we added the third category -*NA*-.	• YES • NO • NA	NO
a7a	**Ex post power calculation:** the statistical power of the study is measured retrospectively, based on the data collected.	• YES • NO	YES
a7c	**Randomised units:** All randomised units. This is usually reported in the 'design' section, but authors might also report it in a table. Not classifiable if not reported.	• NUMBER	YES
a7d	**Number of beneficiaries:** Number of individuals in the treatment(s) group.	• NUMBER	YES
8b	**Type of randomisation:** See Grant et al. (2018) for the item's description. Authors do vary a lot in the extent to which they describe the type of randomisation implemented. Grant (2018) asks for a description of the *type of randomisation* and the reasons for that choice. Since the extent to which authors report the type of randomisation varies a lot, we included a third category which relaxes the Consort's requirements.	• YES_COMP = Precise description of the *type of randomisation,* e.g. single, blocked, stratified, etc.; • YES_INCOMP = Authors report to have randomised the units but insufficiently for replication; • NO = The article does not include any description of the type of randomization process.	NO
9b	**Steps to conceal the randomisation sequence:** See Grant et al. (2018) for the item's description. While double-blind designs are often considered a gold standard in randomized controlled trials (RCTs), their necessity can depend on the nature of the intervention or treatment. It follows that in social sciences' experiments it might be not essential to conceal the randomisation sequence. For this reason we included the -*NA*- category.	• YES = It is described the process by which the allocation sequence was masked; • NO = It is not described how randomization was hidden *although necessary;* • NA = It was not necessary nor possible to conceal the randomization sequence.	NO
11a	**Awareness of intervention assignment after allocation:** See Grant	• YES • NO	NO

	et al. (2018) for the item's description. By masking the assignment of units, authors want to maximise the internal validity of the experiment. The Consort guidelines clearly state that this is an essential requirement. But some types of sociological trials cannot implement any action to mask the awareness of intervention, e.g. trials including incentives or forms of training/mentoring.	• NA	
11b	**Description of similarities of interventions:** See Grant et al. (2018) for the item's description. Third category -*NA*- to exclude those experiments which were not multi-arm trials.	• YES • NO • NA	NO
a13c	**Flow diagram:** Inclusion in the article's body, appendix, or notes of a flow diagram describing the loss of units in any experimental phase.	• YES • NO	YES
17a	**Results for each group: the estimated effect size and its precision:** We decided to code among the complier also those studies that report raw effects and not effect sizes only, because this type of reporting results is less common in Sociology. We included an additional item (a17a1) to report on this aspect.	• YES • NO	NO
a17a1	**Effect size:** Effects sizes including confidence intervals.	• YES • NO	YES
a17a2	**Availability of trial data**: Inclusion in the article's body, appendix, or notes of the possibility to download the data.	• YES • NO	YES
17b	**Binary outcomes**: See Grant et al. (2018) for the item's description. Third category -*NA*- to exclude those experiments which did not measure the outcome in binary terms.	• YES • NO • NA	NO
a17c	**N of outcomes:** The number of outcomes (dependent variables) on which effects are estimated.	• 1 • 2 • 3-5 • 6-10 • 11 or more	YES

