Monetary incentives versus public funding in healthcare research: what matters the most?

Gabriele Letta[a] Luca Salmasi[a] Gilberto Turati[a]

[a] Department of Economics and Finance, Università Cattolica del Sacro Cuore, Milan

May 2025

Abstract

We study the impact of two policy interventions on scientific research productivity in a major private Italian hospital. The first is a performance-based monetary incentive program (a Management-By-Objectives, or MBO, bonus) introduced by the hospital management to reward non-academic physicians for publishing research. The second is the hospital's recognition as a Scientific Institute for Recovery and Research (IRCCS), which allowed its academic medical researchers to access dedicated public research funding. Using detailed panel data on physicians' publications from 2012 to 2022, we employ several difference-in-differences strategies to evaluate each policy's effect. We find that the introduction of monetary incentives did not lead to any significant impact on research output of the previously less research-active (non-academic) physicians, unless they were both treated by the MBO policy and had also access to IRCCS funding ("double-treated"). The IRCCS recognition caused instead a major boost in the publication rates of academic doctors and Medical Directors affected by both policies. We also assess the impact of such policies on the research quality of the hospital, by accounting for citations. We document increased cross-collaboration between the monetary-incentivized groups, indicating the emergence of knowledge spillovers; however, such increase was quite substantial for the non-academic doctors, while being modest for structured researchers in relative terms. Our findings might be able to inform the design of policies to incentivize research in healthcare organizations, highlighting the lower significance of performance-based incentives in absence of adequate research funding means.

Keywords: Health Systems, Medical Research, Incentives, Management-By-Objective, Public Funding.
JEL Classification: I10, I23.

1 Introduction

Over the latest years, the debate on innovation, knowledge production and productivity has strongly focused on the importance of research funding and institutional incentives. Especially in presence of fiscal constraints, both public and private institutions have to deal with complex decision-making processes concerning the allocation of scarce resources across different research activities. This achieves even greater relevance in the light of the budget cuts to research institutions proposed by the American administration for 2025 and 2026, which are deemed as potential triggers of repercussions for the scientific global community. Such cuts involved serious reductions to federal science agencies, including annual decreases in funding to the National Institute of Health - NIH - and to the National Science Foundation - NSF - (Peel et al., 2025). Such tightening, joint with restrictions on international collaborations and indirect cost recovery, caused growing legal opposition and concerns within the scientific community (Garisto, 2025, Terhune, 2025). Budget cuts also harm the execution of large-scale endeavors and innovation programs, while medical research institutions highlight how these actions might dampen the performance of clinical trials (Malakoff, 2025). In such deteriorating research environment, private entities (such as venture capital firms) started mobilizing funds to support researchers, pinpointing a major transition of scientific resources from the public sector towards non-public funding frameworks and incentive structures (Glenza, 2025, UNUMLux, 2025), consolidating a pattern ushered already along previous years and marked by a growing presence of philanthropic associations (Shekhtman et al., 2024).

The European research environment is not devoid of similar issues. In the UK, reductions in research spending over the decade following Brexit, also linked greater difficulties in the access to Horizon Europe, have spread further worries among academic actors (Simpkin and Mossialos, 2017, Fernando et al., 2020, Fahy et al., 2021, Highman et al., 2023, Gregory, 2025, Napolitano, 2025). However, relations with the European world of research has thawed over recent times, especially with the expectation of channeling resources from the drain occurring in the U.S. scientific community (Lawless, 2023, Cookson, 2025). In France and Germany, scientific debate persists on how to reconcile the broad scopes of research equity and autonomy with excellence-driven funding schemes and incentives, specifically in presence of philanthropic and private actors (de Bengy Puyvallée et al., 2025, Marchandot and Morel, 2025). These changes underline an almost fully-fledged transformation in the management of science, as the downsizing of public sectors and the greater involvement of private entities interact with each other, shaping research incentive schemes and scientific priorities. Such transforming ecosystem, which is

jointly affected by political budget shocks, institutional reconfigurations and the emergence of diverse and interacting means of financing, frames the relevance of the present work. By examining performance-based incentives to research and comparing them to public funding granting, within a large European private healthcare institution in Italy, we aim at contributing to the broader debate on how structural incentives shape scientific performance, particularly in sectors where the boundaries between public scopes and private aims are becoming more blurred with time; the healthcare sector in our case, as it has been facing even major challenges over the recent years in most developed countries not only due to public and political transitions, but also given the on-going demographic transformations.

As a matter of fact, the increasing pressure borne by European healthcare systems is a relevant issue nowadays. The aging population is driving up the utilization of healthcare and its costs (Tang et al., 2022), bringing about major consequences in terms of macroeconomic growth (Acemoglu and Restrepo, 2017; Aksoy et al., 2019). As health impairs with age, chronic diseases become more prevalent, which boosts further long-term care expenses (D. E. Bloom et al., 2020; S. Chen et al., 2024; Hacker, 2024; Ye et al., 2023). In addition to such structural issues, the COVID-19 outbreak highlighted the need for a better resilience of health systems to epidemics (D. E. Bloom et al., 2022; S. Chen et al., 2021; Wu and Wang, 2024) which, joint with the other mentioned factors, contribute to a constant increase in health expenditures, in doing so challenging the sustainability of the systems.

In such regards, scientific research and innovation in medical practices turn out to be crucial for the enhancement of the efficiency and effectiveness of healthcare. Whereas the short-term impact of new treatments on efficiency is not to be taken for granted (Grant and Buxton, 2018), as innovative health technologies may initially raise costs (Chandra and Skinner, 2012), a steady and consistent research activity is expected to enhance the quality of care, by improving the long-run health outcomes of the affected population, fostering prevention, and triggering cost savings in a lengthened perspective. This is the reason why a high-standard medical research activity is usually favored by institutions, and healthcare facilities are often ranked and assessed by third-parties in accordance to the quality of the research they are able to carry over.

At the hospital level indeed, research may yield multiple benefits. Research-active institutions are better able to attract and retain high-quality medical professionals (AMS, 2020; Maynou et al., 2024); in addition, if they are able to "translate" research outcomes into medical practices, they tend to improve clinical standards of care (Barrenho et al., 2021, 2025). Such hospitals can also develop comparative advantages that ultimately benefit patient outcomes (N. Bloom, Propper, et al., 2015; N. Bloom et

al., 2020; Ghandour et al., 2022). These potential benefits notwithstanding, conducting research in healthcare settings requires strong incentives and huge resources, increasingly lacking in the current research ecosystems, and also due to the fact that clinical duties often take priority over the rest. Which raises a key policy question: how should individual institutions incentivize research engagement?
In academic environments, physicians and researchers are inherently motivated to publish by career-oriented motives (tenure, promotions, prestige, Checchi et al., 2021), as well as altruistic reasons or due to scientific curiosity (Rousseau et al., 2021). These *academic incentives* mean that even without a direct monetary rewarding of scientific contributions, academic clinicians conduct research "by definition" to progress in their job and, by doing that, they tend to improve care to a broad extent, although not necessarily for the nearer environment they are involved with, like the patients with whom they have direct interaction. This may raise even further concerns regarding the unintended consequences of placing too much weight on research activity. Achieving a higher research productivity in academic institutions is also simplified by the availability of funding and resources, usually granted through public funds. Outside academia, such incentives may be weaker, which is why additional mechanisms might be necessary to trigger research activities among practitioners.

We provide with empirical evidence on two different approaches to foster research in healthcare: performance-based monetary incentives versus increased public funding. We leverage the case study of a large, private, Italian teaching hospital, where both types of policies were implemented in quick succession. First, in 2017 the hospital introduced a performance-based Management-By-Objectives (MBO) bonus program, aimed at financially rewarding the non-academic hospital physicians for publishing papers within an yearly time-span. One year later, in early 2018, after succeeding in a recognition procedure ushered concomitantly with the MBO introduction, the hospital obtained the IRCCS status (Scientific Research Hospital designation), which led to additional government funding for research, which mainly benefited the hospital's academic doctors. Our research design provides with a unique comparison between a direct monetary incentive for individuals and a broader institutional research funding intervention, within the same organization and time frame.

The present work also contributes to several strands of literature. First, being the MBO a form of corporate policy devised by the hospital management, we relate to the literature about management quality and incentives in imperfect markets and public services. Prior studies underlined the importance of management practices in hospitals (N. Bloom, Propper, et al., 2015; N. Bloom et al., 2020; Goodall, 2011) and schools (Muralidharan and Sundararaman, 2011, N. Bloom, Lemos, et al., 2015), as well as

the impact of higher competition and incentive reforms in the healthcare sector (N. Bloom, Propper, et al., 2015; Gaynor et al., 2012, 2013; Kessler and McClellan, 2000; Longo et al., 2017; Propper et al., 2004). As a matter of fact, performance-based incentives in the public sector have been shown to improve relevant outcomes in some contexts (Burgess et al., 2017; Dal Bó et al., 2013). In the medical field, numerous studies examine how physicians respond to either financial incentives (Bertoli and Grembi, 2019; Brosig-Koch et al., 2024; Gruber and Owings, 1996; Molitor, 2018; Shurtz, 2013, 2014) or ranking-based evaluations (Huesmann et al., 2025). We add to this literature by evaluating a performance-pay scheme for hospital physicians aimed at research outputs, rather than clinical outcomes and practices, which is the commonest metrics for the above literature.

Second, our study contributes to the literature on research funding and scientific productivity. Several studies have investigated the returns to public R&D grants and funding on academic research output and innovation. For instance, prior scholars have inquired on how grant funding affects scientists' subsequent publications, citations, entrepreneurship, and mobility. While additional funding often correlates with higher research productivity, even causally (Azoulay et al., 2011, Benavente et al., 2012, Ganguli, 2017, Baruffaldi et al., 2020, Babina et al., 2023, Ghirelli et al., 2023), other evidence display mixed findings, as some authors find weak or non-existent effects of grants on individual publications and productivity (Jacob and Lefgren, 2011, Banal-Estañol et al., 2023). We fit into this research setting by examining the effect of a new modality of granting public funding to medical research (i.e., the *IRCCS* recognition) on the research output of academic doctors and more prolific physicians, and by directly comparing it to the effect of the incentive-based policy rewards.

Third, we refer to the literature on knowledge spillovers and collaboration networks in science. In the economic field, academic outcomes has been found to be indeed shaped by peer effects, like exposure to productive researchers or co-authorship networks (Azoulay et al., 2010; Bosquet et al., 2022; Brogaard et al., 2014; Colussi, 2018; Waldinger, 2010). In medicine, such phenomenon has been studied mainly within the context of innovative practice diffusion among clinicians (Barrenho et al., 2021, 2025). By focusing on the interactions between the two different groups of physicians (those eligible for the MBO bonus and those with direct access to the IRCCS public funding), we can study whether incentives for one group spill over to the other. The intuition behind it is that the MBO scheme might lead doctors not affiliated to universities to seek for collaborations with academic physicians, in doing this affecting their output indirectly, and vice versa.

We provide evidence on three main points. First, we compare the effectiveness of private performance-

based monetary incentives versus public, career-oriented funding incentives in stimulating research output. Second, we examine individual determinants of research productivity in a private hospital operating in an "imperfect" market (non-profit healthcare), where academic motivations may be less dominant. Third, we assess potential research spillovers resulting from the combination of these two policy interventions, including increased cross-group co-authorship. To our knowledge, this is the first study to jointly evaluate a direct monetary reward program and a public research funding boost within the same setting. To preview our findings, we find that the MBO-monetary incentive did not have any significant impact on the publication productivity of non-academic physicians, without being able to narrow the gap between them and their university counterparts. On the other side, the IRCCS recognition and associated funding increased research output among the hospital's academic physicians by more than 100%. A significant, positive effect of the MBO is observed only when both incentives applied (i.e. non-academic physicians included in the group of IRCCS-funds' recipients), which would suggest the effects were cumulative. However, the impact is relevant only with respect to individuals subject to the incentive scheme only or subject to no scheme at all, turning negative and statistically significant when comparing such group to academic physicians only subject to the public-funding improved access. We also retrieve that the policies led to greater collaboration between the two groups of researchers, with the non-academic physicians seeking for collaboration with IRCCS-recipients, while no effect is observed on citations overall. A battery of robustness checks validate the fact that results are not driven by pre-existing trends, by the COVID-19 shock, or by compositional changes and spillovers.

The remainder of the paper is organized as follows. Section 2 focuses on the institutional framework, describing the hospital and the two implemented policies. Section 3 dwells on the data sources and sample construction. Section 4 presents the analysis of the MBO incentive, including the empirical strategy, the main results, and some additional findings achieved through a triple-difference estimation. Section 5 focuses on the IRCCS funding and its effects. Section 6 reports several robustness and validation checks. Section 7 examines other outcomes such as collaborative publications and citation counts. Section 8 concludes. Additional figures and results are provided in the Appendix.

2 Institutional Framework

The study is set in a leading private hospital in Rome, Italy, affiliated with a prestigious private Italian university. The hospital is a major healthcare hub which offers both clinical services and medical

education. It is also renowned for research: in recent years it has ranked among the top hospitals in Italy for research output. As a matter of fact, it has been consistently listed among the top 50 hospitals worldwide over the recent years, and more than 75 of its affiliated researchers are included in the top 2% ranking of scientists globally (according to Standord's standardized citation metrics). These characteristics make it an ideal environment for the studying of policies aimed at boosting research activity.

In 2017–2018, the hospital underwent two significant research-related policy changes. First, in 2017 (in anticipation of a forthcoming evaluation for obtaining research status), the hospital management implemented a performance-based MBO policy to encourage publication by hospital physicians who did not have university affiliations ("non-academic" physicians). This MBO program (which has been active every year since 2017) has offered monetary bonuses for publications: each physician would receive a payout for each peer-reviewed journal article published in a given year, with the amount of the payment being proportional to the journal's Impact Factor. Specifically, for every Impact Factor point of a published article (as indexed by the Web of Science website), the physician earned €500, up to a maximum bonus of €10,000 per year (equivalent to 20 points). Thus, publishing in higher-impact journals yielded larger rewards. The total annual budget allocated to the MBO bonus fund was €1 million. The bonus is also weighted according to who participates to the publication: first-authors of papers get the whole computed bonus if there are not internal collaborators in the publication (100%), and 60% of it if they co-author with internal affiliates. Non-first author researchers who co-author with external colleagues only get 50% of the calculated reward, which goes down to 40% if the publication involves at least another internal affiliated member. A comprehensive scheme of the rewarding policy is presented in Table A1, in the Appendix. Only physicians employed as medical doctors (clinicians) without a university faculty affiliation were eligible for this scheme. Academic faculty physicians were excluded from MBO by definition. The goal of this program was to foster research activity among practitioners who traditionally focused on patient care and had lower research output. Second, in February 2018, the hospital achieved the status of IRCCS (*Istituto di Ricovero e Cura a Carattere Scientifico*, or Scientific Institute for Research and Healthcare). IRCCS is a special designation conferred by the Italian Ministry of Health to institutes that excel in biomedical and healthcare service-related research while also delivering high-quality healthcare. The designation process is rigorous and explicited by the law: after preparing the process over the previous years, the hospital had to submit documentation in early 2017 to the regional government to receive approval; once documentation was approved in August

2017, it underwent expert evaluations and on-site visits by the Ministry late in 2017. Upon recognition in 2018, the hospital was officially accredited as an IRCCS in two specialty research areas (Personalized Medicine and Innovative Biotechnologies). Even though IRCCS are an Italian particularity, there are several international comparable institutions that may be associated in features to IRCCS, as they too are nationally accredited to perform research. Institutes like IHU (*Instituts hospitalo-universitaires*, France), *Universit¨atskliniken* (Germany), AMC (*Academic Medical Centers*, U.S.), IIS (Institutos de Investigaciòn Sanitaria, Spain) are alike to IRCCSs, albeit regulated to a lesser extent.

As an IRCCS, the hospital under matter gains access to dedicated public research funds – notably an annual "Current Research" fund available exclusively to IRCCS institutions, and preferential direct channels to compete for "Targeted Research" grants (which, otherwise, are granted to other hospitals only by means of being appointed by their Regions of reference, upon being awarded the funds themselves in public tenders). The IRCCS status also carries obligations: the hospital must maintain high research standards, and it is subject to occasional re-evaluation by experts. The quality standards foreseen by the government in order to obtain the IRCCS recognition are presented in Table A2.

Concurrent with IRCCS recognition, the hospital management compiled an official list of researchers (the IRCCS research staff "perimeter") who would be eligible to use the newly available public research funds (updated every year)[1]. This IRCCS staff list consisted mainly of academically affiliated physicians (university-employed doctors working at the hospital), but it was also made by some non-academic hospital physicians who were active in research. Importantly, all academic physicians, whether included in the IRCCS list or not, were not part of the MBO bonus program (as noted above), and were appointed by the university. Although such appointment was of course based on preemptive characteristics, it was not a matter of direct self-selection. Meanwhile, non-academic physicians on the IRCCS list were still eligible for the MBO bonus (because they were hospital-employed clinicians). In other words, there was a subset of "double-treated" individuals: a few non-academic doctors who qualified for the IRCCS research staff list (thus benefiting from the IRCCS public funding) while also being eligible for the MBO monetary incentive. We summarize the groups and timing of these interventions below. In 2017 (the introduction of MBO), the relevant physician groups can be categorized as follows: 0 = those with no MBO and not included in the IRCCS perimeter (pure control group), 1 = those who eventually only

[1]As discussed later in the paper, the IRCCS list is updated every year. However, the updating is a formal recognition of individuals already involved in the IRCCS' activities due to their productivity and research activity, even without having official access to the funds until they are acknowledged by the Ministry as formally taking part to the research group. In reality, it would be more accurate to consider such group of individuals as featured by a time-varying size pattern. We take this into account when we discuss the validity of our approach.

benefit from IRCCS funding (academic physicians not eligible for MBO), 2 = those who receive the MBO incentive only (non-academic physicians not included in IRCCS list), and 3 = those who receive both MBO and later IRCCS (non-academic physicians who were included in the IRCCS research staff list). In 2018 (the IRCCS recognition), we can instead categorize individuals as: 0 = neither IRCCS nor MBO (pure control), 1 = MBO-only, 2 = IRCCS-only, and 3 = both IRCCS and MBO. In the empirical analysis we leverage such categorizations to define treatment and control groups for each policy.

MBO-levels of treatment

	Description
Policy	*Management-By-Objective (2017)*
0	Full control (neither MBO nor IRCCS)
1	IRCCS-treated only
2	MBO-treated only
3	MBO+IRCCS double treatment

IRCCS levels of treatment

	Description
Policy	*IRCCS recognition (2018)*
0	Full control (neither MBO nor IRCCS)
1	MBO-treated only
2	IRCCS-treated only
3	MBO+IRCCS double treatment

Table 1: Description of the two sets of treatment tiers.

3 Data

We construct a panel dataset of the professionals working in the hospitals, joint with their research output spanning 2012 to 2022. The core personnel data come from the hospital's administrative records, which include all individuals employed in a professional capacity (physicians, researchers, and other healthcare staff) from 2017 onward. From that, we identify our population of interest: physicians and medical researchers continuously employed at the hospital between 2017 and 2022. For each person, the HR data provide demographic information (age, gender, place of birth), employment details (job title or role, department and unit, contract type, and hiring date), and indicate whether the individual has an academic affiliation (university-employed physician) or is a hospital-only employed clinician. We exclude medical interns, residents, PhD students, and external collaborators, as consistent data on these categories were not available. Because the administrative records begin in 2017, we supplemented them by benchmarking the administrative information on the starting dates with hire dates and career details from online CVs and institutional websites, enabling us to infer each physician's presence at the hospital

in earlier years. Using the hire date information, we retrospectively extend the panel back to 2012 for all individuals who were active within the hospital between 2012 and 2017. We resort to this choice as we are not being able to identify workers who were active before 2017 and left the hospital before that year, hence not allowing us to have any information about their career. Thus, in order not to build an unbalanced panel with attrition focused only on the post-treatment part of the sample, we keep only the units who are balanced for the whole longitudinal dataset. The resulting balanced panel covers 579 physicians and researchers, each observed annually from 2012 through 2022, yielding 6,369 person-year observations. This balanced structure ensures that we can track research output for a fixed cohort over time, including several years before and after two policy changes. We create indicators for each person's group status (Medical directors vs. faculty members, MBO vs. non-MBO, IRCCS list member vs. not, etc.) which remain fixed for analysis in the main empirical strategy, reflecting their initial category. Some individuals (3.4%) switch position in the time series under consideration (e.g., a physician became an academic or vice-versa during the period); we handle such cases in robustness checks by excluding switchers. In addition, some individuals are aggregated to the IRCCS perimeter with one or few years' lag compared to the initial timing (2018). While this would make the setting more adequate to a staggered adoption analysis, in the main estimates we considered such *ever-treated* individuals as treated from the start, due to their preemptive involvement with research activity and with the offical IRCCS-research group, thus considering them as informally treated individuals expecting formal recognition at some point. We corroborate the validity of such design by keeping in the sample only those who receive the IRCCS-status from the beginning, and by performing some robustness Staggered DiD estimates as well, confirming the robustness of our main estimates.

To measure research output, we gathered data on scientific publications from Clarivate's Web of Science database. We retrieved all publications (articles, reviews, etc.) from 2012 to 2022 for the authors in our panel (included the ones excluded from our main dataset due to the attrition focused on the post-2017 period). We accurately crafted such process by ensuring the recovered authors were the ones affiliated to the hospital, by matching names and surnames, fields of research, and by double checking we were not dealing with homonyms. This removes the portion of sampling errors that other studies, like Jacob and Lefgren, 2011, have to account for due to their matching based only on researchers' surnames. Each publication record provides the publication year, journal, number of co-authors, citations (as of the time of data collection, which is October 2024), and some other bibliometric details. We then matched authorship on these papers to the individuals in our hospital staff panel by name. From this,

we computed the number of publications per person per year, which is our main outcome variable. In addition, we recorded the total citations each person's publications received (to evaluate research impact).

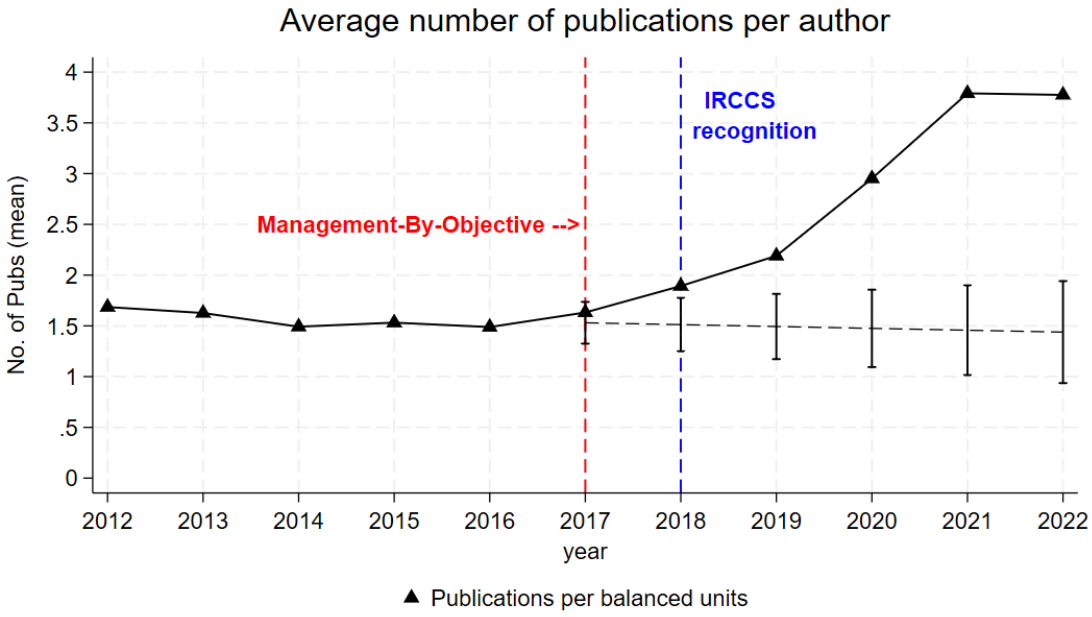
	Mean	SD	Min	Max
Age	51.184	7.859	31	70
Publications	2.187	4.579	0	102
Times cited (WoS)	57.867	204.886	0	7037
Times cited (All outlets)	61.494	218.530	0	7509
Female	0.418	0.493	0	1
Male	0.582	0.493	0	1
Medical Director (hospital only)	0.409	0.492	0	1
Healthcare Professions' Manager	0.019	0.135	0	1
Sanitary Director	0.050	0.218	0	1
Faculty Member with Clinical Functions	0.523	0.500	0	1
MBO-only treated publications	0.452	1.039	0	12
IRCCS-only treated publications	4.978	6.919	0	102
Double-treated publications	2.258	2.776	0	20
Pure control publications	0.452	1.039	0	12
MBO-only treated citations	8.127	32.219	0	1009
IRCCS-only treated citations	150.415	359.049	0	7509
Double-treated citations	66.689	145.036	0	2102
Pure control citations	8.127	32.219	0	1009
MBO-only treated units	88.808	50.987	1	177
IRCCS-only treated units	97.183	56.238	1	194
Double-treated units	32.724	19.206	1	66
Pure control units	88.808	50.987	1	177
Total units of the balanced panel	290	167.156	1	579
Observations	6369			

Table 2: Descriptives

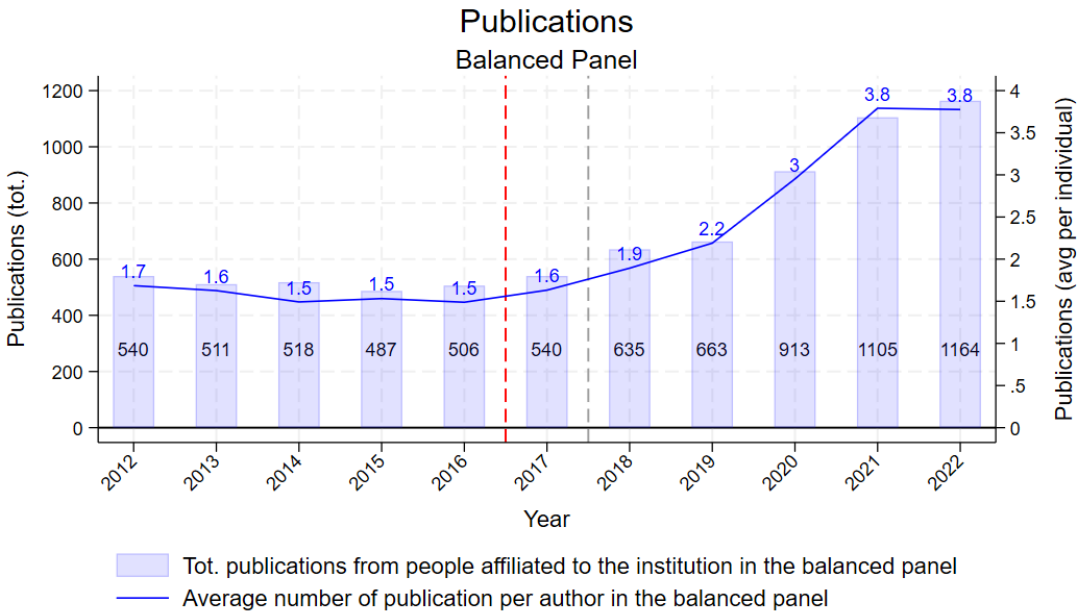
Table 2 reports the summary statistics for the selected sample. About 52% of the 579 researchers in the panel are academic physicians (as in, they hold a university faculty appointment while also performing clinical activity), while the remainder are non-academic hospital physicians or other research/healthcare staff. In terms of treatment groups, it is clear how the units differentiated by status (IRCCS-only, MBO-only, double-treated) do not sum up to the total number of individuals in the sample: this is due to the 20 switching units, which are taken care of in the robustness checks. On average, academic physicians in the IRCCS perimeter have higher scientific output than non-academics, reflecting different incentive structure and attitudes towards research. The average number of publications per person per year is around 5 versus less than 0.5 publication for non-IRCCS, non-academics. Their joint subset is instead in the middle way with respect to prolificacy, reporting almost 2.3 average yearly publications. The pure control group, which includes academic physicians and other professionals non

included in either policy, presents a higher mean annual publication number (0.85). If we look at the times the works published in a given year were reportedly cited in October 2024, the period when data were collected, the statistics follow, intuitively, the same pattern. The variability of the reported variables is quite substantial in all groups. With respect to the size of the groups, the individuals ever being officially included into the IRCCS research staff list starting from 2018, and not belonging to the non-academic MD category, are 194 individuals (approximately 33.5% of the sample), of whom a less than half but still relevant subset (66 individuals) are non-academic physicians (thus eligible for MBO as well). The non-academic non-IRCCS physicians are instead 177 (30.6% of the sample), while the pure control group is made by 162 individuals (28%).

