

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

Gabriele Letta<sup>a</sup> Luca Salmasi<sup>a</sup> Gilberto Turati<sup>a</sup>

<sup>a</sup> Department of Economics and Finance, Università Cattolica del Sacro Cuore, Milan

August 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.

---

\*We kindly thank Paolo Candio, Vincenzo Carrieri, François Lafont, Matteo Lippi Bruni, Domenico Lisi, Fabio Montobbio, Francesco Nicolini, Carmine Ornaghi, Alessandro Pompa Pacchi, Rahmada Thaariq, Vrinda Sharma and the participants to the UPF Student Seminar Series, the III Sapienza Ph.D. Conference in Social Sciences and Economics, the 4<sup>th</sup> APHEC Conference, the Workshop on IPR, Science and Innovation at DEP - Unicatt, the XXXVII SIEP Conference for their helpful comments and suggestions.

# 1 Introduction

How should healthcare institutions incentivize physicians to engage in research? This question recently became relevant in light of the rising costs of healthcare systems, the demographic transition, and the tightening of public research budgets. In this paper we provide empirical evidence on two different approaches to foster research in healthcare: performance-based monetary bonuses (i.e., feeding individual incentives), and government funding allocations (i.e., addressing infrastructure and resources). We exploit the setting of a large private Italian teaching hospital, where both policies were subsequently implemented. In 2017 the hospital introduced a performance-based Management-By-Objectives (MBO) scheme, financially rewarding non-academic physicians for publishing papers. After that, in early 2018, the hospital was granted the IRCCS status (Scientific Research Hospital designation), thus allowing for additional public funding, mostly benefitting academic doctors. Through such combination we aim at comparing an explicit individual monetary reward to a wide institutional infrastructural intervention within the same organization and timespan.

The study can be framed within the general debate on the transformation of research funding. In the U.S., proposed budget cuts for 2025–26 have raised strong concerns among scientists, as federal agencies like the NIH and NSF have been facing impairment in their resources (Peel et al., 2025; Garisto, 2025; Terhune, 2025; Malakoff, 2025). On the other side, it is widely reckoned how private actors, like philanthropic associations and venture capital firms, increasingly stepped in (Glenza, 2025; UNUMLux, 2025; Shekhtman et al., 2024). This context has been marking a shift in funding, further away from the public institutions. The European research environment is all but alien to such issues: after Brexit, the UK experienced a decline in scientific spending (Simpkin and Mossialos, 2017; Highman et al., 2023; Gregory, 2025), whereas France and Germany saw the beginning of a debate concerning how to balance excellence, equity and independence in the presence of fiscal constraints and philanthropic actors (de Bengy Puyvallée et al., 2025; Marchandot and Morel, 2025). Such dynamics highlight a relevant change in science management, with a remarkable fading of the boundaries between private and public incentives.

Healthcare is particularly salient for this debate. Aging populations bring about greater utilization of healthcare (Tang et al., 2022), and also trigger major repercussions on economic growth and market conditions (Acemoglu and Restrepo, 2017, Aksoy et al., 2019). Healthcare cost utilization are also led by the growing prevalence of chronic illnesses and comorbidities (D. E. Bloom et al., 2020; Hacker, 2024,

Ye et al., 2023), whereas the COVID-19 shock have boosted pressure on health systems, calling for additional resilience to epidemic (S. Chen et al., 2021, D. E. Bloom et al., 2022, Wu and Wang, 2024). Innovation, research and their diffusion are major elements for healthcare sustainability, quality of care and long-term efficiency, but they require strong incentives and, often scarce, resource. While 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.

At the hospital level, research-active institutions may not only improve clinical standards, depending on innovation diffusion and translationality (Barrenho et al., 2021, 2025), but also attract high-quality professionals (AMS, 2020; N. Bloom, Propper, et al., 2015). Such institutes could also develop comparative advantages, ultimately favoring patient outcomes (N. Bloom, Propper, et al., 2015, N. Bloom et al., 2020, Ghandour et al., 2022). 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). Yet, outside academia, such *academic incentives* are missing, physicians are more oriented towards clinical task and other activities, thus facing weaker career incentives to conduct research. While this may avoid short-term concerns about research tasks potentially crowding out quality of care, this also dampens the effectiveness of translating research innovations into clinical practices and immediate interlinks between academia and patient care, therefore raising the policy question of how to effectively address the topic.

Beyond contributing to the thick debate on hospital management and incentive schemes in imperfect markets, as the one of education (Muralidharan and Sundararaman, 2011, N. Bloom, Lemos, et al., 2015) and healthcare (N. Bloom, Propper, et al., 2015, N. Bloom et al., 2020, Goodall, 2011), our study aims at shedding some light on whether performance-based pay schemes can effectively stimulate research productivity among non-academic physicians in comparable settings, and to what extent may institutional public funding generate greater or diverse effects. We also try to explore potential spillovers between groups of physicians—whether incentives targeting one group indirectly affect the research output and collaborations of the other.

Our work contributes to three strands of literature. First, we relate to the economics of management and incentives in public services (Goodall, 2011, Dal Bó et al., 2013, Burgess et al., 2017), by evaluating

a hospital-level policy that rewards research rather than clinical performance. As a matter of fact, there is a wide literature strand on fiscal incentives to physicians' activity, although focused on care tasks only (Gruber and Owings, 1996, Shurtz, 2013, 2014, Molitor, 2018, Bertoli and Grembi, 2019, Brosig-Koch et al., 2024). Second, we contribute to the literature on research funding and scientific productivity, assessing how public funds versus private incentives shape outputs in medical research. Such literature involves mixed findings, as some scholars display how 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); by contrast, others find weak or non-existent effects of grants on individual publications and productivity (Jacob and Lefgren, 2011, Banal-Estañol et al., 2023). Third, we engage with the literature on knowledge spillovers and peer effects in publications (Azoulay et al., 2010, Colussi, 2018, Bosquet et al., 2022), exploring whether incentives applied to one group of physicians trigger collaborative responses across groups.

Previewing our results, we find that the MBO scheme did not significantly increase research productivity among non-academic physicians, whereas the IRCCS recognition and associated public funds boosted the publication output of academic doctors by about  $\approx 60\%$ . The combination of the two policies appears cumulative, as non-academic physicians included in IRCCS funding are the only responding positively (and sizeably) to the MBO scheme. We also document that 1) such effects are feasibly driven mostly by funding allocated to operational and ordinary research activities; 2) that there is null increase in collaborative publications across groups (highlighting the relevance of within-group collaboration instead); 3) that there is an ambiguous effect on papers' citation, with academics increasing the bulks of their citations due to higher productivity, but with the MBO-eligible individuals showing a positive effect of citations per paper (the opposite displayed for the academics); this suggests that, although not focusing on raising productivity, non-academic physicians may value quality over quantity, as we cannot document the presence of gaming behaviors.

The remainder of the paper is structured as follows. Section 2 describes the institutional framework and the two policies. Section 3 presents data sources and sample construction. Section 4 presents dynamic estimates of both schemes, while Section 5 focuses on the impact of the IRCCS recognition. Section 6 reports robustness checks. Potential mechanisms are discussed in Section 7, while Section 8 treats the impact on networks, spillovers and research quality. Section 8 concludes.

## 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 Standard'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 whom 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 explicated 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)<sup>1</sup>. 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

---

<sup>1</sup>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.

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 who will not be included in the IRCCS perimeter (pure control group), 1 = those who eventually only 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		IRCCS levels of treatment	
	Description		Description
Policy	<i>Management-By-Objective (2017)</i>	Policy	<i>IRCCS recognition (2018)</i>
0	Full control (neither MBO nor IRCCS)	0	Full control (neither MBO nor IRCCS)
1	IRCCS-treated only	1	MBO-treated only
2	MBO-treated only	2	IRCCS-treated only
3	MBO+IRCCS double treatment	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 structurally employed in a professional capacity (physicians, researchers, and other healthcare staff management) 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. There were not available data on medical interns, residents, PhD students, and external collaborators, who are therefore excluded by definition. 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, institutional websites, LinkedIn pages and the official website of Italian physicians (*FNOMCeO*), 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, in our main analyses we keep the units who are balanced for the whole longitudinal dataset only. However, for the sake of robustness, we also constructed an unbalanced panel by collecting online information<sup>2</sup>. The resulting balanced panel covers 584 physicians and researchers, each observed annually from 2012 through 2022, yielding 6,424 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 the 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

---

<sup>2</sup>Retrieving online information was necessary to construct the unbalanced panel. As a matter of fact, the attrition in the panel after 2017 was easily obtainable by observing which professionals were yearly absent or present in the official records provided by the institution. But to trace their affiliation before 2017, while for most of them the hire date was an available information, for those no longer or not yet employed in either 2020 or 2023, such variable was missing. We hence collected online CVs, LinkedIn self-reported information, or data from institutional websites to reconstruct the affiliation. In some cases, especially for the oldest doctors, no CV or online info was available. We thence looked at their profile on the official online portal of Italian physicians (*FNOMCeO*), which reports date and place of birth, and the dates and places of their medical degree, their habilitation as physicians, their registration in the local medical books, and (if they did it), of their medical residency. We assume that MDs completing their last residency in the university affiliated to the hospital under study, or getting their habilitation immediately after completing their degree in said university if they did not specialize themselves, could be considered as enrolled in the institutional ranks that same year, as for most of our attrition subjects these events were reported happening decades before the starting year of our panel. In the few cases of inconsistency (no official CV information online, no LinkedIn profiles, a degree/residency completed in an other university or hospital), we sought for additional online information via scraping from Italian websites for reviewing physicians (*TopDoctors*, *IlMioDottore*, etc.) or online articles. The source material of this operation has been stored and may be made available upon request.

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 official IRCCS-research group, thus considering them as informally treated individuals expecting formal recognition at some point. We later 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 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 (and also on those in the unbalanced dataset). 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 on researchers' surnames only. Each publication record provides the publication year, journal, number of co-authors, citations (as of the time of data collection, which is Fall 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).

Table 2 reports the summary statistics for the selected sample. About 52% of the 584 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 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 0.46 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 not included in either policy, presents a slightly higher mean annual publication number than doctors eligible for MBO only (0.85). If we look at the times the works published in a given year were reportedly cited

in Fall 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 198 individuals (approximately 34% of the sample). Within the IRCCS group, 69 individuals are non-academic physicians (thus eligible for MBO as well). The non-academic non-IRCCS physicians are instead 178 (30.5% of the sample), while the pure control group is made by 162 individuals (27.7%).

	(1)			
	Mean	SD	Min	Max
Age	51.127	7.857	31	70
Publications	2.186	4.564	0	102
Times cited (WoS)	57.833	204.140	0	7037
Times cited (All outlets)	61.463	217.740	0	7509
Female	0.420	0.494	0	1
Male	0.580	0.494	0	1
Medical Director (hospital only)	0.409	0.492	0	1
Healthcare Professions' Manager	0.018	0.134	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	0.461	1.054	0	12
IRCCS-only treated publications	4.929	6.880	0	102
Double-treated publications	2.258	2.768	0	20
Pure control publications	0.848	1.683	0	23
MBO-only treated citations	8.537	33.919	0	1009
IRCCS-only treated citations	148.799	356.584	0	7509
Double-treated citations	66.691	144.069	0	2102
Pure control citations	15.540	37.684	0	485
MBO-only treated units				178
IRCCS-only treated units				198
Double-treated units				69
Pure control units				162
Total units of the balanced panel				584
Observations	6424			

Table 2: Descriptives

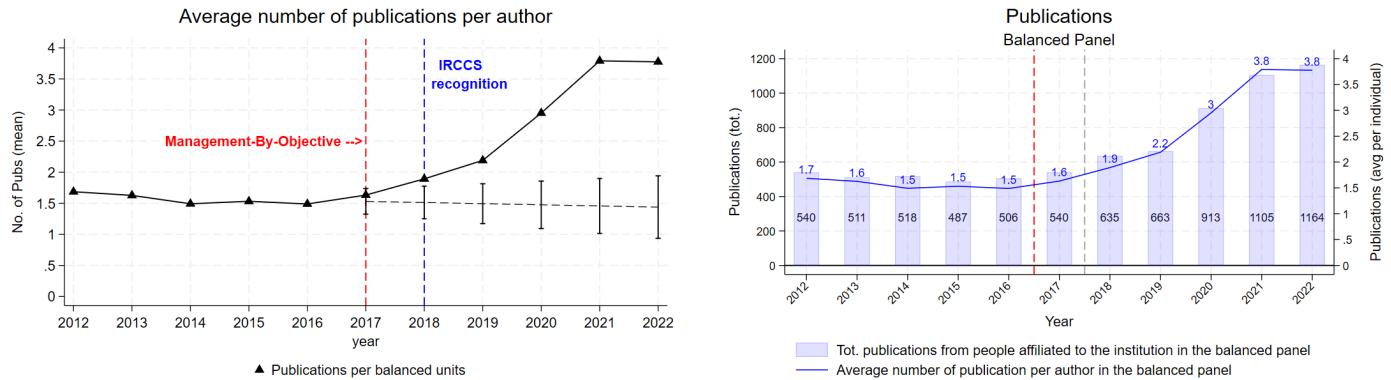
Graphs (a), (b), and (c) in Figure 1 plot the average number of publications per physician counted at the end of each year, from 2012 to 2022. In Graphs (a) and (b) we track the individuals in the balanced panel only, while in Graph (c) we compare the average trends of the balanced panel with several types of datasets with attrition, by removing from the panel all the individuals who are not counted for an increasing number of years. We observe a relatively flat trend in the years up until 2017, followed by a noticeable increase starting around 2018. This pattern is basically invariant in the comparing of

the balanced panel with the various possibility of attrition-flawed datasets the we can build with the available information (Graph (c)). The timing also aligns with the IRCCS recognition, suggesting, at a first glance, a possible aggregate effect of these policies on the overall research productivity of 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 less in the medical field, lowering to even a coupe-of-months time length 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).

On the other side, two 2017-published papers with two 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, one can even argue 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.

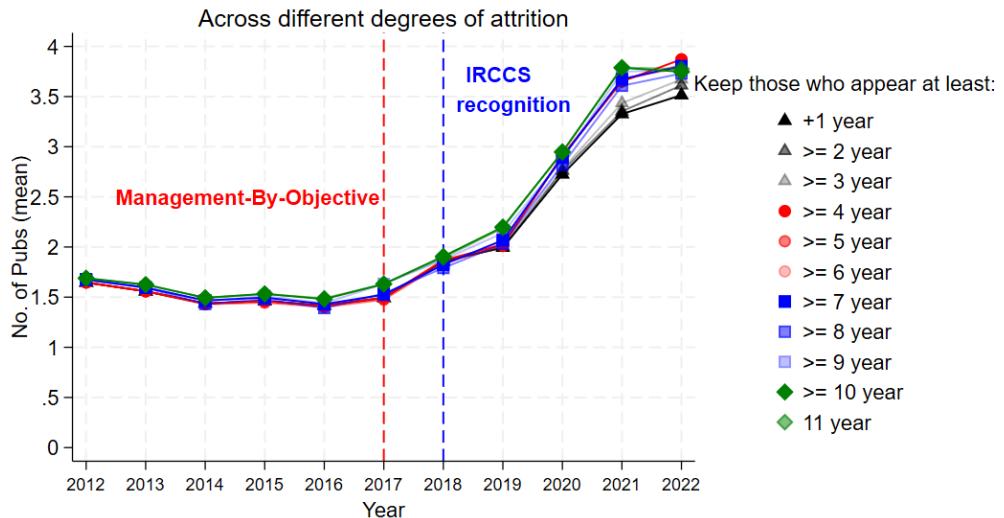
While other secular trends could also contribute to the increase (e.g., growing institutional emphasis on research), the positive, diverging pattern beginning in 2018 is quite suggestive. In the analysis below, we exploit the individual-level variation in exposure to the policies to identify their causal impact.



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

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

### Average number of publications per author (unbalanced)



c) Average publications per researcher per year (2012–2022) - Unbalanced Panel. Each line represents the mean trend observed by analyzing the dataset after having removed from the panel all the units who are not reported in the data for an increasing number of years (from 0, as in, the full dataset, to 11, which is the strongly balanced panel).

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

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 made by 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.

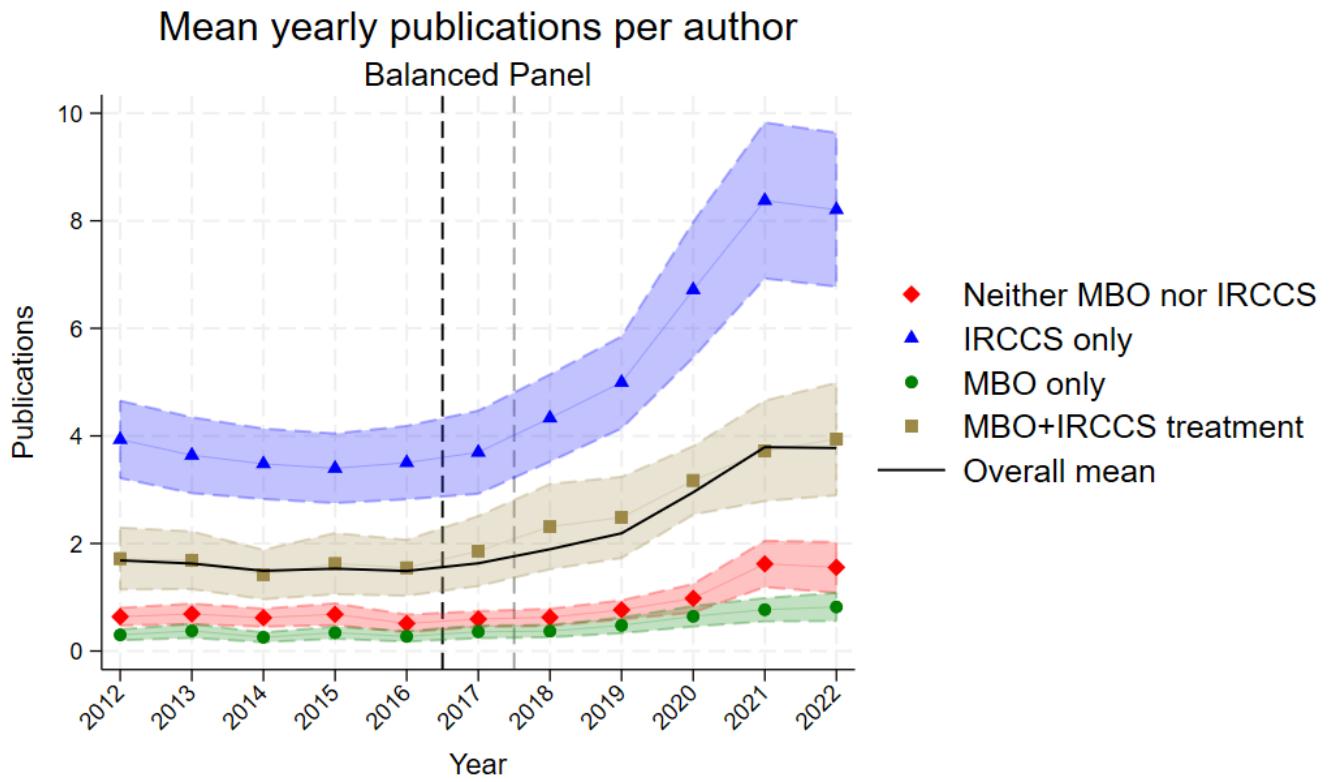


Figure 2: Average yearly publications across different groups of healthcare professionals, defined according their treatment status after 2017 and 2018.

