The Impact of Causal Evidence on Innovation: Lessons from Randomized Controlled Trials in Surgery

Wesley H. Greenblatt
Harvard Medical School and
Boston Children's Hospital
300 Longwood Avenue
Boston, MA 02115

Soomi Kim
Columbia Business School
665 W 130th St, 974 Kravis
New York, NY 10027

August 31, 2025

Abstract

Randomized controlled trials (RCTs) are often considered the gold standard for providing causal evidence, but they are not without limitations. In addition to their high costs, RCTs may produce results that are not useful or generalizable for follow-on research. Employing a difference-in-differences approach, we study how evidence generated by RCTs shapes research activities in surgery. We find that follow-on research appears to decline after the publication of an RCT, with the remaining research being of lower impact and clinical relevance. This dampening effect of RCTs is especially pronounced for high-quality trials and those that evaluate interventions with lower barriers to adoption. We conclude with a discussion of whether these effects have positive or negative implications for innovation.

Keywords: Randomized controlled trials, Evidence, Innovation

JEL Classification: O33, O31, I19

Acknowledgements: Address correspondence to sk5261@columbia.edu. We thank Daniel Creech and Tiantian Zhao for their helpful research assistantship. We also thank seminar participants at the NYU Economics of Strategy Workshop, Temple University, and Columbia University for their feedback.

1. Introduction

Randomized controlled trials (RCTs) are widely regarded as the gold standard for establishing causal evidence. From the first modern RCT of streptomycin for tuberculosis in 1948 (Medical Research Council 1948), RCTs have come to play an increasingly important role not only in medicine, but also in economics (Banerjee et al. 2020; Heckman 2020), strategy (Chatterji et al. 2016), entrepreneurship (Camuffo et al. 2019; Koning et al. 2022), and other social science disciplines. Few policymakers, business leaders, or physicians would argue against the benefits of having the high-quality causal evidence provided by RCTs to inform their decision making, and articles in popular press continue to advocate for additional RCTs (Adgate 2020; Finkelstein 2021; Gallo 2016). This has led some to even worry that the prevalence and emphasis on RCTs might be crowding out other valuable research efforts (Aston 2023; McKenzie 2016).

But beyond just providing causal evidence to inform a particular policy or healthcare decision, the evidence provided by RCTs may also impact downstream innovation. Knowledge production is cumulative, and the evidence produced by RCTs is an input into the production of new knowledge (Romer 1990; Scotchmer 1991). Yet for all of the ubiquity of RCTs in generating new knowledge, little is known about how the arrival of causal evidence shapes the direction and dynamism of follow-on innovation.

While it might seem natural to think that producing high-quality causal evidence will lead to better downstream research, there are reasons to believe the relationship may be more complex. On one hand, RCTs can spur innovation by establishing a reliable foundation to build upon, encouraging future research. They can legitimize new lines of inquiry and settle debates in the field with conclusive findings or signal follow-on research as being important with high economic, social, or reputational returns.

On the other hand, RCTs may have a limited or even negative impact on innovation. Although the strength of RCTs stems from their high internal validity, a key disadvantage is that RCT findings may not always generalize outside of their experimental conditions, such that follow-on researchers may dismiss the findings as having limited applicability. Researchers may also be reluctant to build on an RCT in cases where the adoption cost of research or implementation is high. Perhaps worst of all, an RCT might prematurely shut down a promising line of inquiry if its findings are perceived to be definitive, even when there is scope for further exploration.

On balance, it is not clear which of these two countervailing dynamics will dominate, and the benefits of RCTs must be balanced against their costs. This includes both the often very

substantial direct costs of running RCTs (Aston 2023; Mullard 2018). It also includes the opportunity cost of producing other types of evidence, including high-quality observational studies. Understanding the dynamics of how knowledge workers respond to the arrival of causal evidence—and when and why they do not respond—can help improve the design of RCTs and allocation of resources across different study types.

Empirically, estimating the impact of RCTs on both the amount and type of follow-on innovation is difficult. First, we need to both comprehensively observe the arrival of RCTs to subfields of innovation as well as characterize follow-on research. Second, which subfields receive RCTs is often endogenous, limiting the ability to isolate the causal effect of RCTs on subsequent research. Finally, to explore not only the strengths but also the potential limitations of RCTs, we require a setting in which those limitations are likely detectable.

The field of surgery addresses many of these empirical challenges. First, defining the set of problems eligible for RCTs and tracking the universe of RCTs is difficult in most fields. In biomedicine, however, diseases delineate a clear problem space, with ClinicalTrials.gov providing a registry of clinical trials, including those in surgery. Moreover, most innovations in biomedicine are published. Together, this provides a robust paper trail we can follow and characterize.

Second, the key empirical concern is that subfields that receive RCTs may have different underlying research trajectories than those without RCTs. To mitigate this concern, we define treated subfields as those that received an RCT with published results, while control subfields are those that received an RCT that did not generate any publications. All subfields in our analysis are therefore "RCT-feasible," and this enables us to compare how follow-on innovation in treated subfields builds on RCT-generated evidence, relative to control subfields that could have received such evidence but did not. In addition, published and unpublished RCTs are unlikely to be systematically different in terms of their ex-ante risk or scientific promise. A distinctive feature of many clinical trials is equipoise: a trial is considered ethical only if the investigator is genuinely uncertain about which intervention is better. Equipoise reduces the concern that unpublished RCTs are riskier or less promising since both published and unpublished RCTs must pass similar standards set by institutional review boards.

Finally, surgery is a field where the limits of RCTs may be particularly salient. Due to the difficulties with standardizing experimental conditions, generalizability is a first-order concern in surgery. Furthermore, surgeons face high switching costs especially for tacit surgical techniques, which creates barriers to adopting and building on RCT evidence. Because there is wide variation

in the quality and type of RCTs, we empirically leverage the heterogeneity to understand when RCTs shape follow-on innovation and when their impact is limited.

Using data from ClinicalTrials.gov from 1991 to 2021, we identify 907 completed surgical trials with published results and 3,701 control trials without such publications. To construct the surgical subfields affected by these RCTs, we capitalize on a text similarity algorithm to delineate subfields of research articles that share intellectual content (Azoulay et al. 2019; Greenblatt 2022). After matching treated and control subfields on observable characteristics, we employ a difference-in-differences estimator to study the within-subfield change in follow-on research after the arrival of causal evidence embodied in the publication of an RCT's results.

We find subfields which receive causal evidence from an RCT have a sustained decrease in follow-on innovation. The subfields experience an 18% decline in the flow of articles after an RCT, but this dampening effect of RCT is concentrated among only high-quality trials, while low-quality trials do not appear to have an impact. This pattern suggests that RCTs—when they are high quality and provide rigorous results—may be able to facilitate innovation by resolving key questions in the field and enable researchers to move onto other areas. Indeed, we find that the research that continues in the subfield after an RCT has lower citation impact and clinical relevance, potentially indicating that research priorities may have shifted elsewhere.

However, the fact that most RCTs leave no noticeable impact on follow-on innovation leads to questions around quality control of trial designs and whether investing in RCTs yields valuable returns, given the high costs of conducting them. We also find that barriers to adopting new evidence can mute the effect of RCTs. When the RCT evaluates interventions that directly alter surgical techniques, surgeons appear to not respond, compared to RCTs that evaluate other aspects of surgical care (e.g., preoperative management) that involve lower switching costs.

The remainder of the paper proceeds as follows. Section 2 examines surgical innovation as a setting to study RCTs. Section 3 presents the data, empirical strategy, and descriptive statistics. Section 4 reports our results, while Section 5 discusses the implications of our study.

2. Surgical Innovation and the Study of RCTs

In this section, we first provide a brief overview on innovations in surgery and describe the empirical features of surgery that make it an attractive setting for this paper.

2.1 Overview of Innovations in Surgery

Surgical innovations consist of a broad mix of breakthroughs in three dimensions: basic science, the craft of operative techniques, and technological tools. Major milestones in the history of surgery reflect this heterogeneity (Gawande 2012; Schneider 2020; Rutkow 2022). For example, the birth of modern surgery was propelled by basic science discoveries, with the advent of anesthesia (1846) and antisepsis (1867). Before these advances, surgery was limited to crude, rapid interventions on a narrow set of organs, constrained by the need to minimize patient pain and the high rates of postoperative infections. Surgery has also been shaped by refinements in operative techniques, such as the introduction of minimally invasive surgeries in the 1970s, which enabled operations through smaller incisions and reduced complications compared to traditional open procedures. Finally, the development of new tools has frequently supported surgical innovations, with robotic surgery a notable recent advancement.

Furthermore, surgery faces a dramatically different innovation regime than pharmaceuticals. First, surgery has weak intellectual property rights (Dranove et al. 2022). In drug development, patents play critical roles in incentivizing innovation in exchange for monopoly rights. However, with the exception of medical devices, surgical procedures are not protected by patents.[1] As a result, surgeons are driven primarily by the motivation to improve patient outcomes and by reputational rewards, rather than by direct economic returns from their inventions (Strandburg 2014).