Graphs (a) and (b) in Figure 1 plots the average number of publications per physician counted at the end of each year, from 2012 to 2022. We observe a relatively flat trend in the years up until 2017, followed by a noticeable increase starting around 2018. This timing aligns with the introduction the IRCCS recognition, suggesting, at a first glance, a possible aggregate effect of these policies on the overall research productivity at the hospital, mediating by the supposed lagged effect that would involve the execution of a scientific work before achieving publication. It must be noted that the time-lapse between submission to acceptance/publication are usually quite short in the medical field compared to other fields, like the economic one. Whereas to publish in the best ranked economic journals the lag has been longer than 2 years for decades, even reaching more than a 40 months-lag from submission to publications for papers ranked in the upper quantiles of the distribution (Yohe, 1980, Hadavand et al., 2024), such time-lapse is on average 8 months or in the medical field, lowering to even a coupe-of-months time lengths for systematic reviews or literature reports (T.-A. Chen et al., 2024). In addition, a top-journal in the medical field like *JAMA Network Open*, states in its official address to authors that the median time from submission to publication is 94 days, i.e. 3 months (JAMA, 2025).



a) Average publications per researcher per year (2012–2022).



b) Total number of publications attributed to the employees of the institution for the balanced sample (2012–2022).

Figure 1: Research productivity: average trend and total output (2012–2022).

On the other side, two 2017-published papers with two number of citations each as in October 2024 (if using the WoS citation metrics), are "Reverse Time-Dependent Effect Of Alphafetoprotein And Disease Control On Survival Of Patients With Barcelona Clinic Liver Cancer Stage C Hepatocellular Carcinoma" (Ponziani et al., 2017, *World Journal of Hepatology*, first received in August 2017 and published in December 2017) and "'The chromosome analysis of the miscarriage tissue. Miscarried embryo/fetal crown rump length (CRL) measurement: A practical use" (D'Ippolito et al., 2017, *PLos One*, first received in March 2017 and published in June 2017). They display similar if not even shorter publication lags. In addition to that, it must be noted that the effect observed in a timely fashion with respect to the two policy implementations may reflect the fact that some *pipeline projects*, already on-going before 2017, may have been strategically accelerated or finalized in order to obtain the bonus or enabled by the IRCCS additional funds.

While other secular trends could also contribute to the increase (e.g., growing institutional emphasis on research), the positive, diverging pattern beginning in 2018 is quite suggestive. In the analysis below, we exploit the individual-level variation in exposure to the policies to identify their causal impact. In defining treatment and control groups for the analysis, we use the categorizations described in Section 2. For the MBO policy, the "treated" group consists of non-academic physicians (those subject to the MBO bonus) and the control group consists of those not eligible (academic physicians and other staff). For the IRCCS policy, the treated group consists of those included in the IRCCS researcher list (primarily academic physicians) and the control group is those not on the list. Additionally, we will examine subgroups such as the double-treated individuals. The next sections outline our empirical strategies in detail.

A sketch of the average evolution of publication patterns across groups is presented in Figure 2. A first visual inspection seems to suggest us that the bulk in the jump of yearly publications is led by researchers affected by the IRCCS recognition, with apparently no impact brought about by the performance-based policy, as the only MBO-subject individuals who evidently display a trigger in productivity overall, are the ones who are double-treated.

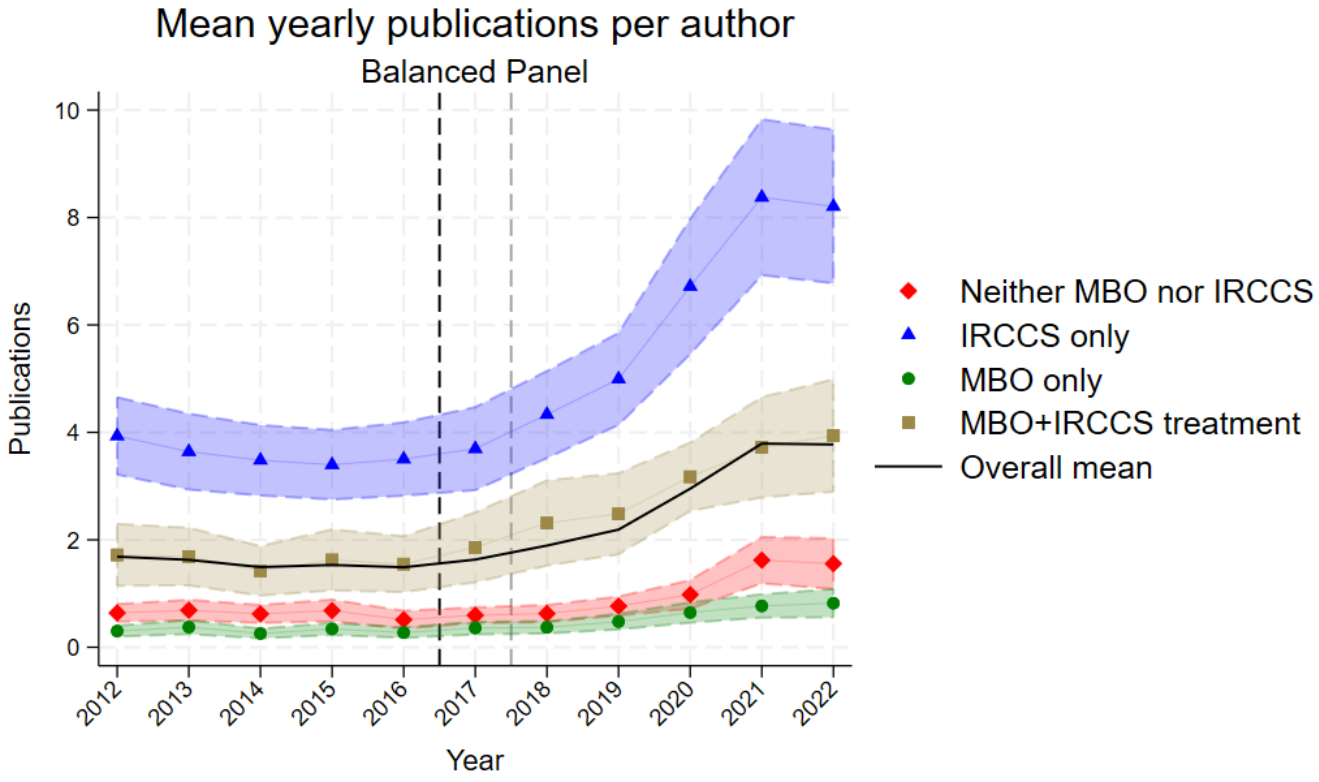


Figure 2: Average yearly publications across different groups of healthcare professionals, defined according their treatment status after 2017 and 2018.

4 MBO Policy

The first policy we analyze is the performance-based monetary incentive (MBO) implemented in 2017. The goal is to estimate the causal effect of this incentive on research output (publications) of the targeted physicians. We employ a TWFE difference-in-differences (DiD) framework, by comparing the change in publications across different groups of individuals within the same model specification, as individuated by a binary indicator, before and after 2017 and according to their appurtenance to the policy treatment. In the baseline specification, we include the group corresponding to 0 and 1 (pure control, and never MBO-treated researchers who become part of the IRCCS perimeter in 2018) as the control group, while groups 2 and 3 are the treated groups (MBO-treated only, double treated after 2018). Then, we repeat the DiD analysis with a the dummy variable assignment varying across definitions, by sub-setting the dataset each time in order to compare different groups. We explain this identification strategy better in the next subsection.

4.1 Empirical Strategy

Our first baseline specification for the MBO policy is a standard static specification for a two-way fixed effects DiD model, estimated via OLS:

$$Y_{it} = \alpha + \beta Post2017_t \times MBO_i + \delta_i + \tau_t + \varepsilon_{it} \tag{1}$$

where Y_{it} is the research output of individual i in year t (measured as the number of publications), $Post2017_t$ is an indicator for the post-intervention period (2017 and later), MBO_i is a binary indicator for the assignment of the MBO-policy treatment. In the baseline, MBO_i is expressed by referencing to Table 1, with $MBO_i = 0$ if $Group_{MBO} = \{0, 1\} = \{\text{Full control, IRCCS-only}\}$, and $MBO_i = 1$ if $Group_{MBO} = \{2, 3\} = \{\text{MBO-only, Double-treated}\}$. Such indicator allows us to compare the effect on the categories of units who received the performance incentive (all non-academic doctors) with respect to the baseline category, as in those who never were eligible for the management-based bonus (faculty members and other non-physician professionals). δ_i are individual fixed effects, τ_t are year fixed effects. The coefficient β, on the interaction terms $Post2017_t \times MBO_i$ capture the DiD estimates of the effect of the MBO incentive on annual publications. By including δ_i, we control for the time-invariant differences between physicians, while τ_t for common shocks in each period. Standard errors are clustered at the individual level. The key identification assumption is that, absent the MBO policy, the research output of the treated and control groups would have followed parallel trends. While such assumption is not directly testable, we present event-study evidence to assess whether the different groups exhibited similar trends prior to 2017. However, as many academic doctors were, around the same time, exposed to the IRCCS intervention in 2018, we recover that in our baseline DiD, their post-2017 effect might partly reflect anticipation or effects caused by the IRCCS recognition. We address this potential overlap in two ways: first, by reporting dynamic estimates for the event-study, which also displays the immediate effect of the treatment before the IRCCS implementation of 2018. Second, by explicitly accounting for IRCCS in a triple-difference model (presented in Section 4.3). In addition, we implement a more granular approach that exploits the multiple treatment groups defined earlier. While in the Equation (1), the MBO eligibility was dealt without accounting for the two compound treatments interacting with each other (MBO and IRCCS), in a second set of OLS estimations, we perform various binary DiD comparisons between specific groups to directly assess the magnitude of treatment effects. Such estimations are structured in order to follow Equation 1, with slight variations; as in, the interaction of

interest involves yet the binary dummy MBO_i, assuming the value 0 or 1 depending on the appurtenance of unit i to the designed (and varying) treatment (1) or control (0) group. Specifically, the estimation is performed several times with different compared units, by re-classifying the treatment and control groups in five different ways: (1) MBO_0: double-treated (MBO+IRCCS) vs. MBO-only; (2) MBO_1: double-treated vs. IRCCS-only; (3) MBO_2: double-treated vs. neither (full control); (4) MBO_3: MBO-only vs. IRCCS-only; and (5) MBO_4: MBO-only vs. full control. For each comparison, we restrict the sample to the units involved in that very comparison only. This provides insight into whether the MBO effect was larger than the IRCCS effect by comparing the different categories. We interpret these results alongside the main specification.

To corroborate our causal claim, as anticipated, we estimate event-study models which allow for the treatment effect to vary in each year relative to the policy change, by substituting the $Post2017 \times MBO_i$ with a series of interactions of MBO_i with time-dummies for each year before and after 2017 (by excluding the one immediately before the treatment year to avoid multicollinearity issues). We do this by estimating the following equation:

$$Y_{it} = \alpha + \sum_{h=2}^{H} \beta_H (Lag\ h)_{it} + \sum_{g=1}^{G} \beta_g (Lead\ g)_{it} + \delta_i + \tau_t + \varepsilon_{it} \tag{2}$$

The reference period being $t := 2017$. $(Lag\ h)_{it}$ are the pre-treatment period dummies associated to the treated units $(Lag\ h)_{it} = \mathbb{1}[t = 2017 - h]\ for\ h \in \{2, ..., H\}$. The post- period dummies interacted with the treatment binary variable are instead indicated by $(Lead\ g)_{it}$ $(Lead\ g)_{it} = \mathbb{1}[t = 2017 + g]\ for\ g \in \{1, ..., K\}$. This yields coefficients for the differential trend between treated and control groups in each year, which we plot to visually inspect pre-trends and dynamic effects. We present these event-study results as figures accompanying the main table.

4.2 Results

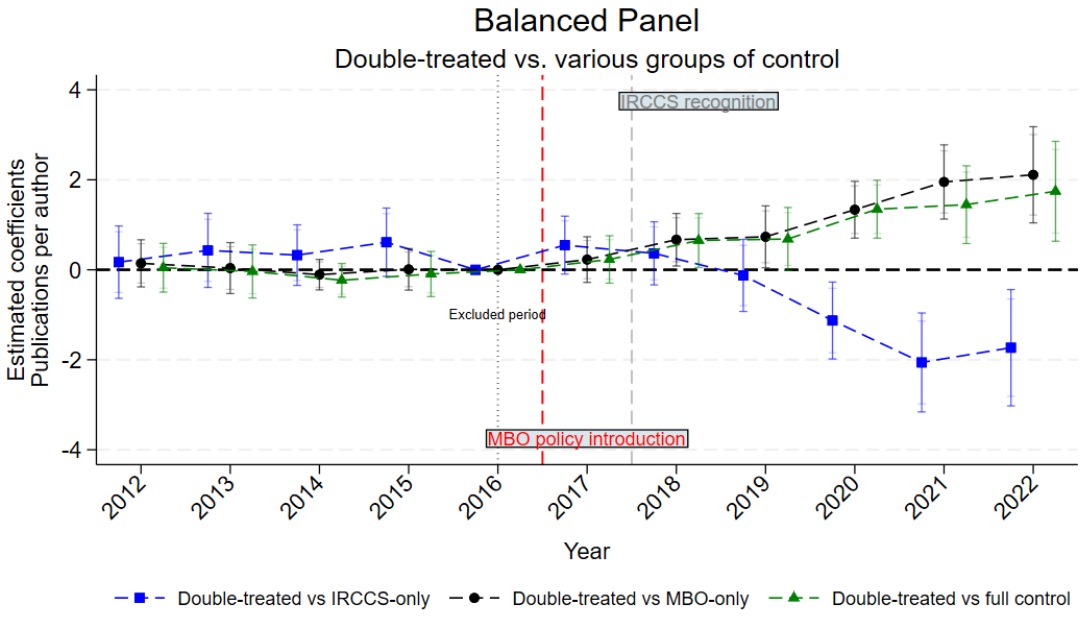
Table 3 presents the difference-in-differences estimates of the effect of the MBO monetary incentive on publications. The table includes different specifications, according to the different strategies described above (Cols. (1-6)), with the outcome staying fixed (first row); e.g., the raw number of individual yearly publications recovered from Web Of Science. Overall, we observe a positive and statistically significant effect of the MBO policy only when the recipients are also part of the IRCCS perimeter (double-treated, i.e. DT), and only when they are compared with similar MBO-subject units or with the pure control. The baseline DiD estimate (comparing MBO-eligible to non-eligible physicians, in Col. (1)) indicates

that after 2017, non-academic physicinas subject to the bonus displayed a decrease in their annual publications by about 150% of the pre-treatment mean, compared to academic physicians and other professionals. We cannot however deem such decrease as a negative effect of the MBO, as the estimate shall be offset by the compound impact of the IRCCS recognition. On the other side, MBO-eligible MDs who were never part of the IRCCS perimeter (Col. (6)) did not experience any significant effect after the policy, compared to non-eligible, non-IRCCS individuals (PC). MBO-physicians who were also involved in the IRCCS policy increase their publications by 75% circa of the pre-treatment average, compared to PC. Differentiating across control grouops indeed, we find that the post-2017 jump in the publications of the DT group is positive and significant only with respect to the MBO group (+78%) and the pure control (+72%), turning negative when compared to the IRCCS researchers (−73%).

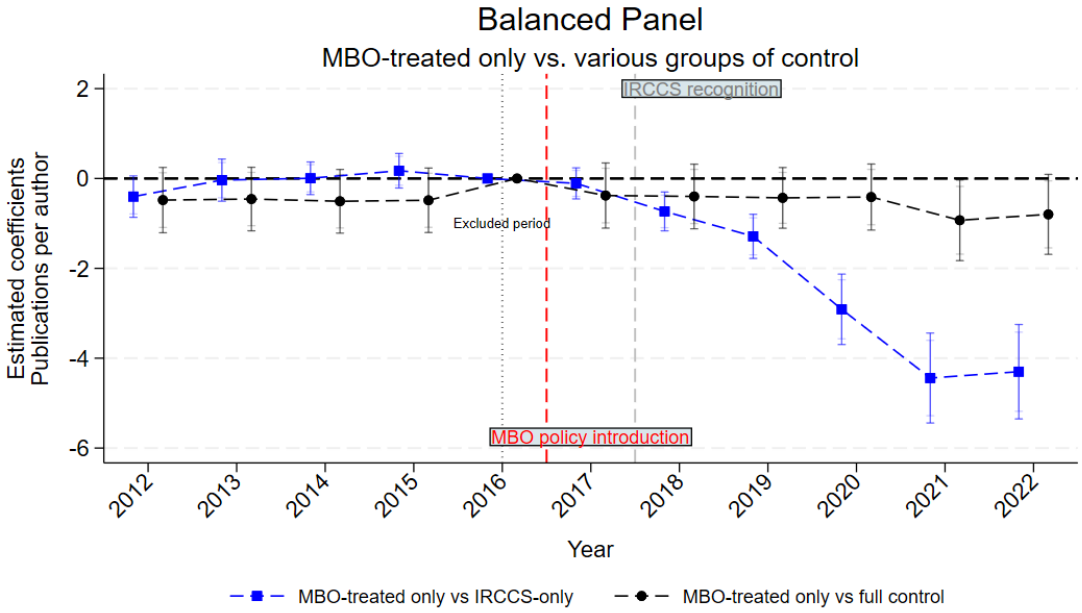
The DT group seems, quite counterintuitively, performing better compared to the MBO units than the control; by looking at the clustered standard errors though, the difference between the two coefficients does not appear to be statistically significant. By contrast, MBO substantially underperform compared to the $IRCCS$-group (Column 5, with a striking −880%, which extreme relative magnitude is possibly driven by the difference in row numbers' scales between the two groups of physicians), while displaying no relevant differential in comparison to the pure controls.

	(1) MBO vs. non-MBO	(2) DT vs IRCCS	(3) DT vs MBO	(4) DT vs control	(5) MBO vs IRCCS	(6) MBO vs control
Publications	-0.9804***	-1.1782***	1.2552***	1.1521***	-2.6635***	-0.2183
(SE)	(0.1742)	(0.3433)	(0.2317)	(0.2380)	(0.3142)	(0.1332)
N	6369	2695	2600	2418	3944	3674
R^2	0.763	0.764	0.597	0.580	0.794	0.438
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Mean	0.65	1.60	1.60	1.60	0.31	0.31
Panel	Full	DT and IRCCS	DT and MBO	DT and control	MBO and IRCCS	MBO and control
Time Range	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022

Table 3: Impact of MBO Incentive on Annual Publications (Difference-in-Differences).



a) Event-study of MBO policy effect on publications of the double-treated units compared to the various control groups. The persistent decline is sensibly driven by the confounding effect of the subsequent IRCCS recognition.



b) Event-study of MBO policy effect on publications of the MBO-only treated units compared to the various control groups. The persistent decline is sensibly driven by the confounding effect of the subsequent IRCCS recognition.

Figure 3: Event-study of MBO policy effect on publications for comparisons across different groups.

Henceforth we pass to the dynamic setting. We omit the event-study for the non-clear cut comparison which involves the analysis of MBO-treated vs the rest, as it involves a spurious effect lead by the compound effect of the IRCCS treatment. Therefore, we estimated a set of different dynamic models (referencing to Equation 2) which compare first the double-treated units (Graph (a) of Figure 3), and then the *MBO*-only individuals (Graph (b) of Figure 3), to the various control groups. While both graphs show the validation of the common trend assumption for all the groups of comparison, the evidence that the MBO policy affects individuals only joint with IRCCS funding is corroborated even more, with the negative and significant gap between double-treated individuals and MBO-treated only ones widening significantly over time, starting from 2018 (the IRCCS recognition year, which is the year when the effect starts kicking in for the double-treated units), and stabilizing one year after the pandemic outbreak.

4.3 Triple-Difference Analysis

Through the results above we attempted at disentangling the effect of the MBO policy under reasonable assumptions and across different groups. However, the almost simultaneous recognition of the hospital as IRCCS in 2018 raises concerns of compound treatment: hence, we also try to address the issue of interaction between the two policies. We therefore estimate a difference-in-differences-in-differences (DiDiD) model which embeds both policies.

$$\begin{aligned} Y_{it} = \alpha + \beta_1 Post2017_t + \beta_2 IRCCS_i + \beta_3(Post2017_t \times MBO_i) + \\ + \beta_4(Post2017_t \times IRCCS_i) + \beta_5(Post2017_t \times MBO_i \times IRCCS_i) + X'_{it}\gamma + \varepsilon_{it}, \end{aligned} \tag{3}$$

In Equation 3 $IRCCS_i$ is a binary indicator assuming value 1 if i would ever be part of the IRCCS staff at a certain point after 2018, to which the associated coefficient to is β_2. By contrast, β_3 captures the effect of the MBO policy on those neither in the IRCCS staff (double-treated include) nor the pure control, as it interacts the $Post2017_t$ time dummy with the binary for being subject to the MBO. β_4 is the coefficient linked to the effect of the IRCCS recognition on those not subject to MBO nor part of the control group, starting from when the MBO is implemented (as the dummy $IRCCS_i$ is interacted with $Post2017_t$). Eventually, the triple interaction coefficient β_5 bears major relevance: it recovers the additional effect, starting from 2017, for individuals who are considered as double-treated (MBO-eligible and included in the IRCCS perimeter after); it hence helps us to disentangle the compound impact of monetary incentives and public funding on the affected individuals. Note that while $Post2017_t$ assumes

unitary value in any year after 2017 included, the IRCCS change kicks in only in 2018. Not to avoid the rising of collinearity issues, we avoid including two way fixed effects, which we account for via a set of time-varying controls: X_{it} (age, age squared and dummies for gender, age cohort, birth province, and department) aims at accounting for differential, individual trends correlated with observables. The triple-difference strategy relies on the assumption that, absent the policies, the difference in outcomes between, say, non-academic and academic doctors would have evolved similarly over time for IRCCS-listed and non-listed groups. We also estimate further specifications, including some of the interactions with or without controls, including two way fixed effects when covariates are ruled out. Eventually, we integrate this set of estimates with an additional "complete" and covariate-less triple DiD, where TWFE are included at the expense of the terms $Post2017$ and $IRCCS_i$, necessarily omitted to avoid collinearity. This approach could appear a bit complex; however, provided with the parallel trend checks we performed in the previous sub-section, we are somewhat able to assume that no other unobserved shocks differentially affected the subgroups' trends.

Findings are reported in Table 4, with the most complete specifications in Cols. (7) and (8). Differently from Table 3, here the outcome is reported in the columns, and the relevant interactions in the first term of the Tables' rows. Column (7) excludes time-varying controls, the $Post2017_t$ and the $IRCCS_i$ terms, but includes units and year fixed effects, while Column (8) uses individual controls but omits the fixed effects due to including the treatment dummies mentioned above. First, we observe how, in Col. (8), there is no significant effect after the policy implementation within the pure control group. Then, the second row of the same column show that the individuals affiliated to the IRCCS (both the IRCCS-researchers only and the compounded treated units) have, on average, higher publications than the other two groups (pure control and MBO-only eligible doctors), across the whole sample (thus even before the IRCCS introduction). However, it must be noted that this framework does not account at all for unobservable heterogeneities, which may not be tackled effectively by conditioning on our set of time-varying regressors. Moving on, the coefficient on $Post2017 \times MBO$ results negative and not statistically differing from 0 in both models, confirming our previous hypothesis that MBO-eligible physicians did not increase their publication activity after the performance-based reform, absent a following affiliation with the IRCCS staff. On the other hand, both columns display that the estimate of $Post2017 \times IRCCS$ is large, strongly significant and comparable across both specifications. This validates the intuition that the IRCCS recognition had a substantial and positive effect on publication activity for the IRCCS-staff, which in such coefficient includes also the double-treated physicians.

2012-2022; SEs clustered at individual level

	(1) Publications	(2) Publications	(3) Publications	(4) Publications	(5) Publications	(6) Publications	(7) Publications	(8) Publications
Post2017								-0.2676
								[0.2526]
IRCCS								2.9667 ***
								[0.3796]
Post2017 × MBO	-0.8874 ***			-0.9804 ***			-0.1609	-0.0312
	[0.2865]			[0.1742]			[0.1295]	[0.3271]
Post2017 × IRCCS							2.1005 ***	1.7957 ***
							[0.3014]	[0.3725]
MBO × IRCCS		0.6178			-0.9683 *		0.2334	-1.8149 ***
		[0.3955]			[0.5746]		[0.6752]	[0.4500]
Post2017 × MBO × IRCCS			0.8278 **			0.2456	-0.9252 **	-0.7509
			[0.3741]			[0.2277]	[0.3810]	[0.4633]
Observations	6,228	6,228	6,228	6,369	6,369	6,369	6,369	6,228
R-squared	0.3422	0.3394	0.3397	0.7626	0.7600	0.7599	0.7716	0.4177
Individual FE	NO	NO	NO	YES	YES	YES	YES	NO
Year FE	NO	NO	NO	YES	YES	YES	YES	NO
Covariates	YES	YES	YES	NO	NO	NO	NO	YES
Method	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Time Range	2012-2022	2012-2022	2012-2022	2012-2022	2012-2022	2012-2022	2012-2022	2012-2022
Panel	Full	Full	Full	Full	Full	Full	Full	Full
Mean	2.187	2.187	2.187	2.187	2.187	2.187	2.187	2.187

*** p<0.01, ** p<0.05, * p<0.1

Table 4: Triple Diff-in-diff, comparing different groups.

We observe then how the interaction, which is the time-invariant effect of being a non-academic Medical Doctor (thus eligible for the MBO bonus) and a IRCCS-staff member at the same time, has apparently a negative effect on scientific productivity with respect to the other groups (pure control and especially IRCCS-only researchers). However, such effect is huge and significant only when accounting for time-varying controls, which however are not able to capture time-invariant unobservable characteristics, nor average time trends. When including unit and time Fixed Effects however, the effect reverts its sign, although getting non-significant, possibly suggesting that the coefficient gets its relevance mostly from unobservable heterogeneities across individuals. Eventually, the triple interaction term *Post2017*×*MBO*×*IRCCS*, which captures the differential effect for individuals simultaneously exposed to both policies (double-treated), is negative in both complete estimations. While they are somehow comparable across the two specifications, the term appears significant only in the TWFE model, therefore highlighting the potential presence of either colliding bias in the included covariates, or the ineffectiveness of capturing time-varying heterogeneities of the fixed effect. However, given the absence of any sort of controls for general time trend in the last model, we are keen on tending to-

wards the former hypothesis. Although the coefficient does not reach conventional significance levels, its negative sign suggests that the effect of IRCCS on research output may be attenuated when combined with MBO, potentially due to the reference group which works as comparison in these regards. To sum up, our findings confirm that a positive shift in research productivity is strongly linked to the accreditation of the hospital to receive IRCCS funding, while highlighting the weakness of implementing the management-by-objective scheme alone. The absence of an additional positive interaction effect in the double-treated group suggests that the joint implementation of MBO and IRCCS does not yield additive benefits; however, this statement shall be taken into account carefully so far, as the effect is with respect to a control group composed by all the groups included in the framework (and not only a selected one). In any case, it shows that what matters the most is the funding policy.