## 4 Dynamic Estimates

### 4.1 Empirical Strategy

Although the policies are two different interventions, they are chronologically subsequent and share the same scope, so we assess them together. To observe which policy results to be more effective, we perform dynamic event-studies estimates stemming from a standard TWFE DiD framework. The first policy is the performance-based monetary incentive (MBO) implemented in 2017. The IRCCS recognition, which triggered an allocation of greater funds for research, began in 2018 and is the other policy of interest. 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 draw from Table 1 to identify the comparison groups, by blending into a unique differentiation scheme. We therefore have:

1. **Group 0:** *Full control*, as in, academic physicians, sanitary directors, sanitary professions' directors and other research staff officially affiliated to the hospital and/or the university, but who are not part of the IRCCS perimeter group. None of them is eligible to the MBO scheme either, as all Medical Directors (non-academic physicians) are excluded by this number;
2. **Group 1:** *MBO-only treated*. These are all the non-academic physicians who are not embedded in the IRCCS research staff, thus they benefit from the incentive but have no direct access to the additional public funds;
3. **Group 2:** *IRCCS-only treated*. As in, all academic individuals (faculty members with clinical functions, not Medical Directors) who are included in the IRCCS perimeter. Thus, they have direct access to dedicated funding, but are not rewarded explicitly due to their performance (no MBO);
4. **Group 3:** *Double-treated*. These are the non-academic physicians, hence benefitting from the MBO payments, who are also selected into the IRCCS staff by the hospital management, either because of their research merits or other reasons. Hence, they are rewarded for their productivity in addition to have straight access to IRCCS activities and resources.

We employ a TWFE difference-in-differences (DiD) dynamic framework, by comparing the change in publications across this different groups of individuals within the same model specification, as individuated by an indicator measuring the intensity of the treatment. Our category of reference is the full control group. Relative to time, we are interested into looking at the impact on publications before and after the introduction of the two policies, which occur in 2017 and 2018. Since the main analysis is dynamic, we do not have a time-invariant post-treatment framework where we compare the outcome before and after a fixed time threshold, like in the standard diff-in-diff, as we have two policy events. By contrast, we keep the year immediately preceding both policies (2016) and observe the evolution of the estimates in the years following the implementation of the policy changes.

We estimate the following equation:

$$Y_{it} = \alpha + \sum_{j \in \{1, 2, 3\}} \left\{ \sum_{h=2}^H \beta_{j,h} \mathbb{1}(G_i = j) \mathbb{1}(t = t_0 - h) + \sum_{g=1}^K \gamma_{j,g} \mathbb{1}(G_i = j) \mathbb{1}(t = t_0 + g) \right\} + \delta_i + \tau_t + \varepsilon_{it}. \quad (1)$$

where  $Y_{it}$  is the research output of individual  $i$  in year  $t$ .  $G_i \in \{0, 1, 2, 3\}$  indicates the comparison group to which the unit  $i$  belongs, with  $j = 0$  being the reference category (full control).  $j = 1$  is the MBO-only group,  $j = 2$  the IRCCS-only group, and  $j = 3$  the double-treated one. The binary indicator  $\mathbb{1}(G_i = j)$  takes then unitary value if  $i$  belongs to a given group of treatment, and 0 otherwise. By contrast, the temporal pre-period dummy  $\mathbb{1}(t = t_0 - h)$  indicates whether the  $t$  is a year preceding the policy, thus it equals 1 if  $t < t_0$ , and 0 otherwise. The interaction between the two dichotomous variables yields a further binary indicator which relates the treatment status to the pre-treatment periods, i.e., the *Lags*. The same, but on the opposite direction, holds for  $\mathbb{1}(t = t_0 + g)$ , which are dummy indicators for the post-treatment time periods, associated as well with the variables for the treatment status (the *Leads*). To avoid multicollinearity issues, we normalized the estimates by omitting  $t = 2016$ , as in the first pre-policy period, which is accounted as the reference temporal category. Stemming from the latter,  $h \in \{2, \dots, H\}$ , while  $g \in \{1, \dots, K\}$ . Note that  $t_0 = 2017$ , event though the IRCCS-intervention begins in 2018: this does not matter in the context of the analysis, as the estimates are normalized with respect to the same pre-treatment period, which is 2016.  $\delta_i$  are individual fixed effects,  $\tau_t$  are year fixed effects. By including  $\delta_i$ , we control for the time-invariant differences between physicians, while  $\tau_t$  for common shocks in each period. Standard errors are clustered at the individual level. These element yield coefficients for the differential trends across the different comparison groups in each year ( $\beta_{j,h}$  and  $\gamma_{j,g}$ ), which we plot to visually inspect pre-trends and dynamic effects. We present these event-study

results as figures.

The key identification assumption is that, absent the two policies, the research output of the various treated groups and the control units would have followed parallel trends. The event-study evidence enables to assess whether the different groups exhibited similar trends prior to 2017 (and 2017 itself, for the IRCCS-treated individuals).

## 4.2 Results

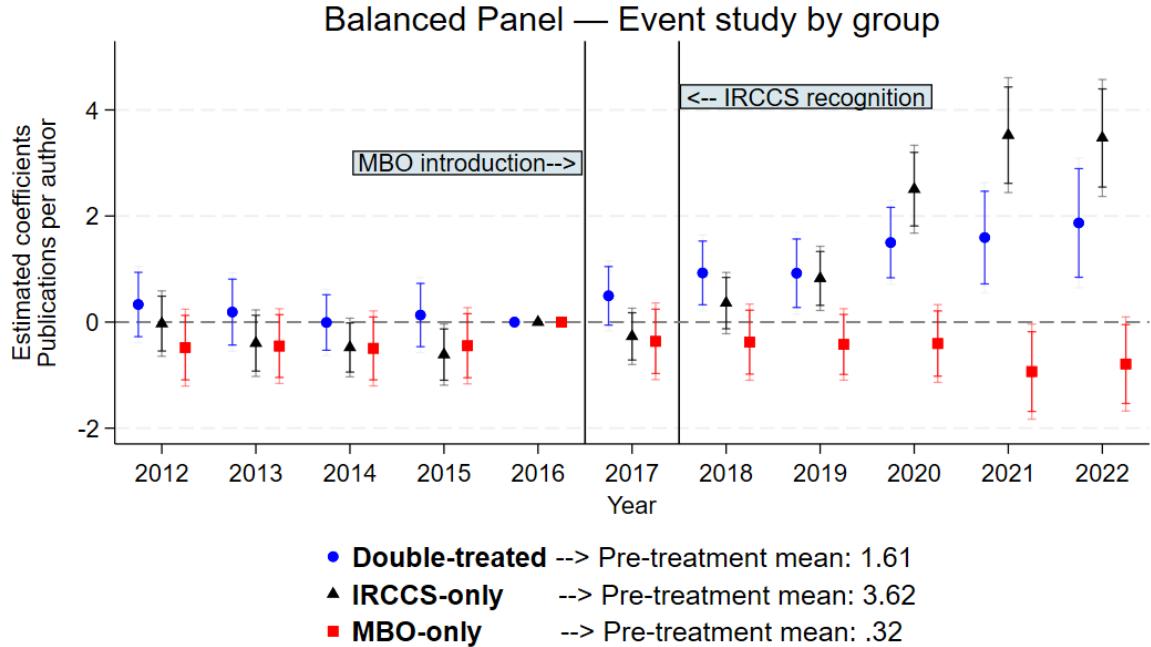


Figure 3: Event-study of the MBO joint with IRCCS recognition policy effect on publications across the different groups, and with respect to the full control group - Balanced panel.

Graph 3 presents the dynamic estimates of the effect of the two policies on publications. The graph reports the coefficients estimated through the Equation 1, distinguishing across the different comparison groups with respect to the reference category (e.g., full control). 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*). In addition to that, the effect seems to be kicking up only after 2018, which is the year of the IRCCS recognition. The most prominent impact is observed with respect to academic doctors who are also part of the IRCCS perimeter, for whom the effect takes off in 2019, one year after the IRCCS recognition, when they increase their outcome by about 1 publication (+28% circa compared to the pre-treatment mean), on par with the increased absolute outcome of the double-

treated individuals, although the relative magnitude of the increase in 2019, for the latter, is greater (+62%). Starting from 2020, in correspondence of the pandemic outbreak, the productivity of academic doctors jump up substantially, overtaking that of the double-treated unit and peaking in 2021, where they register almost a 100% increase. By contrast, the MBO-scheme alone seems to be associated with a slightly negative effect when not paired with the additional funding, although such an effect is not statistically significant.

Although the pre-treatment means are substantially different across all groups, the graph shows that the common trend assumption is validated for all the groups of comparison, and therefore the evidence that the MBO policy affects individuals only joint with IRCCS funding is corroborated even more, with the negative and significant gap relative to the control group between double-treated individuals and MBO-treated only ones widening significantly over time, starting from 2018. To show that the results are not driven by attrition, we report the event-study for the full unbalanced panel in the Appendix: Figure A1 proves that the results in case of attrition are consistent with our main findings, with just some slight difference in magnitude.

## 5 IRCCS Funding Analysis

Given the almost null effect of the MBO policy, and the substantial changed triggered by the additional IRCCS funding, we now assess the IRCCS recognition in a deeper fashion, to provide with more detailed quantitative measures of its impact on the individuals' research productivity. The recognition caused an allocation of greater funds for research, and 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-

### 5.1 Empirical Strategy

To identify the IRCCS recognition's effect, We estimate a baseline specification for a standard two-way fixed effects model, via OLS estimators:

$$Y_{it} = \alpha + \theta Post2018_t \times IRCCS_i + \delta_i + \tau_t + \varepsilon_{it}, \quad (2)$$

where  $Post2018_t$  equals 1 if  $t \geq 2018$  and 0 otherwise,  $IRCCS_i$  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  $2 \times 2$  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 for a more clear-cut identification of the effect, although by doing this our causal quantity of interest consists in an *Intention-to-Treat Effect* (ITT) rather than an *Average Treatment Effect on the Treated* (ATT), which is what is usually estimated in the best-specified Diff-in-Diffs analyses. 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 stays 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 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  $\theta$  captures the average intention-to-treat effect on the IRCCS recognition treatment of those who only receive funding after 2018, with respect to *all* MBO-eligible individuals and the full control group. In such set-up we compare IRCCS-only listed researchers with non-listed individuals and listed-but-non academic individuals eligible for MBO, 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 in the previous section. Since in such framework the issue of self-selection is of paramount relevance (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.

In addition to the main DiD strategy, we implement further sets of estimates. In 2018, the four identified categories (0, 1, 2, 3) are defined as in Equation 1 (0 = control, 1 = MBO-only, 2 = IRCCS-only, 3 = both). After reporting the baseline (findings from Equation 2), 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 2.

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. Note that the baseline comparison in (the one which follows) follows an approach which is quite “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):

$$Y_{it} = \alpha + \sum_{h=2 \wedge h \neq 1}^H \beta_{j,h} \mathbb{1}(IRCCS_i) \mathbb{1}(t = 2018 - h) + \sum_{g=1}^K \gamma_{j,g} \mathbb{1}(IRCCS_i) \mathbb{1}(t = 2017 + g) + \delta_i + \tau_t + \varepsilon_{it}. \quad (3)$$

Although it is common practice to exclude the time period immediately before the policy kicks in for the sake of collinearity, we recover the same approach of the previous event-study, therefore excluding 2016 as our reference category. This is done to enable better comparability between our different sets of estimates.

## 5.2 Results

We report the DiD estimates from Equation 2 in Table 3. 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 2. 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 2.32 additional publications per year (significant at 1%) compared to the other groups. Such increase is quite outstanding, as it accounts for more than 65% 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 compounded publish almost 1.24 more works when compared to the MBO-treated only (+75%), 1.10 fewer works when compared to the IRCCS-only researchers (-66%), and 1.14 more works when compared to the pure control (+69%). 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 3: in Col. (5), the effect of IRCCS recognition on IRCCS vs. MBO is almost +2.65 publications (+74%); in Col. (6), IRCCS publish 2.53 works more than the pure control (+69%).

	(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	2.3200***	1.2398***	-1.0989***	1.1365***	2.6469***	2.5329***
(SE)	(0.2917)	(0.2285)	(0.3337)	(0.2348)	(0.3091)	(0.3148)
N	6424	2623	2739	2430	3987	3794
R <sup>2</sup>	0.774	0.596	0.764	0.582	0.793	0.784
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Mean	3.56	1.65	1.65	1.65	3.56	3.56
Panel (Balanced)	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

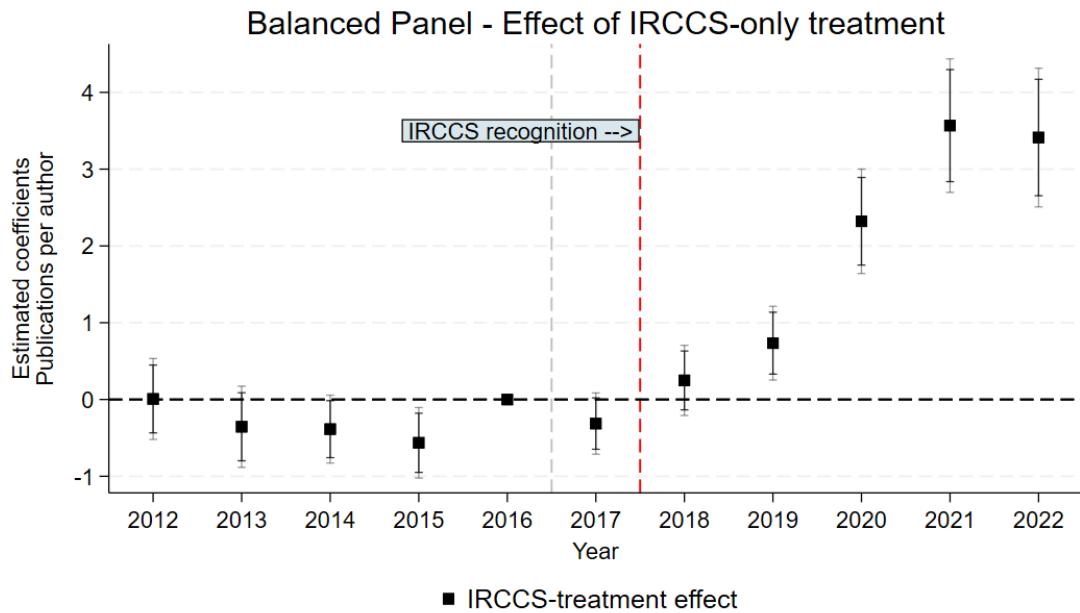
Table 3: Impact of IRCCS recognition on Annual Publications (Difference-in-Differences).

The baseline dynamic specification for the IRCCS effect is represented in the event-study a) 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 are professional researchers, 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 2019, reasonably due to the, existing 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.

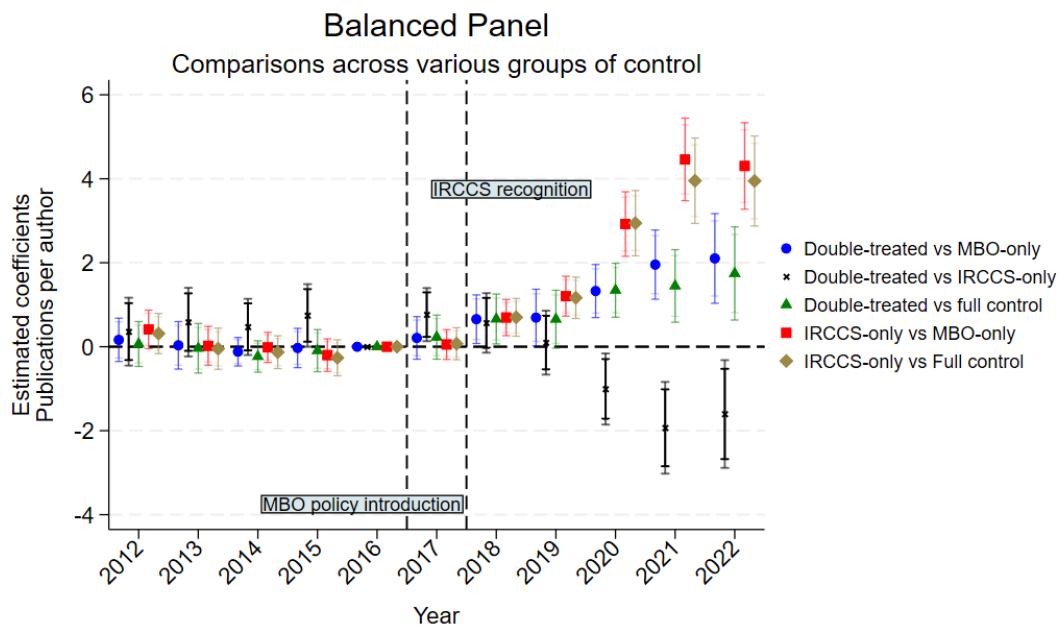
We reckon to have estimated dynamic coefficients based on an identification which was indeed conservative, but not clear-cut: as a matter of fact, some IRCCS' researchers included in the control group were non-academic, hence they were enjoying the performance-based bonus. On top of that, although we 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 estimations again, in doing this showing the graphic representations of the estimates foreseen by Equations 3, in Graph b) of Figure 4. The latter reports the comparisons between double-treated and IRCCS-only treated units and each of the other groups. Both graphs a) and b) 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. As we reported already four out main event studies, we show in the Appendix that the estimates for the balanced panel follow similar patterns (Figure A2).