Second, while drug development is subject to well-established legal rules, surgery faces far less regulatory scrutiny. Much of surgical innovation occurs incrementally, as surgeons modify procedures to address a range of (often unexpected) anatomical variations (Morreim et al. 2006). Surgical innovations which are more significant departures from current practice typically require approval by a local Institutional Review Board (IRB), although this falls short of the rigorous oversight supported by central institutions such as the Food and Drug Administration (FDA). Following experimentation with animal models, surgeons implement their innovation on human patients and disseminate the findings through conferences and publications (Badinski 2020).

Third, surgery is a setting with limited availability of high-quality evidence. Unlike in pharmaceuticals where RCTs serve as the primary method for establishing efficacy and safety, evidence in surgery is dominated by observational studies. Estimates suggest that only 5-20% of

[1] In the US, while it is technically possible to patent surgical procedures, these patents cannot be enforced against medical practitioners. In the EU, procedures are excluded from patentability altogether.

surgical procedures have been evaluated by RCTs (Maier 2006; Dranove et al. 2022). Lamenting the quality of surgical research, the editor of *The Lancet* in 1996 once likened it to "comedic opera performances," a comment that remains infamous even today (Horton 1996; Bagenal et al. 2023). Despite the rise of evidence-based medicine, however, surgical RCTs continue to be scarce.

2.2 Surgery as an Empirical Setting

In order to examine the impact of an RCT on subsequent research trajectory, an ideal empirical setting needs several ingredients: (i) a comprehensive, well-defined set of RCTs, (ii) exogenous variations in which subfields receive an RCT, (ii) clear measures of follow-on research to track how a subfield evolves after the introduction of an RCT, and (iv) institutional features that make both the benefits and drawbacks of RCTs likely observable.

2.2.1 Comprehensive Trial Coverage

Although prior research has extensively discussed the strengths and limitations of RCTs, most of these studies outside of medicine are qualitative reviews, methodological papers, or analyses of select RCTs (Banerjee and Duflo 2009; Pritchett and Sandefur 2015; Chatterji et al. 2016; Frieden 2017; Cartwright and Deaton 2018; Feit and Berman 2019; Heckman 2020).[2] A central challenge is that in many fields—such as management or economics—it is difficult to define the set of problems at risk of being tested by an RCT. An ideal setting requires both a well-defined problem space and a sufficiently large sample of trials.

Medicine provides such a setting: diseases define a clear problem space, and ClinicalTrials.gov provides a comprehensive registry of US clinical trials, including those in surgery.[3] ClinicalTrials.gov data has been used by physicians for meta-analyses of treatment efficacy (e.g., Cochrane) and by social scientists to track developments in pharmaceuticals and medical devices (Krieger 2021; Agha, Kim, and Li 2022; Chandra et al. 2024; Masanneck et al 2024) as well as to study the incentives shaping trial design and reporting (Balasubramanian, Cummings, and Pierce 2022; Guzman, Jelveh, and Kogut 2023; Kao 2025). In this paper, surgery

[2] Some exceptions include DellaVigna, Kim, and Linos (2024), who analyze a comprehensive dataset of RCTs on government communications across 30 US cities and find that RCT evidence adoption is driven by organizational inertia rather than effect size.

[3] ClinicalTrials.gov is maintained by the National Institutes of Health. RCTs testing FDA-regulated interventions are legally required to be registered. While most surgical innovations fall outside the purview of the FDA, institutional review boards and funding agencies typically require registration of surgical trials, such that ClinicalTrial.gov in practice offers a broad coverage of them as well.

offers the empirical features we seek: a well-defined problem space and extensive trial coverage through ClinicalTrials.gov, which allow us to systematically examine RCTs as inputs into subsequent scientific research.

2.2.2. Exogenous Variation in Which Subfields Receive RCTs

The key empirical challenge when studying the impact of RCTs is that subfields that receive RCTs might differ systematically from fields that do not, as surgical trials are especially difficult to conduct. The invasive nature of procedures raises ethical concerns: surgeons may be reluctant to randomize patients to different treatments and patients themselves may be unwilling to consent. It is also logistically challenging to create controlled conditions in surgery, as procedures are performed in teams, individual surgeon skills and hospital facilities vary, and double blinding is often infeasible. Some surgical settings (e.g., emergency trauma surgeries) may have unique ethical and logistical barriers that make running an RCT especially difficult.

If the ability to conduct RCTs is correlated with underlying research activities—such as subfields that are gaining clinical attention or have stronger institutional infrastructure—then subsequent changes in research activity in subfields that receive an RCT may reflect these trends rather than the causal impact of RCTs.

We address this concern in several ways. We first restrict our sample to subfields that are "RCT-feasible." Specifically, we define treated subfields as those with RCTs that generated publications. In contrast, control subfields are those in which RCTs were conducted, but did not result in any publication—that is, no evidence was generated by these trials. This sample selection ensures that all subfields in our sample could have plausibly yielded RCT-based evidence, avoiding comparisons to areas where RCTs may be infeasible. This allows us to examine how research trajectories differ between subfields that received causal evidence and those that could have received such evidence but did not.

Our approach of comparing published (treated) and unpublished (control) trials may raise the concern that the treated trials are systematically different than the control trials if the former yielded publications while the latter did not. Some control trials did not yield publications because they were terminated early. While some of these terminations were plausibly exogenous (e.g., the principal investigator relocated, COVID-19 interruptions), others were due to difficulties with recruiting patients. Some control trials are those that fully completed but were never published, likely due to unfavorable results, insufficient statistical power, or lack of journal interest. These

factors may prompt the concern that unpublished trials—whether terminated early or completed without publications—may have tested interventions that are riskier or of lower scientific promise.

This concern, however, is mitigated by a unique feature of clinical research called equipoise: trials are supposed to only be approved by IRBs if the investigator is genuinely uncertain about which intervention is superior. If the investigator has a prior that one treatment is more effective, the trial cannot proceed as physicians are ethically bound to provide the option they believe to be the most beneficial. Clinical equipoise therefore imposes higher ethical standards than in most other fields. For example, in field experiments run by social scientists, withholding a potentially promising program from a control group would not typically raise objections. Because of equipoise, both published and unpublished RCTs pass similar ethical and institutional review standards. While equipoise does not rule out all differences, it reduces the likelihood that unpublished trials were systematically riskier or less promising ex-ante.

Moreover, the exact timing of RCT completion and publication depends on idiosyncratic factors (e.g., patient enrollment duration and peer review lags) and unlikely to be related to the research trajectory of the subfield. If publication timing is exogenous, and there is a clear break in research activity after an RCT publication, this provides additional credibility to a causal interpretation of our results.

2.2.3 Clear Measures of Follow-on Innovation

The third ingredient of an ideal empirical setting is the ability to track how research evolves after an RCT. Using detailed bibliometric records, we construct several measures of subsequent innovation. First, our baseline outcomes include both the quantity of articles flowing into a subfield after an RCT and their quality, measured by publication impact (e.g., citations).

Second, we measure whether follow-on research after an RCT is incorporated into clinical practice guidelines. These guidelines are typically written by professional medical societies, where a group of expert physicians summarize the state of the evidence in the subfield and provide recommendations for other practitioners (Greenblatt 2022). While we do not have direct measures of clinical adoption, inclusion into clinical practice guidelines serve as a useful proxy for clinical relevance of research produced after an RCT.

2.2.4 A Setting Well-Suited to Detect the Limits of RCTs

To assess not only the strengths but also the drawbacks of RCTs, we require a setting with features that make RCTs' limitations especially visible: contexts where (i) the adoption costs

of new evidence are nontrivial and (ii) concerns about generalizability are salient. Surgery provides such a setting, serving as a "stress test" for understanding the scope and limits of RCTs.

RCTs may have a limited impact on future research if the cost of adopting or building on new evidence is high. Although this issue is not unique to RCTs (all forms of evidence face adoption barriers), the opportunity cost of producing evidence that goes unused is especially high for RCTs given the significant investments they require.

Unlike medical doctors who can more easily adjust drug prescriptions, surgeons face significant barriers when switching surgical techniques (Beane 2018; Badinski 2021; Anthony and Beane 2023). Surgery is often described as a craft—a highly skilled physical activity that requires extensive tacit knowledge developed only through years of training. This creates substantial frictions to incorporating new evidence, even when the evidence is high-quality.

Generalizability concerns of RCTs are also pronounced in surgery. Creating standardized experimental conditions in surgery is difficult, since procedures are conducted as teams, surgeon expertise and institutional capacity differ, and blinding patients and surgeons is often impractical. This stands in contrast to RCTs that test drugs where experimental conditions can be more easily standardized: "Unlike 20 milligram tablets, no two surgical procedures are the same" (Ramchurn and Nundy 2023).