a17d1, a17d2, a17d3	**Outcome type(s):** The type of outcomes on which effects are estimated. We included three items to allow for experiments measuring the impact on multiple outcomes of different types.	• Real world behaviour • Standardized assessment • Simulated behaviour • Reported behaviour • Simulated attitude • Reported attitude • Other = None of the above.	YES
a17e	**Effect observed:** This item assesses whether the authors have reported the presence of a statistically significant treatment effect, even if that effect has been reported only in certain subgroups.	• YES • NO	YES
a17f	**Timing between treatment delivery and outcome -measurement:** The time elapsed from experiment delivery to the first outcome measurement. In case of lengthy treatments (e.g. training), we considered the timing starting from the end of the treatment.	• Immediate; • Up to 3 weeks; • 1-3 months; • 4-11 months; • 1 year or more; • ND; Other	YES
a17g	**Follow ups:** The number of months from the first outcome measurement to the latest follow-up, if multiple follow-ups are conducted.	• NUMBER	YES
19	**Harms or unintended effects:** See Grant et al. (2018) for the item's description. A third category -*NA*- has been included to identify experiments which could not have unforeseen effects.	• YES • NO • NA	NO
20	**Trial limitations:** See Grant et al. (2018) for the item description. We decided to code a study as compliant with this even if it did not mention the problem arising with the multiplicity of outcomes.	• YES • NO	NO
23	**Registration number and name of trial registry:** See Grant et al. (2018) for the item's description. We included a third category -*YES_INCOMP*- to allow for imprecise reporting.	• YES_COMP = There is the registration number/ name of registry; • YES_INCOMP = Authors just report to have preregistered but do not include where nor the number; • NO = The article does not mention any pre-registration.	NO
24	**Where the protocol can be accessed:** See Grant et al. (2018) for the item's description.	• YES_COMP = There is a link, it is easily accessible; • YES_INCOMP = Authors just report the existence of a protocol, but this is not accessible;	NO

		• NO = The article does not include where the protocol can be accessed nor if it has been preregistered.	
26a	**Any involvement of the intervention developer:** See Grant et al. (2018) for the item's description. It is relatively common in the social sciences that a research group is involved both in the creation of a treatment and in its evaluation. This dynamic may give rise to perverse incentives to alter or selectively report results. This item assesses the transparency of the information disclosed to the reader. Given that many papers report this issue implicitly, we added a new category -*YES_INCOMP*- to identify these occurrences.	• YES_COMP = Authors clearly state whether or not researcher have been involved in the design of the treatment; • YES_INCOMP = Authors implicitly address this issue by explicitly stating they are evaluating interventions designed by third parties; • NO = No information about researchers' involvement.	NO
26b	**Other stakeholder involvement:** See Grant et al. (2018) for the item's description. As for the previous items, many researches lack in transparent reporting, but the relevant information can be inferred. We added a new category -*YES_INCOMP*- to identify these occurrences.	• YES_COMP = Authors clearly address the issue; • YES_INCOMP = Authors implicitly address this issue by stating that there has been an involvement in the experiment design or evaluation by stakeholders; • NO = No information about stakeholders' involvement.	NO
a27	**Ethical committee.** Explicit presence of an IRB number.	• YES • NO	YES

Note: for the remaining items, see the CONSORT-SPI checklist (2018)

D. Studies references

Armstrong, T. A. (2002). The effect of environment on the behavior of youthful offenders: A randomized experiment. *Journal of Criminal Justice, 30*(1), 19-28. https://doi.org/10.1016/S0047-2352(01)00119-2

Bancroft, A., Bratter, J., & Rowley, K. (2017). Affirmation effects on math scores: The importance of high school track. *Social Science Research, 64*, 319-333. https://doi.org/10.1016/j.ssresearch.2016.10.001

Barone, C., Assirelli, G., Abbiati, G., Argentin, G., & De Luca, D. (2018). Social origins, relative risk aversion and track choice: A field experiment on the role of information biases. *Acta Sociologica, 61*(4), 441-459. https://doi.org/10.1177/0001699317729872

Barone, C., Fougère, D., & Pin, C. (2021). Social Origins, Shared Book Reading, and Language Skills in Early Childhood: Evidence from an Information Experiment. *European Sociological Review, 37*(1), 18-31. https://doi.org/10.1093/esr/jcaa036

Barone, C., Schizzerotto, A., Abbiati, G., & Argentin, G. (2017). Information Barriers, Social Inequality, and Plans for Higher Education: Evidence from a Field Experiment. *European Sociological Review, 33*(1), 84-96. https://doi.org/10.1093/esr/jcw050