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. We observe indeed a significant shift downwards in the publication output of IRCCS-only treated in comparison to the the other groups in 2015, and a non-significant one in 2017, hence immediately after the MBO implementation (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 the other individuals, with their outcome boosted by the access to new resources. Another relevant patters is that the double treated, when compared to academic physicians, display a slightly significant positive trend even before 2017. This is credibly due to the selection issue of non-academic doctors enrolled in the IRCCS staff due to their above-average productivity: we address this concern later on.

Overall, the IRCCS funding intervention appears to have had a substantial, fast and persistent effect on research output, fed up by the pandemic and reinforced, only for some individuals, by the performance-based scheme.



a) Event-study of IRCCS recognition effect on publications for IRCCS-only treated individuals compared with all the other units.



b) Event-study of IRCCS policy effect on publications for comparisons across different groups.

Figure 4: Event-studies of IRCCS policy effect.

The estimates for the IRCCS-only group suggest that the heterogeneity in incentive structures across the different groups can only modestly be offset by explicitly 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, that 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 other non-academic doctors. In any case, the presence of academics as well in the pure control groups confirms that the heterogeneity in pre-existing incentive behavioral structure may not only exist across groups, but within them as well.

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 monetary rewards, 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

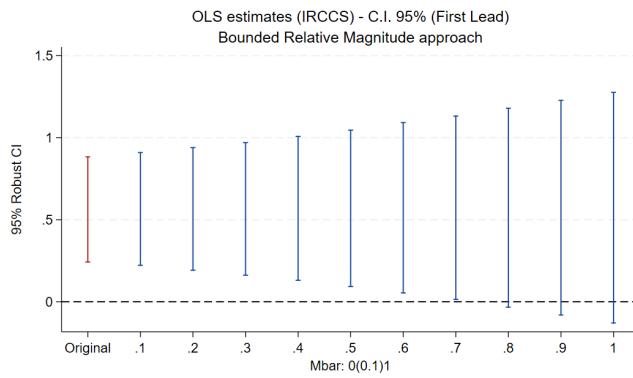
While the main estimates seem to display the absence of deviation from the parallel trends in all the provided specifications, we place additional focus on the pre-trends' issue, given the significance and magnitude of our results. This check is also required due to the fact that the treatment is endogenous 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 a) 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*). In order to facilitate the differentiation between pre- and post- periods, we shift back to the common practice in terms of collinearity issues and account for 2016 (as in,  $t = -1$ ) as the reference category.

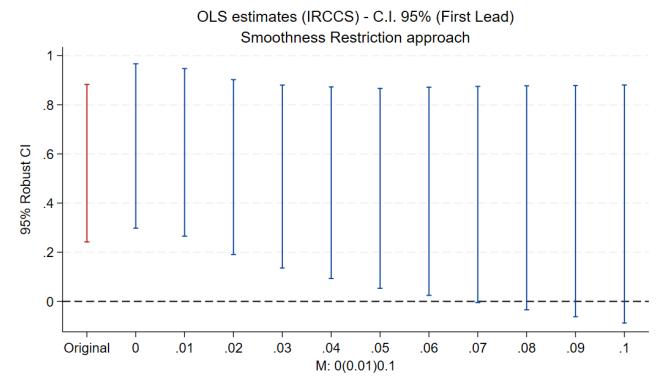
We provide evidence for the *honest* confidence bounds for the baseline estimates (Figure 5), using both approaches (BM in graphs a) and c); SR in graphs b) and d)), and by allowing variation in the common trends able to invalidate the significance of the coefficient on the first lead estimation, which results positive and significant when accounting for 2016 as the omitted year, (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 5. 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 allowed to be biased before becoming invalid. 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

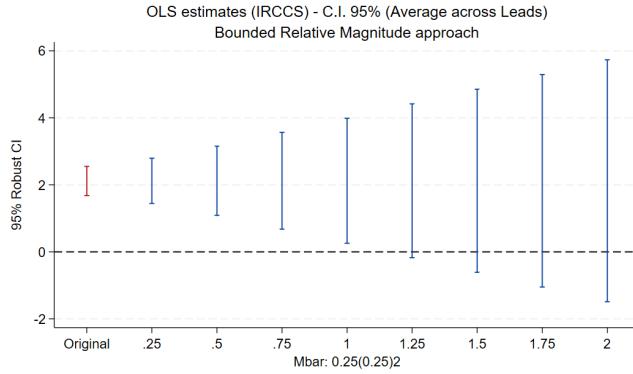
to the first lead (when the effect is lower compared to the rest of the post-treatment period, even non-significant in our plotted estimate, and it has not fully kicked in already), we can consider such results as robust. If we look at the average effect across all leads (as in, the ITT), we observe that the estimated coefficients are even more robust: indeed, they maintain their validity up to almost a 125% bounded relative magnitude violation, and up to a 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 comparison across groups in the Appendix, in Figures from [A3](#) to [A8](#).



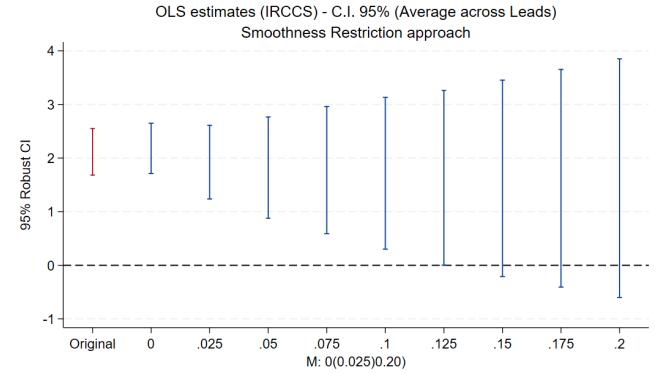
a) BM approach for the significance of the first lead.



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 5: 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 institutional 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 A9), 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)*:

$$\text{ITT} = \mathbb{E}[Y_{it}(1) - Y_{it}(0) \mid \exists k \in \mathcal{T}, k \geq 2018 \wedge D_{ik} = 1] \quad (4)$$

where  $\mathcal{T} = \{2012, 2013, \dots, 2022\}$

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 Effect on the Treated (ATT)*, which would rely on actual entry year:

$$\text{ATT} = \mathbb{E}[Y_{it}(1) - Y_{it}(0) \mid D_{it} = 1] \quad (5)$$

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 A3 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 A10 and Figure A11 in the Appendix; they both corroborate the validity of our simplification procedure by showing the absence of substantial 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 4, 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 4: 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 A4 in the Appendix), in doing this showing the existence of structural differences between the individuals selecting early or late into the IRCCS treatment perimeter.

Furthermore, reliance on our *ITT* 2x2 methodology can also be justified by following the approach by 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 6, 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. There are no significant coefficient estimates lower than 0 either. This supports the reason of opting to the estimation 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 severely biased results.

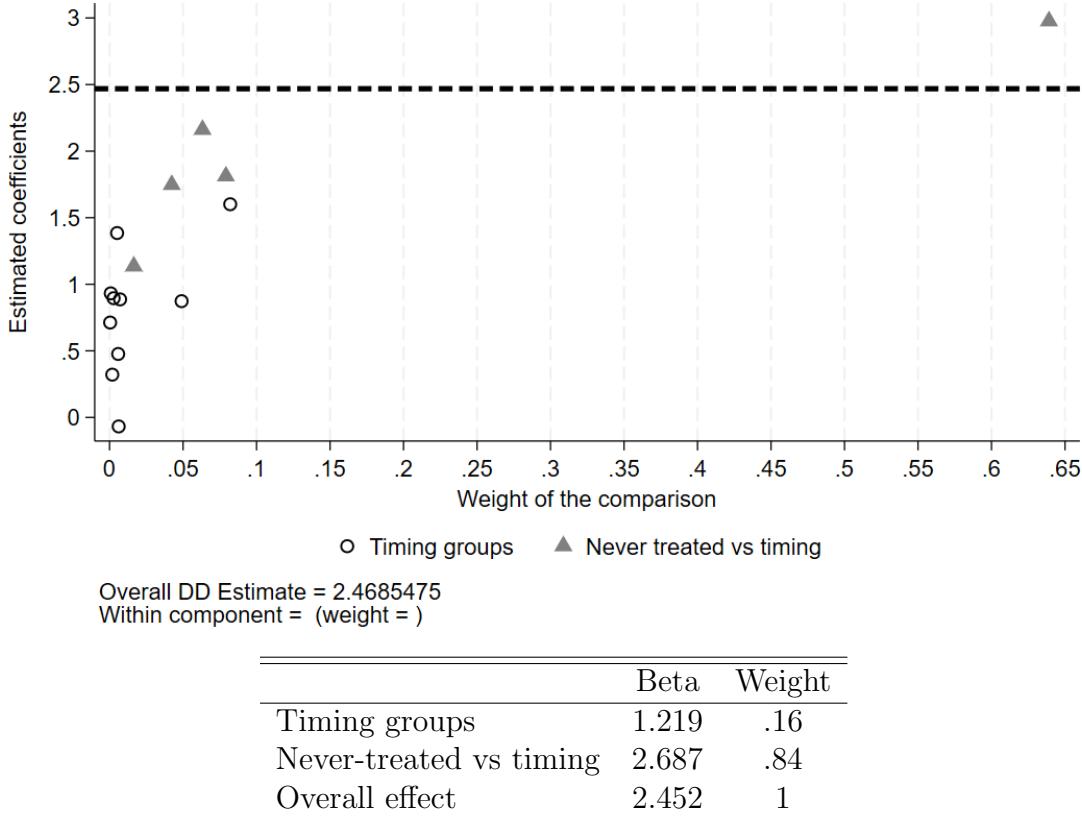


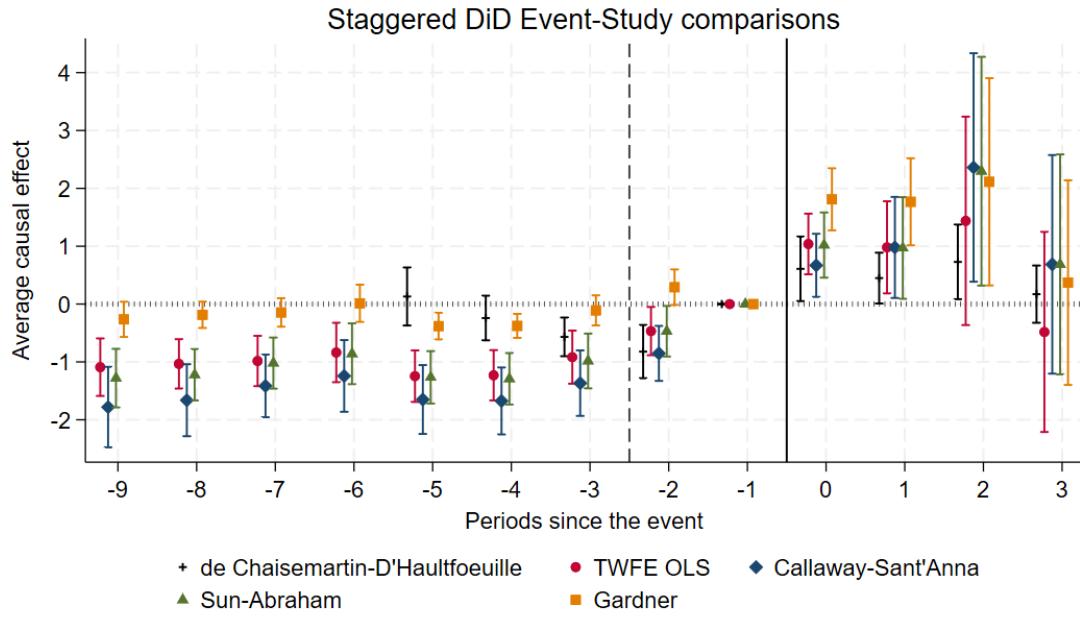
Figure 6: Goodman-Bacon decomposition: weights and estimates.

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 recognition. In the second set, we perform a staggered estimate on

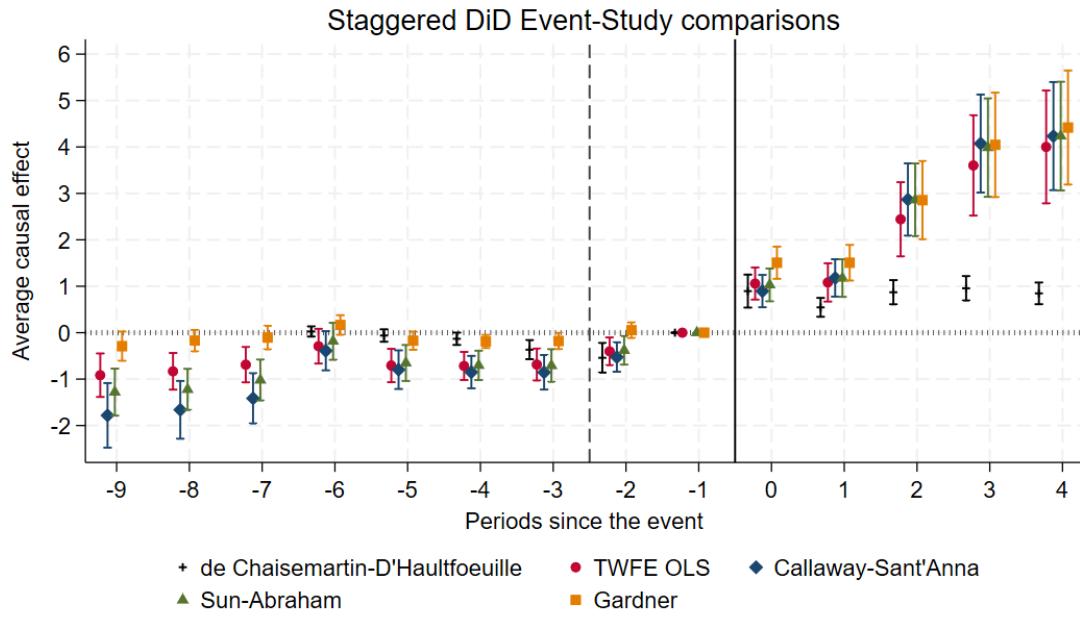
all units, late-enrolled or otherwise. Recent econometric literature on Staggered Difference-in-Difference methodologies displayed already how TWFE OLS estimators may deliver biased estimates (de Chaisemartin and D'Haultfœuille, 2020, Callaway and Sant'Anna, 2021, Sun and Abraham, 2021, Butts and Gardner, 2022), 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:

$$Y_{it} = \delta_i + \tau_t + \sum_{k \neq -1} \beta_k \cdot D_{it}^k + \varepsilon_{it}, \quad (6)$$

where  $Y_{it}$  is the outcome of interest for unit  $i$  in year  $t$ ,  $\delta_i$  and  $\tau_t$  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  $\beta_k$  coefficients are interpreted relative to the year before the IRCCS inclusion. After that, we integrated the estimate by better refined approaches: 1) 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); 2) 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); 3) Sun and Abraham, 2021 (introduce event-study estimators based on interaction terms between time and cohort); 4) Butts and Gardner, 2022 (based on Gardner, 2022: implement a two-stage estimator that residualizes the outcome variable to remove fixed effects, and then estimates dynamic treatment effects). Figure 7 (Graphs a) and b)) shows the dynamic effect of IRCCS using TWFE and the various corrected estimators. The methods relying on linear estimates are normalized at  $k = -1$  to avoid collinearity.



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

All the patterns are very similar, although the different methodologies show strongly heterogeneous findings in terms of magnitude and significance. The TWFE, Callaway-Sant'Anna and Sun-Abraham estimates significantly shift downwards the pre-trend dynamics, showing an evident converging pattern in the pre-treatment period. The same is displayed by the estimates performed via de Chaisemartin-D'Haultfoeuille and Gardner, although their magnitude and significant are lowered.

Such differences in the estimates are reasonably to be credited to the heterogenous nature of the methodologies, as the latter does not rely on linear time normalization, and therefore what is observed as a converging pattern driven by an initially negative difference in trends for coefficients estimated through the other methods, 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 bulk of treated individuals and make up for most of the estimated coefficients. In graph (b) 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'Haultfoeuille, 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 of the MBO (IRCCS).

## Different scales

So far we expressed the outcome in raw numbers, reporting it as the number of yearly publications per authors. However, we re-scale the outcome due to the relevant difference in numbers across heterogeneous types of researchers, with many of them never publishing, whereas others do that even dozens of times in a year. We re-perform the estimates specified in Equation 2, 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. Coefficients on log- and asinh- publications are reported in Table A5, 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 2 as our baseline.

## Excluding COVID-19 Years

It seems clear that the COVID-19 pandemic outbreak in 2020 substantially triggered upwards the pattern of publications. Was such event to affect research productivity, independently on the granting of IRCCS funding to the hospital, then it would be a validity concern. The main confounder could be the reallocation 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 reduced research activity by physicians. On the other hand, the postponement of elective activities might give doctors more time for writing, whereas with the joint national effort put in place in the struggle against the pandemic, COVID-related research at the time could be relatively 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 an existing IRCCS effect, the

first threat to validity would be a major confounder, able to substantially bias the reported estimates. 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<sup>3</sup>. We thus truncate the panel at 2019, and re-run the difference-in-differences analyses. Table A6 in the appendix shows the estimates for the pre-COVID sample. Results are similar in direction to the full-sample findings, although the size of the coefficients is, as expected, reduced by the shortening of the time span.