The most important thing to underline after presenting such estimates is noting that, in the DiDiD model, the post-2017 indicator picks up any overall time shock after 2017 which is common to all groups, hence including, certainly, the *Post*2018 IRCCS recognition. While the inclusion of year fixed effects τ_t rules out common shocks in Col. (7), it does not the same in the final specification. On top of that, even accounting for that, this does not allow to separate the effect heterogenously across individuals, especially due to the fact that the IRCCS recognition occurs one year immediately after the MBO policy. However, to further validate the estimates from the latter framework, we perform a robustness check by sub-setting the sample only between 2016 and 2018, in order to rule out the bulk of the IRCCS' recognition impact, which only confounds the estimates to the extent through which they are triggered immediately. Table A3 in the Appendix shows the results of such analysis, where we can observe how such results are way lower in magnitude and some of them gets non-significant, although the directions of the effects stay the same.

5 IRCCS Funding Analysis

Given the almost contemporaneous overlapping of the two policies, since the second one is introduced just one year after the first one, we cannot really disentangle the effectiveness of either discontinuity without focusing on the second one in details as well. So we now turn to the IRCCS recognition, which triggered an allocation of greater funds for research, and which began in 2018. Such event mainly affected the hospital's academic-affiliated researchers, allowing them to receive dedicated public resources and, in addition, by enhancing the prestige of the institution under the scientific point of view. We use a

difference-in-differences approach to estimate the effect on academic physicians' research output, in doing this exploiting a parallel identification strategy with respect to the assessment of the performance-based bonus.

5.1 Empirical Strategy

Mirroring what we did for the MBO-framework, to identify the IRCCS recognition's effect, We estimate a baseline specification for a stadard two-way fixed effects model, via OLS estimators:

$$Y_{it} = \alpha + \theta Post2018_t \times IRCCS_i + \delta_i + \tau_t + \varepsilon_{it}, \tag{4}$$

where $Post2018_t$ equals 1 if $t \geq 2018$ and 0 othrwise, $IRCCS_i$ indicates if individual i is in the IRCCS perimeter (i.e., directly eligible for IRCCS funding); in the baseline, we consider double-treated individuals as control units, in order to have a lower bound of the estimated effect of the policy. Beyond the double-treated, the baseline control group also includes the other MBO-treated individuals and the pure control group. In such regards, we renovate the aforementioned clarification: while the hospital updates the IRCCS perimeter at the end of every year, we still perform a standard 2×2 TWFE DiD estimation, by considering as treated those who are *ever*, sooner or later, included in the IRCCS perimeter. This is done in order to allow a more clear-cut identification of the effect, although by doing this our causal quantity of interest consists in an ITT rather than an ATT. As a matter of fact, even though solely IRCCS-only individuals should be formally granted access to IRCCS funding, a structured research group, within the personnel working at the institution, takes shape immediately after the recognition of the scientific excellence standards of the hospital under study. Such crowd can be accounted as the set of physicians who actively take part to most IRCCS research activities, and collaborate in order to obtain the privileged funding which is granted to the hospital due to its IRCCS status. In any case, the perimeter is not re-updated yearly from scratch, but there exists a consistent core of researchers which stay constant over time, which is integrated yearly by those fellows who are "*de facto*" IRCCS collaborators to the same funded projects and thus have access to the same resources, even though their status is officially recognized only after few periods; they basically self-select into the treatment after the IRCCS recognition. This notwithstanding, the inclusion into the IRCCS perimeter is always established by the institution (top-down). In addition to that, reflecting the same settings of the MBO policy, the treatment is assumed as an absorbing one. However, and differently from the MBO-policy

which is absorbing by construction, the inclusion in the IRCCS perimeter is not constant until the end of the time-span, as some individuals apparently "switch off" the treatment since are excluded from the list by the management at some point. There is no reason not to maintain the absorbing treatment assumption nonetheless, because while papers are written in a relatively short time frame, the exploitation of granted funds can be actually prolonged in time, so researchers who started projects (or were involved in some) would feasibly keep working to such projects by using the same means even after their IRCCS status has expired. To check for these issues, albeit holding the absorbing assumption as always true, besides the standard test for parallel trends, we also perform additional estimates accounting for the staggered inclusion of some researchers to the core of the IRCCS perimeter, and controlling for the "switchers" across groups.

The coefficient θ captures the average intention-to-treat effect on the IRCCS recognition treatment of those who only receive funding after 2018, with respect to MBO-eligible individuals and the full control group. In such set-up we compare IRCCS-only listed resarchers to non-listed individuals and listed-but-non academic individuals, before *and* after 2018. As before, the main identification assumption is that, without the IRCCS acknowledgment, the scientific productivity across groups would have followed parallel trends in the post- period. While TWFE account for overall temporal trends and time-invariant heterogeneities, we test the assumption through event-study estimations, as we did for the MBO policy. Since in such framework the issue of self-selection is major compared to the MBO-policy setting (as the institution decides whom to include in the perimeter possibly depending on productivity), we also provide the main event-study estimates with additional validity checks which exploit the recent methodology of Rambachan and Roth, 2023. Their so-called *honestDiD* methodology allow for estimates with potential mild violations of the parallel trend assumption (which are plausible in case of selection), by computing confidence sets granting the robustness and significance of estimated dynamic coefficients in the period after the treatment *even with deviation from the common trends*, up to a certain point which is the threshold-level computed by the methodological algorithm itself; more on this later in the paper.

As for the MBO analysis, we implement further sets of estimates. In 2018, the four identified categories (0, 1, 2, 3) are defined slightly differently (0 = control, 1 = MBO-only, 2 = IRCCS-only, 3 = both, as reported already in Table 1). After reporting the baseline, we incorporate the estimates with a methodology to assess the differential effects across different groups. This second strategy entails the following comparisons: (1) double-treated vs. IRCCS-only; (2) double-treated vs. MBO-only; (3) double-treated vs. pure control; (4) IRCCS-only vs. MBO-only; (5) IRCCS-only vs. pure control.

Such comparisons are obtained by estimating a set of equations constructed as follows, by changing the comparison groups for every regression defined in Equation 5.

After this, we compute a set of dynamic coefficients for the IRCCS intervention by means of event-studies for the latter set of models, interacting $IRCCS_i$ with year dummies around 2018, according to the compared groups involved. The baseline comparison does not really reflect Equation 2. As a matter of fact, while in the baseline estimates of Equation 2 the dummy MBO_i takes unitary values if the individual is treated with the incentive policy, irrespective of their double-treatment status (so also IRCCS-individuals are included in the treatment), in (the one which follows) the approach is more "conservative", as the binary variable takes value 1 only if i is included in the IRCCS perimeter *only*, without being subject to the MBO-policy as well (as in, she is an IRCCS-listed academic doctor); the control group consists in the set of groups 3, 1, and 0 combined. Such equation enables the visualization of potential diverging pre-2018 patterns. Here is the estimated equation (via OLS):

$$Y_{it} = \alpha + \sum_{h=1 \ \wedge \ h\neq 2}^{H} \theta_H (Lag\ h)_{it} + \sum_{g=1}^{G} \theta_g (Lead\ g)_{it} + \delta_i + \tau_t + \varepsilon_{it} \tag{5}$$

Note that, while it is common practice to exclude the time immediately *before* the treatment to avoid multi-collinearity issues (as in, $h = 1$), in the latter estimates we opt to exclude $h = 2$ instead (which is $t = 2016$), which is the same excluded period in the event-studies for the MBO-policy treatment. We acted like this in order to allow greater comparability between the event-studies for the different policies.

5.2 Results

We report the DiD estimates from Equation 4 in Table 5. The estimates display a positive and significant impact, confirming the pattern which we already observed previously, showing the importance of IRCCS recognition compared to the performance-based policy. Col. (1) report the coefficient on the interaction $Post2018_t \times IRCCS_i$ described in Equation 4. The other columns display instead the coefficients for the single interaction terms for different groups' comparisons as indicated in the columns themselves (DT vs. MBO, DT vs. IRCCS etc.). The estimated coefficient in Col. (1) implies that, on average, being listed as an IRCCS researcher-only (IRCCS) is linked with a growth of almost 2.25 additional publications per year (significant at 1%) compared to the other groups. Such increase is quite outstanding, as it accounts for more than 73% of the average mean of individuals control units and IRCCS-listed non-academic physicians. When breaking down the effect by category, results are consistent. The units treated compoundedly publish almost 1.26 more works compared to the MBO-treated only (+76%), 1.18 fewer

works compared to the IRCCS-only researchers (−72%), and 1.15 more works compared to the pure control (+70%). Such findings are consistent with the stylized fact that IRCCS contributed substantially to the output of both academics and non-academic MBO-treated MDs who had the opportunity to be subject to both policies, although the researchers involved in scientific activities primarily are the ones who responded the most after the recognition, as they display higher publication rates even when compared to those who are selected into the IRCCS list and allowed to get the performance bonus. We observe this in the final columns of Table 5: in Col. (5), the effect of IRCCS recognition on IRCCS vs. MBO is almost +2.7 publications (+74%); in Col. (6), IRCCS publish almost 2.6 works more than the pure control (+71%).

	(1) IRCCS vs. others	(2) DT vs MBO	(3) DT vs IRCCS	(4) DT vs PC	(5) IRCCS vs MBO	(6) IRCCS vs PC
Publications	2.3562***	1.2552***	-1.1782***	1.1521***	2.6635***	2.5532***
(SE)	(0.2973)	(0.2317)	(0.3433)	(0.2380)	(0.3142)	(0.3199)
N	6369	2600	2695	2418	3944	3762
R^2	0.774	0.597	0.764	0.580	0.794	0.784
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Mean	3.61	1.64	1.64	1.64	3.61	3.61
Panel	Full	DT and MBO	DT and IRCCS	DT and control	IRCCS and MBO	IRCCS and control
Time Range	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022

Table 5: Impact of IRCCS recognition on Annual Publications (Difference-in-Differences).

The baseline dynamic specification for the IRCCS effect is represented in the event-study in Figure 4. As above mentioned, the definition of treated unit here is quite conservative, as we are only comparing academic doctors who receive IRCCS-funding only with the rest of the personnel, included those who receive the performance payment and have access to IRCCS-funding too. Prior to 2018, the raw number of publications for the IRCCS-only group were relatively flat and non-diverging with respect to the rest of the units. While their scientific output was way higher on average (since they were selected for being productive, see Table 2), we notice that, in terms of time-varying pattern, there is no significant

difference from the slope of the publication output of non-treated ones until 2018, thus highlighting the relevant impact of the access to the IRCCS funding and the absence of evident pre-trends. In the graph, we observe a small and slightly significant jump in 2018, reasonably due to the, albeit short, lag in the publishing process. The coefficient keeps increasing in 2019 and displays a greater-than-100% jump in 2020, which could be credited to both the passing of time which reduces the impact of the afore mentioned lag, and to the boost in research activity triggered by the COVID outbreak. The effect seems to stabilize in 2022, after a further substantial growth in 2021. It appears that the IRCCS recognition's impact had already materialized few lags after its implementation, and researchers had possibly already optimized the resources in order to produce output in a timely fashion: so not only new projects have been ushered, but some of them have been completed.

As for the MBO dynamic specification likewise, we estimated dynamic coefficients based on an identification which was quite conservative, but not really clear-cut: indeed, some IRCCS' researchers included in the control group were non-academic, hence they were enjoying the performance-based bonus. On top of that, although we already showed that the MBO-policy effect was quite neutral, it surely contributed to bring about some bias in the baseline estimates, especially for the double-treated individuals. Therefore, we performed a number of diverse dynamic models' estimation again, in doing this showing the graphic representations of the estimates foreseen by Equations 5, in Figure 5. In the latter, Graph a) reports the comparisons between double-treated units and the other groups, while Graph b) the comparison between IRCCS-only researchers and the other individuals. Both graphs keep validating the hypothesis that that the MBO policy does not really bias the estimates as its impact is almost non-existent; again, it slightly affects the research outputs only by shifting up the pattern of publications for the double-treated units; however, such pattern is more than offset by the IRCCS funding granted to academic researchers, whose trends follow along both in comparison with the pure controls and the MBO-treated only individuals.

It is however evident how the compound effect of the performance-based policy and the IRCCS access affects the dynamic comparison of IRCCS-listed academic doctors and the double treated. Although non-significant, we observe an immediate shift downwards in the publication output of IRCCS-only treated in comparison to the compounded ones in 2017, hence immediately after the MBO implementation (c, Graph b)). Anyway, such difference starts reverting back already the year after, with the IRCCS recognition occurrance, after which the IRCCS-only physicians start regaining their comparative advantage with respect to non-academic doctor, with their outcome boosted by the access to new

resources.

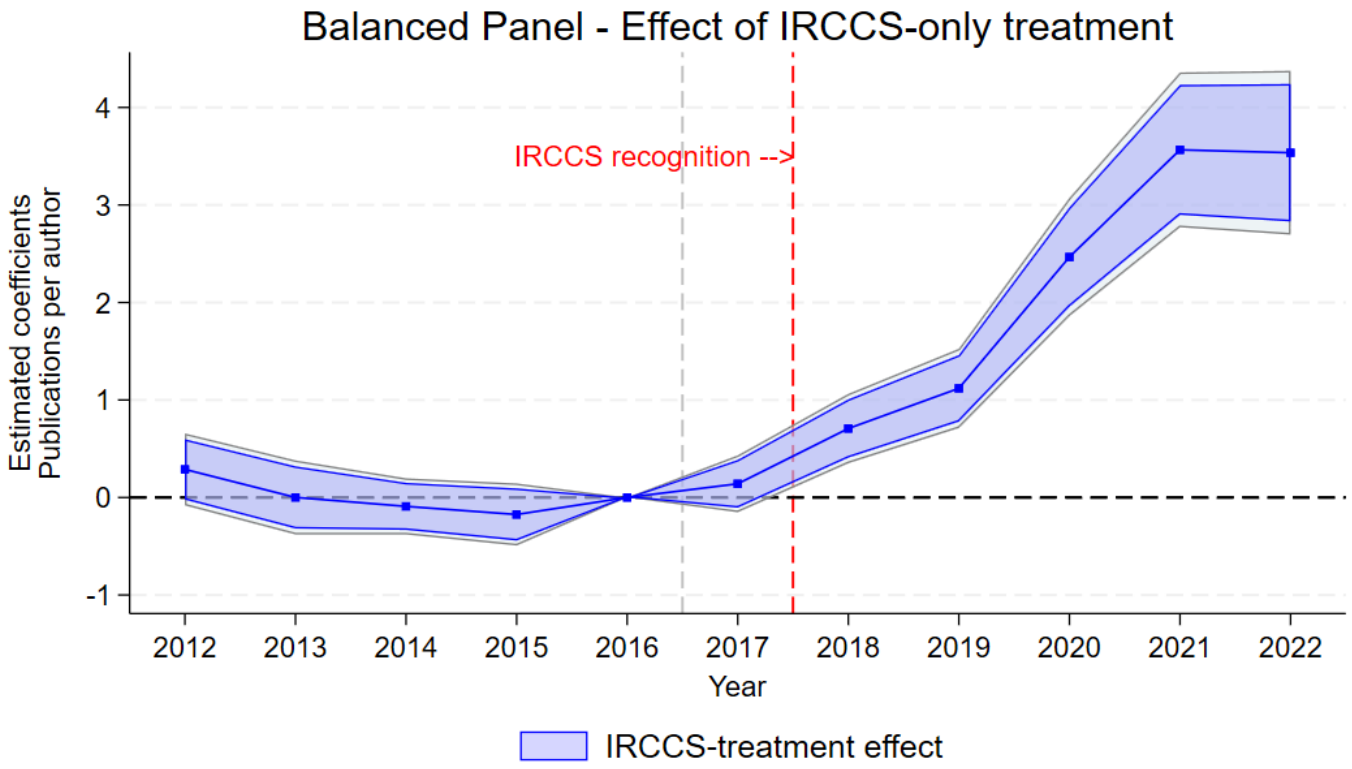
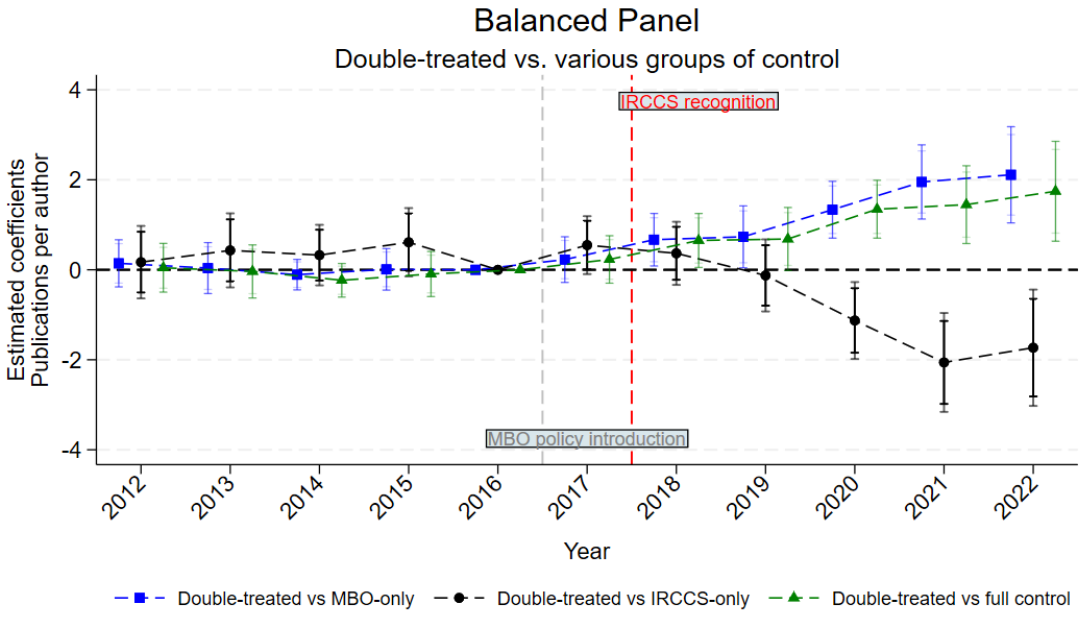
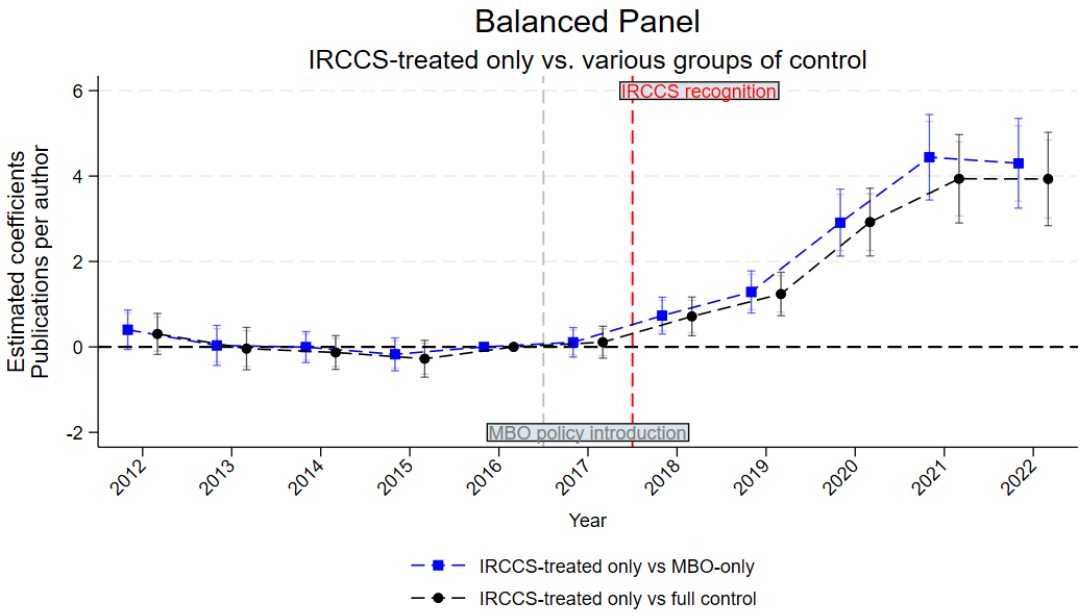


Figure 4: Event-study of IRCCS recognition effect on publications.

Overall, the IRCCS funding intervention appears to have had a substantial, almost immediate and persistent effect on research output, fed up by the pandemic and reinforced, only for some individuals, by the performance-based scheme. The estimates for the IRCCS-only group suggest that the heterogeneity in incentive structures across the different groups can only modestly be offset by straightly rewarding productivity in the short term, if such reward is not paired with additional resources. It must also be noted that additional resources already impact positively those who are selected into the perimeter of those who can actually access them, who are chosen by the management due to their supposed better performance in research-related activities; hence, such individuals possibly already embed a slightly different incentive structure, if compared to the non-academic doctors for instance. In any case, the presence of academics as well in the pure control groups confirms that the heterogeneity in pre-existing incentive behavioral structure is not only present across groups, but also within them.



a) Event-study of IRCCS recognition effect on publications of the double-treated units compared to the various control groups.



b) Event-study of IRCCS recognition effect on publications of the IRCCS-only treated units compared to the various control groups.

Figure 5: Event-study of IRCCS policy effect on publications for comparisons across different groups.

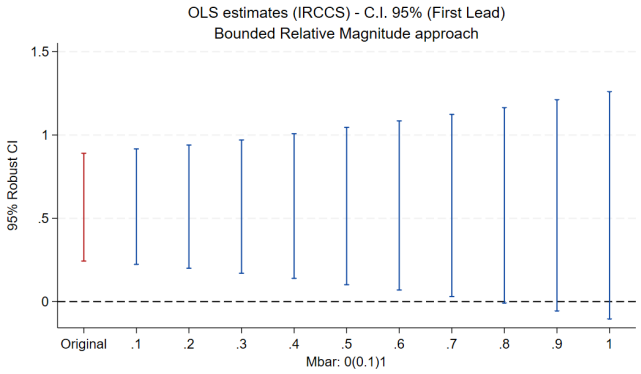
To sum up, it looks like that academic researchers, already incentivized by intrinsic reasons and career motivations, react substantially when provided with more resources, whereas individuals not attracted by the same motifs or similar career progression incentives, show no incremental response to the monetary scheme, or do so only in presence of extra funding. Given the relevance of such results, we next provide with a range of robustness checks to ensure that these findings are not driven by other confounding factors. The first one is included in this very same section, and addresses the concern of potential bias driven by the violation of the parallel trend assumption.

5.2.1 Parallel Trends

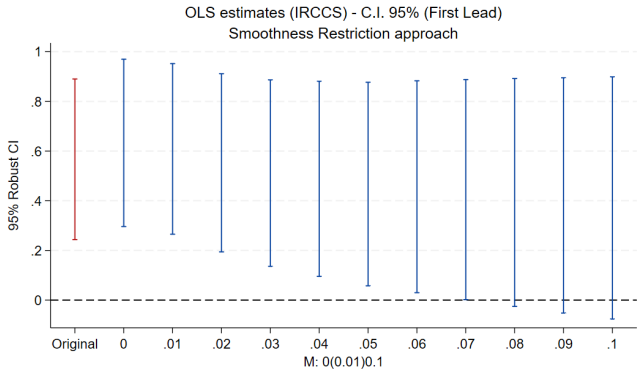
While the main estimates seem to display the absence of deviation from the parallel trends in all the provided specifications, we place additional focus on the pre-trends' issue, given the significance and magnitude of our results. This check is also required due to the fact that the treatment is endogenous and its assignment certainly depends on existing characteristics, which are heterogeneous across groups and units. First of all, notwithstanding the overall absence of statistical significance for the pre-treatment lags interacted with the treatment dummies, we still observe that, especially in our baseline graph in Figure 4, there is a slightly decreasing (albeit non-significant) trend for the IRCCS staff in the periods prior to 2016, which seems to become also statistically significant at 95% in 2015. Although a visual inspection would not assess such deviation as able to undermine the validity of the identification, and even if the trend is falling down (which, if anything, would lead our estimated post-treatment effects to be downward biased), we follow the *honest* methodology by Rambachan and Roth, 2023 to account for plausible deviations from the common trend. In accordance to their approach, we can undertake sensitivity checks even by hypothesizing some significant divergence from parallel trends. We do so by computing *bounds on relative magnitude (BM)*: as in, we replicate their methodology by estimating the entity of the divergence from the parallel patterns after the treatment, up to the size that could be able to invalidate the significance of the post-treatment outcomes, with respect to the assumption of the absence of pre-trends. Such potential *post-* deviation is assumed to have the same size as the one in the period reporting the greatest difference in trends among all of our pre-treatment coefficients (in our baseline case, 2015). We also estimate a second bunch of sensitivity checks, where the assumption regarding the entity of the post-treatment deviation is not made according the maximum size of one of the lag-related differences, but due to a linear extrapolation of the hypothesized pre-trends which would lead the estimated results to be non-consistent (*smoothness restriction - SR*). We provide evidence for

the *honest* confidence bounds for the baseline estimates (Figure 6), using both approaches (BM in graphs a) and c); SR in graphs b) and d)), and by allowing variation in the common trends able to invalidate the significance of the coefficient on the first lead estimation (graphs a) and b)) or the overall average across all post-treatment coefficients (graphs c) and d)). Results are reported, as said before, in Figure 6. In each graph, the y-axis represent the robust 95% confidence interval for the target estimated coefficient. The x-axis, on the other side, represents the varying parameter of interest, *Mbar* in the bounded relative magnitude method and M in the Smoothness one. For the BM approach, it represents the percentage of the size of the post-treatment violation (based on the maximum pre-treatment difference) for which the estimate is allow to be biased before becoming non significant. For the SR approach, the x-axis represent the percentage of the deviation in the slope from the linear extrapolation of the pre-trend divergence (based on the post-treatment evolution) that would make the estimate invalid. In all graphs, the baseline confidence interval (as in, the one with no deviation from common trend assumed) is in correspondence of $Mbar = 0$ (or $M = 0$) and it is indicated by a vertical red C.I. line.

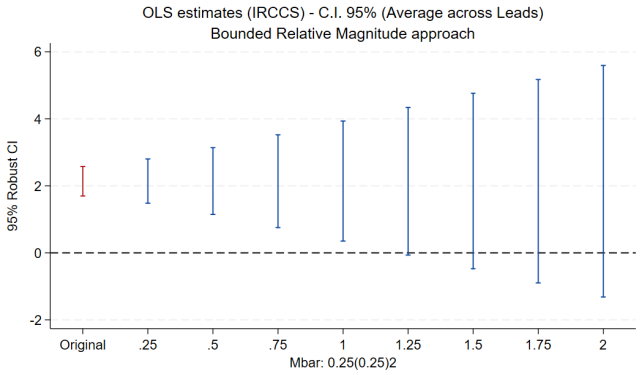
Graphs a) and b) display the robustness confidence set of the estimated coefficient in the first lead for the baseline event-study. According to the BM approach (graph a, the x-axis ranging from 0 to 100%)), the effect maintains its significance up to assuming an actual 80% deviation from the maximum divergence measured in the pre-treatment period. The smoothness restriction approach instead tells that the first lead would be invalid only if the slope divergence in the post-treatment period from the linear extrapolation obtained from the lags' pattern overcame 7%. Considering that we are referring to the first lead (when the effect is lower compared to the rest of the post-treatment period, and it has not fully kicked in already), we can consider such results as robust. If we look at the average effect across all leads (as in, the ATT), we observe that the estimated coefficients are even more robust: indeed, they maintain their validity up to a 125% bounded relative magnitude violation, and up to a greater than 12.5% difference in the linearly extrapolated slopes. This confirms the validity of the parallel trend assumption in our framework. We also report the *honestdid* estimations for the various comparisons across group in the Appendix, in Figures from A1 to A6.