Because of these challenges, a well-designed RCT with high internal validity may be the least generalizable, leading to the risk that the RCT may be dismissed as irrelevant by some surgeons. A surgeon at the Mayo Clinic remarked, "When you read a study, you reflect on whether it is representative of your patient population . . . It is easy to conclude that the answer is 'no.' The mean age in the study is different or 'I do it differently'" (Kolata 2016).

The high cost of adopting new evidence and the generalizability challenges in surgery makes it a natural "stress test" setting to study the limits of RCTs. Although there is a relative paucity of RCTs testing surgeries relative to drug interventions, they have been increasingly prevalent with the rise of evidence-based medicine. This context provides an opportunity to evaluate whether RCTs can meaningfully shape the trajectory of follow-on research in domains like surgery or whether scarce resources would be better allocated elsewhere.

3. Data, Empirical Strategy, and Descriptive Statistics

3.1. Data

3.1.1. Identifying Surgical Trials

To construct our sample, we begin by retrieving all U.S.-based surgical trials in ClinicalTrials.gov from 1991 to 2021. (Appendix B.1 details the search terms we used.) We define surgical procedures as those performed in the operating room under at least regional anesthesia that involves an incision to access the patient's body.[4] We exclude trials related to dentistry, ophthalmology, obstetrics, and cancer as these may have distinct dynamics around the impact of evidence on innovation.[5] Our definition covers trials related to every stage of surgical care, including preoperative preparation, intraoperative practices, and postoperative management. As discussed in Section 2.1, surgical innovations encompass a wide range—from techniques (e.g., incision methods, suturing) to new devices—all of which are captured in our sample.

We consider the full spectrum of surgical RCTs for two reasons. First, surgeons are key decision makers in all aspects of surgical care, and any changes in the surgical process ultimately affect the surgeons. Second, and importantly, this breadth of RCTs generates useful variation in the degree to which the evidence from the RCTs impacts surgeons since some innovations directly reshape the surgeons' practice, while others do so to a lesser extent.

This sample construction yields 907 treated trials (those with published results) and 3,701 control trials (those without published results). While ClinicalTrials.gov provides rich details about these trials, much of the information is non-standardized. We therefore manually reviewed the description of every trial to code a set of key characteristics, such as surgical specialty, intervention being tested, trial endpoints, and the results of the trial.

3.1.2. Identifying Subfields

After compiling our treated and control trials, the next step is to map the research subfields they impact. The goal is to construct fine-grained subfields that reflect the specific clinical question

[4] We exclude procedures that are typically performed by non-surgeons, such as catheter studies.
[5] For reasons of history and tradition, dentistry is considered to be outside of medicine. While ophthalmology is within medicine, ophthalmologic surgery has become highly specialized and is much less tightly connected to the broader discipline of medicine than other surgical fields. Obstetrics, due to the prominence of malpractice concerns and ethical considerations of running trials that could impact unborn infants, may have a distinct relationship with RCTs. Finally, many trials in oncology that include surgery are focused on the chemotherapy and radiation regimen rather than the surgery itself.

that each RCT addresses. To delineate subfields, we begin with the universe of articles in PubMed, a comprehensive database of life science articles maintained by the National Library of Medicine. We then rely on a text similarity algorithm to find the set of articles in PubMed that are the most closely related to each RCT; these clusters of articles define the subfield that the RCT belongs to.

Specifically, we employ the PubMed Related Article (PMRA) algorithm, created by Lin and Wilbur (2007), which calculates text similarity using frequency-based weighting, such that infrequent words contribute more than common terms.[6] A key advantage of PMRA is that it does not rely on citations, which can reflect strategic or social considerations rather than true scientific similarity. In practice, PMRA underpins the "related articles" feature in PubMed, which links each article to other topically related papers. The algorithm has also been used by prior research to determine research subfields and scientific similarity (Azoulay et al. 2019; Myers 2020; Greenblatt 2022), and we use PMRA to generate the subfield for each RCT based on its title and description.

3.1.3. Matching on Observables

After constructing the subfields of our treated and control trials, the final step is to match on key characteristics to ensure covariate balance. We implement coarsened-exact matching (CEM) (Blackwell et al. 2009; Iacus, King, and Porro 2011). For subfields with RCTs published in year t, we restrict the pool of potential control subfields to those that were born by year t. We then perform one-to-one matching on the following features: (i) a set of 16 indicators for trial start years, (ii) a set of 14 indicators for surgical specialty, (iii) whether the trial is multi-site or single-site, (iv) presence of a control arm,[7] and (v) a set of 5 indicators for subfield size at trial completion. We repeat this matching process for every treatment year without replacement. As CEM excludes treated subfields that do not share a common support with potential control subfields, our final sample consists of 521 treated subfields matched to 521 control subfields.

3.2. Econometric Specifications

Our empirical approach leverages a staggered differences-in-differences design, comparing subfields that receive an RCT with published results to matched control subfields that receive an

[6] PMRA tool can be accessed at: https://www.ncbi.nlm.nih.gov/CBBresearch/Wilbur/IRET/TexTool/
[7] Some trials are single-arm, prospective trials, which we include to capture the full range of trial quality.

unpublished RCT. Specifically, our primary estimating equation relates the dependent variable of interest, Y_{it}, of subfield i of RCT j in year t. Formally, we estimate variations of:

$$Y_{it} = \beta_1 Treated_i \times Post\text{-}RCT_t + \beta_2 Post\text{-}RCT_t + f(Age_{it}) + \delta_i + \gamma_t \quad (1)$$

where $Treated_i$ is an indicator variable equal to 1 if a subfield received an RCT with published results and 0 otherwise. The term δ_i represents the individual fixed effects of subfields and controls for time-invariant subfield attributes, while the term γ_t represents calendar year fixed effects to account for changes in publication trends over time. We include a flexible function of subfield age, $f(Age_{it})$, to control for trends in the evolution of subfields as they are born and mature.[8]

We also include an indicator variable, $Post\text{-}RCT_t$, equal to 1 in the years after an RCT's main results were published in the subfield. As control subfields are uniquely linked to treated subfields through the matching process, they inherit a counterfactual date of "treatment" from their paired treated subfield. This, combined with the staggered treatment dates from RCTs, allows us to separately identify $Post\text{-}RCT_t$ from the main effect of interest, $Treated_i \times Post\text{-}RCT_t$, which represents the marginal effect of RCTs. Inclusion of this term addresses the concern that age, year, and subfield fixed effects may not fully account for shifts in subfield activity around the time of an RCT publication (Jaravel, Petkova, and Bell 2018).

3.3. Descriptive Statistics

Table 1 reports the summary statistics and covariate balance between our treated and control subfields. By construction, the subfields are exactly matched on trial years, surgical specialty, multi-site status, presence of a control arm, and pre-period subfield size. Most other characteristics that we did not explicitly match on are also well-balanced, though some differences remain. For example, treated subfields are more likely to receive RCTs that study medications used in surgery. However, these differences in pre-treatment characteristics would only pose a concern if they generate differential pre-trends, which we assess in an event study framework.

Table 1 highlights the diversity in the quality and type of trials, which provides variations for our empirical analyses. For example, there is substantial heterogeneity in terms of trial quality. While most of the surgical trials registered in ClinicalTrials.gov are legitimate randomized

[8] Subfield age is calculated as the difference between the year of interest and the year a subfield is "born." We define birth year as the publication year of the oldest article in the subfield.

controlled trials, 7% are single-arm studies and 13% are not randomized. Blinding is implemented in just 65% of the trials. 29% are multi-site trials, while the remainder are single-site trials conducted at one location.

In terms of surgical specialty, the most common ones are reported in the table; additional matched specialties include thoracic, vascular, urology, and endocrine surgeries. The distribution of specialties reflects our focus on RCT-feasible subfields. For instance, orthopedics is the most represented; this is likely a byproduct of the importance of surgical implants in this subfield, which attracts industry funding from device makers to support RCTs.

About half of the trials evaluate interventions taking place in the operating room. Strikingly, just 15% of the trials test surgical techniques (e.g., incision placement, wound closure methods, comparisons of minimally invasive vs. open surgery, or surgery vs. no surgery). The relative rarity of such trials underscores the challenges of conducting RCTs on surgical techniques, while also providing valuable variations in the extent to which trials directly affect surgeons' work.

Finally, Appendix Table A1 focuses on the treated trials and reports summary statistics on the quality of their result publications. Most are highly cited and published in highly ranked journals, consistent with the medical community's recognition of RCTs as the most rigorous form of evidence. 27% of the trials are eventually incorporated into a clinical practice guideline, on average four years after trial publication.

4. Results

Our first set of results examine how the number of research articles flowing into the subfields changes after the publication of an RCT. As shown in Table 2, we find that research activities into treated subfields appear to dampen after the RCT relative to control subfields, with the decline being more pronounced for higher quality RCTs. Columns 1 and 2 report estimates for the full sample of all subfields; while the point estimates are negative, they are imprecisely estimated. In contrast, when the sample is restricted to the subfields of only top-quality trials,[9] the coefficients are precisely estimated under OLS; the number of articles falls by 1.1 paper per year, which translates to an 18% decrease relative to the baseline mean of 6 articles per year (Column 3). The corresponding Poisson specification indicates a reduction of 10% and is marginally significant.