Barton, A. W., Beach, S. R. H., Lavner, J. A., Bryant, C. M., Kogan, S. M., & Brody, G. H. (2017). Is Communication a Mechanism of Relationship Education Effects Among Rural African Americans? *Journal of Marriage and Family, 79*(5), 1450-1461. https://doi.org/10.1111/jomf.12416

Beausoleil, V., Renner, C., Dunn, J., Hinnewaah, P., Morris, K., Hamilton, A., Braithewaite, S., Hunter, N., Browne, G., & Browne, D. T. (2016). The effect and expense of redemption reintegration services versus usual reintegration care for young African Canadians discharged from incarceration. *Health & Social Care in the Community, 25*(5), 590-601. https://doi.org/10.1111/hsc.12346

Bellé, N. (2014). Leading to make a difference: A field experiment on the performance effects of transformational leadership, perceived social impact, and public service motivation. *Journal of Public Administration Research and Theory, 24*(1), 109-136. https://doi.org/10.1093/jopart/mut033

Berlin, L. J., Dunning, R. D., & Dodge, K. A. (2011). Enhancing the transition to kindergarten: A randomized trial to test the efficacy of the "Stars" summer kindergarten orientation program. *Early Childhood Research Quarterly, 26*(2), 247-254. https://doi.org/10.1016/j.ecresq.2010.07.004

Bhatti, Y., Dahlgaard, J. O., Hansen, J. H., & Hansen, K. M. (2017). Moving the campaign from the front door to the front pocket: Field experimental evidence on the effect of phrasing and timing of text

messages on voter turnout. *Journal of Elections, Public Opinion and Parties, 27*(3), 291-310. https://doi.org/10.1080/17457289.2016.1270288

Brander, M., Bernauer, T., & Huss, M. (2021). Improved on-farm storage reduces seasonal food insecurity of smallholder farmer households – Evidence from a randomized control trial in Tanzania. *Food Policy, 98*, 101891. https://doi.org/10.1016/j.foodpol.2020.101891

Bratter, J. L., Rowley, K. J., & Chukhray, I. (2016). Does a Self-Affirmation Intervention Reduce Stereotype Threat in Black and Hispanic High Schools? *Race and Social Problems, 8*(4), 340-356. https://doi.org/10.1007/s12552-016-9187-4

Buckley, L., Reveruzzi, B., Dingli, K., Chapman, R. L., & Sheehan, M. (2020). Skills for Preventing Injury in Youth: Evaluating Change in Adolescent Alcohol Use, Violence, and Road-Related Injuries. *Youth & Society, 53*(6), 894-912. https://doi.org/10.1177/0044118X20914171

Cheavens, J. S., Feldman, D. B., Gum, A., Michael, S. T., & Snyder, C. R. (2006). Hope therapy in a community sample: A pilot investigation. *Social Indicators Research, 77*(1), 61-78. https://doi.org/10.1007/s11205-005-5553-0

Coman, D. C., Bass, M. P., Alessandri, M., Ghilain, C. S., & Llabre, M. M. (2018). Effect of Equine Assisted Activities on Social and Sensory Functioning of Children with Autism. *Society & Animals, 26*(6), 551-575. https://doi.org/10.1163/15685306-12341479

Cotterill, S., John, P., & Richardson, L. (2012). The Impact of a Pledge Request and the Promise of Publicity: A Randomized Controlled Trial of Charitable Donations. *Social Science Quarterly, 94*(1), 200-216. https://doi.org/10.1111/j.1540-6237.2012.00896.x

Cowan, P. A., Cowan, C. P., Pruett, M. K., Pruett, K., & Wong, J. J. (2009). Promoting Fathers' Engagement With Children: Preventive Interventions for Low-Income Families. *Journal of Marriage and Family, 71*(3), 663-679. https://doi.org/10.1111/j.1741-3737.2009.00625.x

Davidov, E. (2007). Explaining habits in a new context the case of travel-mode choice. *Rationality and Society, 19*(3), 315-334. https://doi.org/10.1177/1043463107077392