## Spillover Effects

A further concern for our identification strategy is the potential violation of SUTVA (*Stable Unit Treatment Value Assumption*), as in, the policies might have indirect effects on those who are not targeted directly, 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. Because we observe that the IRCCS-funded researchers are the ones responding more significantly to the shock (whereas the MBO does not have any impact), if there was such a spillover effect, it would be 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 physician. It is worth noting that such flags are identified on the whole unbalanced panel by checking annually who is in force to the hospital, in order not to exclude those who do not appear continuously over the full longitude of the dataset. Such construction calls for the comparison between findings in the balanced and in the unbalanced panel. Then, we perform two Diff-in-diff estimations for each exercise; one comparing the groups pre- and post- the MBO introduction (2017), and the other pre- and post- the IRCCS recognition (2018). In the first set of estimates we compare IRCCS-only and MBO-only treated units to the full control group (as in, a static version of Equation 1, but without the compound

---

<sup>3</sup>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. We therefore avoid reporting the event-studies with the reduced sample, which may be made available upon request.

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 a second time, though 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. Findings are shown in Table A7 in the Appendix for the balanced panel.

The results of these two exercises offer interesting insights. In absence of *Type A researchers*, who are the MBO-treated collaborating with IRCCS, results are consistent with the main analysis in both the unbalanced and balanced samples, showing a high response for IRCCS-researchers and a negative one for the MBO ones (which is however significant in the balanced panel only). On the other side, removing *Type B researchers* (IRCCS researchers coauthoring with MBO-only units) decreases the differential effect across the two groups in both samples, reducing the impact of the policy on faculty members. In the balanced panel, the negative role of the MBO on the productivity of MDs turns non-significant; it is curious to observe how such coefficient turns positive and different from zero in the unbalanced panel (0.25 and 0.29, which is sizable in magnitude relative to the treated units' pre-event mean), which so far is the only relevant difference we retrieved between our balanced findings and the unbalanced ones. Such outcomes show how the role of collaborating MDs is negligible in the broader context of the hospital research activity; on the other side, co-authoring IRCCS researchers significantly bring about spillover effect within their own group. As a matter of fact, their removal improves the performance of MBO-eligible group individuals relative to the control group, lowering that of their own peers. This implies that research cross-group research networks are feasibly extended to within-group collaborations as well, whereas faculty members, as a group, benefit more from the networks of more prolific and open-to-collaboration individuals than their clinical non-IRCCS counterparts, at least relative to the full control group. Therefore, there seems not to be cross-group spillover effects biasing the results; by contrast, such spillover effects appear to be relevant in a within-group context, where a bunch of possibly "superstar" researchers drive up the performance of the whole IRCCS perimeter. Which sets the fundamentals for the next robustness check. The related dynamic estimates of the spillover checks are reported in the Appendix as well, in Figure A12, Graphs a) (balanced) and b) (unbalanced) respectively.

## 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 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<sup>4</sup>. To guarantee that these “switchers” are not affecting our results to a great extent, we perform a robustness dynamic check starting from our main event study (1), 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 3.9% of the dataset. The findings (Appendix, Figure A13) 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 the extent to which the two groups have different incentive structures driving them. We identify those individuals who may shift upwards the overall publication numbers of their respective group, especially with regards to 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. We then re-estimate both the baseline dynamic specifications: they are reported in the Appendix, in Figures A14 (balanced) and A15 (unbalanced). As expected, the differential impact of the policies decreases in size when an

---

<sup>4</sup>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.

increasingly greater units in the upper parts of the yearly publications distribution are removed from the sample. However, the directions of the effects stay the same even when eliminating the 15% of the sample; it must be noted however, that when the top 15% is excluded, the shift in the double-treated units' publications results slightly greater than the one for the IRCCS-only individuals. While this is a slight reversion of our main result, it is worth reckoning that the double-treated change takes off with a lag after the IRCCS-recognition, which leads to suggest that as the cause for the variation, rather than the performance payment implementation. Nevertheless, such reversion does not take shape in the event study for the balanced panel (Figure A15). Therefore, it is evident that the huge effect we observe is strongly driven, but not entirely caused. by a sub-group of scientific superstars among IRCCS researchers.

## 7 Mechanisms

Whereas the IRCCS recognition seemingly boosted research productivity, we ought to test whether this is actually due to improved access to funds and, in such case, how. To do so, we recover additional information on the institutional research performed by the hospital under study, drawing open data from the Research Workflow of the Italian Ministry of Health, which tracks all the funded research lines of Italian IRCCSs. As specified earlier, access to additional funds can be granted either through Current Research (*Ricerca Corrente*) or Targeted Research (*Ricerca Finalizzata*). The first one is the ordinary research activity to which the Ministry annually allocates funding, through a thorough process of assessment of the research criteria required to provide said funds. The second one is tied to the awarding of targeted grants tied to specific projects, whose financing depends on project-related criteria and upon the will of researchers to apply to such grants.

### Access to IRCCS funds - Current Research

Current Research fund are granted annually to the hospital, in accordance to the research lines defined by the institution itself upon the recognition as IRCCS. In 2018, the hospital was awarded the recognition for the two broad macro-areas of Personalized Medicine and Innovative Biotechnologies. Such broad categories are further sub-divided into smaller ones, called Research Lines (*Linee di Ricerca*). According to the research lines in which the institutions are specialized, the Ministry allocates funds after elaborating a Three-year Plan, previous assessment of the standards required by the law. This

framework relates to *ordinary research*, as funds are granted on a yearly basis upon withstanding excellence criteria in the areas where institutions perform research continuously; therefore, they are not competitive and are only assigned to IRCCS institutes -and not to other Italian healthcare institutions. Such funding has ranged annually between €150m and €175m overall from 2018 to 2023, and for a single IRCCS the allocation can be of few hundreds thousands of € to about €10m, according to their past performance and excellence standards (MdS, 2025). After 2020, for instance, the institution under question received, annually, €7m circa for Current Research activities.

The official website of the Workflow of the Ministry provides with micro-data of all funded annual research lines, together with the publications recognized as being part of such lines, for each institution. Such information can be exploited to assess whether the increased research productivity can be linked to the Current Research areas, to establish a more robust causal connection between the IRCCS recognition and the improvement in productivity. To do this, we recover our main analysis's framework (Equation 1), integrating it with information about research lines and publications. The data availability, however, constrains our estimates. As a matter of fact, research lines were not a thing before 2018, hence we are not able to obtain an adequate counterfactual in the years prior to 2018, dampening our faculty to put in place a proper triple DiD analysis. In order to make up for this, we devised two separate strategies. First, we identify all publications ascribable to the IRCCS Current Research lines after 2018, by merging the publications in our Web of Science-obtained dataset with the ones recovered from the Ministry Workflow<sup>5</sup>, and flagging the managed matchings. Therefore, we first identified all authors who published merged publications in the research lines, assuming their continuous appurtenance to said ordinary research activities. By doing this, we obtain a further individual identification of the treatment status, on top of the MBO-eligibility and mere belonging to the IRCCS staff. In a second methodology, we elaborate a balanced panel where the longitudinal units are the journals where all the articles in our data are published, after removing duplicates<sup>6</sup>. The outlets where at least one article belonging to the research lines are published are flagged as part of the research line. Then, we compare those with said flag with those which do not publish Research Line featured articles. Both strategies employ 2x2

---

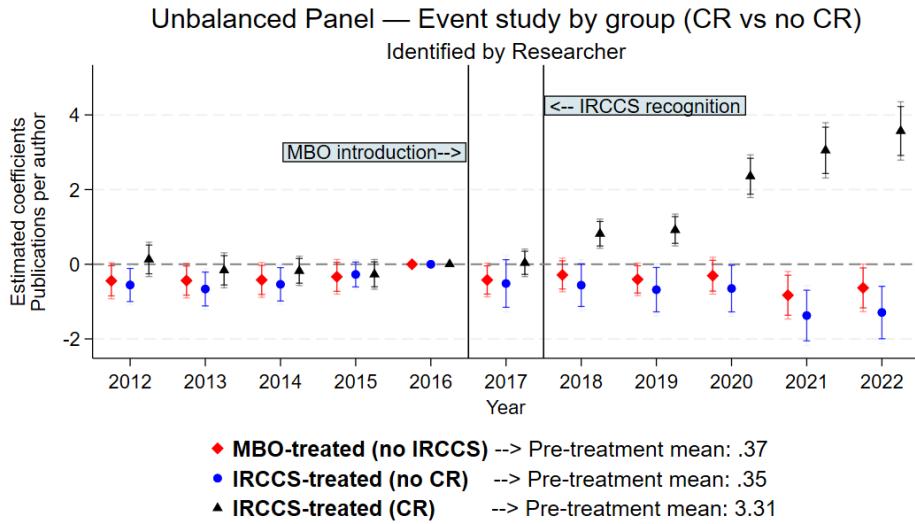
<sup>5</sup>We perform a two-step merging process: an exact merge based on the publication title, and then a fuzzy merge based on correspondence between the scientific outlets, reported in both sources of data, and on similarity of the publication titles (through the Blasnik, 2007 probabilistic matching procedure, by fixing a minimum score of 0.995). Note that the procedure is not able, in any case, to match all the publications provided by the Workflow, as many publications are authored by individuals who are not in our panel (external affiliates, PhD students, research fellows, collaborator and non-employees who still the affiliation anyway). For this reason, we keep our matching score extremely high to avoid false positives, as our outcome can at most be a lower bound nonetheless.

<sup>6</sup>As a matter of fact, the individual panel is based on counts per researcher, so co-authored articles are numbered more than once.

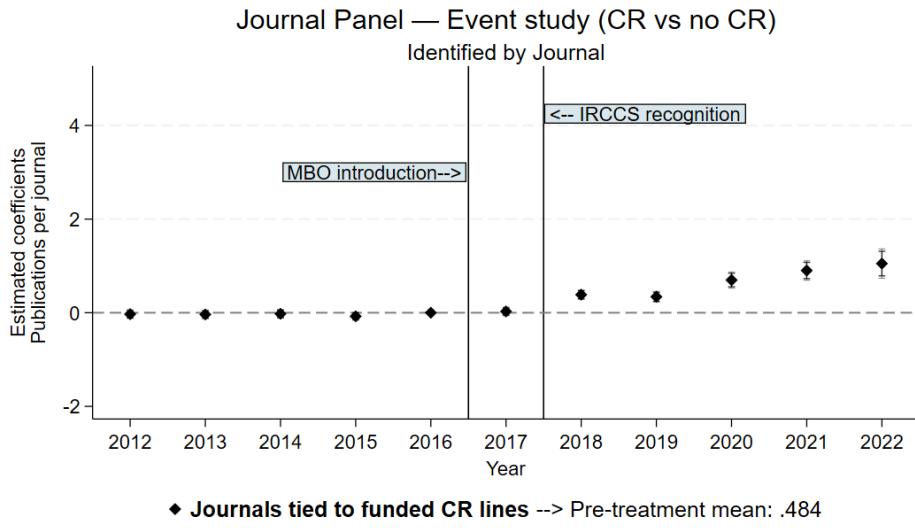
TWFE Diff-in-Diffs methodologies: the first one on the unbalanced panel of all researchers, the second one on a balanced panel of journals. The outcome is the yearly number of publications per researcher in the former, and the annual number of publications present in our dataset per journal in the latter. Results are reported in Graphs a) and b) of Figure 8.

Graph a) mirrors 3: the control group is the full control one, which is compared to the group of MBO-eligible researchers and two groups of IRCCS-list members. However, in this case the further subdivision is not across double-treated and IRCCS-only researchers, but across IRCCS-researchers who published at least once within a flagged Current Research line, and those who did not. Although the number of the latter is way smaller (they are less than 50), we can observe how being part of the IRCCS perimeter, but not operating in any research areas, does not have any improving effect on research productivity, if not even a negative one. The effect on MBO-eligible non-IRCCS MD is the same as the one shown in 3. On the other side, we observe how the relevant boost in the activity of IRCCS researchers is basically entirely driven by individuals ascribable to the Ministry-identified research lines. Therefore, the improvement can be tracked to IRCCS ordinary funding. In the journal panel (b), whose coefficients are extremely precisely estimated, the direction of the effect is the same, although reduced in magnitude due to the different construction of the dataset. However, the hospital publications in journals flagged under the research lines increased continuously by almost 100% the pre-treatment mean from 2020. Both graphs display an evident absence of pre-trend.

While this shows the impact of research lines on productivity, we cannot say much on how funds are actually exploited. CR resources are provided based on ordinary activities, hence it is impossible to link them to specific projects or grants. However, according to the Ministry Three-year Plans of 2017-19 and 2020-22, such funds can be used for research activity defined overall, involving clinical trials, personnel costs and technological transfer, like patents (MdS, 2017, 2020). However, the autonomy of the institutions in utilizing ordinary funding may also imply the utilization of such resources for micro-expenses which, however, significantly boost incremental research activity like publication fees, biobanks, software purchases, conference and workshop funding etc.



- a) Event Study of the impact of IRCCS recognition + inclusion in Current Research lines on publications, across different comparison groups - by researchers.



- b) Event Study of the impact of IRCCS recognition + inclusion in Current Research lines on publications - by journals.

Figure 8: Event Study of the impact of IRCCS recognition + inclusion in Current Research lines on publications, identifying treatment groups by researchers (a) and journals (b).

## Access to IRCCS funds - Targeted Research

The second set of IRCCS resources researchers have access to are the grants for Targeted Research. These funds are annually issued by the Ministry in the shape of public competitions. Such competitions are addressed to 4 types of entities: the Italian Regions (administrative entities, NUTS-2), the National Institute of Health (*Istituto Superiore di Sanità*, ISS), the Experimental Zooprophylactic Institutes (*Istituti Zooprofilattici Sperimentali*, IZS), and the IRCCS. All non-IRCCS Italian healthcare institutions can access targeted funds, conditional on being allocated resources at a second step after another entity (usually the Regions) has been awarded the financing. Due to their institutional nature, IRCCSs (and IZSs) can take part to the competitions without the intermediation of the Regions or the ISS.

To obtain TR funds, the Ministry must first issue a public call, inviting institutions to apply for a three-year term financing for projects aimed at achieving scientific output useful to steer the National Health Service policies. The projects are submitted at the hospital/institutional level, but they are coordinated and managed by a Principal Investigator (PI) who, at times, is supported by one or more Co-PIs. The PI is a researcher of the institution applying for funds. Contrarily to the Current Research funds, TR funds are conditioned on stricter requirements of management provided in the official documentation of the Three-year Ministerial Plans. The Workflow of the Ministry reports year, awarded sum, title and additional information on all Targeted Research funds allocated yearly to the various entities. We managed to recover all funds awarded to the institution under study from 2018 to 2022 (included a project awarded in 2016, upon allocation to another institution). They are 58 projects, awarded a sum (not always entirely paid to the institution) usually ranging between €300k and €500k; funds awarded in 2022, within the scope of the post-Covid Recovery and Resilience Plan, amount all to €1m or little less. Unfortunately, the Workflow does not disclose the PIs' and Co-PIs' names. To recover them, we searched for publications acknowledging the funds as the source of financing; for the majority of cases, the financing disclosure also reported the name of the person to whom the fund was awarded; however, we double-check by looking at the researchers' CVs or at news outlet on-line<sup>7</sup>. Out of 58 projects, we identified PIs and Co-PIs (when present) for 56 projects. Among them, 12 were awarded to PIs/Co-PIs absent in our dataset (possibly due to not being -yet- stably employed at the institution). We eventually flagged the individuals in our unbalanced panel in according to being or having been a PI

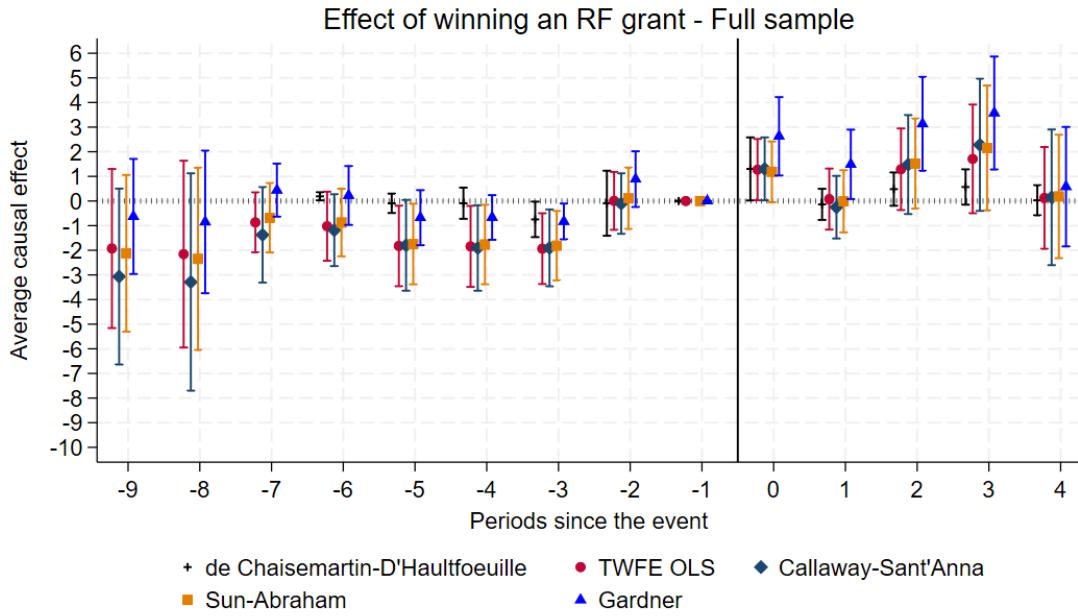
---

<sup>7</sup>For year 2019, the Workflow reports the CUP, Project Unique Code (*Codice Unico di Progetto*), the identifier required by the law for all public investments; for Targeted Funds on such projects, we just typed the identifier on the OpenCuP database of the Italian Presidency of the Council of Ministries, and obtained the PIs. For other years, such procedure was not possible. In any case, the data gathering sources have been stored and may be provided upon request.