[9] Top-quality trials are defined as treated trials that were published in journals with journal impact factor above 90th percentile and their corresponding control trials.

Taken together, it appears that follow-on research only responds when the trial is of high quality and generates particularly strong evidence. The absence of effects in the full sample suggests that most surgical trials do not substantially impact subsequent research activities. For the rest of the analyses, we focus on the sample of subfields with top-quality trials.

To assess pre-trends, we show an event studies version of Equation 1, replacing the single $post\text{-}RCT_t$ indicator with indicators for every year before and after the introduction of RCT. Figure 1 plots the dynamic effects of RCT on the quantity of follow-on research.[10] The decline in research is more prominent in Panel B (focusing on subfields of top-quality trials), compared to Panel A (subfields of all trials). Importantly, there appears to be no difference in pre-trends between treated and control subfields. Moreover, the impact of RCT persists, with the decline emerging around in year three, which is consistent with the lag required to conduct and publish follow-on research.

We next turn to how RCTs may affect the quality of subsequent research. In Table 3, we decompose the number of articles entering the subfield by their citation impact: those in the top quartile of citation distribution (Column 1) versus those in the bottom quartile (Column 2). We find that research with high citation impact declines by 32% in the subfield after an RCT, while lower-cited research remains unaffected.

We also assess how RCT affects the clinical relevance of subsequent research, as proxied by whether the follow-on article is included in a clinical practice guideline. Column 3 reports that the number of clinically relevant articles declines by 45%. Overall, these results suggest that the research that continues in the subfield after the RCT tends to be less impactful.

Finally, we examine whether the effect of RCTs varies by trial type. Column 1 of Table 4 focuses on subfields of trials that directly test surgical techniques, and Column 2 focuses on subfields of trials that evaluate other aspects of surgical care (e.g., preoperative management). The decline in follow-on research is only pronounced among subfields where the trial did *not* test a surgical technique.

[10] Figure 1 plots estimates with OLS. Appendix Figure A2 provides event studies figures with Poisson specifications.

5. Discussion

What is the overall impact of RCTs on innovation? Our results suggest that RCTs dampen research and that this decline is stronger for higher quality RCTs. This pattern is consistent with the interpretation that when RCTs provide conclusive evidence, they can be beneficial for innovation by settling debates in the field, allowing researchers to redirect their research priorities to new areas. At the same time, the finding that most trials do not appear to meaningfully shape subsequent research raises questions about the returns to conducting them, given the resources required to run them. Moreover, RCT's impact seems more limited in settings with high adoption costs. We find no evidence of a decline in research for RCTs that target core surgical techniques. This non-response is consistent with the notion that surgeons are more resistant to RCTs that directly challenge their skills. This emphasizes that, while RCTs are the gold standard in providing causal evidence, when they are seen as unreliable or imposing high adoption costs, their impact on downstream research may be limited.

A strength of our study is its ability to use the paper trail provided by innovation to systematically measure the response to causal evidence. This overcomes the challenges of measuring changes to clinical practice directly, especially since surgical trials frequently test interventions too narrow or detailed to be picked up in coarse billing codes. We emphasize that the gap between findings from clinical research entering clinical practice can be large and has been the subject of much prior research (Oliver et al. 2014; Cabana et al. 1999). Reasons for this gap include organizational inertia (DellaVigna et al. 2024), financial incentives (Dranove et al. 2022), psychological biases (Vivalt and Coville 2023), and marketing and awareness (Azoulay 2004; Cabana et al. 1999). However, unlike the distinction between research and practice in many areas of economics and public policy, most surgical researchers are also practicing surgeons who often research what they practice.

How much research remaining in subfields after the arrival of high-quality causal evidence is optimal to maximize scientific progress? Our study does not allow us to make firm conclusions here. After a high-quality RCT, the scientific frontier expands, and scientific attention should concomitantly shift with the frontier. At the same time, there is an important role for further research within subfields that receive causal evidence, both for continued incremental progress,

and because even high-quality trials can sometimes support conclusions later found to be erroneous in light of subsequent research.[11]

We specifically study contexts within surgery where an RCT could be launched—meaning there was at least some belief that randomization was ethical and the study feasible. In causal inference terms, our estimates are the effect of RCT publication on the subset of fields that can (and do) run RCTs. From a policy perspective, this is precisely where RCTs have the ability to shift follow-on research. But it also raises the question of to what extent a rigorous, well-designed study—even if not an RCT—would produce a similar treatment effect on follow-on innovation as the RCTs in our study. This is especially relevant for those questions for which an RCT might be considered unethical, for which the disease rarity or clinical time frame makes a trial impractical, or for which the expected impact of the research does not justify the cost of an RCT. While this remains an open question, our findings suggest any impact would be proportional to the perceived quality of the evidence provided.

While highlighting the strength of high-quality RCTs in shaping follow-on innovation, our study also speaks to their limitations and how RCTs may be utilized more efficiently within the broader innovation system to advance the scientific frontier. First, in cases where the RCT intervention or setting has a high degree of tacit knowledge, such as surgical techniques, even well-executed trials may have only a small impact. Trial consumers may be skeptical that they would get a similar result if the trial were to be reproduced in their own setting, or that there would be very high costs of learning the required tacit knowledge. For such tacit, hard-to-standardize interventions, investing in observational studies may be a better allocation of resources than relying on RCTs. This is especially the case in light of evidence that well-designed observational studies may produce unbiased estimates when compared to RCTs of the same topic (Anglemyer et al. 2014; Benson and Hartz 2000; Concato et al. 2000; Waddington et al. 2023).

Second, we empirically focus on the effect of RCTs on their individual subfields. Our study does not capture the indirect effect of how a scientific reward system that might prioritize RCTs over other types of studies could have on the direction of innovation. Were RCTs to be

[11] It is difficult to systematically measure whether an RCT may have prematurely terminated a field's research activities, but prior evaluations of RCTs suggest it is not infrequent. A study of RCTs in three leading medical journals found that 13% of RCTs led to a reversal of medical practice, some of which were medical practices supported by prior RCTs (Herrera-Perez et al. 2019). More directly addressing the possibility of premature termination, an editor of a leading medical journal found that 8% of RCTs submitted to his journal were "zombie" trials which were fatally flawed and 14% contained false data (Carlisle 2021).

overemphasized relative to a societally optimal amount, this would shift research towards those subfields for which an RCT is ethical and feasible and away from those that are not amenable to RCTs. Some scholars have voiced concern that the "near-religious zeal" for RCTs might lead researchers away from "asking and trying to answer big and important questions ... in pursuit of clean answers to small questions of little policy consequence" (Heckman 2020). To the extent policymakers or leading academics want to shift the balance between the amount of RCTs and other forms of evidence, they may do so both via push mechanisms (e.g., additional grants tied to methodology) and pull mechanisms (e.g., tenure, leading journal publications).

Finally, our sample includes many low-quality trials and trials which did not complete. These efforts expose patients to potential risk with little advancement in knowledge. When combined with our findings that follow-on innovation is much less responsive to evidence from low-quality trials, this suggests having fewer, higher quality trials may more efficiently advance science and protect patients. In contrast to surgical RCTs, many drug RCTs face the disciplining force of review by the FDA which helps to maintain trial quality. A centralized institution reviewing surgical clinical trial proposals could help maintain high trial standards, enhancing the impact of surgical RCTs on follow-on innovation and patient care.

References

Adgate, Bard. 2020. "Randomized Control Testing: A Science Being Tested to Measure an Advertiser's ROI. *Forbes*.

Agha, Kim, and Li. 2022. "Insurance Design and Pharmaceutical Innovation." *American Economic Review: Insights*, **4**(2): 191-208.

Anglemyer, Andrew, Hacsi T. Horvath, and Lisa Bero. 2014. "Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials." *Cochrane Database of Systematic Reviews,* **2014**(4):MR000034.

Aston, Thomas. 2023. "Randomista mania." *Medium.* https://thomasmtaston.medium.com/randomista-mania-1e2befca89e9.

Azoulay, Pierre. 2004. "Do Pharmaceutical Sales Respond to Scientific Evidence?" *Journal of Economics & Management Strategy*, **11**(4):551-594.

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S. Graff Zivin. 2019. "Does Science Advance One Funeral at a Time?" *American Economic Review*, **109**(8):2889-2920

Badinski, Ivan. 2020. "Surgical Innovation and the Challenge of Changing Demographics." *Working Paper*.

Badinski, Ivan. 2021. "The Age of Modern Medicine: Physician Learning Costs and Medical Procedure Adoption." *Working Paper*.

Bagenal, Jessamy, Naomi Lee, Adesoji O. Ademuyiwa, Dmitri Nepogodiev, Antonio Ramos-De la Medina, Bruce Biccard, Marie Carmela Lapitan, and Wangari Waweru-Siika. 2023. "Surgical research—comic opera no more." *The Lancet*, **402**(10396): 86-88.