Desai, R. M., & Olofsgård, A. (2019). Can the poor organize? Public goods and self-help groups in rural India. *World Development, 121*, 33-52. https://doi.org/10.1016/j.worlddev.2019.04.009

Destin, M. (2016). An Open Path to the Future: Perceived Financial Resources and School Motivation. *Journal of Early Adolescence, 37*(7), 1004-1031. https://doi.org/10.1177/0272431616636480

Dinaj-Koci, V., Chen, X. G., Deveaux, L., Lunn, S., Li, X. M., Wang, B., Braithwaite, N., Marshall, S., Gomez, P., & Stanton, B. (2012). Developmental Implications of HIV Prevention During Adolescence: Examination

of the Long-Term Impact of HIV Prevention Interventions Delivered in Randomized Controlled Trials in Grade Six and in Grade 10. Youth & Society, 47(2), 151-172. https://doi.org/10.1177/0044118X12456028

Doabler, C. T., Clarke, B., Kosty, D., Smolkowski, K., Kurtz-Nelson, E., Fien, H., & Baker, S. K. (2019). Building number sense among English learners: A multisite randomized controlled trial of a Tier 2 kindergarten mathematics intervention. *Early Childhood Research Quarterly, 47*, 432-444. https://doi.org/10.1016/j.ecresq.2018.08.004

Dušek, L., Pardo, N., & Traxler, C. (2022). Salience and Timely Compliance: Evidence from Speeding Tickets. *Journal of Policy Analysis and Management, 41*(2), 426-449. https://doi.org/10.1002/pam.22387

Edwards, R. C., Vieyra, Y., & Hans, S. L. (2020). Maternal support for infant learning: Findings from a randomized controlled trial of doula home visiting services for young mothers. *Early Childhood Research Quarterly, 51*, 26-38. https://doi.org/10.1016/j.ecresq.2019.08.001

Ehlert, M., Finger, C., Rusconi, A., & Solga, H. (2017). Applying to college: Do information deficits lower the likelihood of college-eligible students from less-privileged families to pursue their college intentions? Evidence from a field experiment. *Social Science Research, 67*, 193-212. https://doi.org/10.1016/j.ssresearch.2017.04.005

Gamoran, A., Miller, H. K., Fiel, J. E., & Valentine, J. L. (2021). Social Capital and Student Achievement: An Intervention-Based Test of Theory. *Sociology of Education, 94*(4), 356-376. https://doi.org/10.1177/00380407211040261

Gamoran, A., Turley, R. N. L., Turner, A., & Fish, R. (2012). Differences between Hispanic and non-Hispanic families in social capital and child development: First-year findings from an experimental study. *Research in Social Stratification and Mobility, 30*(1), 97-112. https://doi.org/10.1016/j.rssm.2011.08.001

Gassman-Pines, A., & Yoshikawa, H. (2005). Five-year effects of an anti-poverty program on marriage among never-married mothers. *Journal of Policy Analysis and Management, 25*(1), 11-30. https://doi.org/10.1002/pam.20154

Goldrick-Rab, S., Kelchen, R., Harris, D. N., & Benson, J. (2016). Reducing Income Inequality in Educational Attainment: Experimental Evidence on the Impact of Financial Aid on College Completion. *American Journal of Sociology, 121*(6), 1762-1817. https://doi.org/10.1086/685442

Gurantz, O., Howell, J., Hurwitz, M., Larson, C., Pender, M., & White, B. (2020). A National-Level Informational Experiment to Promote Enrollment in Selective Colleges. *Journal of Policy Analysis and Management, 40*(2), 453-479. https://doi.org/10.1002/pam.22262

Guryan, J., Christenson, S., Cureton, A., Lai, I., Ludwig, J., Schwarz, C., Shirey, E., & Turner, M. C. (2021). The Effect of Mentoring on School Attendance and Academic Outcomes: A Randomized Evaluation of the Check & Connect Program. *Journal of Policy Analysis and Management, 40*(3), 527-551. https://doi.org/10.1002/pam.22264