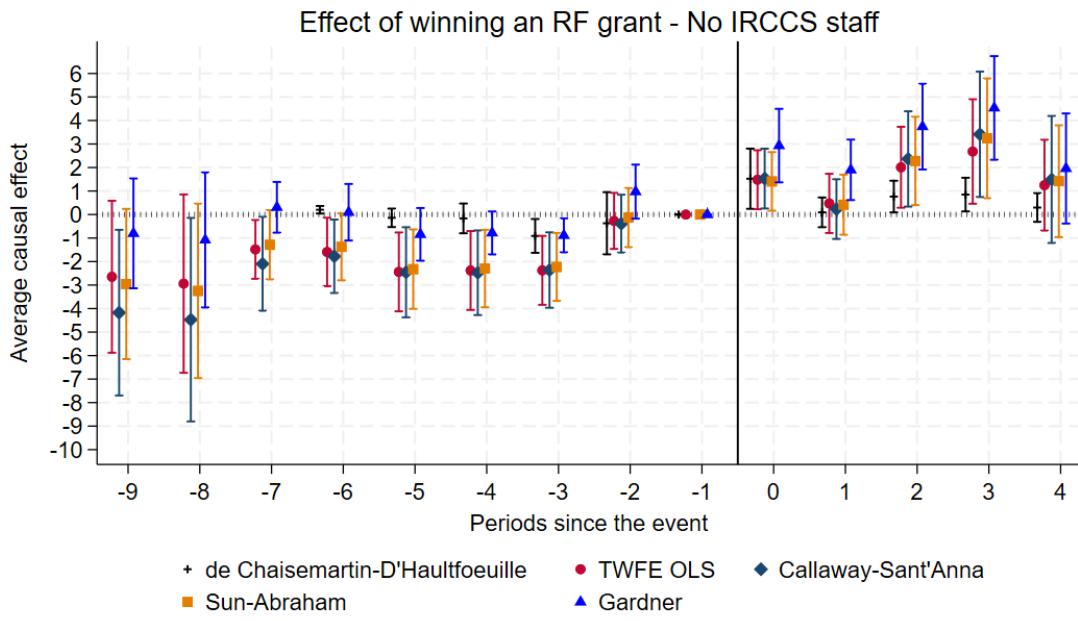
or Co-PI in one of the 42 recovered projects, and the year of their awarding. Then, we performed two different staggered dynamic DiD estimates exploiting different methodologies on the impact of IRCCS recognition, in the same fashion as in Equation 6. We devised two different group-comparisons: in the first, we estimated the impact of receiving TR funds by comparing those who win them against all the rest of individuals who never do so; in the second one, we compared the same individuals with a truncated sample obtained by excluding individuals in the IRCCS-perimeter who never receive a TR fund directly. Results are in Graphs a) and b) of Figure 9.

When comparing treated units with the rest of the sample (a), we can observe a slightly positive effect, similar in magnitude to the one observed in the staggered Event-Studies of Figure 7, although in most specifications is only significant at the treatment period (0) and after 3 lags. When excluding the IRCCS individuals who are not winning awards (b), the estimates recover precision at lag 2 as well. In both cases, the effects revert back to 0 at the 4th period, possibly consistent with the 3 years duration of the finance scheme. The inconsistent and low precision of the estimates possibly signal a not necessary overlapping between PIs and more productive researchers; in addition, the significant effect observed in the main analyses may be more robustly driven by Current Research activities, for which the PI/co-PI figure is non-existent, and by within-group coauthorship, able to boost research output of expert researchers irrespective of their direct awarding of funds as Principal Investigators.

In terms of pre-trend, the pattern mirrors the pre-treatment behavior of estimates displayed in the main staggered event-study (7), with a slow convergence towards a common trend starting from a downward biased pre-treatment path: as already argued, this is likely traceable to selection. By contrast, the presence of a higher number of non-significant coefficient possibly mirrors the lower power of the estimates for the treated individuals in the present framework. Such findings suggest that, while the awarding of a grant is evidently relevant in shaping research productivity, in the IRCCS context the publications' boost is feasibly triggered by Current Research ordinary activities mostly. This is even meaningful by considering that those who do not obtain Targeted Research funds may always apply for other competitive research grants, whereas the stability, safety and lack of constraints of the annual Current Research allocations cannot be compensated via alternative solutions.



a) Event-study of the staggered awarding of Targeted Research funds to P.I.s of the IRCCS perimeters (full unbalanced sample).



b) Event-study of the staggered awarding of Targeted Research funds to P.I.s of the IRCCS perimeters (excluding IRCCS researchers without access to TR funds).

Figure 9: Event-studies of the staggered awarding of Targeted Research funds to P.I.s of the IRCCS perimeters.

## 8 Networks and Research Quality

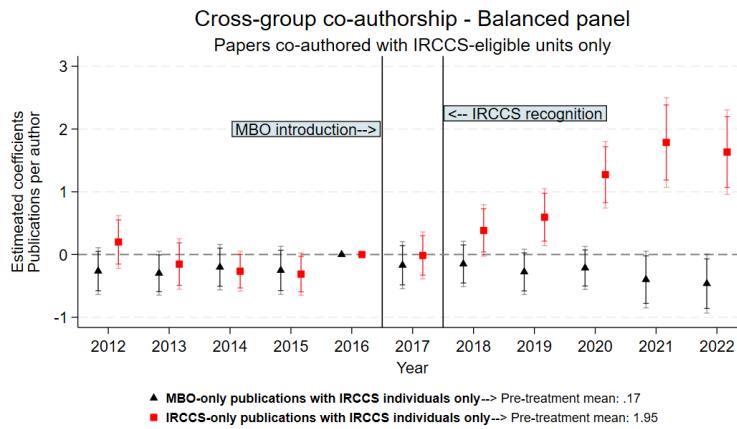
It is valuable to assess whether the policies had a significant impact on the nature of the research collaborations stemmed after their implementation, and the impact/quality of the output. We therefore perform two sets of additional estimates in order to tackle two questions mostly: (1) whether and to what extent the policies affected research collaborations across and within the two groups, and (2) how our results relate to research quality of the members of the institution under study, proxied by citations.

### Cross- and within-groups Co-authorship

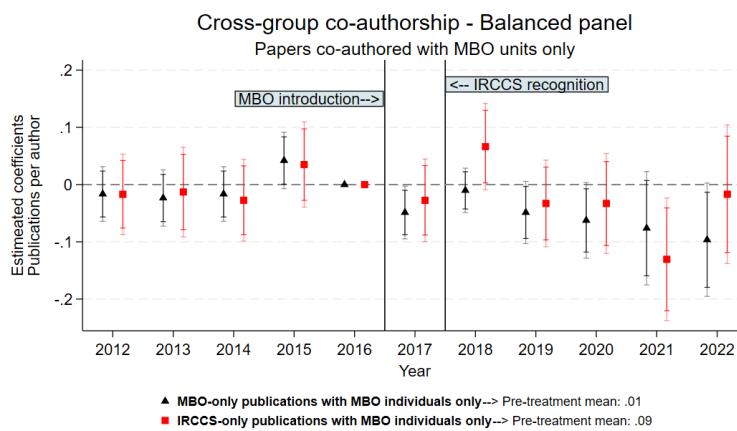
We begin this section by defining all papers listing, among the co-authors, both a non-academic doctor (subject to MBO) and an IRCCS researcher of the same institution as “networked” publications (elaborated on the whole unbalanced panel sample); this is a decent representation of network effects between the two categories of hospital employees. We make use of the same categorization we employed when controlling for the validity of *SUTVA*, by subdividing networked publications according to the groups’ perspective. *Type A* are those for which at least an IRCCS-eligible individual is a co-author, and there are no MBO-eligible co-authors. The opposite holds for *Type B* papers: at least an IRCCS physician is a co-author, and there are no IRCCS co-authors. Finally, *Type AB* are publications where there are at least both a MBO-eligible and an IRCCS researcher. Therefore, after excluding the overlapping category of the double-treated, we dynamically compare across the remaining group (as in 1) the impact of the policies on papers published 1) joint with MBO-eligible MDs only; 2) joint with IRCCS-researchers only; 3) joint with at least one MBO-eligible MD and one IRCCS researcher. The reference group is always the full control one.

Graphs a), b), and c) in Figure 10 reports visually the dynamic estimates for networked publications. Panel (a) shows the event-study for *Type A* (IRCCS co-authorship), Panel (b) for *Type B* ones (MBO co-authorship), and c) for *Type AB* (mixed co-authorship). All three graphs indicate little to no difference in cross-group collaboration rates between the two groups prior to the policies. What is shown is that cross-group co-authorship is limited, as IRCCS-treated groups improve their co-authorship rates within their own group (graphs a) and b)), while MBO-eligible individuals do not substantially increase their co-authored output, neither between- nor within-groups. As already foreseen in Section 6, the most relevant spillover effect is brought about by IRCCS researchers within-group, with null or even negative effect (compared to the control group) led by collaboration with MDs. Results are consistent when

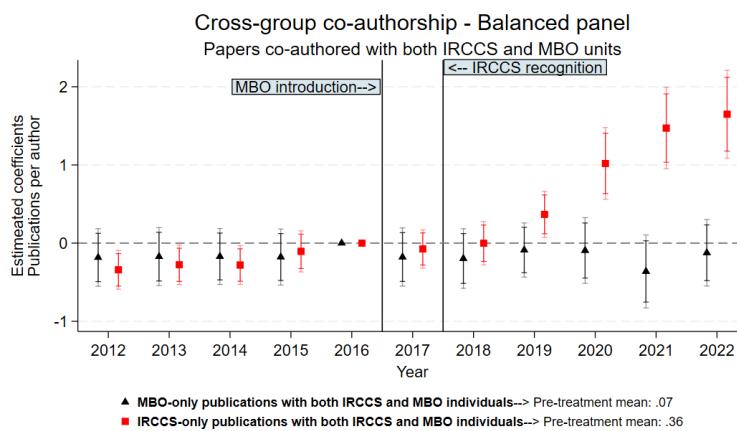
accounting for the full panel with attrition ([A16](#) in the Appendix).



a) Outcome: publications co-authored with IRCCS units only.



b) Outcome: publications co-authored with MBO-eligible units only.



c) Outcome: publications co-authored with both MBO-eligible units and IRCCS researchers.

Figure 10: Dynamic impact of the policy implementation on annual cross-group publications in the balanced panel.

## Citation Impact

A further relevant question is whether a greater number of publications also triggered a positive shift in the influence and quality of works, or it came at the expense of it. We analyze the policy effect on citations, employed as proxy of publications' impact. We do that through the re-modeling the IRCCS dynamic equation (Equation 3) 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 were gathered, which is Fall 2024 ( $[2024 - t]^{-1}$ ). After doing that, we normalize each paper by time, via 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. 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 above. 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-eligible to the MBO, while the MBO-treatment also considers the double-treated.

These results are displayed in dynamic terms in the event-studies in Figure 11, displaying no significant anticipatory behavior or serious PT violations. Apparently, by (inversely) weighting publications by the number discounted citations (a), the effect of the IRCCS' recognition is amplified for more structured researcher, although the effect of the policy is lagged, and takes off in 2020 only, although they range between 5 and 7 citations. By looking at the sum of (discounted) citations as outcome (c), the IRCCS recognition seems to have boosted the number of quotes for IRCCS physicians, by numbers ranging around 20 citations in 2020-2022. However, the IRCCS group's post-2018 average number of (discounted) citations per publications shows a major and significant decrease from 2019 (e).

If we look at the MBO-scheme, the picture is even more insightful. We observe indeed a negative, slightly significant effect on the number of MBO-eligible individuals' publications in the IPW-weighted regression, although this occurs only after the IRCCS recognition (clearly due to the bias induced by the comparison with IRCCS individuals), as in the first year of the MBO implementation the impact seems

to be positive and significant (b). 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, and increasing over time (f).

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 researchers 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 could be linked to the fact that the performance-based rewards are not only based on quantity alone, but also on the impact factor of the journal where publications are issued.

## Strategic behavior of MDs

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 they 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.

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. 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 3, with varying comparison groups and all the rest, but instead of the interaction between *Post2018* and *IRCCS*<sup>i</sup>, we just keep the time-invariant treatment dummy flagging the belonging to the MBO scheme. 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 their 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 5. 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.

	(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 (€)	-1037.37939***	-1466.69556***	1780.15076***	1628.63623***	-3556.25537***	-269.15292**
(SE)	(389.74774)	(223.09830)	(188.85347)	(200.98660)	(200.08897)	(133.17281)
R <sup>2</sup>	0.610	0.668	0.639	0.635	0.702	0.483
Mean	1196.52	2216.09	2216.09	2216.09	520.66	520.66
Sum of weighted IF	-0.05247	-11.46735***	3.66103***	3.92988***	-17.96440***	0.14813
(SE)	(2.09297)	(3.33887)	(0.93505)	(0.66406)	(4.41293)	(0.76541)
R <sup>2</sup>	0.336	0.348	0.372	0.613	0.339	0.191
Mean	3.16	5.68	5.68	5.68	1.48	1.48
Eligible Papers	-1.25746	-4.68687***	4.03192***	3.58589***	-9.94389***	-0.69990***
(SE)	(0.98802)	(0.83610)	(0.34259)	(0.38195)	(0.94491)	(0.25696)
R <sup>2</sup>	0.581	0.615	0.687	0.669	0.637	0.571
Mean	2.82	4.88	4.88	4.88	1.46	1.46
Delta (Pubs-Eligible)	0.17500	3.65820***	-2.78376***	-2.37526***	7.48813***	0.54775***
(SE)	(0.69514)	(0.66451)	(0.27954)	(0.30812)	(0.73891)	(0.18284)
R <sup>2</sup>	0.406	0.477	0.586	0.553	0.471	0.520
Mean	-1.03	-1.88	-1.88	-1.88	-0.47	-0.47
N	8651	4186	4224	3612	5017	4465
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 5: Regressions between enrollment into IRCCS perimeter and several factors.

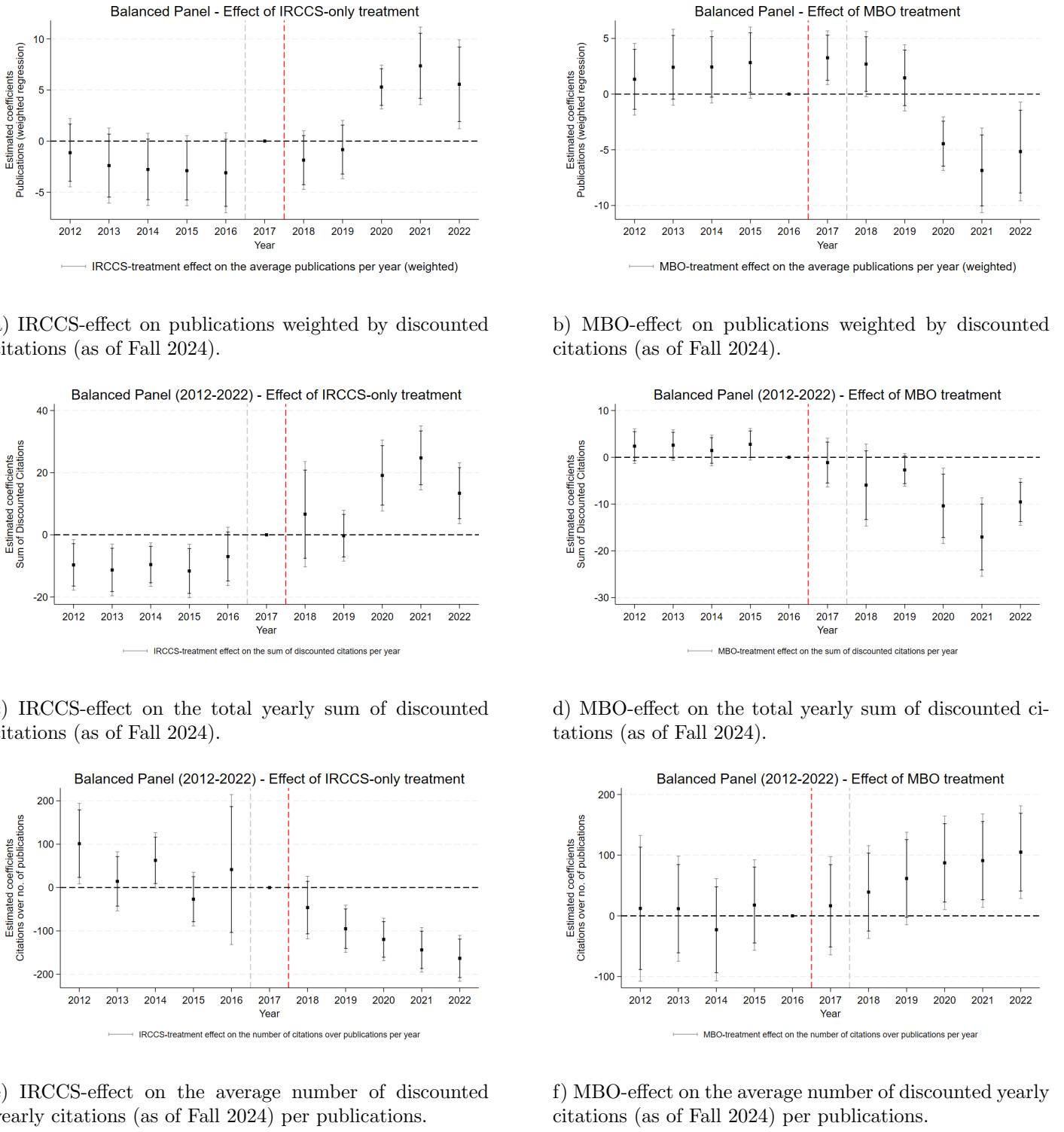


Figure 11: Event-Studies representing the dynamic estimates of the IRCCS recognition and MBO-policy impacts on the impact of published papers, proxied with yearly citations measured as of Fall 2024.

## 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 the 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 were also 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 respond 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 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; in our case, however, to be triggered is mostly within-group collaboration, rather than the one between groups. The absence of gain from such collaborations ought not to be underestimated in the long-run, as they could lack the scope of the synergies required to foster 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, but only 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 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.