Balasubramanian, Parasuram, Trey Cummings, and Lamar Pierce. 2022. "Research Validity Across Organizational Forms: Evidence from Phase 2 Oncology Clinical Trials." *Working Paper*.

Banerjee, Abhijit V. and Esther Duflo. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics*, **1**: 151-178.

Banerjee, Abhijit, Esther Duflo, and Michael Kremer. 2020. "The Influence of Randomized Controlled Trials on Development Economics Research and on Development Policy." in Kaushik Basu, David Rosenblatt, and Claudia Sepulveda (eds.). *The State of Economics, the State of the World.* MIT Press.

Beane, Matthew and Callen Anthony. 2023. "Inverted Apprenticeship: How Senior Occupational Members Develop Practical Expertise and Preserve Their Position When New Technologies Arrive." *Organization Science*, **35**(2):405-431.

Beane, Matthew. 2018. "Shadow Learning: Building Robotic Surgical Skill When Approved Means Fail." *Administrative Science Quarterly*, **64**(1): 87-123.

Benson, Kjell, and Arthur J. Hartz. 2000. "A Comparison of Observational Studies and Randomized, Controlled Trials." *New England Journal of Medicine,* **342**(25):1878-1886.

Blackwell, Matthew, Stefano Iacus, Gary King, and Giuseppe Porro. 2009. "CEM: Coarsened Exact Matching in Stata." *The Stata Journal,* **9**(4):524-546

Cabana, Michael D., Cynthia S. Rand, Neil R. Powe, Albert W. Wu, Modena H. Wilson, Paul-Andre C. Abboud, and Haya R. Rubin. 1999. "Why Don't Physicians Follow Clinical Practice Guidelines? A Framework for Improvement." *Journal of the American Medical Association*, **282**(15):1458-1465.

Camuffo, Arnaldo, Alessandro Cordova, Alfonso Gambardella, Chiara Spina. 2019. "A Scientific Approach to Entrepreneurial Decision Making: Evidence from a Randomized Control Trial." *Management Science*, **66**(2):564-586.

Carlisle, J. B., 2021. "False individual patient data and zombie randomized controlled trials submitted to *Anaesthesia*." *Anaesthesia*, **76**(4):472-479.

Chandra, Amitabh, Jennifer Kao, Kathleen L. Miller, Ariel D. Stern. 2024. "Regulatory Incentives for Innovation: The FDA's Breakthrough Therapy Designation." *The Review of Economics and Statistics*.

Chatterji, Aaron K., Michael Findley, Nathan M. Jensen, Stephan Meier, and Daniel Nielson. 2016. "Field Experiments in Strategy Research." *Strategic Management Journal*, **37**(1):116-132.

Concato, John, Nirav Shah, and Ralph Horwitz. 2000. "Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs." *New England Journal of Medicine*, **342**(25):1887-1892.

Deaton, Angus and Nancy Cartwright. 2018. "Understanding and Misunderstanding Randomized Controlled Trials." *Social Science & Medicine*, **210**: 2-21.

DellaVigna, Stefano, Woojin Kim, Elizabeth Linos. 2024. "Bottlenecks for Evidence Adoption." *Journal of Political Economy*, **132**(8):2748-2789.

Dranove, David, Craig Garthwaite, Christopher Heard, Bingxiao Wu. 2022. "The economics of medical procedure innovation." *Journal of Health Economics*, **81**:102549.

Feit, Elea McDonnell and Ron Berman. 2019. "Test & Roll: Profit-Maximizing A/B Tests." *Marketing Science*, **38**(6):1038-1058.

Finkelstein, Amy. 2021. "Bringing Rigorous Testing to Health Care Policy." *The New York Times*.

Frieden, Thomas R. 2017. "Evidence for Health Decision Making — Beyond Randomized, Controlled Trials." *New England Journal of Medicine*, **377**(5): 465-475.

Gallo, Amy. 2016. "A Refresher on Randomized Controlled Experiments." *Harvard Business Review*.

Gawande, Atul. 2012. "Two Hundred Years of Surgery." *The New England Journal of Medicine*, **366**(18): 1716-1723.

Guzman, Jorge, Zubin Jelveh, and Bruce Kogut. 2024. "The unfolding SPIRIT of science: Evidence from p-hacking in academic clinical trials." Working paper.

Greenblatt, Wesley. 2022. "Anchoring Exploration: Knowledge Certification and the Dynamics of Innovation." *Working Paper*.

Heckman, James J. 2020. "Randomization and Social Policy Evaluation Revisited" in Florent Bedecarrats, Isabelle Guerin, and Francois Roubaud (eds.). Randomized Control Trials in the Field of Development: A Critical Perspective. Oxford University Press.

Herrera-Perez, Diana, Alyson Haslam, Tyler Crain, Jennifer Gill, Catherine Livingston, Victoria Kaestner, Michael Hayes, Dan Morgan, Adam S. Cifu, Vinay Prasad. 2019. "A comprehensive review of randomized clinical trials in three medical journals reveals 396 medical reversals." *eLife*, **8**:e45183.

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2011. "Causal Inference without Balance Checking: Coarsened Exact Matching." *Political Analysis,* **20**(1):1-24.

Jaravel, Xavier, Neviana Petkova, and Alex Bell. 2018. "Team-Specific Capital and Innovation." *American Economic Review,* **108**(4–5):1034-1073.

Kao, Jennifer. 2025. "Information Disclosure and Competitive Dynamics: Evidence from the Pharmaceutical Industry." *Management Science*, **71**(7): 5419-6318.

Kolata, Gina. 2016. "Why 'Useless' Surgery Is Still Popular." *The New York Times*. https://www.nytimes.com/2016/08/04/upshot/the-right-to-know-that-an-operation-is-next-to-useless.html.

Koning, Rembrand, Sharique Hasan, Aaron Chatterji. 2022. "Experimentation and start-up performance: Evidence from A/B testing." *Management Science*, **68**(9):6434-6453.

Krieger, Joshua L. 2021. "Trials and Terminations: Learning from Competitors' R&D Failures." *Management Science*, **67**(9): 5525-5548.

Lin, Jimmy, and W. John Wilbur. 2007. "PubMed Related Articles: A Probabilistic Topic-Based Model for Content Similarity." *BMC Bioinformatics,* **8**: 423 (2007).

Maier, Ronald V. 2006. "What the Surgeon of Tomorrow Needs to Know About Evidence-Based Surgery." *JAMA Surgery*, **141**:(3):317-323.

Masanneck, Lars, Pauline Gieseler, William J. Gordon, Sven G. Meuth, and Ariel D. Stern. 2024. "Evidence from ClinicalTrials.gov on the Growth of Digital Health Technologies in Neurology Trials." *npj Digital Medicine*. **6**(23).

McKenzie, David. 2016. "Have RCTs taken over development economics?" World Bank Blogs. https://blogs.worldbank.org/en/impactevaluations/have-rcts-taken-over-development-economics.

Medical Research Council. 1948. "Streptomycin Treatment of Pulmonary Tuberculosis." *British Medical Journal* **2**:769-782.

Morreim, Haavi, Michael J. Mack, Robert M. Sade. 2006. "Surgical Innovation: Too Risky to Remain Unregulated?" *The Annals of Thoracic Surgery*, **82**(6): 1957-1965.

Myers, Kyle. 2020. "The Elasticity of Science." *American Economic Journal: Applied Economics*, **12**(4):103-134.

Mullard, Asher. 2018. "How much do phase III trials cost?" *Nature Reviews Drug Discover*, **17**(777).

Pritchett, Lant, and Justin Sandefur. 2015. "Learning from Experiments When Context Matters." *American Economic Review*, **105**(5): 471-475.

Oliver, Kathryn, Simon Innvar, Theo Lorenc, Jenny Woodman, and James Thomas. 2014. "A systematic review of barriers to and facilitators of the use of evidence by policymakers." *BMC Health Services Research*, **14**:2.

Ramchurn, Tetraj P. and Nundy, Samiran. 2023. "Randomised Clinical Trials in Surgery: Are We at a Crossroads?" *Annals of Medicine and Surgery,* **86**(1), 3–6.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy*, **98**(5):S71-S102.

Rutkow, Ira. 2022. *Empire of the Scalpel: The History of Surgery*. New York, NY: Scribner.

Schneider, David. 2020. *The Invention of Surgery: A History of Modern Medicine: From the Renaissance to the Implant Revolution*. New York, NY: Pegasus Books.

Scotchmer, Suzanne. 1991. "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law." *Journal of Economic Perspectives*, **5**(1):29-41.