Hanselman, P., Bruch, S. K., Gamoran, A., & Borman, G. D. (2014). Threat in Context: School Moderation of the Impact of Social Identity Threat on Racial/Ethnic Achievement Gaps. *Sociology of Education, 87*(2), 106-124. https://doi.org/10.1177/0038040714525970

Hertel-Fernandez, A., & Porter, E. (2021). Why Public Sector Union Members Support Their Unions: Survey and Experimental Evidence. *Social Forces,* 100(1), 375–399. https://doi.org/10.1093/sf/soaa087

Hess, D. R., Hanmer, M. J., & Nickerson, D. W. (2016). Encouraging Local Compliance with Federal Civil Rights Laws: Field Experiments with the National Voter Registration Act. *Public Administration Review, 76*(1), 165-174. https://doi.org/10.1111/puar.12473

Heyman, B., Harrington, B., & Heyman, A. (2010). A randomised controlled trial of an energy efficiency intervention for families living in fuel poverty. *Housing Studies, 26*(1), 117-132. https://doi.org/10.1080/02673037.2010.512787

Howell, W. G., Wolf, P. J., Campbell, D. E., & Peterson, P. E. (2002). School Vouchers and Academic Performance: Results from Three Randomized Field Trials. *Journal of Policy Analysis and Management, 21*(2), 191-217. https://doi.org/10.1002/pam.10023

Huis, M., Lensink, R., Vu, N., & Hansen, N. (2019). Impacts of the Gender and Entrepreneurship Together Ahead (GET Ahead) training on empowerment of female microfinance borrowers in Northern Vietnam. *World Development, 120*, 46-61. https://doi.org/10.1016/j.worlddev.2019.04.001

Jackson, A. P., & Preston, K. S. J. (2019). Self-Efficacy Among Unmarried Black Mothers and Fathers of Young Children: A Pilot Study. *Race and Social Problems, 11*, 112-118. https://doi.org/10.1007/s12552-018-9232-6

Jagannathan, R., & Camasso, M. J. (2003). Family cap and nonmarital fertility: The racial conditioning of policy effects. *Journal of Marriage and Family, 65*(1), 52-71. https://doi.org/10.1111/j.1741-3737.2003.00052.x

Jakobsen, M., & Andersen, S. C. (2013). Coproduction and equity in public service delivery. *Public Administration Review, 73*(5), 704-713. https://doi.org/10.1111/puar.12094

Jakobsen, M., & Andersen, S. C. (2013). Intensifying Social Exchange Relationships in Public Organizations: Evidence from a Randomized Field Experiment. *Journal of Policy Analysis and Management, 32*(1), 60-82. https://doi.org/10.1002/pam.21662

Jilke, S., Lu, J., Xu, C., & Shinohara, S. (2019). Using Large-Scale Social Media Experiments in Public Administration: Assessing Charitable Consequences of Government Funding of Nonprofits. *Journal of Public Administration Research and Theory, 29*(4), 627-639. https://doi.org/10.1093/jopart/muy021

Karimli, L., Lecoutere, E., Wells, C. R., & Ismayilova, L. (2021). More assets, more decision-making power? Mediation model in a cluster-randomized controlled trial evaluating the effect of the graduation program on women's empowerment in Burkina Faso. *World Development, 137*, 105159. https://doi.org/10.1016/j.worlddev.2020.105159

Keller, T., Takacs, K., & Elwert, F. (2022). Yes, You Can! Effects of Transparent Admission Standards on High School Track Choice: A Randomized Field Experiment. *Social Forces, 101*(1), 341-368. https://doi.org/10.1093/sf/soab111

Kelly, E. L., Moen, P., Oakes, J. M., Fan, W., Okechukwu, C., Davis, K. D., Hammer, L. B., Kossek, E. E., King, R. B., Hanson, G. C., Mierzwa, F., & Casper, L. M. (2014). Changing Work and Work-Family Conflict: Evidence from the Work, Family, and Health Network. *American Sociological Review, 79*(3), 485-516. https://doi.org/10.1177/0003122414531435