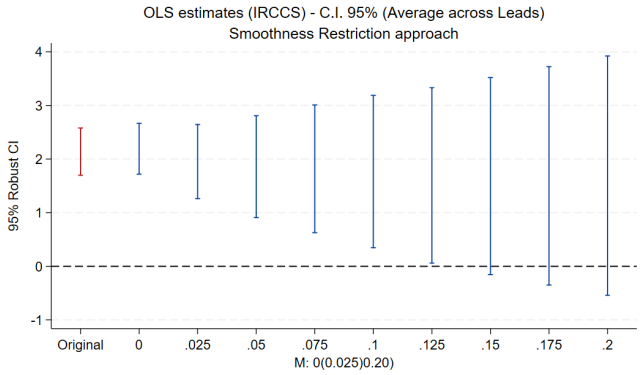
a) BM approach for the significance of the first lead.



b) SR approach for the significance of the first lead.



c) BM approach for the significance of the average across all leads.



d) SR approach for the significance of the average across all leads.

Figure 6: Honest DiD robust confidence sets for overall IRCCS effect estimated with different methodologies (Bounded Relative Magnitude - BM: a) and c). Smoothness Restriction - SR: b) and d)) and for different quantities of interest (coefficient on the first lead of the event study, a) and b). Average across all post-treatment leads, c) and d)).

5.2.2 Validity of the TWFE 2x2 DiD identification

The IRCCS recognition happened at once at the institution level, and thus there exists a bulk of core researchers who stay in the IRCCS list for the whole post-treatment time-span; however the perimeter is re-updated annually by the hospital management. Therefore, the inclusion of some individuals in the IRCCS research staff could be considered the proper setting for a *partially staggered adoption roll-out*. In such context, we have a *core group* of approximately 140–150 researchers in the triennium 2018-2020 (more specifically, 152, 140, 141, as shown in Figure A7), and a growing number of physicians formally added in subsequent years (so that they become 181 in 2021, and over 200 in 2022), especially after the COVID outbreak. While the late additions often consist of previously MBO-treated physicians included for their research engagement, also some faculty members are newly added up to the IRCCS staff *officially*.

In the main analysis, treatment is defined based on the *"ever-treated" criterion*, meaning that we deem all individuals who were ever listed in the IRCCS registry as treated from 2018 onward. This is coherent with our interest towards the causal estimand studied throughout the paper, which is the *Intention-to-Treat (ITT)*:

$$\begin{aligned} \text{ITT} =& \mathbb{E}\left[Y_{it}(1) - Y_{it}(0) \mid \exists\, k \in \mathcal{T},\ k \geq 2018 \wedge D_{ik} = 1\right] \\ & \text{where } \mathcal{T} = \{2012, 2013, \ldots, 2022\} \end{aligned} \tag{6}$$

where $D_{ik} = \{0, 1\}$ indicates the treatment status assignment, and it implies that an individual can be deemed as treated as long as she actually receives the access to the IRCCS perimeter *at a certain point in time starting from 2018*. This is different from the *Average Treatment on the Treated (ATT)*, which would rely on actual entry year:

$$ATT = \mathbb{E}[Y_{it}(1) - Y_{it}(0) \mid D_{it} = 1] \tag{7}$$

where $D_{it} = 1$ indicates the assignment of the treatment in the year in which an individual i formally enters the IRCCS perimeter. This approach can be considered valid under *partly* staggered entry. This holds consistently within our institutional settings, as individuals are enrolled in the IRCCS perimeter by the hospital management according to their research commitment, already ascribable within the boundaries of an "unofficial" IRCCS core. In terms of empirical appraisal, we support our claim with several strategies. First, we observe that, the approximation notwithstanding, all the basic parallel trend

tests show no anticipatory behavior nor divergence in the pre-trends before the IRCCS recognition, which implies that even if selection may have been in place for the late-enrolled, it surely occurred after the IRCCS recognition, and never before. Second, such tests are all corroborated by using the *honest* methodology of Rambachan and Roth, 2023 to check for the validity of the effects in presence of hypothetical deviations from the common trends. Then, to confer more validity to such simplification, we remove from the sample all the "late-adopters", as in those who are not part of the initial core of IRCCS researchers, to assess whether the effect is driven by them or if it is biased upwards by selection issues. Table A4 in the Appendix proves us that is not the case, as the effects are still present (and, indeed, magnified) by excluding those who are selected into the research perimeter with some lag after the IRCCS' recognition, showing that, if anything, what we estimate in our ITT analysis is a lower bound for the true ATT, which is biased downwards by the inclusion of the later-treated. The dynamic estimates derived from such sensitivity checks are also reported in Figure 7 and Figure A8 in the Appendix; they both corroborate the validity of our simplification procedure by showing the absence of pre-trend. The reduced significance of the effect for the double-treated individuals is reasonably due to the lowered statistical power of estimates caused by the sub-setting of the dataset. This statement can be better understood by looking at Table 6, which displays the composition of the group of the IRCCS-enrolled individuals.

	N	Percent		N	Percent
IRCCS First - MBO	20	13.16	IRCCS Late - MBO	33	35.48
IRCCS First - No MBO	132	86.84	IRCCS Late - No MBO	60	64.52
Total	152	100	Total	93	100

Table 6: Composition of the IRCCS group, and distinction between "first-" and "late-" adopters,

Such robustness checks are made necessary by the fact that, if we look ad the characteristics and the composition of the sub-groups of the IRCCS perimeter (2018 vs late adopters), significant heterogeneities are displayed (Table A5 in the Appendix), in doing this showing the existence of structural differences between the individuals selecting early or late into the IRCCS treatment perimeter.

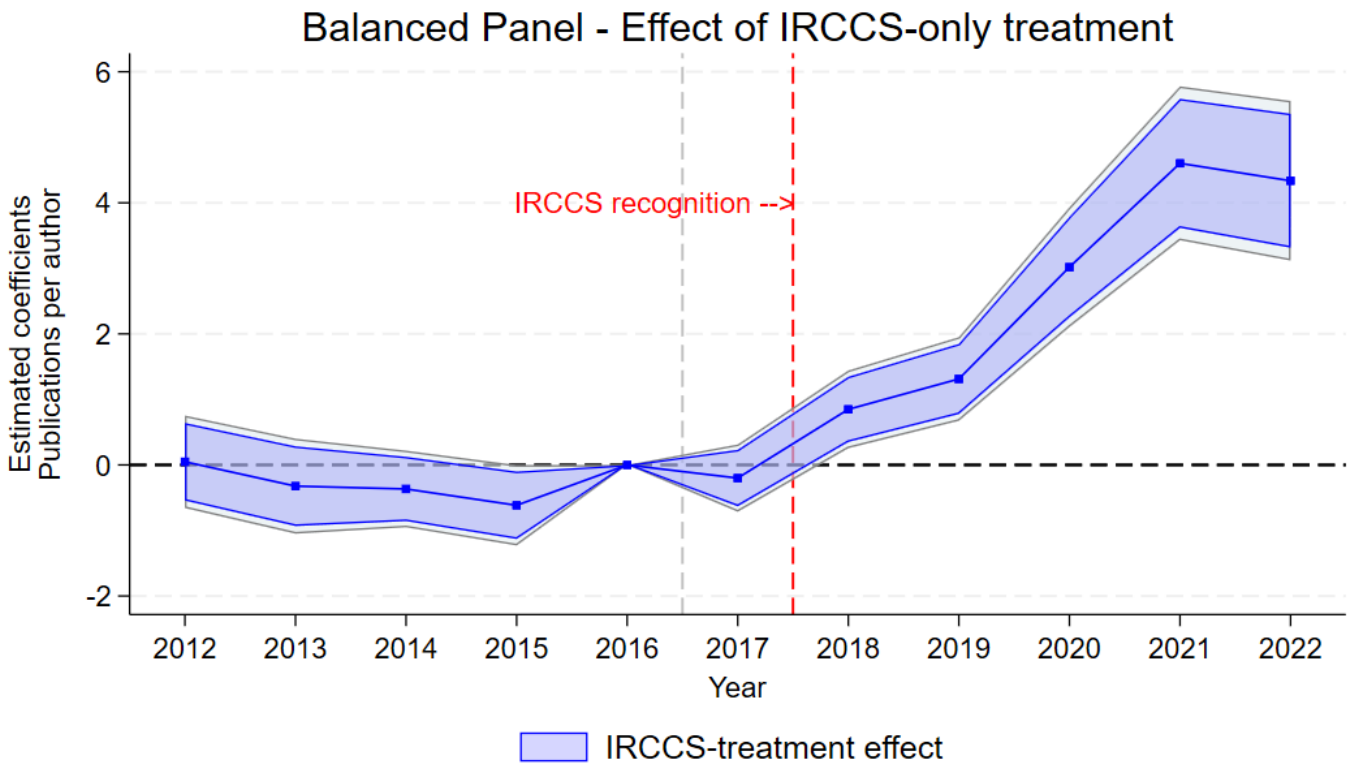
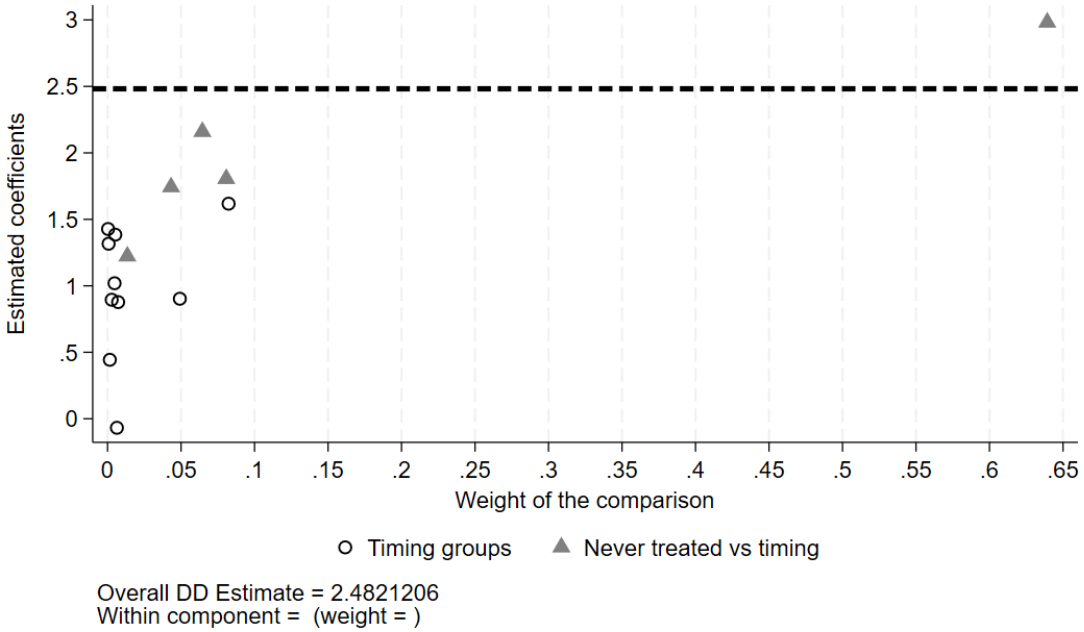


Figure 7: Event-study of IRCCS recognition effect on publications. Units included in the IRCCS perimeter in any year after 2018 are excluded from the estimates.

Furthermore, reliance on our ITT 2x2 methodology is also justified if we follow the approach of Goodman-Bacon, 2021. We hence estimate a DiD equation with variation in treatment timing (depending on when treated researchers are included in the IRCCS perimeter). Starting from the assumption that a staggered TWFE OLS DiD coefficient is a weighted average of 2x2 DiD coefficients obtained from the comparisons of all different cohorts stemming from the heterogeneous roll-out of the treatment, we decompose the DiD coefficient into sub-coefficients comparing the various cohorts, to which a weight is assigned depending on its relative importance credited to the finding in explaining the results. Results are reported in Figure 8, both in the graph and the table below. The decomposition shows that 84% of the estimates stem from the comparison between treated and never-treated, rather than those across heterogeneous treatment-timing cohorts. In addition, the bulk of the overall effect (about 65%) is to be credited to the first sub-coefficient, as in the one stemming from comparing the individuals treated in 2018 to the never-treated. All other coefficients deliver estimates below the overall effect, although the weights assigned to them are extremely low in magnitude, which proves how the existing selection mechanism is actually downward biasing the results. This supports the reason of opting to the estima-

tion of an ITT via a basic 2x2 identification strategy rather than focusing on ATT as our estimand of interest, as the staggered design would be flawed by serious endogeneity issues and it would deliver severly biased results.



	Beta	Weight
Timing groups	1.253	.159
Never-treated vs timing	2.714	.841
Overall effect	2.482	1

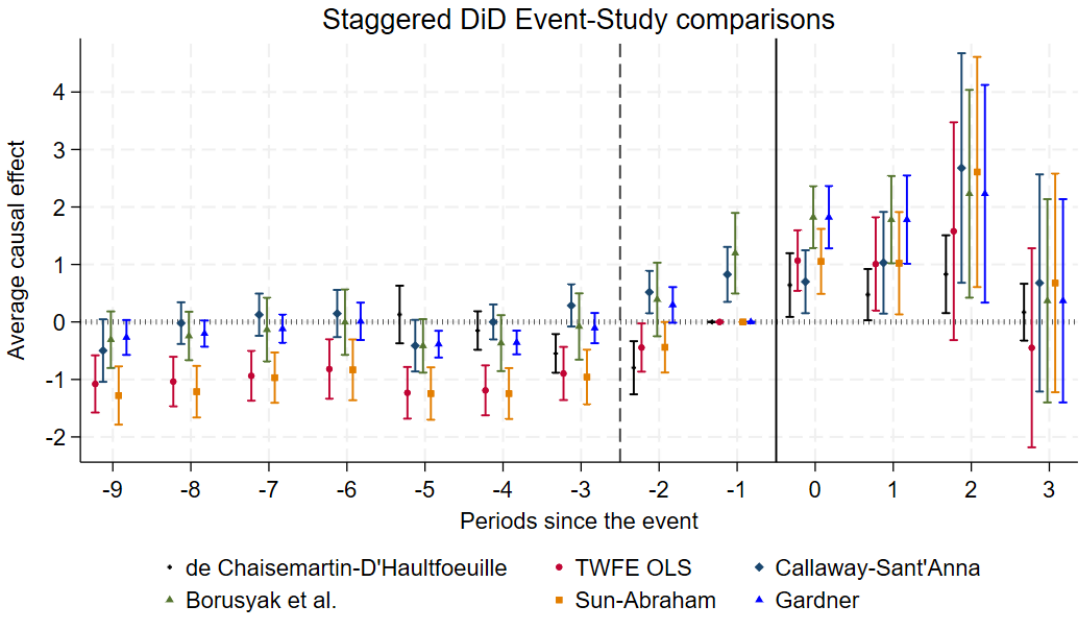
Figure 8: Bacon decomposition: weights and estimates.

Eventually, we report the extent of the selection issue by estimating two sets of event-study equations; in the first set, we aim at estimating the (downward biased) staggered ATT effect on individuals enrolled in the IRCCS perimeter after 2018 excluded, hence without including all the "core" researchers treated in bulk concomitantly with the IRCCS regonition. In the second set, we perform a staggered estimate on all units, late-enrolled or otherwise. Recent econometric literature on Staggered Difference-in-Difference methodologies displayed already how TWFE OLS estimators may deliver biased estimates (de Chaisemartin and D'Haultfœuille, 2020, Callaway and Sant'Anna, 2021, Sun and Abraham, 2021, Butts and Gardner, 2022, Borusyak et al., 2024), especially when dealing with heterogenous timing, dynamically varying effects and endogenous treatment; we hinted such issues in the decomposition from Goodman-Bacon, 2021). Thus, as first step, we estimate a dynamic TWFE OLS equation for staggered

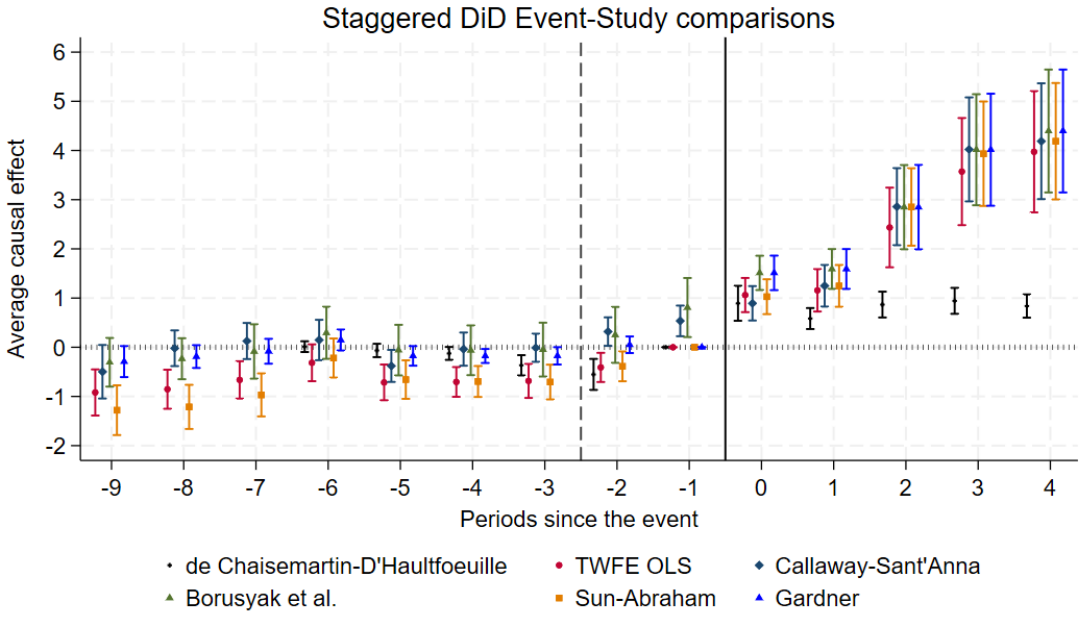
treatment adoption, with the following specification:

$$Y_{it} = \delta_i + \tau_t + \sum_{k \neq -1} \beta_k \cdot D_{it}^k + \varepsilon_{it}, \tag{8}$$

where Y_{it} is the outcome of interest for unit i in year t, δ_i and τ_t are unit and time fixed effects. D_{it}^k is a binary variable taking unitary value if i is k years away from treatment in t, 0 otherwise. We normalize the coefficient for the period immediately before treatment ($k = -1$) to zero, so that all β_k coefficients are interpreted relative to the year before the IRCCS inclusion. After that, we integrated the estimate by better refined approaches: de Chaisemartin and D'Haultfœuille, 2020: correction to the TWFE estimator by excluding "forbidden comparisons" (e.g., already treated units used as controls). The estimator relies on 2x2 comparisons of yearly "joiners" with non-treated units; Callaway and Sant'Anna, 2021: estimate non-parametric group-time ATTs using only never-treated units as controls, allowing for flexible heterogeneity across groups and over time. Borusyak et al., 2024: construct counter-factuals using untreated observations in each period and predict potential outcomes under no treatment assumption. Sun and Abraham, 2021: introduce event-study estimators based on interaction terms between time and cohort. Butts and Gardner, 2022 (based on Gardner, 2022): implements a two-stage estimator that residualizes the outcome variable to remove fixed effects, and then estimates dynamic treatment effects. Figure 9 (Graphs a) and b)) shows the dynamic effect of IRCCS using TWFE and the various corrected estimators. The methods who rely on linear estimates are normalized at $k = -1$ to avoid collinearity. All the patterns are very similar, although the different methodologies used to estimate coefficients show strongly heterogeneous findings in terms of magnitude and significance.The TWFE and Sun-Abraham estimates significantly shift downwards the pre-trend dynamic, showing an evident converging pattern in the pre-treatment period. The same is displayed by the estimates performed via de Chaisemsartin-D'Haultfoeuille and Gardner, although their magnitude is lowered. By contrast, the non-parametric coefficients provided by Borusyak et al. and Callaway-Sant'Anna suggest the presence of an evident anticipatory pattern, who violates the parallel trend assumption, in the immediate lags prior to the IRCCS rollout, while substantially holding the PT assumption as consistent for the previous relative time periods.



a) Event-study of the staggered inclusion in the IRCCS perimeter of late-enrolled only, excluding the "core" researchers treated in 2018.



b) Event-study of the staggered inclusion in the IRCCS perimeter of all treated units, including the "core" researchers treated in 2018.

Figure 9: Event-study of IRCCS policy effect on publications for comparisons across different groups.

Such differences in the estimates are reasonably to be credited to the heterogenous nature of the methodology, as the latter ones do not rely on excluding time periods for the sake of collinearity, and therefore what is observed as a converging pattern driven by an initially negative difference in trends for coefficients estimated through the other method, due to the estimates being with respect to an omitted time unit, for the non-parametric estimates translate into a positive anticipatory behavior emerging in the lags prior to the enrollment. We also observe how the visual inspection of diverging trends displays less evident heterogeneities in the full sample (b) with respect to the one excluding early adopters (a), who are the big bulk of treated individuals and make up for most of the estimated coefficients. In graph (a) indeed, the post-treatment effects are downward biased to a lesser extent and appear to be diverging rather than re-absorbing. It must also be noted how, in graph (b), the estimates yielded by the methodology of de Chaisemartin and D'Haultfœuille, 2020 are sensibly downward shifted compared to the other approaches, which display comparable coefficient magnitudes. This is reasonable, as the weights constructed through their corrected estimation methodology are based on the comparison between "joiners" (as in, late adopters) with the never-treated units. As the majority of the treated units is made up by core researchers, the results delivered by comparing late joiners to the control group are necessarily lower than those observed by estimating the effect via other procedures. This is corroborated by the fact that, in the event-study excluding early adopters, the approach by de Chaisemartin and D'Haultfœuille, 2020 yields less understated dynamic coefficients. This suggests that selective inclusion into IRCCS (which indeed exists and is apparently due on past productivity) is possibly downward-biasing our ITT findings, which are consistently acknowledged as the baseline results or the present study.

6 Robustness Checks

In this section, we present further robustness tests to verify whether the estimated effects are driven by the policies or by other possible confounders. First, we try to assess the effects found already in our baseline model using different scales of the outcome. We then address potential bias from the COVID-19 shock; control for spillover effects in publications across groups; and account for individuals who changed status and for the presence of potential "research superstars" who might be driving down (up) the effect in either estimation setting.

6.1 Different scales

So far we expressed the outcome in raw numbers, reporting it as the number of yearly publications per authors. However, given the relevant difference in numbers across different types of researchers, with many of them who never publish whereas others do that even dozens of times in a year, to re-scale the main outcome could be a compelling idea. We re-perform the estimates specified in Equation 4, changing the outcome from the raw number of articles to: 1) their log-number (re-scaling by adding 1 as constant); 2) their inverse hyperbolic sign. In such a setting, one could be led to suggest using a ML estimation of the coefficient via a Poisson methodology, given the count data nature of the outcome. However, due to the high prevalence of zeros in the outcome, coupled with the relative little dimension of the panel with respect to that, such procedure would bring about over-dispersion and hence we avoid reporting such estimates. Coefficients on log- and asinh- publications are reported in Table A6, in the Appendix. While the direction and significance of all coefficients is consistent with those of the baseline model, their percentage size is definitively lower compared to the one computed on raw numbers relative to the pre-treatment mean, never overcoming the threshold of 40%. However, the high amount of zeros in the outcome and the methodology employed (Diff-in-Diffs) makes the estimates sensitive to scale variations (J. Chen and Roth, 2023), which is proven by the difference in the estimates depending on the kind of outcome transformation performed. This is why we hold such findings ar robustness check while keeping to account on Equation 4 as our baseline.

6.2 Excluding COVID-19 Years

It seems clear that the COVID-19 pandemic outbreak in 2020 substantially triggered upwards the pattern of publications. The main concern would be if such event could have affected research productivity, independently on the granting of IRCCS funding to the hospital. The main confounder would be the re-allocation of time due to the pandemic: as a matter of fact, time and efforts could have been devoted to medical assistance a the expense of research, especially for non-faculty doctors (Mantellini et al., 2020, Ahrendt et al., 2022, Franzoni et al., 2025). In such case, the diverging pattern should not be credited to the take-off of the funding policy, but to the reduced research activity by physicians. On the other hand, the postponement of elective activities might have given some doctors more time for writing; in addition to that, due the joint national effort put in place in the struggle to counteract the pandemic, COVID-related research at the time could have been easy to perform, with lower publishing

lags. While the latter two hypotheses are not necessarily a concern, since they would explain a feasible amplification of a still existing IRCCS effect, the first threat to validity would be a major confounder, able to substantially bias the effect in the reported estimates. First, such concern shall be mitigated by the fact that, notwithstanding that MDs were on the frontline in the struggle against the pandemic, also the assistance activity of faculty members was increased due to the emergency. However, to ensure that our results are not driven by anomalies during the pandemic, we re-estimate the main models using data only up to 2019[2]. We thus truncate the panel at 2019, and re-run the MBO and IRCCS difference-in-differences analyses. Tables A7 and A8 in the appendix show the estimates for the pre-COVID sample, both for the MBO and the IRCCS framework. In both cases, the results are similar in direction to the full-sample results, although the size of the coefficients is, as expected, reduced by the shortening of the time span. We avoid re-performing the event-studies with the reduced sample, as the full dynamic specifications already show that the first leads already hint the overall patterns described insofar even without Covid occurring.

6.3 Spillover Effects

A further concern for our identification strategy is the potential violation of SUTVA (*Stable Unit Treatment Value Assumption*), as in, that the policies might have indirect effects on those who are not directly targeted, mostly due to collaboration. For instance, IRCCS physicians who do not receive MBO-bonuses might however "spill" their gains towards non-academic MDs or even the double-treated ones, by increasing their output and at the same time co-authoring with the performance-based remunerated doctors. Since we observe that the IRCCS-funded researchers are the ones responding more significantly to the shock, if there was such a spillover effect, it would indeed be biasing our estimation downwards, in doing this reducing the differential across groups. However, we test for spillovers by conducting the following two exercises. First, we remove from the sample the double-treated units, who bring about cross-group spillover by definition as they constitute an overlap of the two groups; then, we identify all publications involving a collaboration between an MBO-eligible physician and an IRCCS-ever listed researcher. We call publications of *Type A* those which, in the panel, can be flagged as authored by a MBO-treated individual in collaboration with at least one member of the IRCCS perimeter. *Type B* publications are instead those identifiable as authored by IRCCS staff and co-authored with at least a non-academic

[2]Note that already by looking at the main event-studies we can retrieve that the effect kicks in before the pandemic started, which is already a good validation of our main results.

physician. Then, we perform two Diff-in-diff estimations for each exercise. In the first we compare MBO-treated to the rest of the sample (as in Equation 1, but without the compound treatment), by first excluding authors of *Type A* papers, hence the non-academic MDs who *ever* collaborates with an *ever* non-MBO IRCCS member. Then, we do the same for the IRCCS-treated (mirroring Equation 4 while excluding double-treated again), this time excluding *Type B*, as in IRCCS researchers who *ever* co-author with MBO-treated units. This quite stringent exclusion of *all* authors of cross-written papers ensures that the treatment effects only stem from either within-group or individual publications. Such approach eliminates entire collaboration links between the two groups. We avoid removing both groups from the sample at the same time in order not to lose too much statistical power for the estimates. The results of these two exercises seem to remain consistent with the main findings (Table A9 in the Appendix). The Management-by-Objective impact remains negative and statistically significant when both MBO-treated and IRCCS-treated authors of cross-group joint papers are removed. As a matter of fact, the absolute value of the negative effect is magnified when *Type A* researchers (non-academic MDs collaborating with IRCCS individuals) are removed, while the drop is weakened when excluding *Type B* researchers from the regression (academic IRCCS collaborating with MBO-treated individuals). The same, but in the opposite direction, is retrieved when testing the consistency of the effect of the IRCCS recognition on the productivity of empowered researchers, even when excluding the network groups. The evidence suggests that the observed effect may be actually dampened by the collaboration with IRCCS researchers sought by MBO-only doctors; absent such behavior, our estimates could be actually showing even greater divergent patterns. The related dynamic estimates of such robustness checks are reported in the Appendix as well, in Figure A9, Graphs a) and b) respectively.