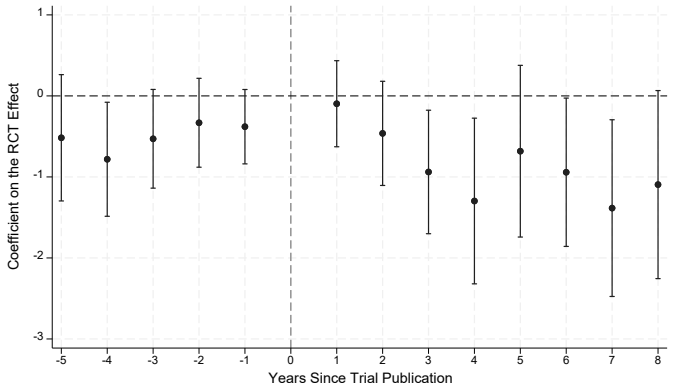
Strandburg, Katherine J. 2014. "Legal but Unacceptable: Pallin v. Singer and Physician Patenting Norms." *Intellectual Property at the Edge: The Contested Contours of IP*. Ed. Rochelle Cooper Dreyfuss and Jane C. Ginsburg. Cambridge: Cambridge University Press. 321–342.

Vivalt, Eva, and Aidan Coville. 2023. "How do policy makers update their beliefs?" *Journal of Development Economics*, **165**:103121.

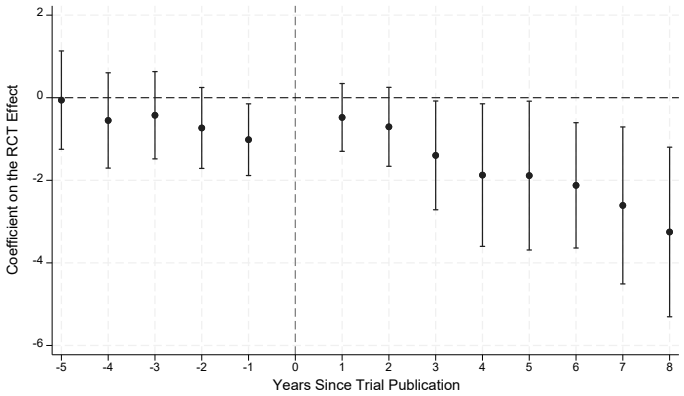
Waddington, Hugh S., Paul V. Villar, and Jeffrey C. Valentine. 2023. "Can Non-Randomised Studies of Interventions Provide Unbiased Effect Estimates? A Systematic Review of Internal Replication Studies." *Evaluation Review*, **47**(3):563-593.

Figure 1. Impact of RCT on the Quantity of Follow-on Research

Panel A. Subfields of All Trials



Panel B. Subfields of Top-Quality Trials



Notes: This figure shows the impact of RCT on the quantity of follow-on research. The figure plots the coefficients and 95% confidence intervals from estimating an event studies version of Equation 1 that replaces the pooled $Post\text{-}RCT_t$ indicator with separate indicators for every year before and after RCT publication. The outcome variable is the number of original research articles flowing into the subfield five years before an RCT publication and eight years after. Panel A reports results for subfields of all trials, while Panel B focuses on subfields of top-quality trials (treated trials that were published in journals with journal impact factor above 90^{th} percentile and their corresponding control trials). The unit of analysis is a subfield $\times$ year. Results are based on OLS estimation.

Table 1. Summary Statistics

	Treated	Control	
Trial start year	2011.9 (4.8)	2011.9 (4.8)	
Trial completion year	2015.2 (4.5)	2014.7 (4.7)	
Trial design			
Has a control arm	93%	93%	
Randomized	87%	87%	
Blinded	65%	60%	*
Multi-site	29%	29%	
Eligibility			
Enrollment size	322.1 (836.2)	374.9 (3916.4)	
Inclusion/exclusion criteria length	173.1 (146.3)	161.8 (164.5)	
Eligibility age	67.5 (17.8)	69.6 (16.6)	**
Both genders	85%	87%	
Industry funded	24%	27%	
Surgical specialty (most freq shown)			
Orthopedics	33%	33%	
General	20%	20%	
Cardiovascular	12%	12%	
Gynecology	10%	10%	
Various	8%	8%	
Spine	5%	5%	
Plastics	3%	3%	
Neurosurgery	3%	3%	
Otolaryngology	2%	2%	
Intervention			
Surgical technique	15%	13%	
Device	21%	29%	**
Medication	47%	34%	***
Anesthesia	8%	8%	
Systems of care/care pathways	4%	5%	
Physical therapy	4%	3%	
Surgical stage			
Preoperative	11%	14%	
Intraoperative	48%	51%	
Postoperative	27%	24%	
Trial endpoints			
Subjective outcomes	46%	46%	
Objective outcomes	37%	33%	
Surgical complications	16%	20%	*
Pain	24%	25%	

Quality of life	15%	17%	
Survival/disease progression	4%	4%	
Subfield			
Subfield birth year	1996.5 (11.2)	1994.9 (11.1)	**
Pre-period subfield size	60.2 (87.2)	59.8 (84.4)	
N of subfields	521	521	

Notes: This table reports the summary statistics and the covariate balance between treated and control subfields. Treated subfields are those that received an RCT with published results, while control subfields are those that received an unpublished RCT. The sample consists of subfields of U.S.-based surgical trials in ClinicalTrials.gov from 1991-2021. Subfields are constructed using the PubMed Related Article algorithm (Lin and Wilbur 2007), which groups each focal RCT with its most topically related articles. For continuous variables, means with standard deviations in parentheses are reported and compared using independent t-tests. For categorical variables, percent shares are reported and compared using Pearson's chi-square tests. Statistical significance of differences between treated and control subfields is indicated as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2. Impact of RCT on the Quantity of Follow-on Research

	Subfields of All Trials		Subfields of Top-Quality Trials	
	(1)	(2)	(3)	(4)
	OLS	Poisson	OLS	Poisson
VARIABLES	N of Papers	N of Papers	N of Papers	N of Papers
Treated × Post-RCT	-0.385	-0.009	-1.144**	-0.105*
	(0.268)	(0.038)	(0.520)	(0.059)
Number of subfields	1,042	1,036	350	349
Number of subfield-years	12,926	12,850	4,444	4,434
Mean of dependent variable	6.123	6.123	6.037	6.037
Adjusted R2 or log likelihood	0.765	-25,325	0.766	-8,659

Notes: This table reports results from estimating Equation 1 and shows the impact of RCT on the quantity of follow-on research. The unit of analysis is a subfield × year. The treatment variable "Treated" is defined as subfields that had a published RCT, while "Post-RCT" turns one years after an RCT publication. The outcome variable is the number of original research articles flowing into the subfield five years before an RCT publication and eight years after. Columns 1-2 report sample using subfields of all trials, while Columns 3-4 are restricted to subfields of top-quality trials (treated trials that were published in journals with journal impact factor above 90th percentile and their corresponding control trials). Columns 1 and 3 report OLS estimates, while Columns 2 and 4 report estimates with conditional quasi-maximum likelihood Poisson specifications. All columns include a set of 32 indicators for calendar years, a set of 42 indicators for subfield age (with age 42 indicator absorbing subsequent indicators), and subfield fixed effects. Standard errors are clustered at the subfield level. Statistical significance is indicated as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3. Impact of RCT on the Quality of Follow-on Research

	Publication Impact of Follow-on Research		Clinical Relevance of Follow-on Research
	(1)	(2)	(3)
VARIABLES	N of Papers in Top Quartile of Citations	N of Papers in Bottom Quartile of Citations	N of Papers included in CPG
Treated × Post-RCT	-0.561***	0.076	-0.300***
	(0.203)	(0.151)	(0.084)
Number of subfields	350	350	350
Number of subfield-years	4,444	4,444	4,444
Mean of dependent variable	1.777	1.004	0.666
Adjusted R2	0.630	0.560	0.461

Notes: This table reports results from estimating Equation 1 and shows the impact of RCT on the quality of follow-on research, focusing on subfields with top quality trials. The unit of analysis is a subfield × year. The treatment variable "Treated" is defined as subfields that had a published RCT, while "Post-RCT" turns one years after an RCT publication. The outcome variable is the number of original research articles flowing into the subfield five years before an RCT publication and eight years after. Columns 1-2 decompose the number of research articles flowing into the subfield by whether they are in the top or bottom quartile of citation impact. Column 3 focuses on the number of research articles that become included in a clinical practice guideline. All columns report OLS estimates and include a set of 32 indicators for calendar years, a set of 42 indicators for subfield age (with age 42 indicator absorbing subsequent indicators), and subfield fixed effects. Standard errors are clustered at the subfield level. Statistical significance is indicated as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4. Impact of RCT on Follow-on Research by Trial Type