Kinzie, M. B., Whittaker, J. V., Williford, A. P., DeCoster, J., McGuire, P., Lee, Y., & Kilday, C. R. (2014). MyTeachingPartner-Math/Science pre-kindergarten curricula and teacher supports: Associations with children's mathematics and science learning. *Early Childhood Research Quarterly, 29*(4), 586-599. https://doi.org/10.1016/j.ecresq.2014.06.007

Kisida, B., Bowen, D. H., & Greene, J. P. (2018). Cultivating interest in art: Causal effects of arts exposure during early childhood. *Early Childhood Research Quarterly, 45*, 197-203. https://doi.org/10.1016/j.ecresq.2017.12.003

Kraft, M. A., & Monti-Nussbaum, M. (2017). Can Schools Enable Parents to Prevent Summer Learning Loss? A Text-Messaging Field Experiment to Promote Literacy Skills. *The ANNALS of the American Academy of Political and Social Science, 674*(1), 85-112. https://doi.org/10.1177/0002716217732009

Lachaud, J., Mejia-Lancheros, C., Nisenbaum, R., Stergiopoulos, V., O'Campo, P., & Hwang, S. W. (2021). Housing First and Severe Mental Disorders: The Challenge of Exiting Homelessness. *The ANNALS of the*

American Academy of Political and Social Science, 693(1), 209-231. https://doi.org/10.1177/0002716220987220

Lambert, S. J., Henly, J. R., Schoeny, M., & Jarpe, M. (2019). Increasing Schedule Predictability in Hourly Jobs: Results From a Randomized Experiment in a US Retail Firm. *Work and Occupations, 46*(2), 176-226. https://doi.org/10.1177/0730888418823241

Lasley, J. (2003). The effect of intensive bail supervision on repeat domestic violence offenders. *Policy Studies Journal, 31*(2), 187-207. https://doi.org/10.1111/1541-0072.00011

Levinson, E. M., Vogt, M., Barker, W. F., Jalongo, M. R., & Van Zandt, P. (2017). Effects of Reading with Adult Tutor/Therapy Dog Teams on Elementary Students' Reading Achievement and Attitudes. *Society & Animals, 25*(1), 38-56. https://doi.org/10.1163/15685306-12341427

Linos, K., Carlson, M., Jakli, L., Dalma, N., Cohen, I., Veloudaki, A., & Spyrellis, S. N. (2022). How Do Disadvantaged Groups Seek Information About Public Services? A Randomized Controlled Trial of Communication Technologies. *Public Administration Review, 82*(4), 708-720. https://doi.org/10.1111/puar.13437

Malmberg-Heimonen, I., & Vuori, J. (2005). Financial incentives and job-search training: Methods to increase labour market integration in contemporary welfare states? *Social Policy & Administration, 39*(3), 247-259. https://doi.org/10.1111/j.1467-9515.2005.00438.x

Martini, A., Azzolini, D., Romano, B., & Vergolini, L. (2020). Increasing College Going by Incentivizing Savings: Evidence from a Randomized Controlled Trial in Italy. *Journal of Policy Analysis and Management, 40*(3), 814-840. https://doi.org/10.1002/pam.22260

Matous, P., & Wang, P. (2019). External exposure, boundary-spanning, and opinion leadership in remote communities: A network experiment. *Social Networks, 56*, 10-22. https://doi.org/10.1016/j.socnet.2018.08.002

McGuire, W., O'Brien, B. G., Baird, K., Corbett, B., & Collingwood, L. (2020). Does Distance Matter? Evaluating the Impact of Drop Boxes on Voter Turnout. *Social Science Quarterly, 101*(5), 1789-1809. https://doi.org/10.1111/ssqu.12853