6.4 Controlling for outliers and switchers

Our main estimates have been performed, so far, by considering group status as fixed and absorbing, based on the initial policy definitions at the moment of the implementation of the given policies. In reality, there could be some variations over time due to changes across different roles. For instance, some physicians shift from non-academic positions to academic role at a certain point in time post-treatment, exiting from the group of those who are eligible in the MBO. On the other side, non-prolific academics may be lose their tenure and become MDs without faculty appointment. In either case, the effect of productive non-academic MDs considered as "fixed" MBOs at the implementation of the MBO switching into the academic group, as the opposite pattern indeed, should again bias down our estimates,

as the publications of these more (less) productive individuals shall be accounted for in the outcome of the control (treatment) group in the baseline analysis[3]. To guarantee that these "switchers" are not affecting our results to a great extent, we perform a robustness check where we exclude any unit whose status changed during the observation treatment window. To be precise, we allow for three scenarios: 1) drop any non-academic physician in 2017 who becomes, at a given point, a faculty member after 2017 (shifting from MBO-eligibility to non-eligibility); 2) any faculty member who left academia after 2017, to sort into non-academic medical practice; 3) any individual who undertook either of such changes. Cross-position switchers account for barely 3.5% of the dataset. The findings (Appendix, Table A10) are extremely similar in all scenarios, suggesting that dynamic transitions are not relevant enough to create distortions in the effects' estimates.

In addition to that, we check whether the two groups have different incentive structures actually moving them. We do this by identifying those individuals who may shift upwards the overall publication numbers of their respective group, especially with regards to the IRCCS-affiliated members. Hence, we re-perform the analysis by removing: 1) outliers (as in, the last percentile of the annual publication distribution); 2) the top 10%; 3) the upper quartile of the annual publication distribution[4]. We do this and re-estimate both the baseline MBO and IRCCS specifications: they are reported in the Appendix, in Tables from A11 to A13 for the MBO-treatment estimates, and in Tables from A14 to A16 for the IRCCS ones. As expected, the negative shift after the MBO-implementation for treated units, as for the positive effect of the IRCCS recognition, decrease in size when an increasingly greater units in the upper parts of the publications distribution are removed from the sample. However, the directions of the effects stay the same even when eliminating (almost) one fourth of the sample; in addition, even if some coefficients lose some degrees of significance, the ones who were statistically different from zero in the main analysis never lose such feature. Hence, it is evident that the huge effect we observe is driven but not entirely made by a sub-group of scientific superstars among IRCCS researchers.

[3]Note that one might argue, for the baseline comparison, that switching out of the MBO-treatment group should be accounted as a violation of the absorbing treatment assumption, as MDs who become academics lose the eligibility to the performance-based payment. However, it is reasonable to keep such individuals in the MBO-treatment group, as the purpose of the performance-based payment is exactly that of triggering an incentive mechanism for which individuals start publishing after receiving the money in the first place. If after that they keep publishing even while losing the eligibility to the MBO, and if they adjust their incentive structure to such a point that they even switch to an academic position, the policy could be considered as effective. More on this in the next section.

[4]Note that the threshold for the upper quartile is just 3 annual publication, so we drop all those who have published at least 3 publication in any year, but we keep those who have published exactly 3.

7 Research Quality and Networks

It is valuable to assess whether the policies had a significant impact on the nature of the research collaborations stemmed after their implementation, and the impact/quality of the output. We therefore perform two sets of additional estimates in order to tackle two questions mostly: (1) whether and to what extend the policies affected research collaborations across the two groups (which so far we have used only as an excluded group to perform robustness checks), and (2) how our results relate to research quality of the members of the institution under study, proxied by citations.

Cross-groups Co-authorship

We begin this section by defining all papers listing, among the co-authors, both a non-academic doctor (subject to MBO) and an IRCCS researcher of the same institution as "networked" publications; this is a decent representation of network effects between the two categories of hospital employees. We make use of the same categorization we employed when controlling for the validity of *SUTVA*, by subdivide networked publications according to the groups' perspective. *Type A* are those for which an MBO-eligible individual is an author (MBO-group outcome count) and at least one co-author is a designated researcher. The opposite holds for *Type B* papers: a IRCCS physician is an author (IRCCS-group outcome count), and at least one co-author comes from the non-academic physicians' group. Therefore, *Type A* and *Type B* label the very same publications, but they are attributed to different groups for our assessments' purpose. In doing this, we consider the number of *Type A* publications for each non-academic individual and that of *Type B* works for the IRCCS units as different outcomes. Table 7 displays the DiD estimates for the effect of the policies on the collaborations as defined above; Cols. (1) and (2) report coefficients for *Type A* and *Type B* outcomes in the MBO-based framework. Cols. (3) and (4) do the same for the IRCCS-based setting. For collaborations of MBO-treated physicians with IRCCS researcher (*Type A*), a significant increase in cross-group collaborations post-2017 (MBO) relative to non-treated MDs is retrieved (23.7 p.p., which amounts to 76.6% of the pre-treatment mean). On the other side, the effect on *Type B* publications, as in those included in the count for IRCCS-researchers (which are double-treated here, as they are control group in the MBO-based DiD) and co-authored with the treated units, is negative and significant (-18.5 p.p., -60%). By contrast, after the IRCCS recognition, we observe a significant decrease in *Type A* publications for the treated (which, again, are double-treated, as in this setting the MBO-only treated individuals are in the control group),

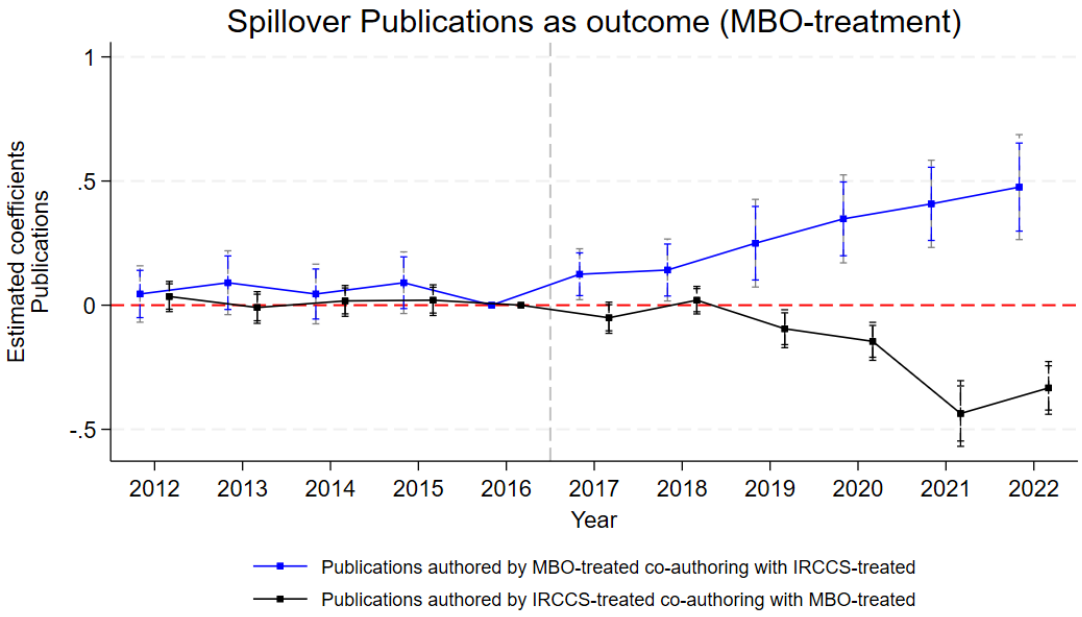
which amounts to -12.8 p.p.. However, such decrease is quite irrelevant relative to the mean of the treatment group prior to the IRCCS recognition (-3.5%). On the other side, the increase in *Type B* papers for IRCCS-treated individual amounts to 38.2 p.p., a bit greater coefficient but still quite negligible relative to the mean ($+10.6\%$). Such findings suggest that non-academic doctors who were prompted to do research, apparently sought out more partnerships with the members of the IRCCS perimeter. Similarly, for the academic IRCCS researchers (Type B outcomes), there is an increase in collaborations with non-faculty MDs after 2018, compared to non-IRCCS academics. While the increase in co-authorship between MBO-treated and IRCCS members is quite substantial relative to the mean of the non-academic physicians, the incentive to collaboration for the IRCCS's individuals is sensibly lower in percentage compared to their pre-2018 average. While the results may suggest that the policies led to better integration between the two groups, we could also foresee the emergence of strategic behavior among MBO-treated individuals aimed at improving their performance indicators by aggregating their activity to that of IRCCS-based researchers.

2012-2022; SEs clustered at individual level

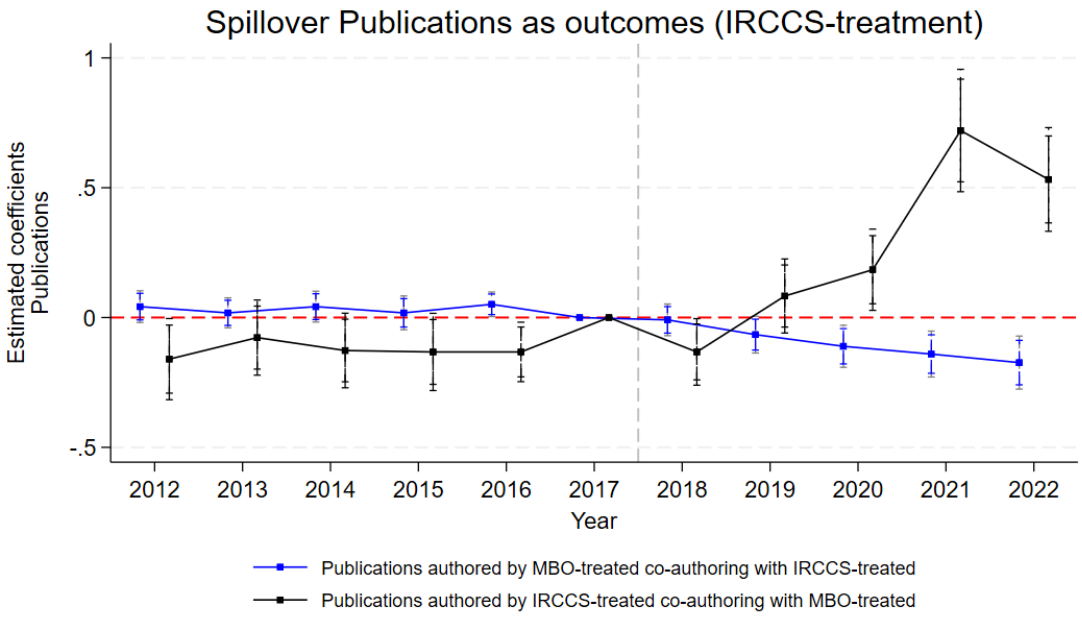
	(1) Publications by MBO-treated co-authored with IRCCS (Type A)	(2) Publications by IRCCS-treated co-authored with MBO (Type B)	(3) Publications by MBO-treated co-authored with IRCCS (Type A)	(4) Publications by IRCCS-treated co-authored with MBO (Type B)
Post 2017*MBO	0.23717 ***	-0.18498 ***		
	[0.04766]	[0.02874]		
Post 2018*IRCCS			-0.12843 ***	0.38241 ***
			[0.02778]	[0.05308]
Observations	5,691	5,691	5,691	5,691
R-squared	0.48304	0.49938	0.47510	0.51355
Individual FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Method	OLS	OLS	OLS	OLS
Time Range	2012-2022	2012-2022	2012-2022	2012-2022
Panel	Full	Full	Full	Full
Mean	0.309	0.309	3.609	3.609

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Impact of MBO-policy and IRCCS recognition on annual cross-group publications (Difference-in-Differences).



(a) MBO physicians collaborating with IRCCS researchers (MBO-treatment effect)



(b) IRCCS researchers collaborating with MBO physicians

Figure 10: Event-study of policy impact on cross-group collaborative publications (IRCCS-treatment effect).

Figure 10 reports visually the dynamic estimates for networked publications. Panel (a) shows the event-study for *Type A* (MBO group co-authorship) after 2017, while Panel (b) for *Type B ones* (IRCCS group co-authorship) after 2018. Both graphs indicate little to no difference in cross-group collaboration rates between the two groups prior to the policies. After the two policies implementations, the affected groups show substantial effects. In (a), all MBO physicians start collaborating way more with IRCCS-researchers in 2017, and such increases is perstistent through 2018 and beyond. By contrast, double-treated physicians appear to collaborate less with non-academic MDs, although such pattern starts showing up in 2019 only. In (b), IRCCS academics are featured by a relative increase in collaborations with MBO doctors with a lag after 2018 (the effect seems to be small a negative and barely significant in the first post-treatment period). On the other side, as already showed in graph a) as well, double-treated physicians reduced their collaboration activities with other IRCCS researchers, by displaying a modest in relative terms but still persistent decline. While such effects would suggest an increased interdisciplinary externalities, which would surely benefit research by allowing a easier transfer from science to clinician practices, the fact that the effect is substantial only for MDs who eligible to the MBOs, negative for the double-treated overall and positive and significant but slightly relevant for the IRCCS researchers could underline the presence of free-riding behavior in publications that, although undistinguishable in the presen analysis, could be better assessed by looking at the papers' impact.

Citation Impact

A further relevant question is whether a greater number of publications also triggered a positive shift in the influence and quality of such works, or it came a the expense of it. We hence analyze the policy effect on citations, employed as proxy of publications' impact. We do that through by re-modeling the IRCCS equation (Equation 4) in order to adapt it to a set of different outcomes and weights. First, as discussed already in other studies such as the ones by Azoulay et al., 2011 or Waltman and van Eck, 2013, we cannot estimate the effect on citations without accounting for the publications' age. Thus, we build a yearly discounting term to control for the period when the citations when gathered, which is Fall 2024 ($[2024 - t]^{-1}$). After doing that, we normalize each paper for time by multiplying its count of citations by the discount term, and build a yearly aggregate individual impact weight by summing up all yearly normalized citations for every researcher. This is the first normalized variable we obtain, which we use first as a Inverse Probability Weight in a regression of the treatment on the count of publications, and then as an outcome. The second outcome is obtained from the sum of all non-discounted yearly

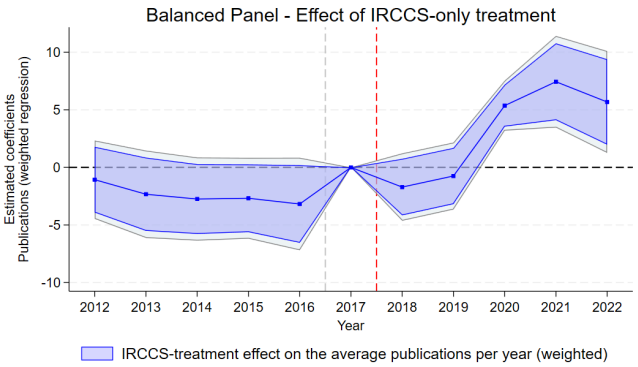
citations received by papers published by a given author in a given year; after aggregating them, we normalize such total by time, making use of the discount term described already. Eventually, we build an averaged measure of publications' quality, by dividing the number of total (discounted) citations of papers published in a given year by an author by the number of total papers published by the very same author. In these regressions, as in the latest ones, the IRCCS-treatment is assigned only to IRCCS researchers non-eligbile to the MBO, while the MBO-treatment also considers the double-treated.

2012-2022; SEs clustered at individual level						
	(1) Publications (IPW) (by no. of discounted 2024 citations)	(2) Sum of discounted 2024 citations	(3) 2024 Citations (over yearly publications)	(4) Publications (IPW) (by no. of discounted 2024 citations)	(5) Sum of discounted 2024 citations	(6) 2024 Citations (over yearly publications)
Post2018*IRCCS	5.26598 ***	20.70414 ***	-142.44386 ***			
	[1.88683]	[3.41649]	[23.02236]			
Post2017*MBO				-2.98268 *	-9.53533 ***	60.70126 ***
				[1.53918]	[2.01651]	[18.88641]
Observations	3,258	6,369	6,369	3,258	6,369	6,369
R-squared	0.85726	0.44474	0.27837	0.85488	0.43467	0.27060
Individual FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Method	OLS	OLS	OLS	OLS	OLS	OLS
Time Range	2012-2022	2012-2022	2012-2022	2012-2022	2012-2022	2012-2022
Panel	Balanced	Balanced	Balanced	Balanced	Balanced	Balanced
Mean	3.609	16.18	302.8	0.653	2.492	123.7

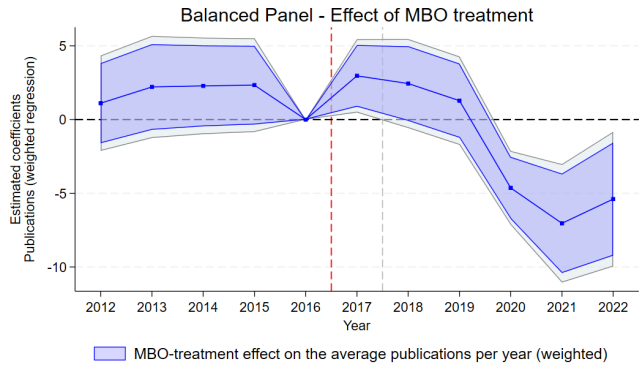
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Impact of IRCCS recognition and MBO policy on Publication quality (Difference-in-Differences).

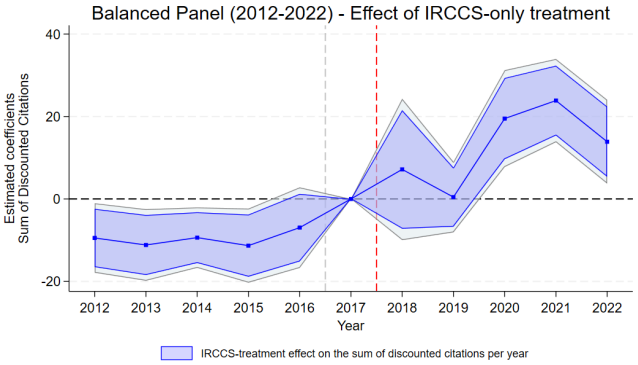
The results are presented in Table 8 for the IRCCS and MBO effect on citations. Apparently, by (inversely) weighting publications by the number discounted citations, the effect of the IRCCS' recognition is amplified for more structured researchers. By looking at the sum of (discounted) citations as outcome, IRCCS recognition seems to have boosted the number of quotes for IRCCS physicians, by a number of 20.7 citations (+127%). However, the IRCCS group's post-2018 average number of (discounted) citations per publications shows a major and significant decrease: −142.44, amounting to almost −50% of the pre-2018 mean. By contrast, we observe a negative, slightly significant effect on the number of MBO-eligible individuals' publications in the IPW-weighted regression, while the overall drop in the sum of discounted citations is quite striking in relative terms, as it amount to −9.5 citations, as in −382%. Quite surprisingly, and specularly to what emerged for the IRCCS-affected individuals, the impact on the average number of discounted citations over yearly publications is quite substantial: almost +61 average citations per publications, which means an increase of $\approx 50\%$ compared to prior to 2017.



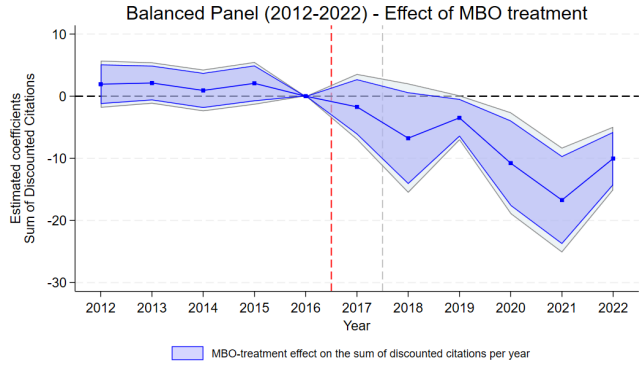
a) IRCCS-effect on publications weighted by discounted citations (as of Fall 2024).



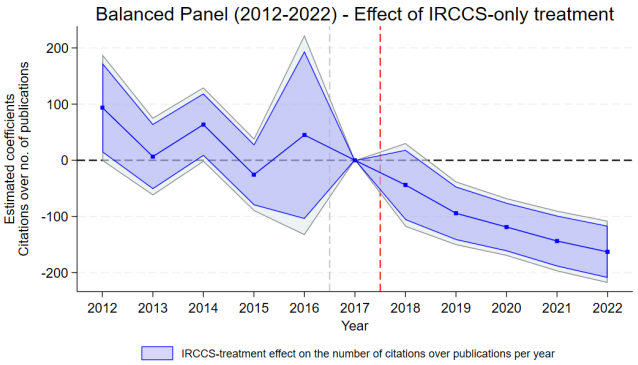
b) MBO-effect on publications weighted by discounted citations (as of Fall 2024).



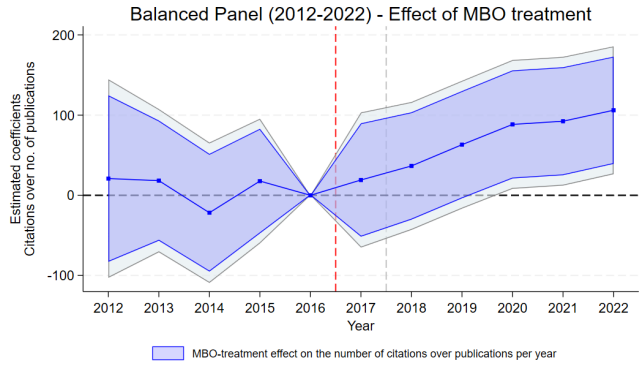
c) IRCCS-effect on the total yearly sum of discounted citations (as of Fall 2024).



d) MBO-effect on the total yearly sum of discounted citations (as of Fall 2024).



e) IRCCS-effect on the average number of discounted yearly citations (as of Fall 2024) per publications.



f) MBO-effect on the average number of discounted yearly citations (as of Fall 2024) per publications.

Figure 11: Event-Studies representing the dynamic estimates of the IRCCS recognition and MBO-policy impacts on the impact of published papers, proxied with yearly citations measured as of Fall 2024.

Such findings leads to the conclusion that, while the two concomitant policies have an established effect on physicians' research productivity, which is positive for the structured selected academics and basically null for the incentivized MDs, the quality impact is basically the opposite. The aggregate increase (decrease) in the IRCCS- (MBO-) treated individuals' normalized citations is clearly driven by the shift in the raw number of publications, while the average quality of the works seems to be neglected by the units who can directly access funding, and maximized by the MBO-incentivized MDs. While prior beliefs would suggest that, due to the embedded incentive structures, core researchers would focus on quality over quantity upon accessing funds, it looks like that, once achieved the possibility to increase their resources, they fuel their scientific output numbers at the expense of the influence brought about by their research. On the other side, MBO-eligible individuals seem not to carry over projects indifferently in bulks, but even without significantly raising their publication rates, they tend to focus on the quality of their research. This may be due to the fact that the performance-based rewards are not only based on quantity, but also on the impact factor of the journal where publications are issued.

These results are displayed in dynamic terms in the event-studies in Figure 11, along with the evidence of absent driving pre-trends or anticipatory behaviors, .

8 Alternative effects of the MBO

In the analyses reported insofar we came to the conclusion that the MBO policy does not directly impact the scientific productivity of the affected physicians; on the contrary, the compound effect of the IRCCS recognition on the institution seems actually to trigger a relative negative effect of the incentive scheme when combined with the increase of public funding. However, two main evidence we gathered may actually suggest that the MBO was not fully neutral for the MDs' behavior: first, we see that many MDs where subsequently enrolled into the IRCCS perimeter, and the anticipatory behavior in the event-studies performed for the staggered DiD actually highlights the presence of an evident productivity-based selection process. In such regard, the implementation of the MBO might have incentivized doctors with more research-oriented embedded incentive structure to pursue scientific activity, re-directing them towards the IRCCS perimeter and, up to a second moment, triggering a productive pattern boosted by the additional resources. Second, while IRCCSs' researchers clearly display an obvious raise in overall (discounted) citations, due to their more numerous scientific outputs, the average numbers of citations per paper is increased for MBO-affected individuals, while the effect

seems negative for IRCCS-researchers after the IRCCS recognition. This may either mean that MDs value quality over quantity, or that the put in place a strategic publication-oriented research behavior with the purpose of achieving the right amount of output, weighted by the proper Impact Factor and accounting for the reduction due to internal co-authorship, in order to obtain a pre-fixed monetary sum in accordance to their preferences.