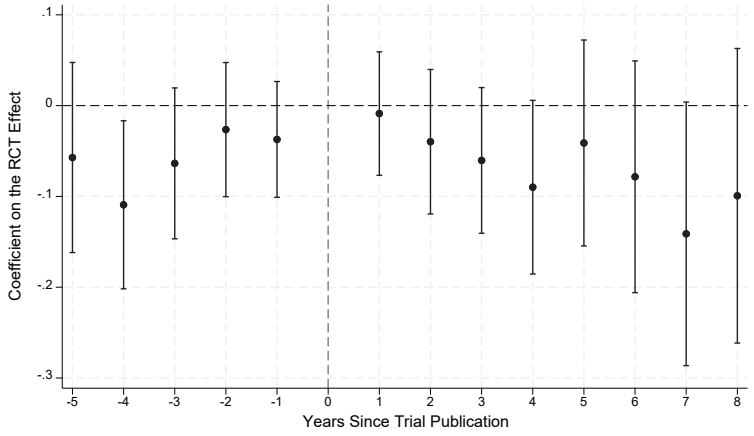
VARIABLES	(1) Surgical Technique Trials N of Papers	(2) Other Surgical Trials N of Papers
Treated × Post-RCT	0.002	-1.279**
	(1.313)	(0.579)
Number of subfields	64	286
Number of subfield-years	820	3614
Mean of dependent variable	6.161	6.009
Adjusted R2	0.761	0.768

Notes: This table reports results from estimating Equation 1 and shows the impact of RCT on the quantity of follow-on research by trial type, focusing on subfields with top quality trials. The unit of analysis is a subfield × year. The treatment variable "Treated" is defined as subfields that had a published RCT, while "Post-RCT" turns one years after an RCT publication. The outcome variable is the number of original research articles flowing into the subfield five years before an RCT publication and eight years after. Column 1 focuses on subfields of trials that evaluate surgical techniques, while Column 2 focuses on subfields of trials that evaluate all other aspects of surgical practice. All columns report OLS estimates and include a set of 32 indicators for calendar years, a set of 42 indicators for subfield age (with age 42 indicator absorbing subsequent indicators), and subfield fixed effects. Standard errors are clustered at the subfield level. Statistical significance is indicated as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

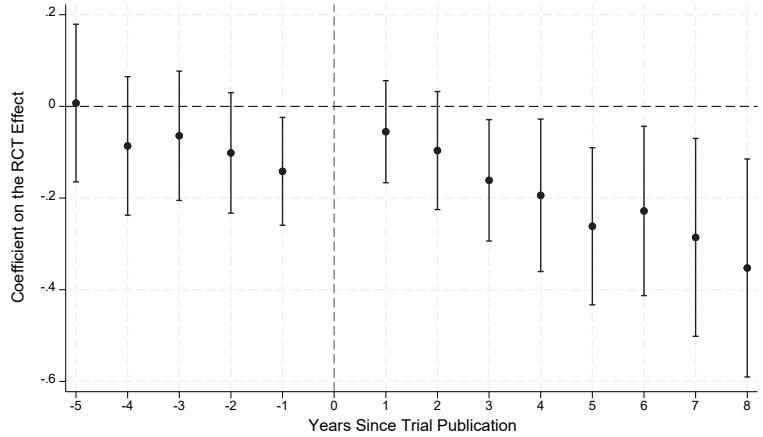
Appendix A
Ancillary Results and Robustness Checks

Figure A1. Impact of RCT on the Quantity of Follow-on Research – Poisson Specification

Panel A. Subfields of All Trials



Panel B. Subfields of Top-Quality Trials



Notes: This figure shows the impact of RCT on the quantity of follow-on research. The figure plots the coefficients and 95% confidence intervals from estimating an event studies version of Equation 1 that replaces the pooled *Post-RCT*$_t$ indicator with separate indicators for every year before and after RCT publication. The outcome variable is the number of original research articles flowing into the subfield five years before an RCT publication and eight years after. Panel A reports results for subfields of all trials, while Panel B focuses on subfields of top-quality trials (treated trials that were published in journals with journal impact factor above 90$^{\text{th}}$ percentile and their corresponding control trials). The unit of analysis is a subfield × year. Results are based on estimates with conditional quasi-maximum likelihood Poisson specifications.

Table A1. Summary Statistics of Treated Trials

	Treated
RCT publication	
Journal impact factor percentile	75.2 (25.0)
Total citations percentile	73.5 (24.6)
Inclusion into clinical practice guidelines	27%
Years taken to be included in clinical guidelines	4.0 (2.9)
RCT results	
Positive	58%
Negative	5%
Null	33%
Mixed	1%
N of trials	521

Notes: This table reports summary statistics for the 521 treated trials (those with publications). Percentiles are calculated relative to the universe of all life science articles in PubMed. Information on trial results—classified as positive (favorable toward the intervention being studied in the trial), negative, null, or mixed—are collected from ClinicalTrials.gov and the trial's result publications. Means with standard deviations in parentheses are reported for continuous variables, while percent shares are reported for categorical variables.

Appendix B
Measurement and Methods

B.1 Identifying Surgical Trials

We retrieved all surgical trials from ClinicalTrials.gov in April 2024. To identify relevant trials, we searched the "intervention/treatment" field on the homepage of ClinicalTrials.gov using both general surgery terms (e.g., surgery, excision, resection) and specialty specific terms. The full list of search terms is available below:

General surgery terms: surgery, excision, resection, anastomosis, reconstruction, repair

Breast: breast implant, breast implantation, lumpectomy, mastectomy, radical mastectomy, segmental mastectomy, simple mastectomy, subcutaneous mastectomy, halsted ii, mastopexy, hadfield′s procedure, hadfield procedure, hadfield operation, hadfield′s operation, breast augmentation, augmentation mammaplasty, breast implant removal, explant surgery, breast implant revision, breast lift, mastopexy, breast reconstruction, breast reduction plasty, breast reduction, reduction mammaplasty

Cardiovascular: aortoplasty, arterial switch operation, arterial switch procedure, rastelli operation, rastelli repair, rastelli technique, mustard repair, mustard procedure, mustard operation, jatene technique, jatene procedure, jatene operation, senning operation, senning technique, senning procedure, rastelli procedure, bentall procedure, bentall operation, bentall′s procedure, bentall′s operation, blalock–hanlon shunt, blalock–hanlon procedure, blalock–hanlon operation, blalock–taussig shunt, blalock–thomas–taussig shunt, blalock-taussig-thomas shunt, blalock–taussig operation, cardiomyoplasty, cardiotomy, commissurotomy, coronary artery bypasses, endarterectomy, fontan procedure, fontan operation, hemi-fontan, glenn procedure, glenn operation, heart transplantation, maze procedure, maze surgery, maze operation, cox-maze procedure, cox-maze operation, cox maze, norwood procedure, norwood operation, pericardiectomy, pericardiotomy, pericardiostomy, pericardial window technique, valvuloplasty, valve surgery, sternotomy

Endocrine: adrenalectomy, parathyroidectomy, thyroidectomy

General: appendectomy, appendicectomy, biliopancreatic diversion, biliopancreatic bypass, bilio-pancreatic diversion, bilio-pancreatic bypass, billorth ii, billrothii gastrojejunostomy, billroth 2, billroth i, billroth 1, cecostomy, cholecystectomy, cholecystostomy, choledochostomy, colectomy, collis gastroplasty, colon resection, colostomy, diverticulectomy, duodenostomy, enterostomy, esophagectomy, ivor lewisesphagectomy, ivor lewis esophagogastrectomy, gastric pull-up, ivor-lewisesphagectomy, ivor lewisoesophagoastrectomy, oesophagoastrectomy, esophagoplasty, esophagostomy, fistulotomy, fundectomy, fundoplication, dor fundoplication, nissen procedure, nissen operation, nissen fundoplication, belsey mark iv repair, belsey procedure, belsey fundoplication, toupet procedure, toupet surgery, toupet operation, toupet fundoplication, ganglionectomy, gastrectomy, subtotal gastrectomy, partial gastrectomy, gastric resection, polyaprocedure, polya-typegastrectomy, reichel–polyaoperation, polya gastrectomy, gastroduodenostomy, gastroenterostomy, gastropexy, gastroplasty, gastrostomy, gavriliu′s operation, dor-gavriliu procedure, gavriliu′s procedure, gavriliu procedure, gavriliu operation, hartmann′s operation, hartmann′s pouch, proctosigmoidectomy, hartmann procedure, hartmann operation, hartmann′s procedure, heller myotomy, myotomy, hemorrhoidectomy, hepatectomy, hepatoportoenterostomy, hepatic portoenterostomy, hernioplasty, hernia repair, herniorraphy, inguinal hernia surgery, halsted i, ileostomy, islets of langerhans transplantation, pancreatic islet transplantation, islet transplantation, jejunoileal bypass, jejuno-ileal bypass, ileojejunal bypass, jejunostomy, kasai procedure, kasai operation, kausch–whipple procedure, whipple′s procedure, whipple procedure, pancreaticoduodenectomy, kausch–whipple pancreaticoduodenectomy, pylorus-preserving pancreaticoduodenectomy, ladd′s procedure, ladd procedure, laparoscopy, laparotomy, liver transplantation, lord′s procedure for haemorrhoids, lord′s procedure, lord′s operation, lord′s procedure for hydrocele, lord′s procedure, lord′s operation, lymphadenectomy, mayo repair, umbilical hernia repair, miles operation, miles′ operation, mile′s operation, abdominoperineal resection, abdominoperineal excision, miles surgery, miles procedure, miles′ procedure, neurectomy, omentopexy, pancreas transplantation, pancreatectomy, pancreaticoduodenectomy, pancreaticojejunostomy, percutaneous endoscopic gastrostomy, proctectomy, rectum excision, rectum resection, proctocolectomy, pyloromyotomy, pyloroplasty, finney pyloroplasty, heineke–mikulicz pyloroplasty, jaboulay pyloroplasty, jaboulay′s pyloroplasty, ramstedt′s

pyloromyotomy, fredet-ramstedt pyloromyotomy, ramstedt′s operation, ramstedt′s procedure, ramstedt pyloromyotomy, fredet ramstedt pyloromyotomy, anderson-hynes pyeloplasty, roux-en-y anastomosis, roux-en-y, roux-en-ygastric bypass, sigmoidostomy, sphincterotomy, lateral internal sphincterotomy, transduodenal sphincterotomy, anal sphincterotomy, spleen transplantation, splenectomy, splenopexy, ureterosigmoidostomy, vagotomy