Merz, E. C., Landry, S. H., Johnson, U. Y., Williams, J. M., & Jung, K. (2016). Effects of a responsiveness-focused intervention in family child care homes on children's executive function. *Early Childhood Research Quarterly, 34*, 128-139. https://doi.org/10.1016/j.ecresq.2015.10.003

Michelson, M. R. (2006). Mobilizing the Latino youth vote: Some experimental results. *Social Science Quarterly, 87*(5), 1188-1206. https://doi.org/10.1111/j.1540-6237.2006.00423.x

Modestino, A. S. (2019). How Do Summer Youth Employment Programs Improve Criminal Justice Outcomes, and for Whom? *Journal of Policy Analysis and Management, 38*(3), 600-628. https://doi.org/10.1002/pam.22138

Montgomery, R. J. V., Kwak, J., Kosloski, K., & Valuch, K. O. (2011). Effects of the TCARE® intervention on caregiver burden and depressive symptoms: Preliminary findings from a randomized controlled study. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 66B*(5), 640-647. https://doi.org/10.1093/geronb/gbr088

Murry, V. M., Brody, G. H., McNair, L. D., Luo, Z., Gibbons, F. X., Gerrard, M., & Wills, T. A. (2005). Parental involvement promotes rural African American youths' self-pride and sexual self-concepts. *Journal of Marriage and Family, 67*(3), 627-642. https://doi.org/10.1111/j.1741-3737.2005.00158.x

Nam, Y., Kim, Y., Clancy, M., Zager, R., & Sherraden, M. (2012). Do Child Development Accounts Promote Account Holding, Saving, And Asset Accumulation for Children's Future? Evidence from a Statewide Randomized Experiment. *Journal of Policy Analysis and Management, 32*(1), 6-33. https://doi.org/10.1002/pam.21652

Neumann, R. (2018). The framing of charitable giving: A field experiment at bottle refund machines in Germany. *Rationality and Society, 31*(1), 98-131. https://doi.org/10.1177/1043463118820894

Nguyen, C. V., & Tarp, F. (2022). Changing male perceptions of gender equality: Evidence from a randomised controlled trial study. *World Development, 158*, 106019. https://doi.org/10.1016/j.worlddev.2022.106019

Ollonqvist, K., Aaltonen, T., Karppi, S.-L., Hinkka, K., & Pöntinen, S. (2007). Network-based rehabilitation increases formal support of frail elderly home-dwelling persons in Finland: Randomised controlled trial. *Health & Social Care in the Community, 16*(2), 115-125. https://doi.org/10.1111/j.1365-2524.2007.00733.x

Pager, D., Goldstein, R., Ho, H., & Western, B. (2022). Criminalizing Poverty: The Consequences of Court Fees in a Randomized Experiment. *American Sociological Review, 87*(3), 529-553. https://doi.org/10.1177/00031224221075783

Pahlke, E., Bigler, R. S., & Green, V. A. (2010). Effects of learning about historical gender discrimination on early adolescents' occupational judgments and aspirations. *Journal of Early Adolescence, 30*(6), 854-894. https://doi.org/10.1177/0272431609361200

Penuel, W. R., Bates, L., Gallagher, L. P., Pasnik, S., Llorente, C., Townsend, E., Hupert, N., Domínguez, X., & VanderBorght, M. (2012). Supplementing literacy instruction with a media-rich intervention: Results

of a randomized controlled trial. *Early Childhood Research Quarterly, 27*(1), 115-127. https://doi.org/10.1016/j.ecresq.2011.07.002

Rusby, J. C., Smolkowski, K., Marquez, B., & Taylor, T. K. (2008). A small-scale randomized efficacy trial of Carescapes: Enhancing children's social development in child care homes. *Early Childhood Research Quarterly, 23*(4), 527-546. https://doi.org/10.1016/j.ecresq.2008.08.003

Seear, K. H., & Vella-Brodrick, D. A. (2013). Efficacy of Positive Psychology Interventions to Increase Well-Being: Examining the Role of Dispositional Mindfulness. *Social Indicators Research, 114*(3), 1125-1141. https://doi.org/10.1007/s11205-012-0193-7