Triggering IRCCS enrollment

Our first objective is to find out whether there exists some form of linkage between the MBO policy implementation and the probability to be enrolled into the IRCCS perimeter. However, the available data constrain our possibility to provide with causal evidence in such regard. As a matter of fact, having the recognition of the IRCCS status been occurred after the implementation of the monetary incentive, there is no pre-treatment counterfactual enabling the identification of the policy via a DiD strategy, like the one we implemented for our main analysis. Therefore, we try to enact some correlational analysis. We perform various sets of estimates. We want to assess whether being part of the MD group (i.e., subject to the MBO) is related a higher probability of being enrolled into the IRCCS perimeter, versus the belonging to the faculty member set. Given that a counterfactual for such enrollment likelihood prior to the implementation of the MBO does not exist, we restrict our panel to the period 2018-2022 and avoid the exploitation of any time discontinuity. To this extent, to merely compare MBO individuals to non-MBO individuals, even by including time FEs and population averages, would be ineffective as the effect would mostly be driven by the composition of the groups; and we know, a priori, that faculty members make up for the bulk of the IRCCS staff. Since we acknowledged that there is selection of individuals into the IRCCS list, we assume such selection to be linked to research productivity. Thus, we estimate the likelihood of being enrolled into the IRCCS perimeter as a function of the following main elements: 1) the belonging to the MBO-treated group (vs. non-MBO); 2) the average number of annual publications in the time-span 2012-2016 and the papers published in 2017; 3) the interaction between MBO and the latter two; 4) the total number of citations (discounted by the distance from 2024) received in 2012-2016 and 2017; 5) the average number of discounted citations for papers published in 2012-2016 and 2017. We initially do this by using all IRCCS enrollments, hence with the core researchers included in 2018. Then, we do this again by excluding such set of units, assuming that the structural differences between the core IRCCS researchers and those included later is too high to enable a reliable comparison. When we do this, we include publications, total discounted citations and discounted citations per paper

in 2018 as well. We include time fixed effects, population averages, and a bunch of controls (sex, age, age squared, department and status). The function can be summed up by the following equation, estimated via MLE both with a Logit and Probit model:

$$
\begin{aligned}
Pr(\text{IRCCS}_{it} = 1) = \Phi(&\beta_0 + \beta_1 \text{MBO}_i + \beta_2 Publications_{i,2012-2016} + \beta_3 Publications_{i,2017} + \\
&+ \beta_4 MBO_i \times Publications_{i,2012-2016} + \beta_5 MBO_i \times Publications_{i,2017} + \\
&+ \beta_6 Tot\_Citations_{i,2012-2016} + \beta_7 Tot\_Citations_{i,2017} + \\
&+ \beta_7 Avg\_Citations_{i,2012-2016} + \beta_8 Avg\_Citations_{i,2017} + \\
&+ \gamma_t + X'_{it}\delta), \quad \text{estimated as Population Averages (PA)}
\end{aligned}
\tag{9}
$$

Standard errors are estimated non-parametrically via bootstrapping. Results are reported in Table 9, expressed as Marginal Effects of the estimated coefficients. Cols. (1) and (3) represent the Logit and Probit models including the bulk enrollment of 2018, while Cols. (2) and (4) restrict the sample to 2019-2022, excluding the bulk enrollment and involving the bibliometric controls for the research outcome of 2018. In presence of such many regressors, the belonging to the MBO-policy group seems not having a positive influence on the enrollment into the IRCCS perimeter, as displayed by the first row, which shows negative and significant coefficients, albeit small, even if with different degrees of statistical relevance. In terms of bibliometric controls, it seems that only research activity prior to the MBO-implementation seems to matter consistently across different specifications: an additional average publication in the 2012-2016 time-span seems indeed to be linked to an increase in the group-wise probability of being enrolled between 2018 and 2022 by 1 p.p., and by 2.5 p.p. starting from 2022. Such link however seems being transversal across both groups, as the interaction with the MBO is significant only for the first set of estimates, which involves 2018's bulk enrollment and it is somehow offset by the decrease in 6 p.p. brought about by the dichotomous treatment indicator alone. Publications in 2017 and 2018, hence the ones potentially triggered by the policy, result statistically comparable to zero. Only publications in 2017, interacted with the treatment dummy, display a positive coefficient of 2 p.p. at 5%. Total citations are significant and negatively linked to IRCCS enrollment, and only those for papers published in 2012-2016. Nevertheless, their marginal effect is smaller than 0.5 p.p.

	(1) Likelihood of IRCCS enrollment (ME)	(2) Likelihood of IRCCS enrollment (ME)	(3) Likelihood of IRCCS enrollment (ME)	(4) Likelihood of IRCCS enrollment (ME)
MBO vs. no MBO	-0.06771	-0.03243	-0.06240	-0.02728
	***	**	***	*
	[0.01653]	[0.01482]	[0.01637]	[0.01635]
Avg Annual Pubs. 2012-2016	0.01284	0.02402	0.01393	0.02655
	***	**	***	**
	[0.00378]	[0.01094]	[0.00326]	[0.01164]
MBO × Avg Annual Pubs. 2012-2016	0.02255	0.00597	0.02245	0.00393
	**			
	[0.00905]	[0.01330]	[0.01366]	[0.01280]
Annual Pubs. 2017	0.00353	-0.00413	0.00327	-0.00432
	[0.00408]	[0.01211]	[0.00398]	[0.00730]
MBO × Annual Pubs. 2017	0.02299	0.01608	0.02079	0.01525
	**			
	[0.01114]	[0.01375]	[0.01429]	[0.01573]
Annual Pubs. 2018		-0.00510		-0.00575
		[0.00700]		[0.00622]
MBO × Annual Pubs. 2018		0.00927		0.00984
		[0.01441]		[0.01497]
Total Cits. 2012-2016	-0.00187	-0.00323	-0.00218	-0.00318
	**	**	***	
	[0.00075]	[0.00151]	[0.00074]	[0.00199]
Total Cits. 2017	-0.00020	-0.00019	-0.00020	-0.00023
	[0.00025]	[0.00072]	[0.00035]	[0.00209]
Total Cits. 2018		0.00078		0.00080
		[0.00134]		[0.00135]
Avg Cits. per paper 2012-2016	0.00688	0.00715	0.00754	0.00707
	***	**	***	
	[0.00249]	[0.00300]	[0.00238]	[0.00468]
Avg Cits. per paper 2017	0.00328	0.00535	0.00339	0.00528
		*	*	
	[0.00240]	[0.00278]	[0.00196]	[0.00559]
Avg Cits. per paper 2018		-0.00037		-0.00023
		[0.00274]		[0.00319]
Observations	2,824	1,634	2,824	1,634
PA	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Method	Logit	Logit	Probit	Probit
Time Range	2018-2022	2019-2022	2018-2022	2019-2022
Panel	Full	No 2018	Full	No 2018
Mean	0.0405	0.0198	0.0405	0.0198

*** p<0.01, ** p<0.05, * p<0.1

Table 9: Regressions between enrollment into IRCCS perimeter and several factors.

Finally, it looks like that the average citations per paper, which is our rough but most effective proxy for quality, display a significant and positive effects for the papers published between 2012 and 2016, but mostly for the enrollment which excludes 2018. While this is a hint that there be a mechanism of valuing quality over quantity in the selection into the IRCCS perimeter, the marginal effect are sensibly low in magnitude, even when significant.

Such results, to be taken with a grain of salt as they do not even embed individual fixed effects as the MLE estimations are performed through PAs, suggest a plausibly underlying existing selection mechanism to be linked to the valuing of quality over quantity for explaining the selection into the IRCCS staff. However, the transversality of the results across groups and the low magnitude of the findings, does not allow us to confirm that a potential driver of such selection shall be credited to the implementation of the MBO policy.

Strategic behavior of MDs

Our last check concerns the possibility, for MDs, to put in place strategic behaviors in order to achieve the exact amount of MBO bonus without really improving their overall research productivity in terms of publications. As a matter of fact, while we observed a lack of effect of the MBO implementation on the general output of the treated ones, we still found significant differences in the average quality of papers published by MDs after the MBO compared to the ones published by the IRCCS staff, with the latter being decreasing after the additional funding and the former being positively affected by the policy. The current check is made necessary in order to assess whether the finding is lead by a trade-off between quantity and quality or by a gaming strategy performed by doctors. The bonus is indeed directly proportional to the Impact Factor of the journals where articles are published. On top of that, even though the bonus is decreased for publications co-authored with internal authors, the positive spillover effects shown in the previous section may be a proof of the fact that individuals could actual free-ride publications in order to achieve a reduced bonus, which may be not reached at all in absence of collaborations, in doing this increasing their payout. In our dataset, we have information about the amount of the yearly payout received by the individuals subject to the policy, the number of the eligible publications, and the total amount of the Impact Factors of the publishing journal discounted by the co-authorship weights. A peculiarity of our dataset is that, even though the bonus is only paid to Medical Directors, the computation of the above metrics is made for all researchers active in the institution, Faculty Members included, even though they do not perceive any related direct benefit.

	(1) MBO vs. non-MBO	(2) DT vs IRCCS	(3) DT vs MBO	(4) DT vs control	(5) MBO vs IRCCS	(6) MBO vs control
Pay Out (€)	-1754.58826**	-1977.42029***	1889.24585***	1577.16296***	-3919.94873***	-382.23447**
(SE)	(704.82104)	(320.82687)	(282.11951)	(290.89667)	(232.46324)	(168.14507)
R^2	0.554	0.663	0.480	0.493	0.687	0.379
Mean	623.69	1292.45	1292.45	1292.45	288.78	288.78
Sum of weighted IF	0.06349	-14.01008***	3.94934***	4.25400***	-19.70594***	0.16595
(SE)	(3.97897)	(4.43107)	(1.50687)	(1.11445)	(5.30614)	(1.16628)
R^2	0.308	0.323	0.168	0.467	0.319	0.142
Mean	1.84	3.22	3.22	3.22	1.15	1.15
Eligible Papers	-3.60687*	-6.13109***	4.32669***	3.43250***	-10.85470***	-1.10835***
(SE)	(2.06476)	(1.12172)	(0.49358)	(0.52927)	(1.12714)	(0.32390)
R^2	0.541	0.596	0.572	0.576	0.611	0.433
Mean	1.58	3.14	3.14	3.14	0.81	0.81
Delta (Pubs-Eligible)	2.01255	4.87741***	-3.15255***	-2.32054***	8.24297***	0.90595***
(SE)	(1.32853)	(0.90433)	(0.42678)	(0.45205)	(0.88717)	(0.22649)
R^2	0.346	0.446	0.409	0.446	0.436	0.351
Mean	-0.33	-0.74	-0.74	-0.74	-0.13	-0.13
N	5181	2563	1900	1982	3193	2618
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Panel	Full (no 0s)	DT and IRCCS (no 0s)	DT and MBO (no 0s)	DT and control (no 0s)	MBO and IRCCS (no 0s)	MBO and control (no 0s)
Time Range	2017–2022	2017–2022	2017–2022	2017–2022	2017–2022	2017–2022

Table 10: Regressions between enrollment into IRCCS perimeter and several factors.

While the ideal setting would be that of a Difference-in-Differences strategy like the ones employed so far by performing estimates on different outcomes, the absence of yearly Impact Factor data for all the journals where paper were published before 2017, constrains our analysis. That means, we do not possess a hypothetical computation metrics for the assignment of the bonus prior to the implementation of the MBO policy itself. On the other hand, we can rely on a TWFE estimation where we can compare the metrics computed for the MDs and those computed for the faculty members, in the time-span for which we have this information available (2017-2022). The underlying assumption is that, absent the bonus, the research output of faculty members should not depend, in terms of bibliometric structure, on the weights foreseen by the incentive scheme, nor the payout. We hence estimate a TWFE equation similar to Equation 1, with varying comparison groups and all the rest, but instead of the interaction between *Post2017* and *MBO˙i* we just keep the time-invariant treatment dummy. We regress 4 different outcomes: the Payout in €, the total sum of discounted points of Impact Factor, the number of papers deemed as eligible for publications (which differ from the whole number retrieved from Web of Science), and the difference between the actual number of publications and the eligible ones. The discrepancy between the latter two numbers is due to different motivations: first, the number of eligible paper

is collected by the company who compute the metrics and assign the bonus. However, such number is double-checked by the individuals themselves, whom are send list of papers to be corrected and verified, in order to add forgotten or more recent papers or remove those erroneously linked to the wrong author. Such self-reporting practice leads to numerous opportunistic behaviors which cannot be entirely controlled by the metrics-assessing firm, which only performs random checks to verify the alignment between what is reported in the payout document and the actual research productivity of individuals. There are, thence, numerous mis-specifications due to either researchers not updating thei publications since they are potentially uninterested in obtaining the bonus (i.e., faculty members) and physicians pumping research output to increase their payout. Some researchers claim indeed that papers written by homonyms are authored by them, as the firm assign the eligible documents on a surname-based criterion. Such behavior is enabled by the controls made randomly, which allows risk-taking behavior. In order to avoid potential estimates to be driven too much by the downward-reviewing attitude of self-reporting negligence of faculty members, we drop from the sample all individuals with zero eligible documents over the whole analyzed sample. Equally, supposing that a gaming strategy still requires a minimum degree of research-oriented attitude, we remove from the sample all individuals with zero actual publications in the time-span. Results are reported in Table 10

We can observe how, for all the payout-related variables, how the MBO-treated individuals perform worse than their unpaid counterpart, as the coefficients all follow the same patterns observed for actual publications in the baseline analysis for the MBO-policy incentive. MBO-treated perform way worse in terms of Payout, Impact factor and eligible papers. Only double-treated, and only when compared to other MDs or to the pure control group, display positive and significant coefficients. The only peculiar pattern is observed for the case of the Delta between actual publications and self-reported eligible documents. There we observe a null effect of being MBO irrespective of the selected comparison, while both MBO and DT units display higher deltas with respect to IRCCS-only and pure control. On the other side, the double-treated ones seem to have lower deltas with respect to both MBO-only and control units. In such regard, it would seem like that the self-reporting bias notwithstanding, MDs still do not put in place a gaming strategy with the aim of perceiving the bonus, and may actually value quality over quantity in the scope of an incentive structure which is less oriented towards research activities.

9 Conclusions

This paper examined two concurrent strategies to trigger research activity in a healthcare institution: direct monetary incentives based on performance for individual, non-academic, physicians versus increased public funding linked to institutional support. Using individual-level data from a major private Italian hospital, we found that the performance-based monetary reward (MBO) did not lead to significant improvements in research output of clinicians, who previously had limited publication activity, and were the main target of said intervention. The MBO seemed to spur a positive impact only on those MDs who also were involved in the research activity financed through the new accessible funding. As a matter of fact, the increase of public research funds through IRCCS recognition produced a huge boost in the productivity of the academic physicians included in the research perimeter. Our findings highlight the relative importance of incentives in driving research productivity, which can be apparently improved through individual rewards only in presence of adequate, material resources. As non-academic doctors did not responded as strongly as expected, when incentivized to do so by monetary gains, our research shows instead how the different incentive structures embedded in the main groups who make up the hospital personnel under question play a major role in shaping research output patterns; indeed, the (substantially) increased number of works published by individuals already affiliated to the academic institution, seem to prove so. The findings seem to demonstrate that lack of proper funding was the main obstacle to academic research engagement, and that mere monetary rewards are not able to narrow the gap between the objectives of non-structured researchers and those who account on that to progress in their professional careers. We also document that the combination of resources and incentives may nonetheless yield interesting outcomes with respect of cross-group collaboration: as better funding allows scientific opportunities, individual incentives should ensure such opportunities to be exploited. However, relatively speaking, it looks like such opportunities have been mostly seized by less prolific researchers with respect to how structured scientists got involved in cross-group collaborations, possibly highlighting a scope for free-riding spillovers. However, the gain from such collaborations ought not to be underestimated in the long-run, as they could improve the scope of the synergies required to fostered the translational impact of medical research (i.e., clinical intuitions shaping academic inquiry and the opposite). Importantly, we find ambiguous effects on research quality. The increase in research output went along by a major boost in overall citation impacts for IRCCS researchers, and a substantial decrease for MBO-treated physicians, even in relevant terms. However, the direction this effect was evidently

driven by the shift in the raw numbers of scientific works: as a matter of fact, we observe a persistent decrease in the number of discounted citations per paper for IRCCS-researchers after the policy, and a positive, a bit lagged but significant trend in the average (discounted) impact of the publications of MBO-incentivized physicians, underlining their greater value for quality over quantity, which may be caused by either improvement of intrinsic attitudes towards valuable research, or by gaming strategies elaborated in concomitance with the bonus scheme designed by the management, which takes into account the impact factor of publishers. Moreover, the policies appeared to foster greater collaboration between clinicians and academics, which could further enhance the translational value of the research . This integration of efforts is particularly valuable in healthcare, where bridging the gap between frontline practice and research can accelerate innovation and implementation of findings.

This paper contributes to policy discussions on how to stimulate innovation in public institutions and healthcare organizations. It also offers evidence that targeted performance incentives can only be an effective tool to unlock research potential, in presence of *actual* resources being provided to those who require them. For hospital administrators and policymakers, the results may imply that investing in a culture of research via both feeding long-term professional incentives relying on career progression and scientific curiosity (properly rewarded) may be more effective than short-run performance-based remuneration. It is indeed true that incentives ought to be aligned with multifaceted targets to avoid strategic behaviors which may possibly bring about detrimental societal repercussions. The main limitation of this study is its focus on a single (albeit large and relevant) institution, over a horizon which is long enough to observe individual behaviors, but not wide enough to capture the full unfolding of long-term research outcomes. Longer-term consequences on research quality, career trajectories, and patient outcomes, although beyond the current scope of this article, should be the main target for for future related study. On top of that, while our quasi-experimental strategy helps disentangling the effects of the policies (and across strongly comparable individuals), broader studies able to better frame the external validity of our estimates to different conditions, such as public healthcare facilities or geographical heterogeneities, should be pursued. In any case, our results offers a compelling case that, within an imperfect market environment like the healthcare one, managerial triggers on incentives can only play an important role in fostering innovation if properly traded off with the necessities required by the need for proper funding.

Appendix

Target	MDs without faculty structured appointments
Object	Full papers published between Jan 1 and Dec 31 (each year)
Payment per IF point	€500 per IF point
Max reimbursable IF	20 IF points per year
Max annual gross	€10,000
Author role (institution affiliation)	**Bonus share**
First or only author from the institution	100%
First author with other institutional co-authors	60%
Co-author on a paper first-authored by an institutional member	40%
Co-author on a paper first-authored externally	50%

- **IF** = Impact Factor of the Journal.

Table A1: Description of the MBO-rewarding scheme.

Requirement	Description	Evaluator
Scientific excellence	Proven research productivity in a specific biomedical area	Scientific Committee (Ministry of Health)
Healthcare quality	High-level and specialized clinical care	Ministry of Health / AGENAS
Integration	Strong integration between research and healthcare activities	Ministry of Health / AGENAS
Qualified personnel	Presence of qualified scientific and healthcare staff	Ministry of Health
Infrastructure	Adequate facilities and technological resources	AGENAS
External evaluation	Periodic assessment of all requirements	Ministry of Health
Recognition validity	Five-year renewable accreditation	Ministry of Health

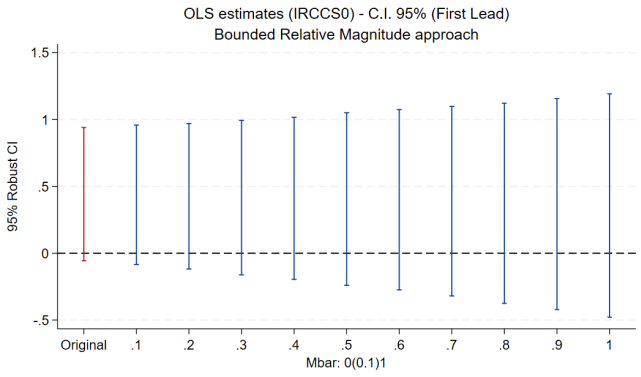
Table A2: Description of the IRCCS' requirement to achieve recognition (Source: DL 288/2003).

2012-2022; SEs clustered at individual level

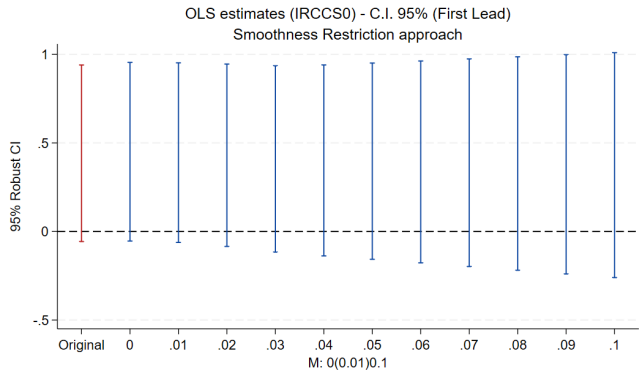
	(1) Publications	(2) Publications	(3) Publications	(4) Publications	(5) Publications	(6) Publications	(7) Publications	(8) Publications
Post2017								0.06839
								[0.17762]
IRCCS								2.97093***
								[0.40128]
Post2017 × MBO	-0.49848***			-0.10817			-0.00719	-0.11682
	[0.19260]			[0.12208]			[0.09933]	[0.27715]
Post2017 × IRCCS							0.41295**	0.28231
							[0.18272]	[0.24031]
MBO × IRCCS		0.72602*						-1.83172***
		[0.42276]						[0.47509]
Post2017 × MBO × IRCCS			0.81178**			0.29829	0.03520	0.30768
			[0.39721]			[0.24884]	[0.30633]	[0.41861]
Observations	1,709	1,709	1,709	1,737	1,737	1,737	1,737	1,709
R-squared	0.37898	0.37871	0.37898	0.91201	0.91195	0.91212	0.91278	0.46120
Individual FE	NO	NO	NO	YES	YES	YES	YES	NO
Year FE	NO	NO	NO	YES	YES	YES	NO	NO
Covariates	YES	YES	YES	NO	NO	NO	YES	YES
Method	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Time Range	2016-2018	2016-2018	2016-2018	2016-2018	2016-2018	2016-2018	2016-2018	2016-2018
Panel	Balanced	Balanced	Balanced	Balanced	Balanced	Balanced	Balanced	Balanced
Mean	1.671	1.671	1.671	1.671	1.671	1.671	1.671	1.671

*** p<0.01, ** p<0.05, * p<0.1

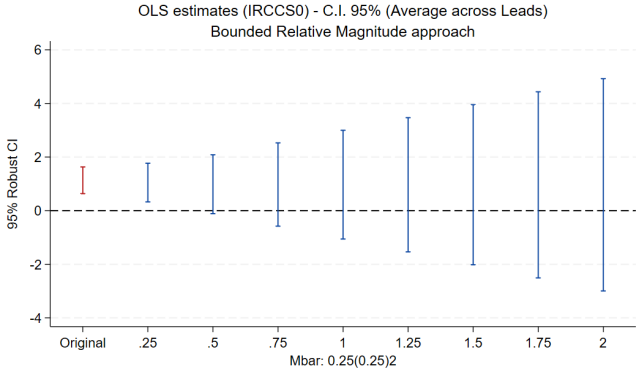
Table A3: Triple Diff-in-diff between 2016-2018, comparing different groups.



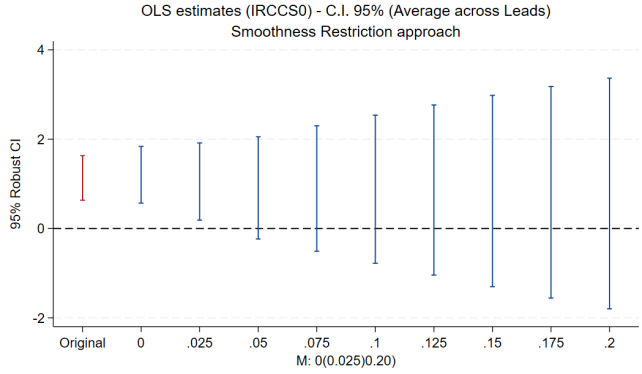
a) BM approach for the significance of the first lead.



b) SR approach for the significance of the first lead.

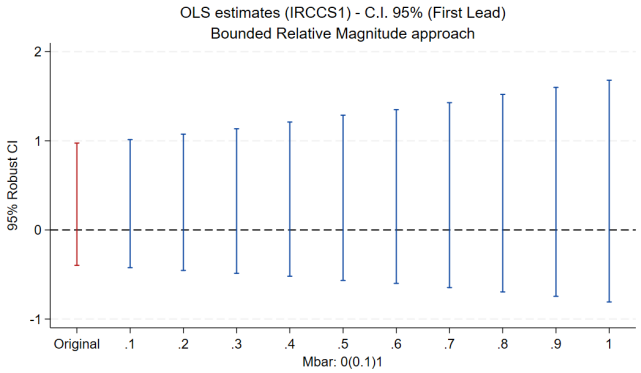


c) BM approach for the significance of the average across all leads.

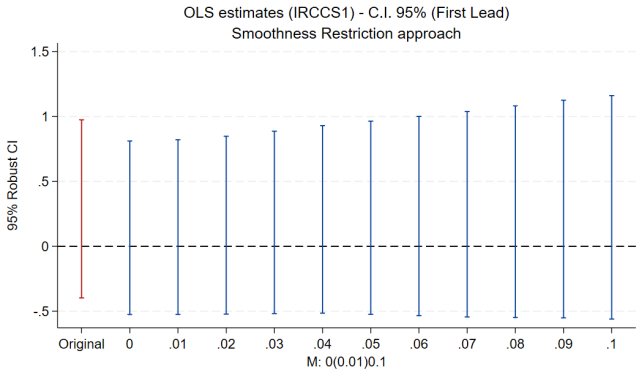


d) SR approach for the significance of the average across all leads.

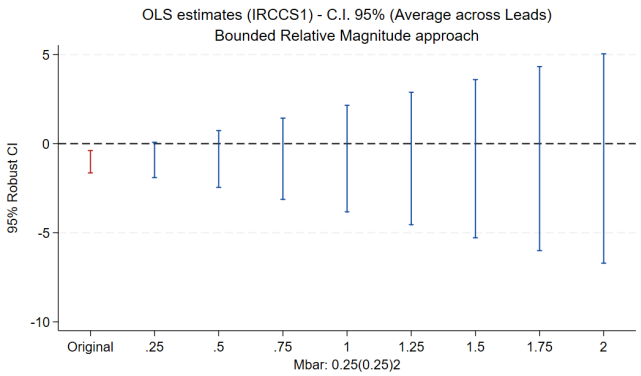
Figure A1: Honest DiD robust confidence sets in the comparison between double-treated and MBO-only units, estimated the or overall IRCCS effect by adopting with different methodologies (Bounded Relative Magnitude - BM: a) and c). Smoothness Restriction - SR: b) and d)) and for different quantities of interest (coefficient on the first lead of the event study, a) and b). Average across all post-treatment leads, c) and d)).



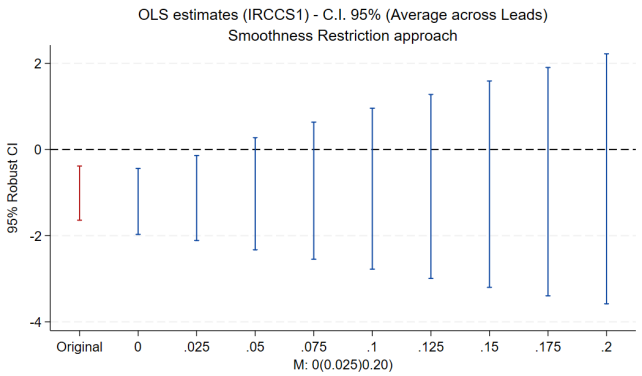
a) BM approach for the significance of the first lead.



b) SR approach for the significance of the first lead.

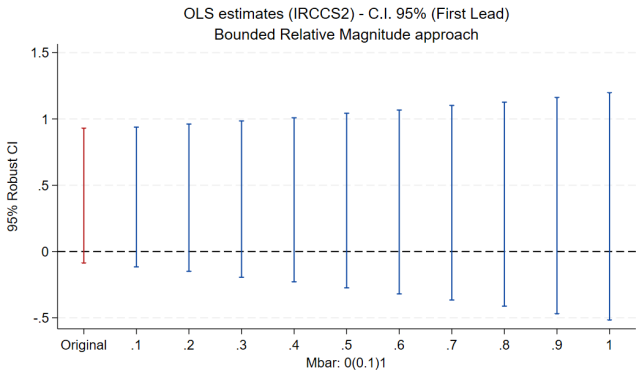


c) BM approach for the significance of the average across all leads.

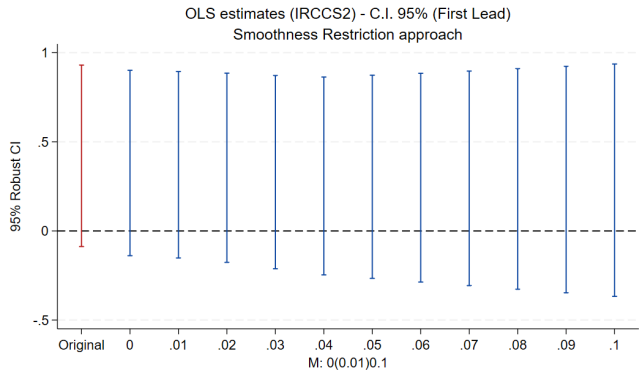


d) SR approach for the significance of the average across all leads.