Gynecology: amniotomy, cervicectomy, clitoridectomy, clitoridotomy, clitoroplasty, colporrhaphy, colpotomy, vaginotomy, fimbrioplasty, hymenorrhaphy, hymenotomy, hysterectomy, wertheim′shysterectomy, wertheimhysterectomy, wertheim–meigsoperation, wertheim′s operation, hysterotomy, labiaplasty, myomectomy, fibroidectomy, oophorectomy, ovariectomy, salpingectomy, salpingoophorectomy, ovariectomy, salpingostomy, sex reassignment surgery, symphysiotomy, synchondrotomy, pelviotomy, trachelectomy, tubal ligation, tube sterilization, fimbriectomy, madlener technique, pomeroy technique, kroener fimbriectomy, noda′s method, aldridge′s procedure, tubal reversal, tubal anastomosis, tuboplasty, vaginectomy, vaginoplasty, vestibulectomy, vulvectomy, vulvoplasty

Neurosurgery: acoustic neuroma surgery, amygdalohippocampectomy, arteriovenous malformation surgery, cerebral decortication, cerebral cortex decortication, hemidecortication, chiari malformation surgery, chiari decompression surgery, skull base surgery, cingulotomy, craniectomy, craniotomy, deep brain stimulation, hemispherectomy, hypophysectomy, lobectomy, lobotomy, microvascular decompression surgery, nerve transfer, neurotization, nerve graft, nerve transplant, pallidotomy, split-brain procedure, corpus callosotomy, thalamotomy, trephination, trepanation, trepanning, trephining, ventriculoperitoneal shunt, ventriculo-peritoneal shunt, ventriculo peritoneal shunt, cerebrospinal fluid shunt, ventriculostomy, ventriculocisternostomy, pinealectomy, hippocampectomy, cranioplasty

ORL: auriculectomy, bronchotomy, cochlear implantation, cricothyroidotomy, cricothyrotomy, endolymphatic shunt, frenectomy, frenuloplastyof tongue, tongue frenuloplasty, gillies lift, gillies′ lift, glossectomy, killian′s operation, killian operation, killian procedure, killian′s procedure, labyrinth fenestration, laryngectomy, hemilaryngectomy, laryngoplasty, mastoidectomy, tympanomastoidectomy, myringoplasty, myringotomy, tympanostomy, ear tubes, nasal valve reconstruction, nasal valve surgery, nasal valve repair, open nasal fracture reduction, ossicular reconstruction, ossicular implantation, ossicular replacement, parotidectomy, pharyngectomy, laryngopharyngectomy, pharyngostomy, pneumonectomy, rhinectomy, rhinotomy, septoplasty, sinusotomy, sistrunk procedure, sistrunksurgery, sistrunkoperation, stapedectomy, stapes mobilization, mobilization of stapes, submandibular gland removal, submandibular gland excision, thyrotomy, tonsillectomy, adenotomy, adenoidectomy, turbinate reduction, turbinectomy, conchotomy, turbinoplasty, tympanoplasty, uvulopalatoplasty, uvulopalatopharyngoplasty, uvulotomy, attico-antrotomy, labyrinthectomy, radical maxillary antrotomy, caldwell-luc, aryotenoidectomy, thyreoplasty, isshiki procedure, thyroplasty, isshiki thyroplasty

Orthopedics: acetabuloplasty, acromioplasty, amputation, disarticulation, anterior cruciate ligament reconstruction, arthrodesis, arthroplasty, arthroscopy, arthrotomy, astragalectomy, talectomy, whitman operation, bankart repair, bankart procedure, arthroscopic capsulolabral repair, brostrom procedure, bursectomy, chopartamputation, chopart′samputation, coccygectomy, epiphysiodesis, facetectomy, fasciotomy, hemiarthroplasty, hemicorporectomy, hemipelvectomy, ilizarov technique, ilizarov method, internal fracture fixation, osetosynthesis, intramedullary fracture fixation, intramedullary nailing, joint capsule release, keller′s excision arthroplasty, keller′s arthroplasty, keller resection arthroplasty, keller procedure, keller arthroplasty, keller operation, keller′s operation, keller′s procedure, keller excision arthroplasty, limb lengthening surgery, limb-lengthening surgery, limg salvage surgery, limb-salvage surgery, limb sparing surgery, lisfranc′s amputation, lisfranc amputation, meniscectomy, meniscal resection, open fracture reduction, ostectomy, osteotomy, paul′s operation, proximal abducting ulna osteotomy, paul′s surgery, paul′s procedure, paul procedure, paul operation, paul surgery, phemister graft, phemister bone graft, phemister procedure, phemister-charnley, posterior cruciate ligament reconstruction, rotationplasty, syme′s amputation, syme amputation, synovectomy, synovium resection, tendon transfer, tenodesis, tenotomy, tendon release, tendon lengthening, heel-cord release, tommy john surgery, ulnar collateral ligamentreconstruction, ulnar collateral ligament reconstruction

Plastics: abdominoplasty, tummy tuck, lipoabdominoplasty, blepharoplasty, brachioplasty, burrow′s triangle excision, buttock lift, gluteal augmentation, gluteal lift, cervicoplasty, cheek augmentation, cheek reduction, buccal fat removal, escharotomy, facial implant, facial implants, forehead lift, brow lift, frontalis lift, frontalis suspension, frontalis sling, genioplasty, mentoplasty, grafting, hair transplantation, liposuction, aspiration lipectomy, lipoplasty, lipectomy, neck lift, rhytidectomy, otoplasty, otoplasty, palatoplasty, panniculectomy, pectus excavatum repair, rhinoplasty, rhytidoplasty, face lift, submental lipectomy, neck lift, thigh lift, thighplasty, tissue expansion, surgical flaps, z-plasty, homans-miller procedure, homans-miller operation

Spine: corpectomy, diskectomy, discectomy, foraminotomy, hemilaminectomy, khyphoplasty, vertebral augmentation, vertebroblasty, laminectomy, laminoplasty, laminotomy, laminectomy, spinal fusion, spondylosyndesis, spondylodesis, total disc replacement, disc arthroplasty, vertebrectomy, cordotomy, myelotomy, rhizotomy

Thoracic: nuss procedure, nuss pectus excavatum repair, sympathectomy, diaphragmatic plication, esophageal reconstruction, lobectomy, lung decortication, lung transplantation, mediastinoscopy, pleurodesis, pneumonectomy, lung volume reduction surgery, pneumotomy, segmentectomy, thoracoplasty, thoracoscopy, thoracostomy, thoracotomy, thymectomy, thymus transplantation, trendelenburg′s operation, trendelenburg operation, video-assisted thoracoscopic surgery, wedge resection, zenker's diverticulectomy, zenker's diverticulum repair, zenker's diverticulum surgery, thoracoplasty, mediastinotomy

Urology: jaboulay–winkelmann operation, jaboulay procedure, winkelmann procedure, winkelmann operation, jaboulay operation, jaboulay–winkelmann′s technique, jaboulay′s technique, winkelmann′s technique, cystectomy, transurethral resection of bladder, cystostomy, foley operation, foley′syvpyeloplasty, foleyyvpyeloplasty, foreskin restoration, kidney transplantation, meatotomy, nephrectomy, nephrolithotomy, nephropexy, nephrostomy, percutaneous nephrostomy, nephrotomy, orchidectomy, orchiectomy, orchiopexy, penectomy, peotomy, penile implantation, phalloplasty, prostatectomy, transurethral resection of prostate, pyeloplasty, scrotoplasty, ureterostomy, urethropexy, urethroplasty, urostomy, vasectomy, vas ligation, vasoepididymostomy, vasovasostomy, vasectomy reversal

Vascular: axillofemoral bypass, axillo-femoral bypass, endarterectomy, carotid endarterectomy, homans operation, homans′ procedure, homans′ operation, homans procedure, kondoleon procedure, kondoleon′s procedure, kondoleon′s operation, kondoleon operation