Seroczynski, A. D., Evans, W. N., Jobst, A. D., Horvath, L., & Carozza, G. (2016). Reading for Life and Adolescent Re-Arrest: Evaluating a Unique Juvenile Diversion Program. *Journal of Policy Analysis and Management, 35*(3), 662-682. https://doi.org/10.1002/pam.21916

Seron, C., Van Ryzin, G., Frankel, M., & Kovath, J. (2001). The impact of legal counsel on outcomes for poor tenants in New York City's Housing Court: Results of a randomized experiment. *Law and Society Review, 35*(2), 419-433. https://doi.org/10.2307/3185408

Serrano Corkin, D. M., Ekmekci, A., & Fisher, A. (2020). Integrating Culture, Art, Geometry, and Coding to Enhance Computer Science Motivation Among Underrepresented Minoritized High School Students. *The Urban Review, 52*, 950–969. https://doi.org/10.1007/s11256-020-00586-8

Sheely, R. (2015). Mobilization, Participatory Planning Institutions, and Elite Capture: Evidence from a Field Experiment in Rural Kenya. *World Development, 67*, 251-266. https://doi.org/10.1016/j.worlddev.2014.10.024

Tanyel, S. S., & Kiralp, F. S. S. (2021). Tolerance for Sustainable Peace Culture in a Divided Society: The effect of Peace Education on Tolerance Tendency and Human Values. *Social Indicators Research, 156*, 223-246. https://doi.org/10.1007/s11205-021-02630-w

Tjaden, J., & Dunsch, F. A. (2021). The effect of peer-to-peer risk information on potential migrants – Evidence from a randomized controlled trial in Senegal. *World Development, 145*, 105488. https://doi.org/10.1016/j.worlddev.2021.105488

Trukeschitz, B., Eisenberg, S., Schneider, C., & Schneider, U. (2022). Exploring the effectiveness of a fitness-app prototype for home care service users in Austria and Italy. *Health & Social Care in the Community, 30*(5), e2884-e2895. https://doi.org/10.1111/hsc.13733

Turley, R. N. L., Gamoran, A., McCarty, A. T., & Fish, R. (2017). Reducing children's behavior problems through social capital: A causal assessment. *Social Science Research, 61*, 206-217. https://doi.org/10.1016/j.ssresearch.2016.06.015

Wang, H., Chu, J., Loyalka, P., Xin, T., Shi, Y., Qu, Q., & Yang, C. (2016). Can Social-Emotional Learning Reduce School Dropout in Developing Countries? *Journal of Policy Analysis and Management, 35*(4), 818-847. https://doi.org/10.1002/pam.21915

Wang, J. S.-H., Ssewamala, F. M., Neilands, T. B., Bermudez, L. G., Garfinkel, I., Waldfogel, J., Brooks-Gunn, J., & You, J. (2018). Effects of Financial Incentives on Saving Outcomes and Material Well-Being: Evidence From a Randomized Controlled Trial in Uganda. *Journal of Policy Analysis and Management, 37*(3), 602-629. https://doi.org/10.1002/pam.22065

Wood, R. G., Moore, Q., Clarkwest, A., & Killewald, A. (2014). The Long-Term Effects of Building Strong Families: A Program for Unmarried Parents. *Journal of Marriage and Family, 76*(2), 446-463. https://doi.org/10.1111/jomf.12094

Zhu, A. Y. F., Yu, C. W. M., & Chou, K. L. (2019). Improving Financial Literacy in Secondary School Students: An Randomized Experiment. *Youth & Society, 53*(4), 539-562. https://doi.org/10.1177/0044118X19851311

Zvoch, K., & Stevens, J. J. (2013). Summer school effects in a randomized field trial. *Early Childhood Research Quarterly, 28*(1), 24-32. https://doi.org/10.1016/j.ecresq.2012.05.002