## 10 Appendix

<b>Target</b>	MDs without faculty structured appointments
<b>Object</b>	Full papers published between Jan 1 and Dec 31 (each year)
<b>Payment per IF point</b>	€500 per IF point
<b>Max reimbursable IF</b>	20 IF points per year
<b>Max annual gross</b>	€10,000
<b>Author role (institution affiliation)</b>	<b>Bonus share</b>
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.

<b>Requirement</b>	<b>Description</b>	<b>Evaluator</b>
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).

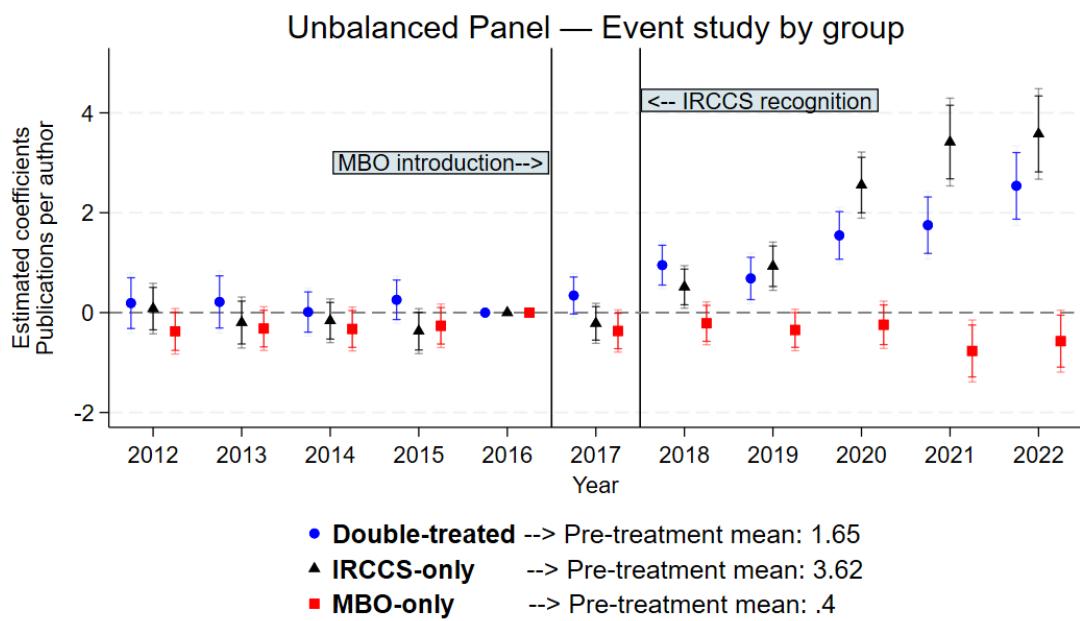
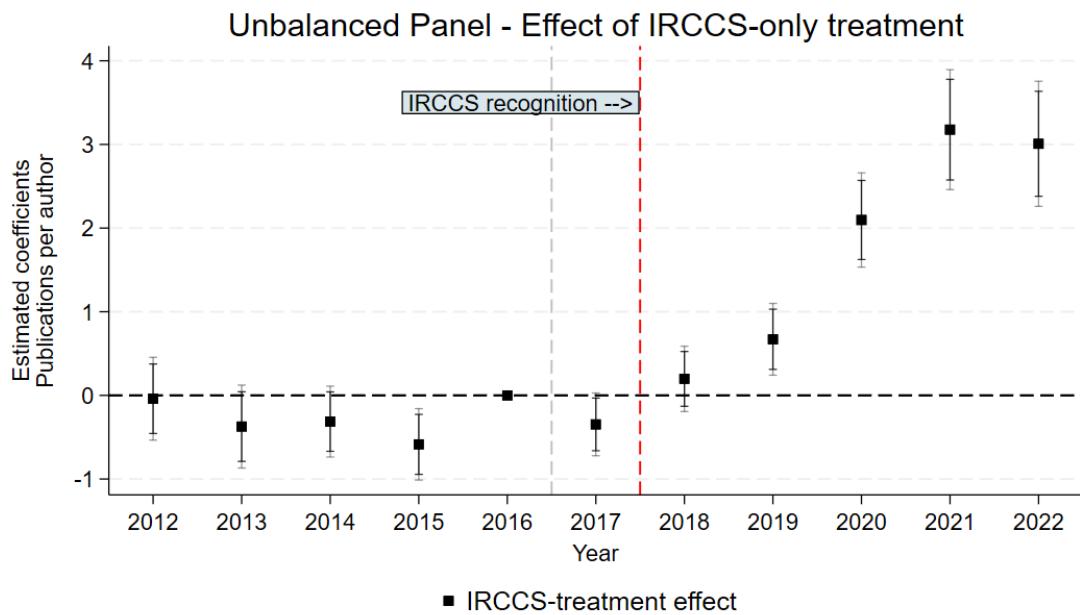
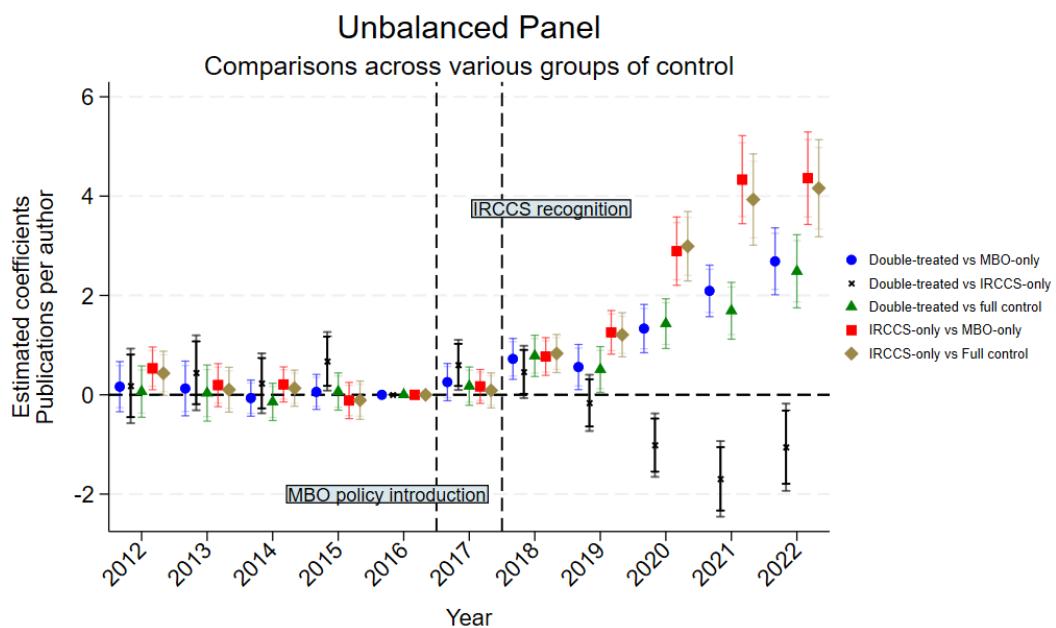


Figure A1: Event-study of the MBO joint with IRCCS recognition policy effect on publications across the different groups, and with respect to the full control group. - Full unbalanced panel.

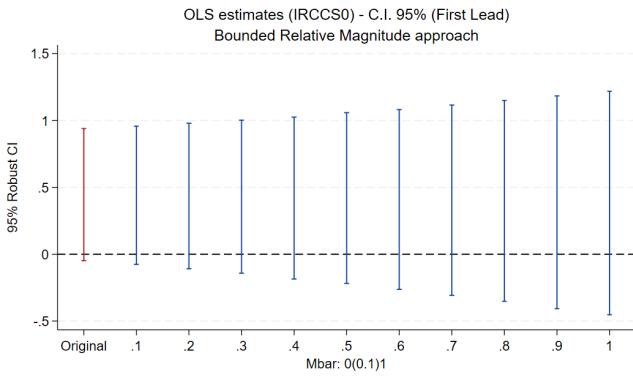


a) Event-study of IRCCS recognition effect on publications for IRCCS-only treated individuals compared with all the other units.

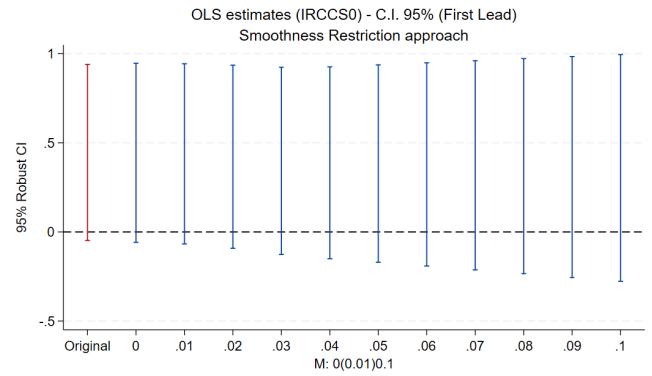


b) Event-study of IRCCS policy effect on publications for comparisons across different groups.

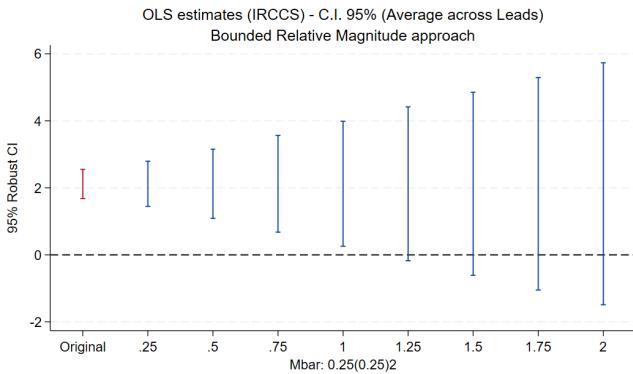
Figure A2: Event-studies of IRCCS policy effect for the unbalanced panel.



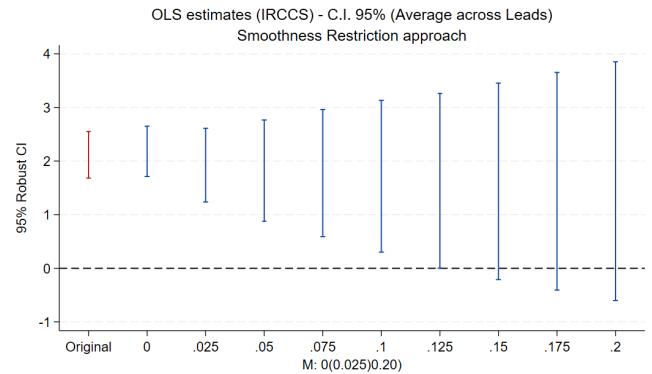
a) BM approach for the significance of the first lead.



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

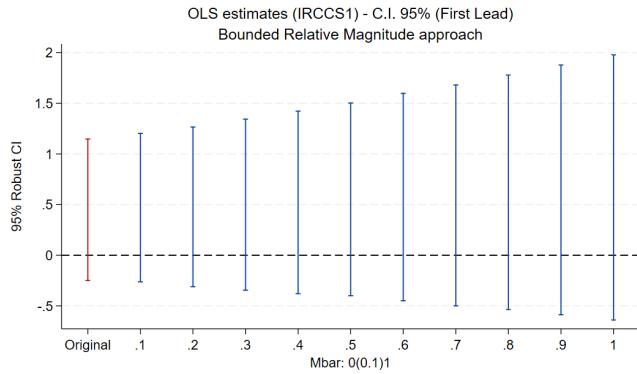


c) BM approach for the significance of the average across all leads.

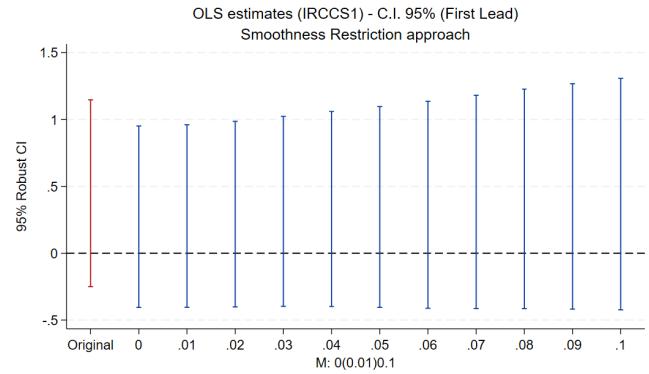


d) SR approach for the significance of the average across all leads.

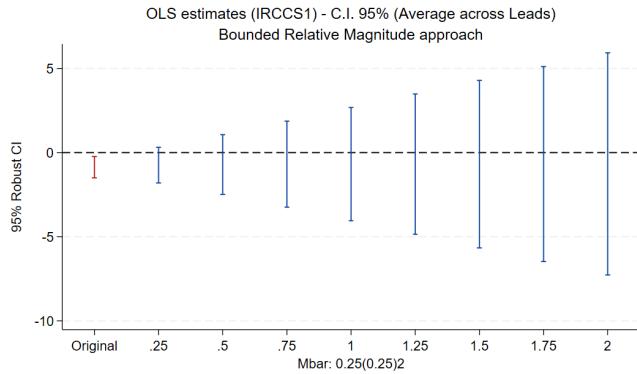
Figure A3: Honest DiD robust confidence sets in the comparison between double-treated and MBO-only units, estimated the overall IRCCS effect by adopting different methodologies (Bounded Relative Magnitude - BM: a) and c). Smoothness Restriction - SR: b) and d)) and for different quantities of interest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



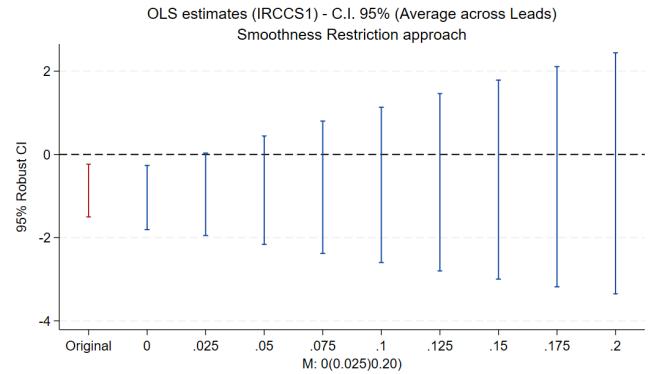
a) BM approach for the significance of the first lead.



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

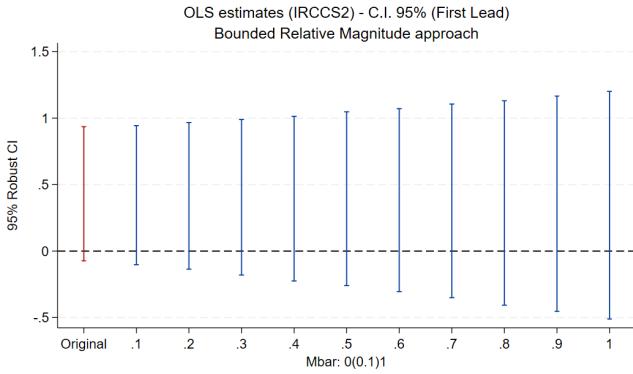


c) BM approach for the significance of the average across all leads.

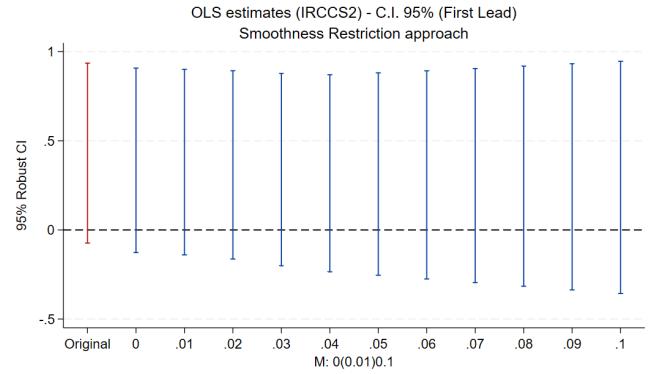


d) SR approach for the significance of the average across all leads.

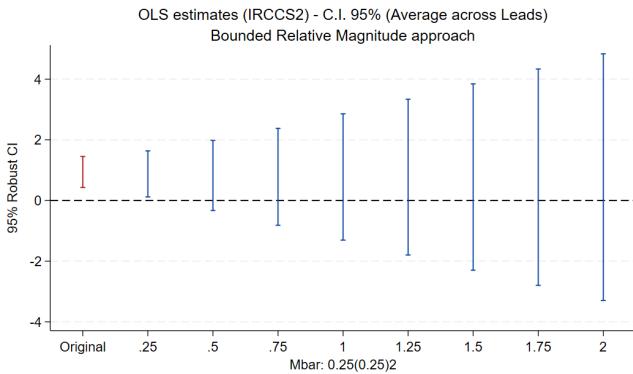
Figure A4: Honest DiD robust confidence sets in the comparison between double-treated and IRCCS-only units, estimated the overall IRCCS effect by adopting different methodologies (Bounded Relative Magnitude - BM: a) and c). Smoothness Restriction - SR: b) and d)) and for different quantities of interest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



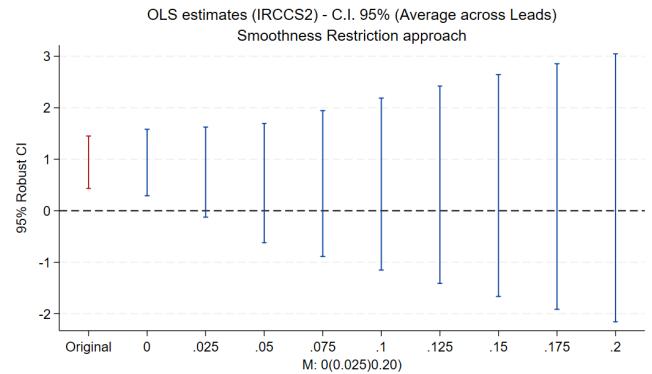
a) BM approach for the significance of the first lead.



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

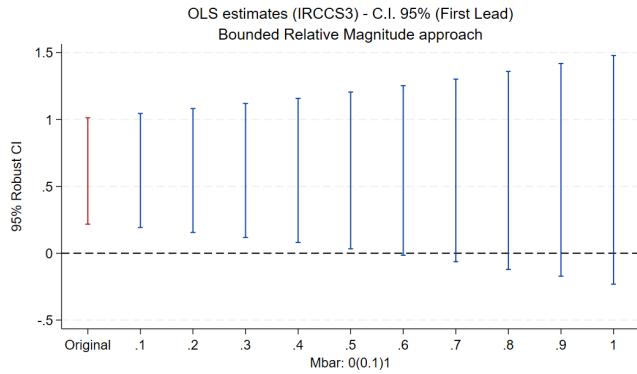


c) BM approach for the significance of the average across all leads.