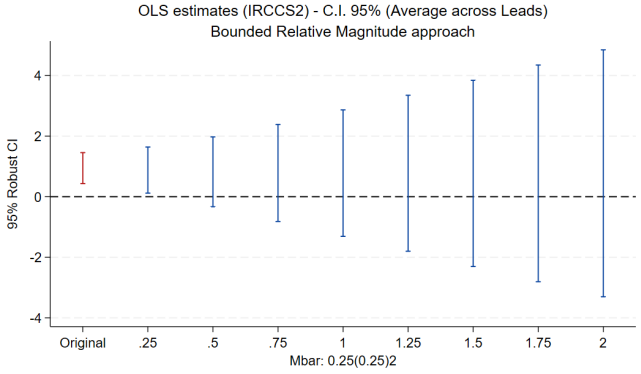
Figure A2: Honest DiD robust confidence sets in the comparison between double-treated and IRCCS-only units, estimated the or overall IRCCS effect by adopting with different methodologies (Bounded Relative Magnitude - BM: a) and c). Smoothness Restriction - SR: b) and d)) and for different quantities of interest (coefficient on the first lead of the event study, a) and b). Average across all post-treatment leads, c) and d)).



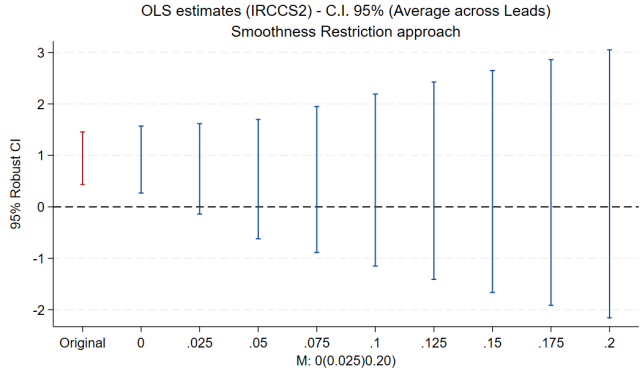
a) BM approach for the significance of the first lead.



b) SR approach for the significance of the first lead.

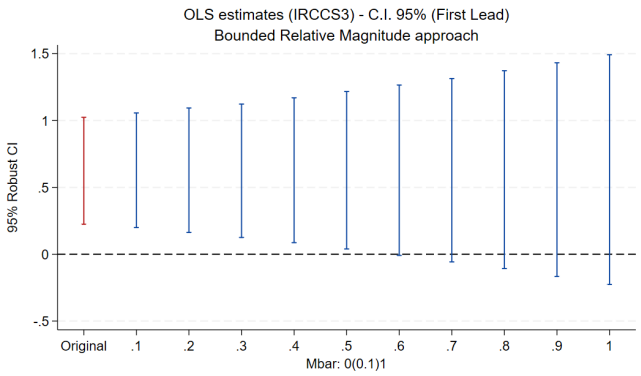


c) BM approach for the significance of the average across all leads.

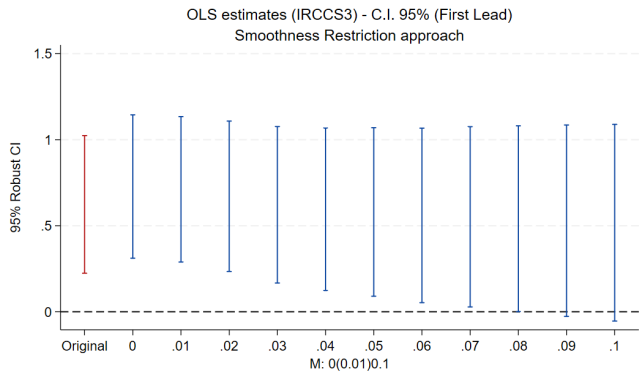


d) SR approach for the significance of the average across all leads.

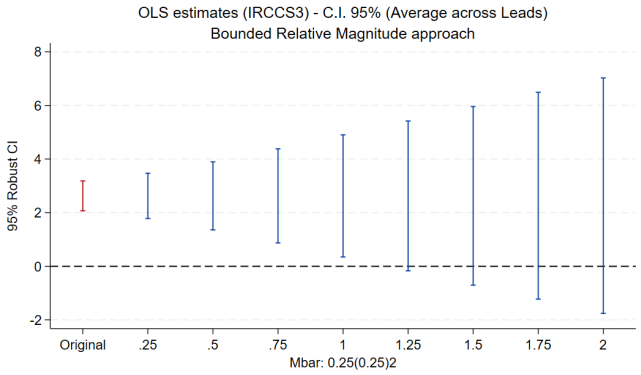
Figure A3: Honest DiD robust confidence sets in the comparison between double-treated and pure control units, estimated the or overall IRCCS effect by adopting with different methodologies (Bounded Relative Magnitude - BM: a) and c). Smoothness Restriction - SR: b) and d)) and for different quantities of interest (coefficient on the first lead of the event study, a) and b). Average across all post-treatment leads, c) and d)).



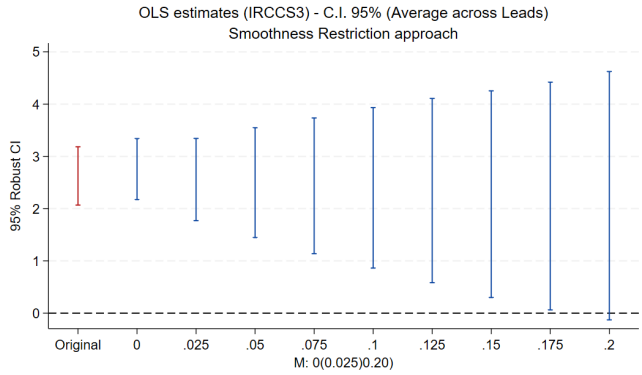
a) BM approach for the significance of the first lead.



b) SR approach for the significance of the first lead.

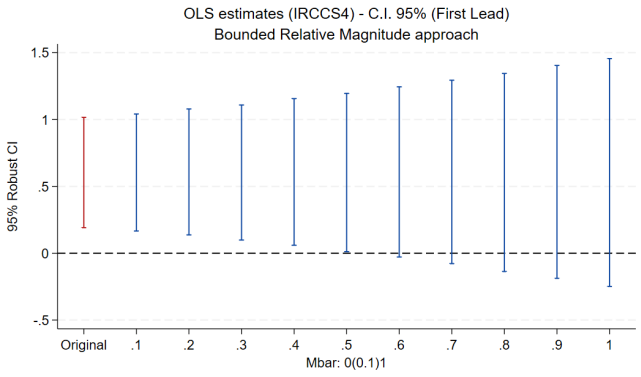


c) BM approach for the significance of the average across all leads.

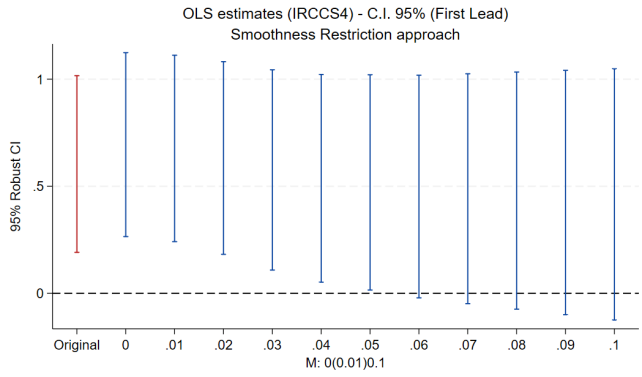


d) SR approach for the significance of the average across all leads.

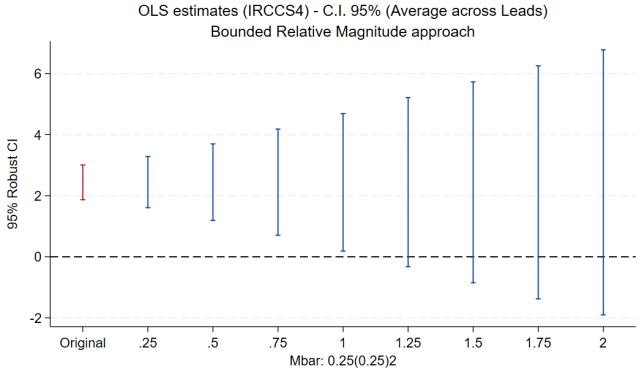
Figure A4: Honest DiD robust confidence sets in the comparison between IRCCS-only and MBO-only units, estimated the or overall IRCCS effect by adopting with different methodologies (Bounded Relative Magnitude - BM: a) and c). Smoothness Restriction - SR: b) and d)) and for different quantities of interest (coefficient on the first lead of the event study, a) and b). Average across all post-treatment leads, c) and d)).



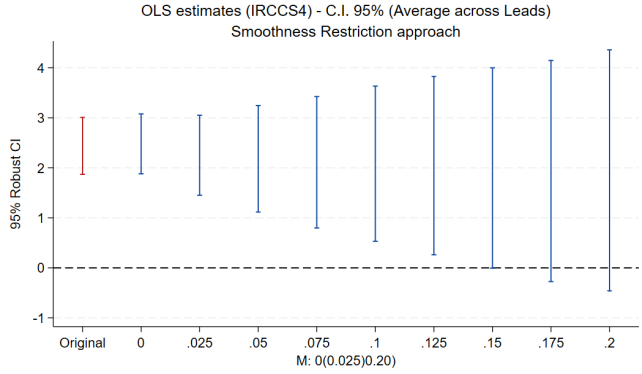
a) BM approach for the significance of the first lead.



b) SR approach for the significance of the first lead.

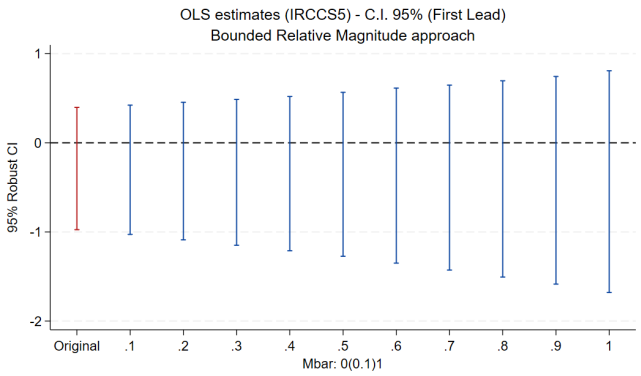


c) BM approach for the significance of the average across all leads.

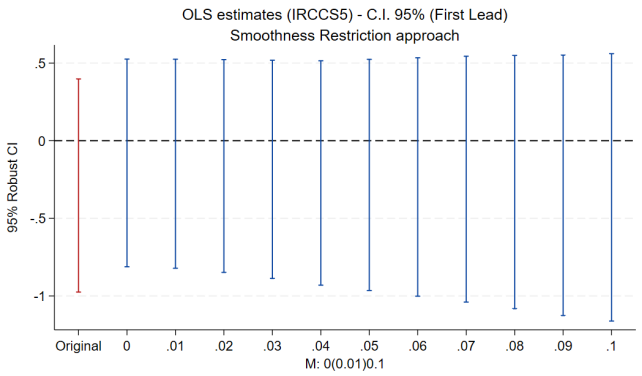


d) SR approach for the significance of the average across all leads.

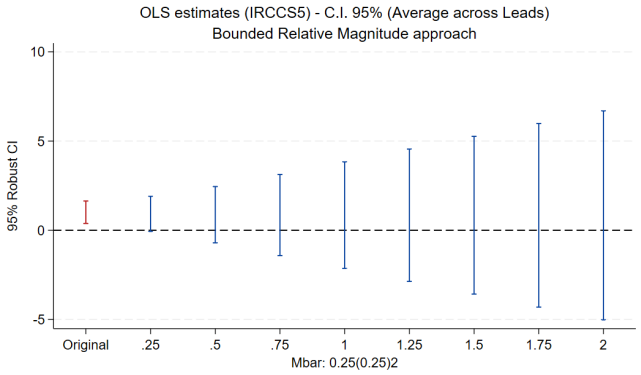
Figure A5: Honest DiD robust confidence sets in the comparison between IRCCS-only and pure control units, estimated the or overall IRCCS effect by adopting with different methodologies (Bounded Relative Magnitude - BM: a) and c). Smoothness Restriction - SR: b) and d)) and for different quantities of interest (coefficient on the first lead of the event study, a) and b). Average across all post-treatment leads, c) and d)).



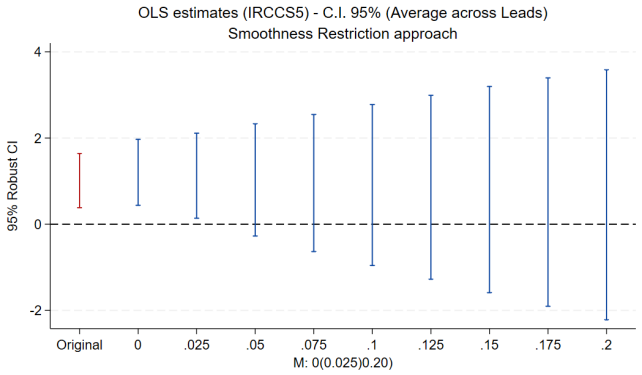
a) BM approach for the significance of the first lead.



b) SR approach for the significance of the first lead.



c) BM approach for the significance of the average across all leads.



d) SR approach for the significance of the average across all leads.

Figure A6: Honest DiD robust confidence sets in the comparison between IRCCS-only and double-treated units, estimated the or overall IRCCS effect by adopting with different methodologies (Bounded Relative Magnitude - BM: a) and c). Smoothness Restriction - SR: b) and d)) and for different quantities of interest (coefficient on the first lead of the event study, a) and b). Average across all post-treatment leads, c) and d)).

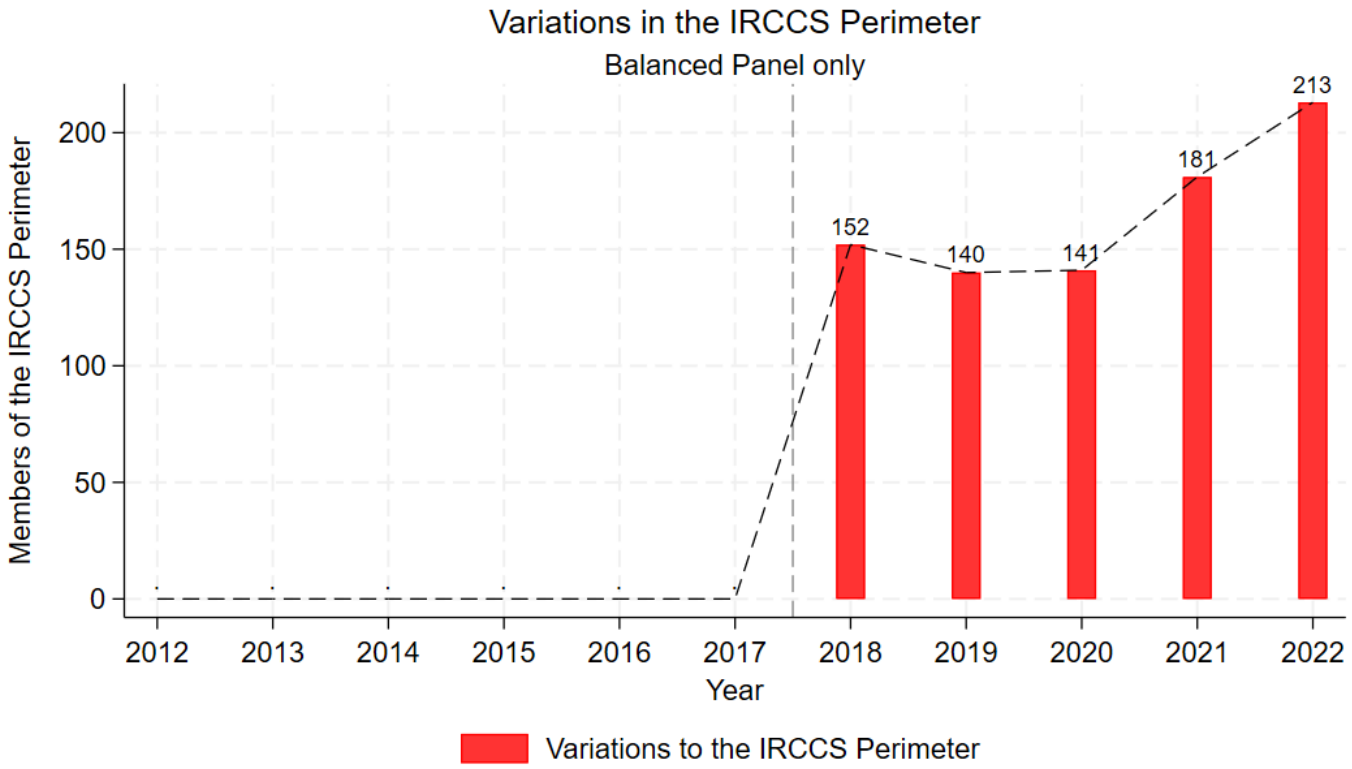


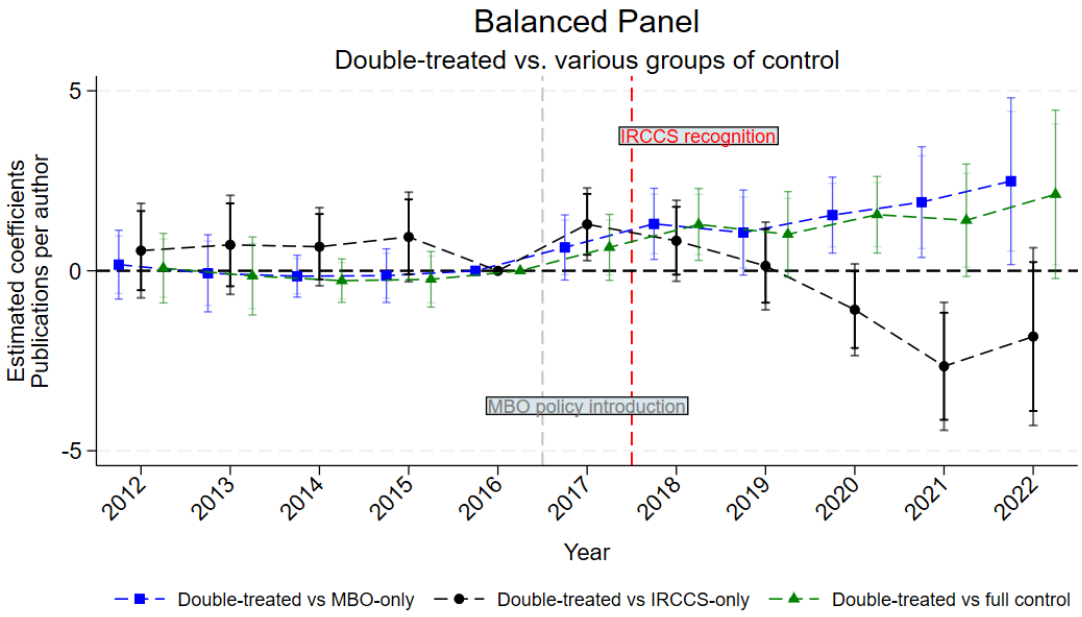
Figure A7: Evolution of the units "lately selected" into the IRCCS perimeter.

	(1) IRCCS vs. others	(2) DT vs MBO	(3) DT vs IRCCS	(4) DT vs PC	(5) IRCCS vs MBO	(6) IRCCS vs PC
Publications	3.0995***	1.4839***	-1.2850**	1.3989***	3.3732***	3.2621***
(SE)	(0.4064)	(0.4249)	(0.5220)	(0.4285)	(0.4363)	(0.4409)
N	5346	2229	1672	2047	3293	3111
R^2	0.790	0.621	0.768	0.598	0.809	0.799
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Mean	4.36	2.52	2.52	2.52	4.36	4.36
Panel	No late inclusion	No late inclusion	No late inclusion	No late inclusion	No late inclusion	No late inclusion
Time Range	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022

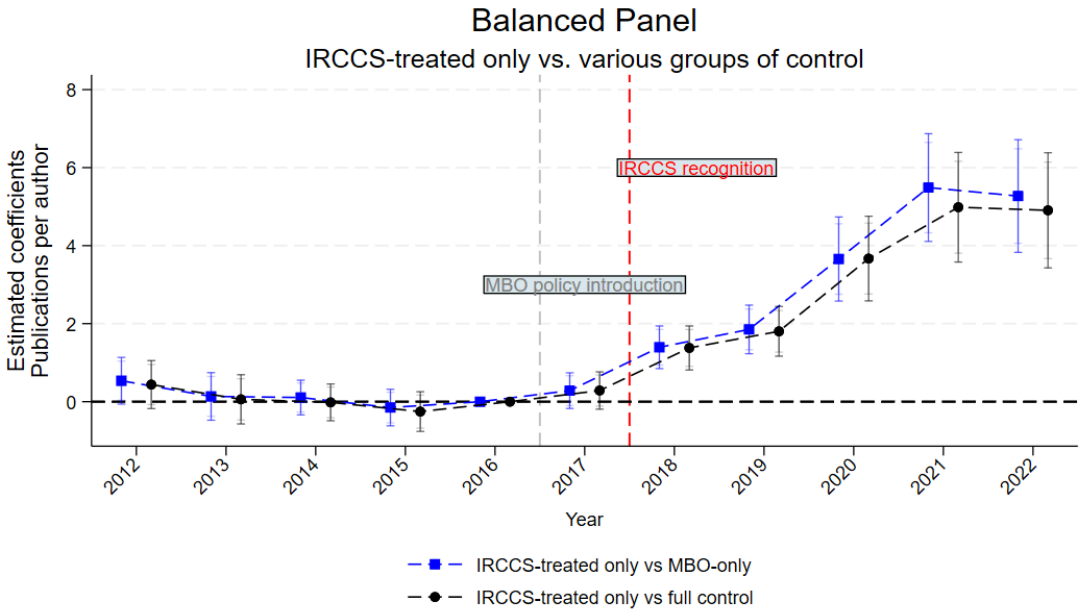
Table A4: Impact of IRCCS recognition on Annual Publications in the sample obtained after excluding units later included in the IRCCS perimeter (Difference-in-Differences).

	Core IRCCS	Late IRCCS	p-value
Age	53.0 (7.6)	50.7 (7.0)	0.020**
Publications	4.3 (5.6)	1.5 (2.1)	0.000***
Citations	185.3 (523.3)	48.5 (136.5)	0.014**
	Categorical variables (column %)		
Gender			
Female	29.6%	47.3%	
Male	70.4%	52.7%	
Status			
Medical Director	0.0%	0.0%	
Healthcare Professions Manager	19.7%	36.6%	
Sanitary Director	0.0%	0.0%	
Faculty Clinician	2.6%	3.2%	
Department			
Diagnostic Imaging	20.4%	7.5%	
Emergency & Anesthesiology	7.2%	7.5%	
Cardiovascular Sciences	9.9%	10.8%	
Women, Children & Public Health	13.2%	16.1%	
Lab Sci & Infectiology	7.9%	9.7%	
Gastro/Nephro/Endo	21.1%	11.8%	
Aging & Neurosciences	14.5%	35.5%	
Directorate General	0.0%	0.0%	
Clinical Governance	2.0%	0.0%	
Health Governance	3.9%	1.1%	

Table A5: Difference in characteristics (mean, SD in parentheses, and p-value of the t-test of the diff. in means) and composition (%) of the group of researchers included in the IRCCS perimeter in 2018 (early) and those enrolled later.



a) Event-study of IRCCS policy effect on publications of the double-treated units compared to the various control groups.



b) Event-study of IRCCS policy effect on publications of the IRCCS-only treated units compared to the various control groups.

Figure A8: Event-study of IRCCS policy effect on publications for comparisons across different groups. Units included in the IRCCS perimeter in any year after 2018 are excluded from the estimates.

	(1) IRCCS vs. others	(2) DT vs MBO	(3) DT vs IRCCS	(4) DT vs PC	(5) IRCCS vs MBO	(6) IRCCS vs PC
Publications (log)	0.2710***	0.3078***	-0.0063	0.2949***	0.3300***	0.3153***
(SE)	(0.0341)	(0.0595)	(0.0589)	(0.0602)	(0.0391)	(0.0402)
Publications (asinh)	0.3220***	0.3908***	0.0129	0.3751***	0.3964***	0.3784***
(SE)	(0.0429)	(0.0764)	(0.0750)	(0.0772)	(0.0495)	(0.0509)
N	6369	2600	2695	2418	3944	3762
R^2	0.750	0.604	0.722	0.586	0.801	0.771
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Panel	Full	DT and MBO	DT and IRCCS	DT and control	IRCCS and MBO	IRCCS and control
Time Range	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022

Table A6: Impact of IRCCS recognition on Annual Publications, re-scaled via log. and asinh.

	(1) MBO vs. non-MBO	(2) DT vs IRCCS	(3) DT vs MBO	(4) DT vs control	(5) MBO vs IRCCS	(6) MBO vs control
Publications	-0.1830*	-0.2991	0.6491***	0.6824***	-0.9483***	0.0333
(SE)	(0.1100)	(0.2742)	(0.2271)	(0.2299)	(0.1667)	(0.0740)
N	4632	1960	1920	1776	2856	2672
R^2	0.842	0.833	0.661	0.621	0.865	0.462
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Mean	0.65	1.60	1.60	1.60	0.31	0.31
Panel	Full	DT and IRCCS	DT and MBO	DT and control	MBO and IRCCS	MBO and control
Time Range	2012–2019	2012–2019	2012–2019	2012–2019	2012–2019	2012–2019

Table A7: Impact of MBO Incentive on Annual Publications in the sample truncated to exclude Covid years (Difference-in-Differences).

	(1) IRCCS vs. others	(2) DT vs MBO	(3) DT vs IRCCS	(4) DT vs PC	(5) IRCCS vs MBO	(6) IRCCS vs PC
Publications	0.8571***	0.6491***	-0.2991	0.6824***	0.9483***	0.9815***
(SE)	(0.1673)	(0.2271)	(0.2742)	(0.2299)	(0.1667)	(0.1704)
N	4632	1920	1960	1776	2856	2712
R^2	0.844	0.661	0.833	0.621	0.865	0.854
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Mean	3.61	1.64	1.64	1.64	3.61	3.61
Panel	Full	DT and MBO	DT and IRCCS	DT and control	IRCCS and MBO	IRCCS and control
Time Range	2012–2019	2012–2019	2012–2019	2012–2019	2012–2019	2012–2019

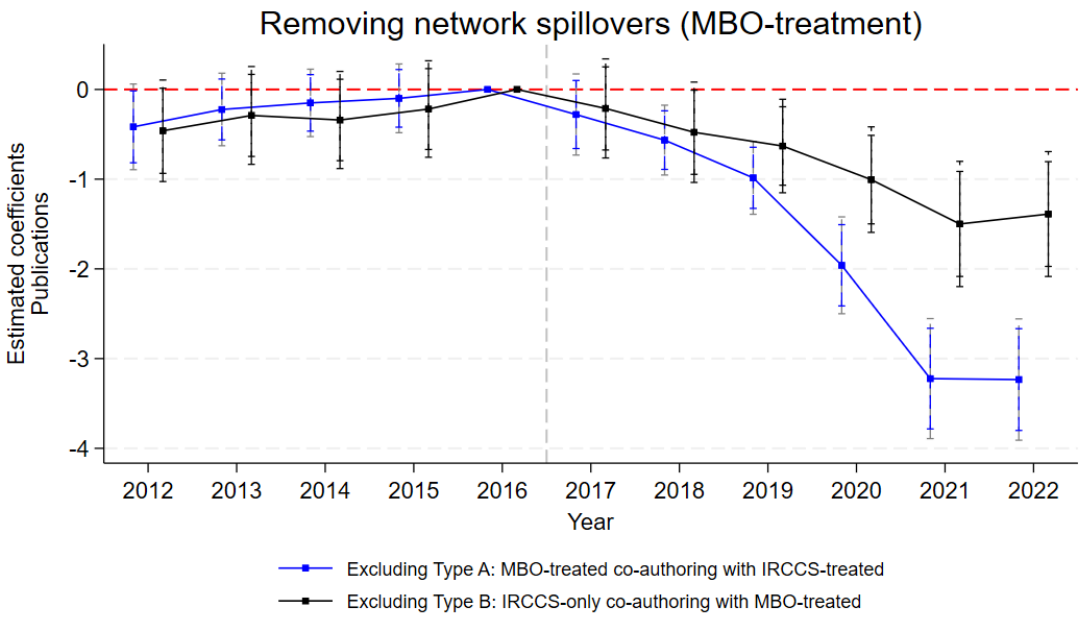
Table A8: Impact of IRCCS recognition on Annual Publications in the sample truncated to exclude Covid years.