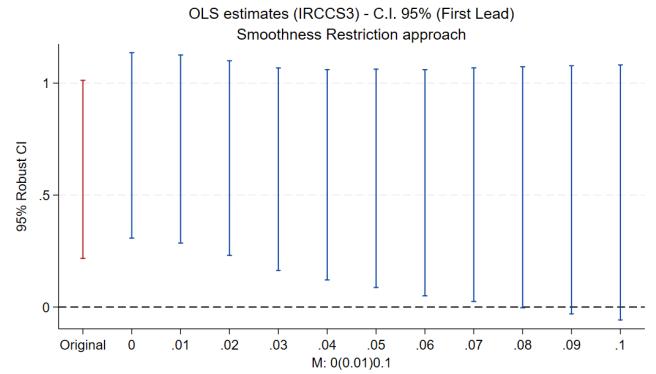


d) SR approach for the significance of the average across all leads.

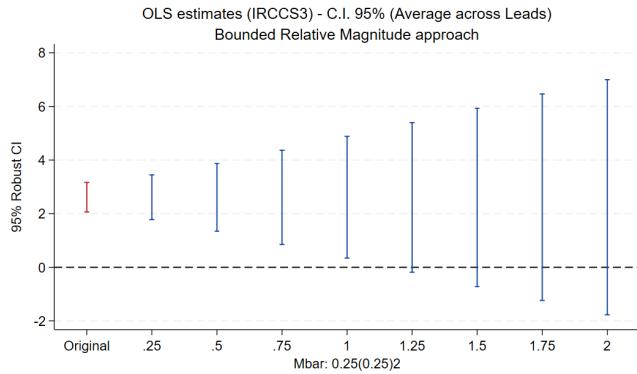
Figure A5: Honest DiD robust confidence sets in the comparison between double-treated and pure control units, estimated the overall IRCCS effect by adopting different methodologies (Bounded Relative Magnitude - BM: a) and c). Smoothness Restriction - SR: b) and d)) and for different quantities of interest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



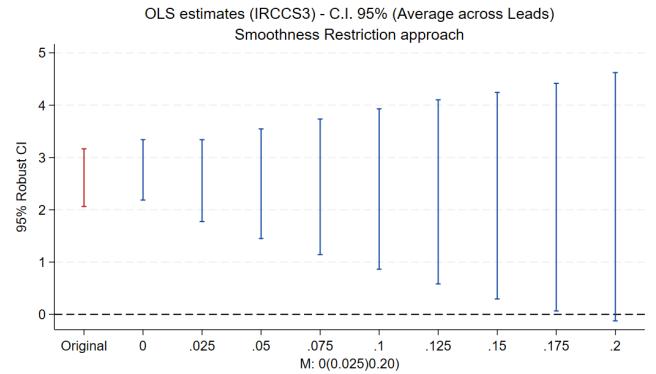
a) BM approach for the significance of the first lead.



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

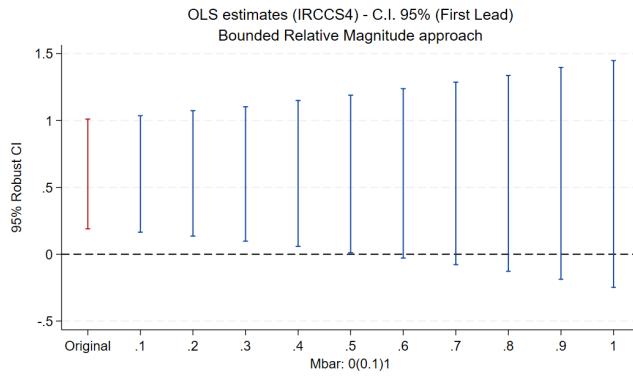


c) BM approach for the significance of the average across all leads.

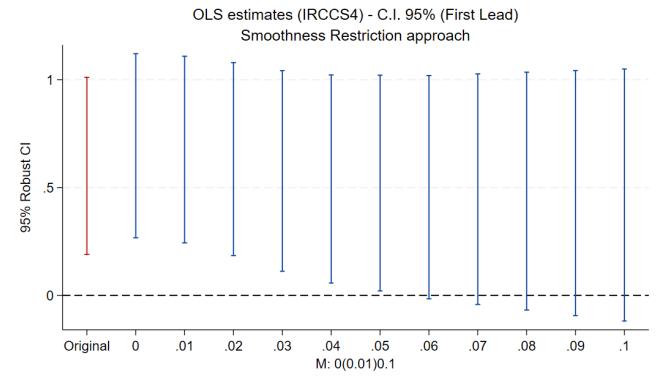


d) SR approach for the significance of the average across all leads.

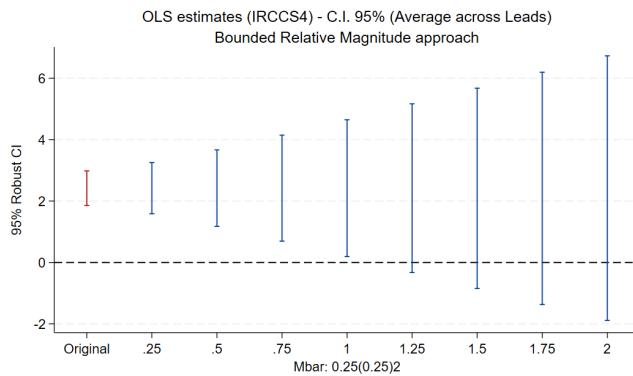
Figure A6: Honest DiD robust confidence sets in the comparison between IRCCS-only and MBO-only units, estimated the or overall IRCCS effect by adopting with different methodologies (Bounded Relative Magnitude - BM: a) and c). Smoothness Restriction - SR: b) and d)) and for different quantities of interest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



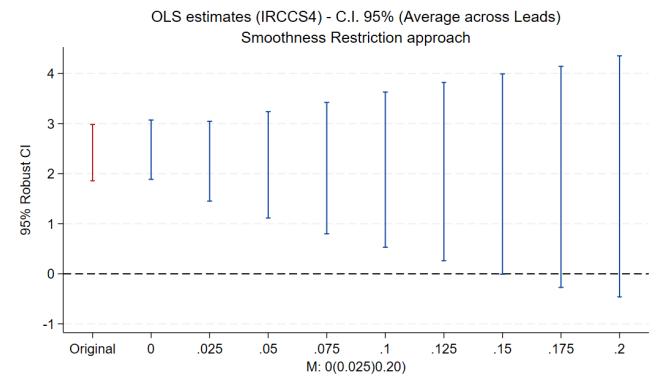
a) BM approach for the significance of the first lead.



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

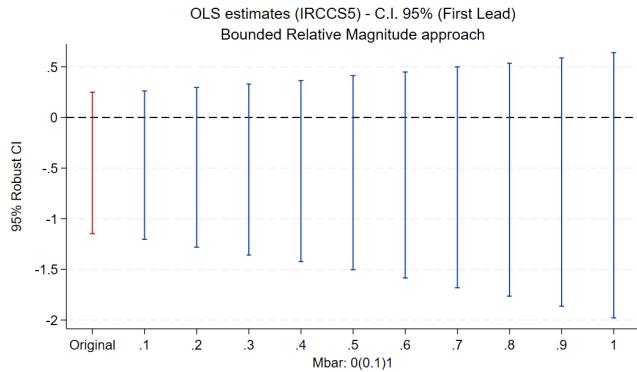


c) BM approach for the significance of the average across all leads.

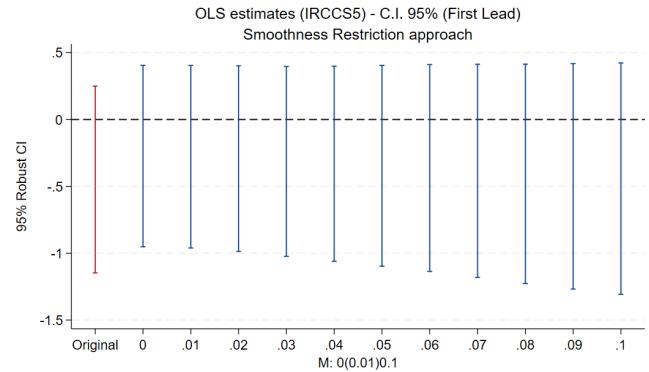


d) SR approach for the significance of the average across all leads.

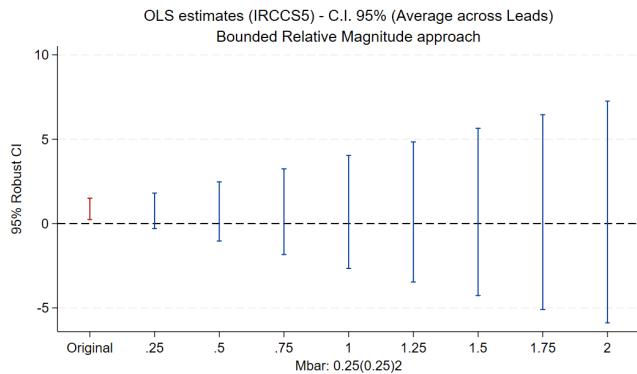
Figure A7: Honest DiD robust confidence sets in the comparison between IRCCS-only and pure control units, estimated the or overall IRCCS effect by adopting with different methodologies (Bounded Relative Magnitude - BM: a) and c). Smoothness Restriction - SR: b) and d)) and for different quantities of interest (coefficient on the first lead of the event study, a) and b). Average across all post-treatment leads, c) and d)).



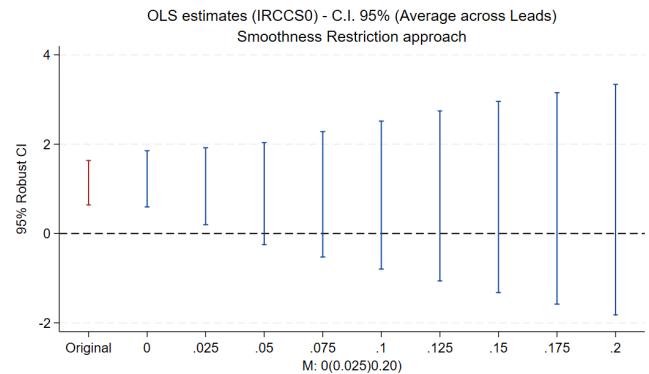
a) BM approach for the significance of the first lead.



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 A8: Honest DiD robust confidence sets in the comparison between IRCCS-only and double-treated units, estimated the overall IRCCS effect by adopting 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)).

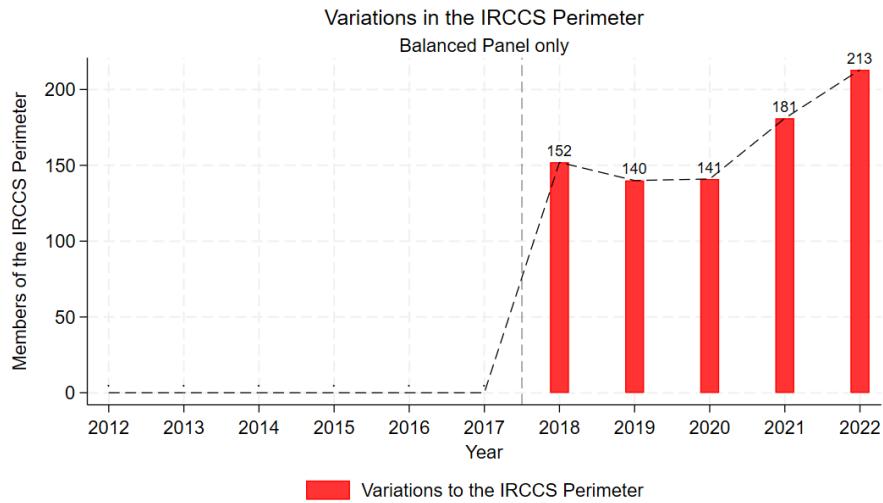


Figure A9: Evolution of the units “lately selected” into the IRCCS perimeter.

	(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	3.0454*** (0.3986)	1.4331*** (0.4110)	-1.1422** (0.4977)	1.3508*** (0.4145)	3.3460*** (0.4292)	3.2307*** (0.4338)
N	5390	2252	1705	2059	3325	3132
R <sup>2</sup>	0.789	0.620	0.768	0.600	0.808	0.798
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Mean	4.30	2.51	2.51	2.51	4.30	4.30
Panel (Balanced)	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 A3: 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	52.8 (7.6)	50.6 (7.0)	0.025**
Publications	4.2 (5.6)	1.5 (2.1)	0.000***
Citations	182.8 (518.5)	48.5 (135.8)	0.015**
<i>Categorical variables (column %)</i>			
<i>Gender</i>			
Female	29.7%	47.9%	
Male	70.3%	52.1%	
<i>Status</i>			
Medical Director	0.0%	0.0%	
Healthcare Professions Manager	21.3%	36.2%	
Sanitary Director	0.0%	0.0%	
Faculty Clinician	2.6%	3.2%	
<i>Department</i>			
Diagnostic Imaging	20.0%	7.4%	
Emergency & Anesthesiology	7.1%	7.4%	
Cardiovascular Sciences	10.3%	10.6%	
Women, Children & Public Health	12.9%	16.0%	
Lab Sci & Infectiology	7.7%	10.6%	
Gastro/Nephro/Endo	21.9%	11.7%	
Aging & Neurosciences	14.2%	35.1%	
Direktorate General	0.0%	0.0%	
Clinical Governance	1.9%	0.0%	
Health Governance	3.9%	1.1%	

Table A4: 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.

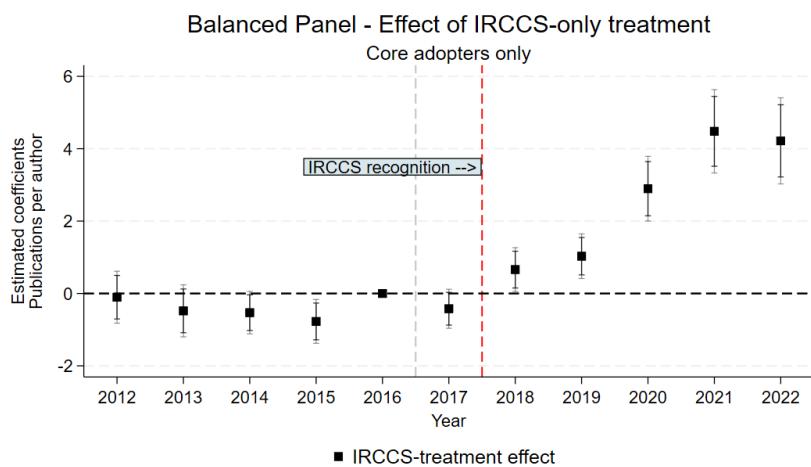


Figure A10: Event-study of IRCCS recognition effect on publications. Units included in the IRCCS perimeter in any year after 2018 are excluded from the estimates.

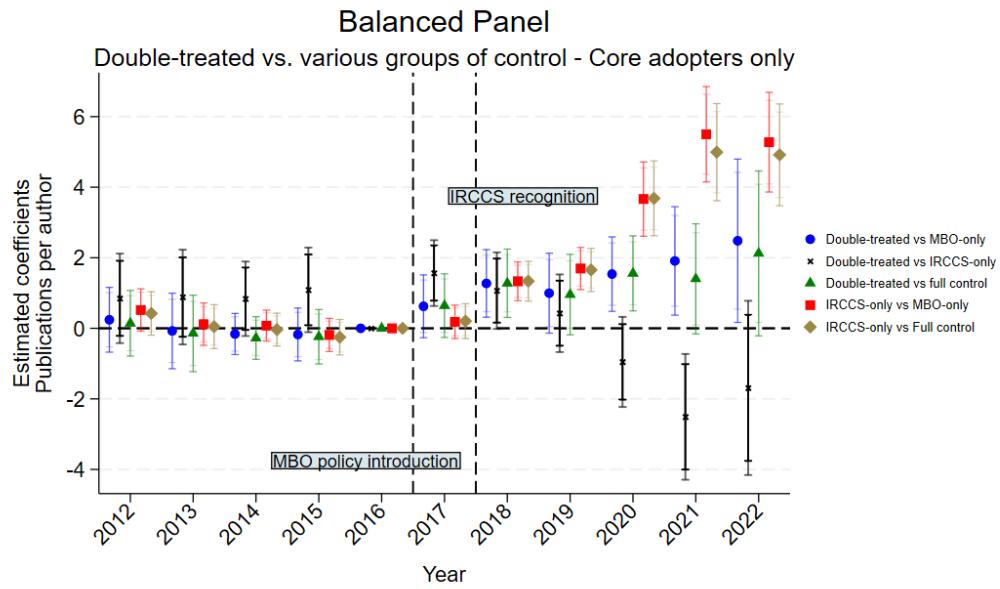


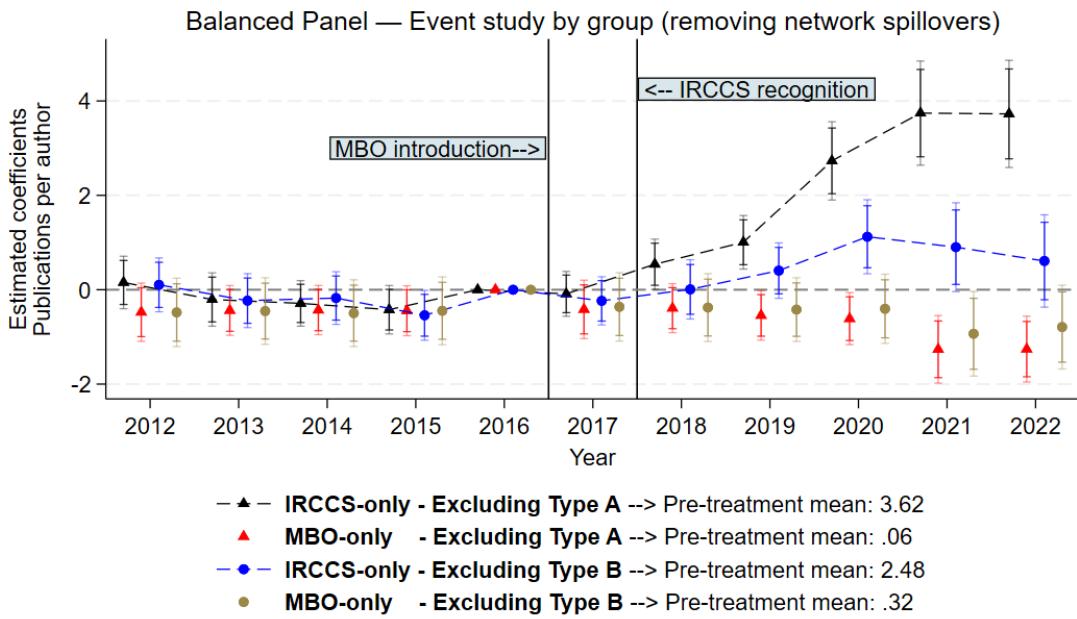
Figure A11: Event-study of IRCCS policy effect on publications for comparisons across different groups. Units included in the IRCCS perimeter in any year after 2018 are excluded from the estimates.