2012-2022; SEs clustered at individual level

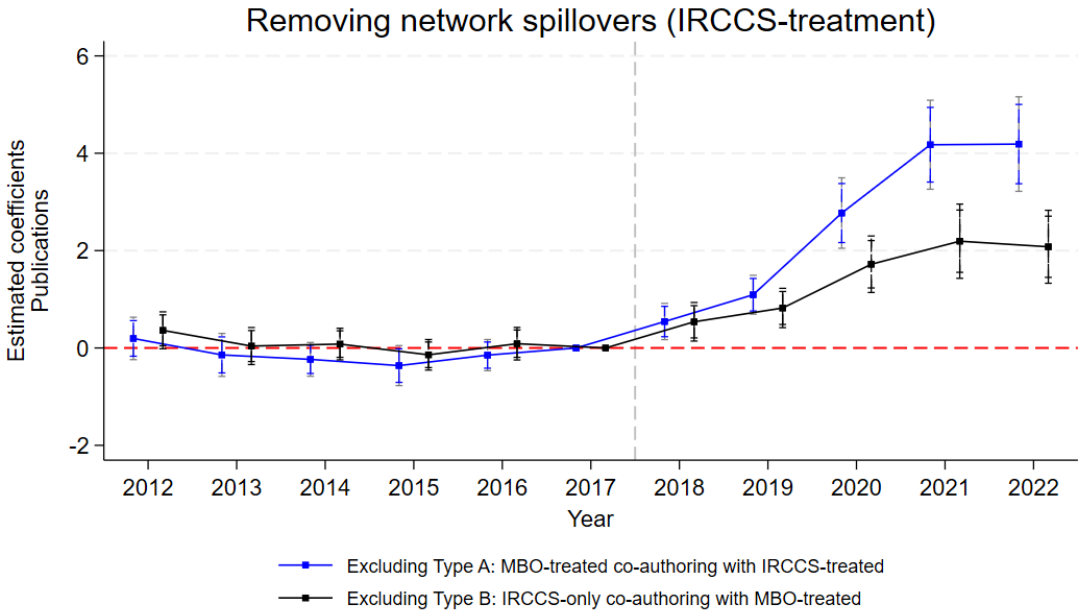
	(1) Publications	(2) Publications	(3) Publications	(4) Publications
Post 2017*MBO	-1.46107 ***	-0.59558 ***		
	[0.16033]	[0.11642]		
Post 2018*IRCCS			2.65333 ***	1.30822 ***
			[0.29579]	[0.17934]
Observations	5,410	5,176	5,410	5,176
R-squared	0.77285	0.69057	0.78582	0.70015
Individual FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Method	OLS	OLS	OLS	OLS
Time Range	2012-2022	2012-2022	2012-2022	2012-2022
Panel	No Type A	No Type B	No Type A	No Type B
Mean	0.0800	0.309	3.367	2.573

*** p<0.01, ** p<0.05, * p<0.1

Table A9: Impact of MBO-implementation and IRCCS recognition on Annual Publications in the sub-samples obtained by excluding collaborating units.



a) Event-study of the MBO policy effect on publications once excluded double-treated units and *Type A* researchers (MBO-treated units who *ever* co-authored with IRCCS researchers) in the first case, and double-treated units and *Type B* researchers (IRCCS-treated units who *ever* co-authored with non-academic physicians) in the second case.



a) Event-study of the IRCCS recognition effect on publications once excluded double-treated units and *Type A* researchers (MBO-treated units who *ever* co-authored with IRCCS researchers) in the first case, and double-treated units and *Type B* researchers (IRCCS-treated units who *ever* co-authored with non-academic physicians) in the second case.

Figure A9: Event-study of for the dynamic effects of the MBO policy (a) and the IRCCS recognition (b) effect on publications in different samples subset based on collaboration dynamics. Double-treated units are always excluded.

2012-2022; SEs clustered at individual level

	(1) Publications	(2) Publications	(3) Publications	(4) Publications	(5) Publications	(6) Publications
Post 2017*MBO	-0.99410 ***	-0.97558 ***	-0.98844 ***			
	[0.18123]	[0.17501]	[0.18214]			
Post 2018*IRCCS				2.45143 ***	2.36095 ***	2.45662 ***
				[0.31386]	[0.29975]	[0.31666]
Observations	6,182	6,336	6,149	6,182	6,336	6,149
R-squared	0.76641	0.76227	0.76607	0.77908	0.77413	0.77879
Individual FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Method	OLS	OLS	OLS	OLS	OLS	OLS
Time Range	2012-2022	2012-2022	2012-2022	2012-2022	2012-2022	2012-2022
Panel	No switchers	No switchers	Without	No switchers	No switchers	Without
Panel	out of MBO	into MBO	switchers	out of MBO	into MBO	switchers
Mean	0.547	0.653	0.547	3.609	3.591	3.591

*** p<0.01, ** p<0.05, * p<0.1

Table A10: Impact of MBO-implementation and IRCCS recognition on Annual Publications in the sub-samples obtained by excluding those who switch across groups.

	(1) MBO vs. non-MBO	(2) DT vs IRCCS	(3) DT vs MBO	(4) DT vs control	(5) MBO vs IRCCS	(6) MBO vs control
Publications	-0.5634***	-0.6030**	1.1212***	1.0332***	-1.8209***	-0.1068
(SE)	(0.1113)	(0.2510)	(0.2136)	(0.2189)	(0.1782)	(0.1001)
N	6116	2464	2573	2385	3724	3652
R^2	0.667	0.636	0.593	0.552	0.713	0.453
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Mean	0.62	1.53	1.53	1.53	0.30	0.30
Panel	No 99th pct	No 99th pct	No 99th pct	No 99th pct	No 99th pct	No 99th pct
Time Range	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022

Table A11: Impact of MBO-implementation on Annual Publications in the sub-sample obtained by excluding authors whose number of annual publication figures at least once in the top 1% of the yearly publication distribution.

	(1) MBO vs. non-MBO	(2) DT vs IRCCS	(3) DT vs MBO	(4) DT vs control	(5) MBO vs IRCCS	(6) MBO vs control
Publications	-0.0035	0.1259	0.5867***	0.6361***	-0.5265***	0.0338
(SE)	(0.0592)	(0.1666)	(0.1426)	(0.1444)	(0.1086)	(0.0624)
N	4290	924	2208	1889	2398	3366
R^2	0.453	0.448	0.421	0.438	0.487	0.388
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Mean	0.36	0.79	0.79	0.79	0.27	0.27
Panel	No 10th dc	No 10th dc	No 10th dc	No 10th dc	No 10th dc	No 10th dc
Time Range	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022

Table A12: Impact of MBO-implementation on Annual Publications in the sub-sample obtained by excluding authors whose number of annual publication figures at least once in the top 10% of the yearly publication distribution.

	(1) MBO vs. non-MBO	(2) DT vs IRCCS	(3) DT vs MBO	(4) DT vs control	(5) MBO vs IRCCS	(6) MBO vs control
Publications	0.0210	0.1888	0.4270***	0.4354***	-0.2724**	0.0012
(SE)	(0.0510)	(0.1704)	(0.1389)	(0.1409)	(0.1242)	(0.0539)
N	3476	462	1935	1527	1948	3014
R^2	0.363	0.438	0.356	0.356	0.383	0.329
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Mean	0.27	0.51	0.51	0.51	0.24	0.24
Panel	No 4th qt	No 4th qt	No 4th qt	No 4th qt	No 4th qt	No 4th qt
Time Range	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022

Table A13: Impact of MBO-implementation on Annual Publications in the sub-sample obtained by excluding authors whose number of annual publication figures at least once in the top 25% of the yearly publication distribution.

	(1) IRCCS vs. others	(2) DT vs MBO	(3) DT vs IRCCS	(4) DT vs PC	(5) IRCCS vs MBO	(6) IRCCS vs PC
Publications	1.6113***	1.1212***	-0.6030**	1.0332***	1.8209***	1.7253***
(SE)	(0.1426)	(0.2136)	(0.2510)	(0.2189)	(0.1782)	(0.1849)
N	6116	2573	2464	2385	3724	3536
R^2	0.686	0.593	0.636	0.552	0.713	0.686
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Mean	2.35	1.55	1.55	1.55	2.66	2.66
Panel	No 99th pct	No 99th pct	No 99th pct	No 99th pct	No 99th pct	No 99th pct
Time Range	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022

Table A14: Impact of IRCCS recognition on Annual Publications in the sub-sample obtained by excluding authors whose number of annual publication figures at least once in the top 1% of the yearly publication distribution.

	(1) IRCCS vs. others	(2) DT vs MBO	(3) DT vs IRCCS	(4) DT vs PC	(5) IRCCS vs MBO	(6) IRCCS vs PC
Publications	0.5603***	0.5867***	0.1259	0.6361***	0.5265***	0.5768***
(SE)	(0.0862)	(0.1426)	(0.1666)	(0.1444)	(0.1086)	(0.1108)
N	4290	2208	924	1889	2398	2079
R^2	0.465	0.421	0.448	0.438	0.487	0.481
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Mean	0.94	0.80	0.80	0.80	1.04	1.04
Panel	No 10th dc	No 10th dc	No 10th dc	No 10th dc	No 10th dc	No 10th dc
Time Range	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022

Table A15: Impact of IRCCS recognition on Annual Publications in the sub-sample obtained by excluding authors whose number of annual publication figures at least once in the top 10% of the yearly publication distribution.

	(1) IRCCS vs. others	(2) DT vs MBO	(3) DT vs IRCCS	(4) DT vs PC	(5) IRCCS vs MBO	(6) IRCCS vs PC
Publications	0.3449***	0.4270***	0.1888	0.4354***	0.2724**	0.2841**
(SE)	(0.0927)	(0.1389)	(0.1704)	(0.1409)	(0.1242)	(0.1261)
N	3476	1935	462	1527	1948	1540
R^2	0.370	0.356	0.438	0.356	0.383	0.388
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Mean	0.51	0.55	0.55	0.55	0.47	0.47
Panel	No 4th qt	No 4th qt	No 4th qt	No 4th qt	No 4th qt	No 4th qt
Time Range	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022	2012–2022

Table A16: Impact of IRCCS recognition on Annual Publications in the sub-sample obtained by excluding authors whose number of annual publication figures at least once in the top 25% of the yearly publication distribution.

	(1) MBO vs. non-MBO	(2) DT vs IRCCS	(3) DT vs MBO	(4) DT vs control	(5) MBO vs IRCCS	(6) MBO vs control
Pay Out (€)	-2224.22900***	-1919.22046***	2051.05811***	1726.63208***	-4006.25513***	-383.18332***
(SE)	(655.32159)	(309.17627)	(262.29660)	(272.93503)	(214.52034)	(129.73976)
R^2	0.539	0.659	0.482	0.492	0.690	0.376
Mean	496.20	1296.96	1296.96	1296.96	215.85	215.85
Sum of weighted IF	-2.56335	-13.43816***	4.57402***	4.50302***	-19.76748***	-0.19042
(SE)	(3.50900)	(4.16684)	(1.22695)	(1.04295)	(5.03401)	(0.78501)
R^2	0.306	0.323	0.169	0.466	0.320	0.141
Mean	1.44	3.21	3.21	3.21	0.82	0.82
Eligible Papers	-3.98309**	-5.97766***	4.63950***	3.71497***	-11.06847***	-1.09332***
(SE)	(1.90916)	(1.07426)	(0.46722)	(0.49784)	(1.07856)	(0.26383)
R^2	0.532	0.595	0.569	0.569	0.616	0.429
Mean	1.26	3.12	3.12	3.12	0.61	0.61
Delta (Pubs-Eligible)	2.36909*	4.79941***	-3.38431***	-2.56286***	8.40499***	0.87501***
(SE)	(1.21334)	(0.86698)	(0.40610)	(0.42853)	(0.84863)	(0.18826)
R^2	0.336	0.453	0.414	0.448	0.440	0.354
Mean	-0.34	-0.86	-0.86	-0.86	-0.16	-0.16
N	6369	2695	2600	2418	3944	3674
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Panel	Full	DT and IRCCS	DT and MBO	DT and control	MBO and IRCCS	MBO and control
Time Range	2017–2022	2017–2022	2017–2022	2017–2022	2017–2022	2017–2022

Table A17: Regressions between enrollment into IRCCS perimeter and several factors.

References

Acemoglu, D., & Restrepo, P. (2017). Secular Stagnation? The Effect of Aging on Economic Growth in the Age of Automation. *American Economic Review*, *107*(5), 174–179.

Ahrendt, D., Consolini, M., Mascherini, M., & Sàndor, E. (2022). *Fifth round of the living, working and covid-19 e-survey: Living in a new era of uncertainty.* (tech. rep.). Eurofound - Living conditions and family life.

Aksoy, Y., Basso, H. S., Smith, R. P., & Grasl, T. (2019). Demographic Structure and Macroeconomic Trends. *American Economic Journal: Macroeconomics*, *11*(1), 193–222.

AMS. (2020). *Transforming health through innovation: Integrating the nhs and academia* (tech. rep.). The Academy of Medical Sciences.

Azoulay, P., Graff Zivin, J. S., & Manso, G. (2011). Incentives and creativity: Evidence from the academic life sciences. *The RAND Journal of Economics*, *42*(3), 527–554.

Azoulay, P., Graff Zivin, J. S., & Wang, J. (2010). Superstar Extinction. *The Quarterly Journal of Economics*, *125*(2), 549–589.

Babina, T., He, A. X., Howell, S. T., Perlman, E. R., & Staudt, J. (2023). Cutting the Innovation Engine: How Federal Funding Shocks Affect University Patenting, Entrepreneurship, and Publications*. *The Quarterly Journal of Economics*, *138*(2), 895–954.

Banal-Estañol, A., Jofre-Bonet, M., Iori, G., Maynou, L., Tumminello, M., & Vassallo, P. (2023). Performance-based research funding: Evidence from the largest natural experiment worldwide. *Research Policy*, *52*(6), 104780.

Barrenho, E., Gautier, E., Miraldo, M., Propper, C., & Rose, C. (2025). Innovation Diffusion Among Coworkers: Evidence from Senior Doctors. *Management Science*.

Barrenho, E., Miraldo, M., Propper, C., & Walsh, B. (2021). The importance of surgeons and their peers in adoption and diffusion of innovation: An observational study of laparoscopic colectomy adoption and diffusion in England. *Social Science & Medicine*, *272*, 113715.

Baruffaldi, S. H., Marino, M., & Visentin, F. (2020). Money to move: The effect on researchers of an international mobility grant. *Research Policy*, *49*(8).

Benavente, J. M., Crespi, G., Figal Garone, L., & Maffioli, A. (2012). The impact of national research funds: A regression discontinuity approach to the Chilean FONDECYT. *Research Policy*, *41*(8), 1461–1475.

Bertoli, P., & Grembi, V. (2019). Malpractice risk and medical treatment selection. *Journal of Public Economics*, *174*, 22–35.

Bloom, D. E., Kuhn, M., & Prettner, K. (2022). Modern Infectious Diseases: Macroeconomic Impacts and Policy Responses. *Journal of Economic Literature*, *60*(1), 85–131
Cited By :75.

Bloom, D. E., Chen, S., Kuhn, M., McGovern, M. E., Oxley, L., & Prettner, K. (2020). The economic burden of chronic diseases: Estimates and projections for China, Japan, and South Korea. *The Journal of the Economics of Ageing*, *17*, 100163.

Bloom, N., Lemos, R., Sadun, R., & Van Reenen, J. (2015). Does Management Matter in schools? *The Economic Journal*, *125*(584), 647–674.

Bloom, N., Lemos, R., Sadun, R., & Van Reenen, J. (2020). Healthy Business? Managerial Education and Management in Health Care. *The Review of Economics and Statistics*, *102*(3), 506–517.

Bloom, N., Propper, C., Seiler, S., & Van Reenen, J. (2015). The Impact of Competition on Management Quality: Evidence from Public Hospitals. *The Review of Economic Studies*, *82*(2), 457–489.

Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting Event-Study Designs: Robust and Efficient Estimation. *The Review of Economic Studies*, rdae007.

Bosquet, C., Combes, P.-P., Henry, E., & Mayer, T. (2022). Peer Effects in Academic Research: Senders and Receivers. *The Economic Journal*, *132*(648), 2644–2673.

Brogaard, J., Engelberg, J., & Parsons, C. A. (2014). Networks and productivity: Causal evidence from editor rotations. *Journal of Financial Economics*, *111*(1), 251–270.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Kokot, J., & Wiesen, D. (2024). A new look at physicians' responses to financial incentives: Quality of care, practice characteristics, and motivations. *Journal of Health Economics*, *94*, 102862.

Burgess, S., Propper, C., Ratto, M., & Tominey, E. (2017). Incentives in the Public Sector: Evidence from a Government Agency. *The Economic Journal*, *127*(605), F117–F141.

Butts, K., & Gardner, J. (2022). Did2s: Two-Stage Difference-in-Differences. *The R Journal*, *14*(3), 162–173.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230.

Chandra, A., & Skinner, J. (2012). Technology Growth and Expenditure Growth in Health Care. *Journal of Economic Literature*, *50*(3), 645–680.

Checchi, D., De Fraja, G., & Verzillo, S. (2021). Incentives and Careers in Academia: Theory and Empirical Analysis. *The Review of Economics and Statistics*, 1–17.

Chen, J., & Roth, J. (2023). Logs with Zeros? Some Problems and Solutions*. *The Quarterly Journal of Economics*, qjad054.

Chen, S., Cao, Z., Nandi, A., Counts, N., Jiao, L., Prettner, K., Kuhn, M., Seligman, B., Tortorice, D., Vigo, D., Wang, C., & Bloom, D. (2024). The global macroeconomic burden of Alzheimer's disease and other dementias: Estimates and projections for 152 countries or territories. *The Lancet Global Health*, *12*(9), e1534–e1543

Cited By :18.

Chen, S., Prettner, K., Kuhn, M., & Bloom, D. E. (2021). The economic burden of COVID-19 in the United States: Estimates and projections under an infection-based herd immunity approach. *Journal of the Economics of Ageing*, *20*

Cited By :34.

Chen, T.-A., Lin, M.-H., Chen, Y.-C., & Chen, T.-J. (2024). The Time from Submission to Publication in Primary Health Care Journals: A Cross-Sectional Study. *Publications*, *12*(2), 13.

Colussi, T. (2018). Social Ties in Academia: A Friend Is a Treasure. *The Review of Economics and Statistics*, *100*(1), 45–50.

Cookson, C. (2025). Crick Institute says it needs extra cash to capitalise on US brain drain. *Financial Times*.

Dal Bó, E., Finan, F., & Rossi, M. A. (2013). Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service*. *The Quarterly Journal of Economics*, *128*(3), 1169–1218.

de Bengy Puyvallée, A., Storeng, K. T., & Rushton, S. (2025). The Gates Foundation's network diplomacy in European donor countries. *Globalization and Health*, *21*(1), 22.

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *The American Economic Review*, *110*(9), 2964–2996.

Fahy, N., Hervey, T., Dayan, M., Flear, M., Galsworthy, M., Greer, S., Jarman, H., & McKee, M. (2021). Assessing the potential impact on health of the UK's future relationship agreement with the EU: Analysis of the negotiating positions. *Health Economics, Policy, and Law*, 1–18.

Fernando, B., Brown, G., Thomas, E., Head, M., Nurse, P., & Rees, M. (2020). COVID-19 shows UK–EU collaborations are irreplaceable. *Nature*, *586*(7828), 200–200.

Franzoni, F., Lucifora, C., Russo, A. G., & Vigani, D. (2025). In the eye of the storm. Long-term impact of the Covid-19 pandemic on healthcare utilization in Lombardy. *Economics & Human Biology*, 101497.

Ganguli, I. (2017). Saving Soviet Science: The Impact of Grants When Government R&D Funding Disappears. *American Economic Journal: Applied Economics*, *9*(2), 165–201.

Gardner, J. (2022). Two-stage differences in differences. *arXiv*.

Garisto, D. (2025). Trump moves to slash NSF: Why are the proposed budget cuts so big? *Nature*.

Gaynor, M., Moreno-Serra, R., & Propper, C. (2012). Can competition improve outcomes in UK health care? Lessons from the past two decades. *Journal of Health Services Research & Policy*, *17 Suppl 1*, 49–54.

Gaynor, M., Moreno-Serra, R., & Propper, C. (2013). Death by Market Power: Reform, Competition, and Patient Outcomes in the National Health Service. *American Economic Journal: Economic Policy*, *5*(4), 134–166.

Ghandour, Z., Siciliani, L., & Straume, O. R. (2022). Investment and quality competition in healthcare markets. *Journal of Health Economics*, *82*, 102588.

Ghirelli, C., Havari, E., Meroni, E., & Verzillo, S. (2023). The Long-Term Causal Effects of Winning an ERC Grant. *IZA Discussion Paper No. 16108*, (JRC133001).

Glenza, J. (2025). Private firms are trying to fill research gaps, but their 'puny' budgets are no match for federal funds. *The Guardian*.

Goodall, A. H. (2011). Physician-leaders and hospital performance: Is there an association? *Social Science & Medicine (1982)*, *73*(4), 535–539.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277.

Grant, J., & Buxton, M. J. (2018). Economic returns to medical research funding. *BMJ Open*, *8*(9), e022131.

Gregory, A. (2025). NHS cancer patients denied life-saving drugs due to Brexit costs, report finds. *The Guardian*.

Gruber, J., & Owings, M. (1996). Physician Financial Incentives and Cesarean Section Delivery. *The RAND Journal of Economics*, *27*(1), 99–123.

Hacker, K. (2024). The Burden of Chronic Disease. *Mayo Clinic Proceedings: Innovations, Quality & Outcomes*, *8*(1), 112–119.

Hadavand, A., Hamermesh, D. S., & Wilson, W. W. (2024). Publishing Economics: How Slow? Why Slow? Is Slow Productive? How to Fix Slow? *Journal of Economic Literature*, *62*(1), 269–293.

Highman, L., Marginson, S., & Papatsiba, V. (2023). Higher education and research: Multiple negative effects and no new opportunities after Brexit. *Contemporary Social Science*, *18*(2), 216–234.

Huesmann, K., Ndiaye, Y. S., Waibel, C., & Wiesen, D. (2025). How the Design of Ranking Systems and Ability Affects Physician Effort. *Management Science*.

Jacob, B. A., & Lefgren, L. (2011). The impact of research grant funding on scientific productivity. *Journal of Public Economics*, *95*(9-10), 1168–1177.

JAMA. (2025). *For authors* (tech. rep.). JAMA Network Open.

Kessler, D. P., & McClellan, M. B. (2000). Is Hospital Competition Socially Wasteful?*. *The Quarterly Journal of Economics*, *115*(2), 577–615.

Lawless, J. (2023). The UK is rejoining the European Union's science research program as post-Brexit relations thaw. *AP News*.

Longo, F., Siciliani, L., Gravelle, H., & Santos, R. (2017). Do hospitals respond to rivals' quality and efficiency? A spatial panel econometric analysis. *Health Economics*, *26*(S2), 38–62.

Malakoff, D. (2025). NIH slashes overhead payments for research, sparking outrage and lawsuit. *Science*.

Mantellini, P., Battisti, F., Sassoli de Bianchi, P., Armaroli, P., Battagello, J., Giorgi Rossi, P., Giubilato, P., Senore, C., Ventura, L., Zorzi, M., & Zappa, M. (2020). *Rapporto sulla ripartenza degli screening - settembre 2020* (tech. rep.). Osservatorio Nazionale Screening.

Marchandot, B., & Morel, O. (2025). The Financialization of Healthcare in France: Trends and implications. *Public Health in Practice*, *9*, 100620.

Maynou, L., McGuire, A., & Serra-Sastre, V. (2024). What happens when the tasks dry up? Exploring the impact of medical technology on workforce planning. *Social Science & Medicine*, *352*, 117014.

Molitor, D. (2018). The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration. *American Economic Journal: Economic Policy*, *10*(1), 326–356.

Muralidharan, K., & Sundararaman, V. (2011). Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy*, *119*(1), 39–77.

Napolitano, A. (2025). Five years on: How Brexit changed three scientists' careers. *Nature*.

Peel, M., Bott, I., & Kuchler, H. (2025). The cost of Trump's attack on American science. *Financial Times*.

Propper, C., Burgess, S., & Green, K. (2004). Does competition between hospitals improve the quality of care?: Hospital death rates and the NHS internal market. *Journal of Public Economics*, *88*(7), 1247–1272.

Rambachan, A., & Roth, J. (2023). A More Credible Approach to Parallel Trends. *Review of Economic Studies*, *90*(5), 2555–2591.

Rousseau, S., Catalano, G., & Daraio, C. (2021). Can we estimate a monetary value of scientific publications? *Research Policy*, *50*(1), 104116.

Shekhtman, L. M., Gates, A. J., & Barabási, A.-L. (2024). Mapping philanthropic support of science. *Scientific Reports*, *14*(1), 9397.

Shurtz, I. (2013). The impact of medical errors on physician behavior: Evidence from malpractice litigation. *Journal of Health Economics*, *32*(2), 331–340.

Shurtz, I. (2014). Malpractice Law, Physicians' Financial Incentives, and Medical Treatment: How Do They Interact? *The Journal of Law and Economics*, *57*(1), 1–29.

Simpkin, V. L., & Mossialos, E. (2017). Brexit and the NHS: Challenges, uncertainties and opportunities. *Health Policy*, *121*(5), 477–480.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199.

Tang, B., Li, Z., Hu, S., & Xiong, J. (2022). Economic Implications of Health Care Burden for Elderly Population. *Inquiry: A Journal of Medical Care Organization, Provision and Financing*, *59*, 00469580221121511.

Terhune, C. (2025). NIH scientists speak out over estimated $12 billion in Trump funding cuts. *Reuters*.

UNUMLux. (2025). Our Helpline for American Scientists. *Lux Capital*.

Waldinger, F. (2010). Quality Matters: The Expulsion of Professors and the Consequences for PhD Student Outcomes in Nazi Germany. *Journal of Political Economy*, *118*(4), 787–831.

Waltman, L., & van Eck, N. J. (2013). A systematic empirical comparison of different approaches for normalizing citation impact indicators. *Journal of Informetrics*, *7*(4), 833–849.

Wu, P., & Wang, S. (2024). Examining the Economic Value of Investing in Health. *Journal of Systems Science and Information*, *12*(5), 575–589.

Ye, X., Wang, M., Xia, Y., He, P., & Zheng, X. (2023). Direct economic burden attributable to age-related diseases in China: An econometric modelling study. *Journal of Global Health*, *13*
Cited By :6.

Yohe, G. W. (1980). Current Publication Lags in Economics Journals. *Journal of Economic Literature*, *18*(3), 1050–1055.