	(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)	0.2700*** (0.0337)	0.3027*** (0.0587)	-0.0006 (0.0577)	0.2902*** (0.0594)	0.3333*** (0.0388)	0.3183*** (0.0400)
Publications (asinh)	0.3213*** (0.0425)	0.3841*** (0.0754)	0.0190 (0.0735)	0.3691*** (0.0763)	0.4011*** (0.0492)	0.3828*** (0.0507)
N	6424	2623	2739	2430	3987	3794
R <sup>2</sup>	0.749	0.605	0.721	0.588	0.798	0.770
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Panel	Full (Balanced)	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

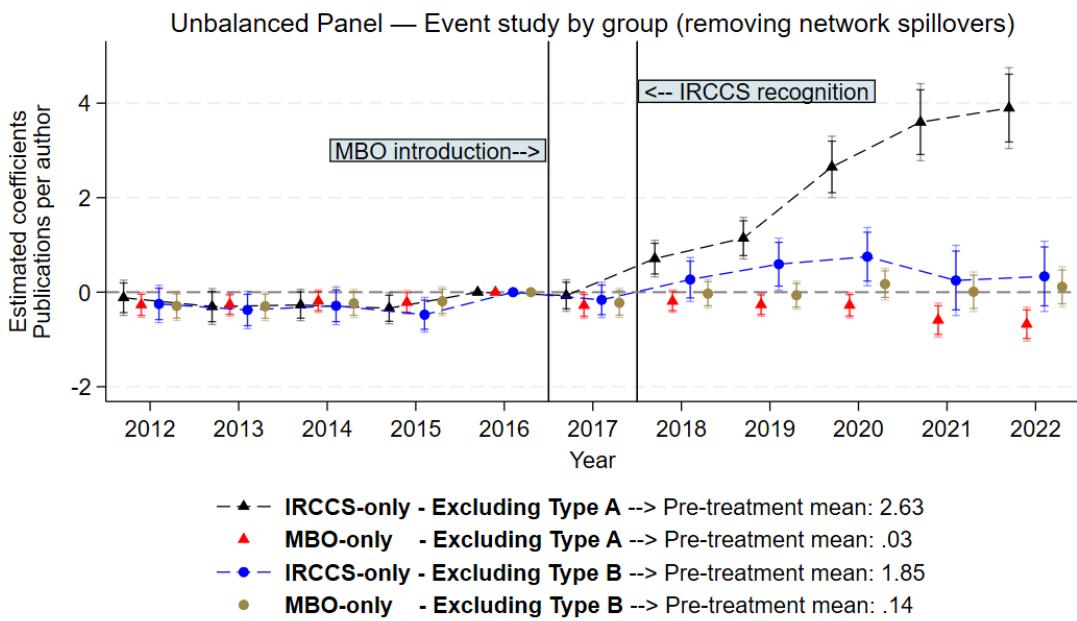
Table A5: Impact of IRCCS recognition on Annual Publications, re-scaled via log. and asinh.

	(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	0.8531*** (0.1666)	0.6286*** (0.2215)	-0.2834 (0.2688)	0.6658*** (0.2244)	0.9494*** (0.1659)	0.9867*** (0.1696)
N	4672	1940	1992	1788	2884	2732
R <sup>2</sup>	0.842	0.661	0.833	0.625	0.864	0.854
Time FE	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Mean	3.56	1.65	1.65	1.65	3.56	3.56
Panel (Balanced)	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 A6: Impact of IRCCS recognition on Annual Publications in the sample truncated to exclude Covid years.



a) Balanced panel - Event-study of the two policies' effect on publications of the IRCCS-treated and MBO-treated only groups 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.



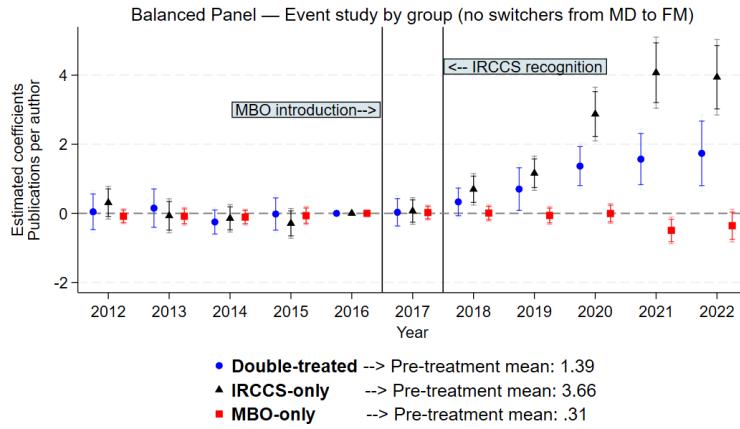
a) Unbalanced panel - Event-study of the two policies' effect on publications of the IRCCS-treated and MBO-treated only groups once excluded double-treated units and *Type A* researchers (MBO-treated units who *ever* co-authored with IRCCS researchers) in the first case, and double-treated units and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

Figure A12: Event-study of for the dynamic effects of the MBO policy and the IRCCS recognition effect on publications in different samples subset based on collaboration dynamics. Double-treated units are always excluded. a) is the Balanced panel, b) the unbalanced one.

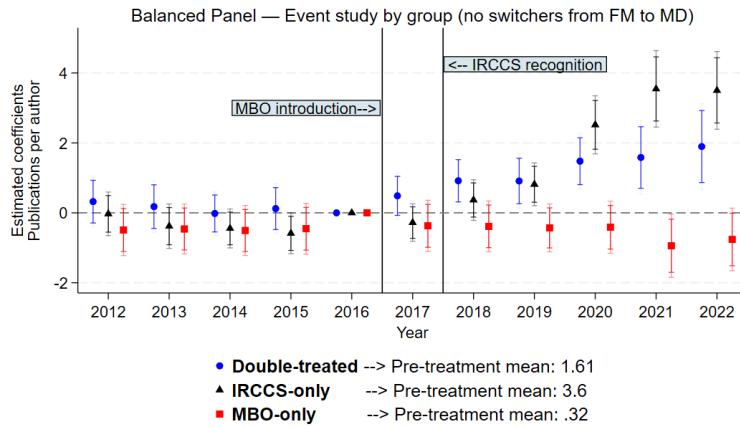
	2012-2022; SEs clustered at individual level							
	(1) Pubs	(2) Pubs	(3) Pubs	(4) Pubs	(5) Pubs	(6) Pubs	(7) Pubs	(8) Pubs
Post 2017*MBO-only	-0.33454 ***	-0.16009			-0.06752	0.25386 ***		
Post 2017*IRCCS-only	[0.08826] 1.81748 ***	[0.10775] 0.50939 ***			[0.06888] 2.26424 ***	[0.07945] 0.65571 ***		
Post 2018*MBO-only	[0.25013]	[0.14837]		-0.37688 ***	-0.18098	[0.23770]	[0.16236]	-0.09643 0.29371 ***
Post 2018*IRCCS-only			[0.08684] 2.24115 ***	[0.11162] 0.68130 ***			[0.07356] 2.63295 ***	[0.08541] 0.75051 ***
			[0.27949]	[0.17094]			[0.25929]	[0.18266]
Observations	6,985	6,717	6,985	6,717	12,195	12,794	12,195	12,794
R-squared	0.79416	0.68426	0.79865	0.68591	0.71270	0.51740	0.71924	0.51824
Individual FE	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	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 I	No Type A	No Type B	No Type A	No Type B	No Type A	No Type B	No Type A	No Type B
Panel II	Balanced	Balanced	Balanced	Balanced	Unbalanced	Unbalanced	Unbalanced	Unbalanced
Mean across treated	2.035	1.269	2.035	1.260	1.011	0.583	1.033	0.597

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

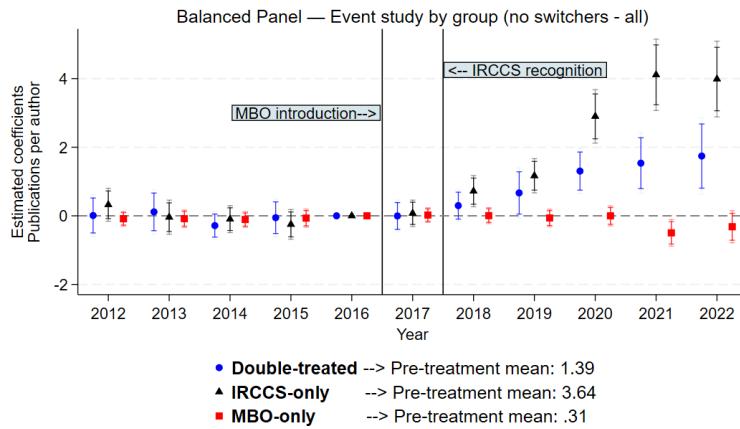
Table A7: Impact of MBO-implementation and IRCCS recognition on Annual Publications in the sub-samples obtained by excluding collaborating units.



a) Removing switchers from Medical Direction to Faculty Membership.

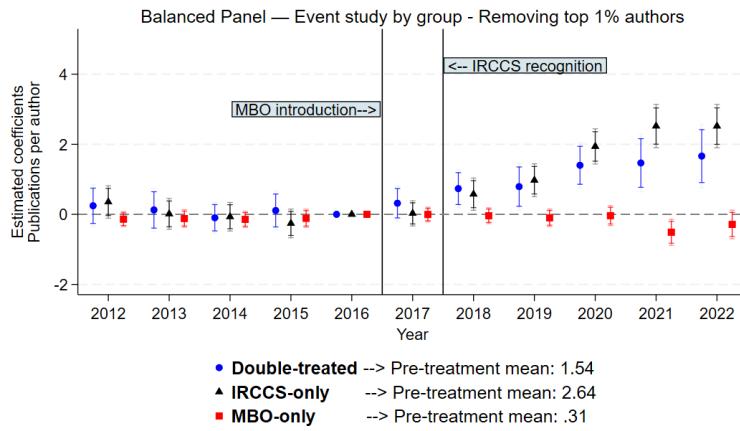


b) Removing switchers from Faculty Membership to Medical Direction.

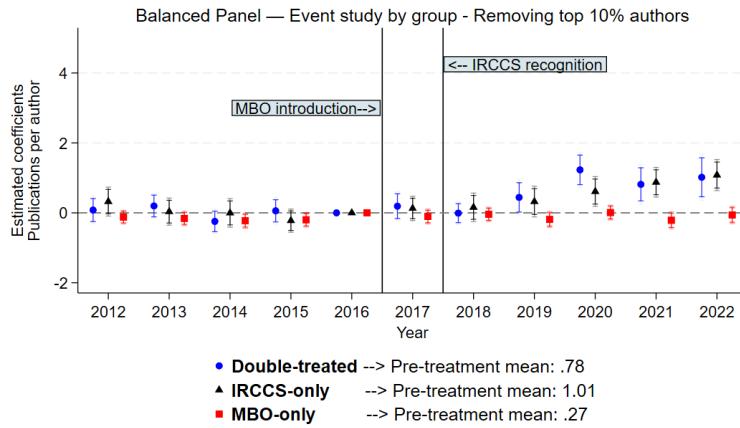


c) Removing all switcher.

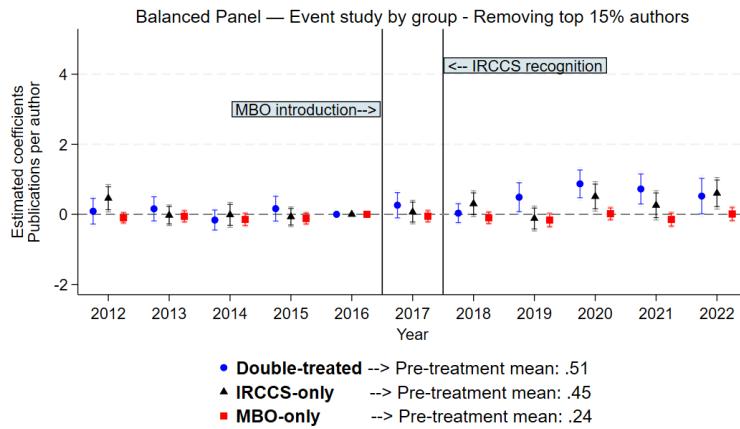
Figure A13: Dynamic impact of MBO-implementation and IRCCS recognition on Annual Publications in the sub-samples obtained by excluding those who switch across groups.



a) Removing top 1% researchers.

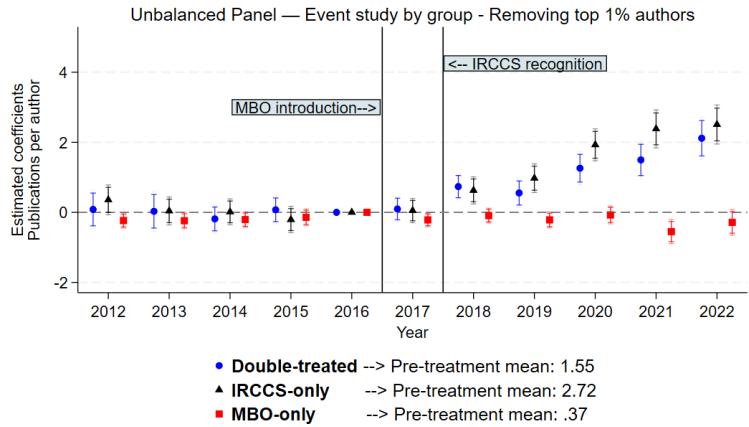


b) Removing top 10% researchers.

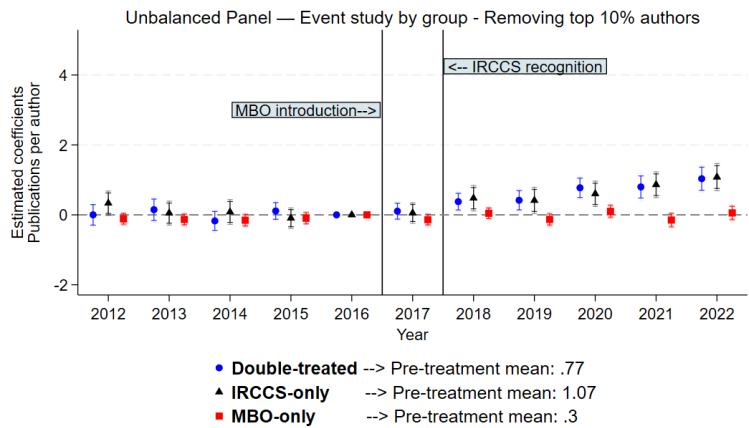


c) Removing top 15% researchers.

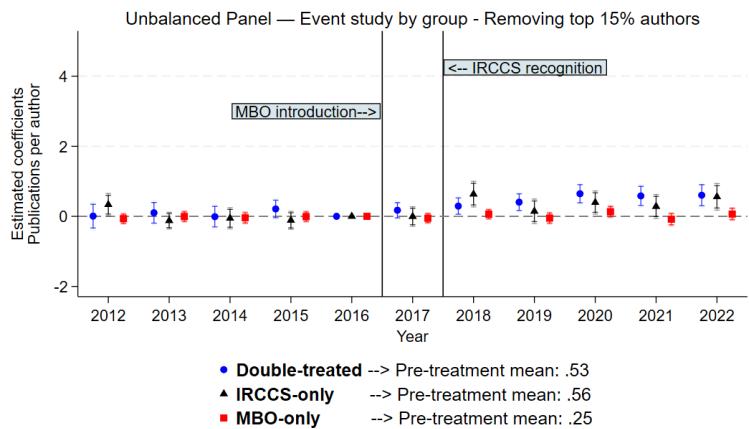
Figure A14: Balanced panel - Dynamic impact of MBO-implementation and IRCCS recognition on annual publications in the sub-samples obtained by excluding most prolific authors.



a) Removing top 1% researchers.

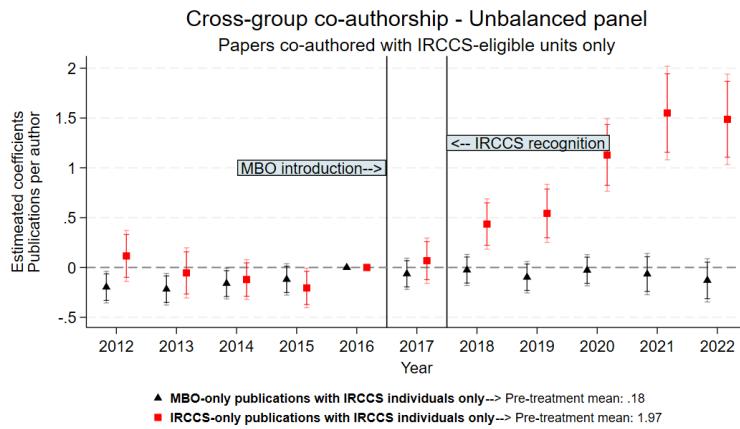


b) Removing top 10% researchers.

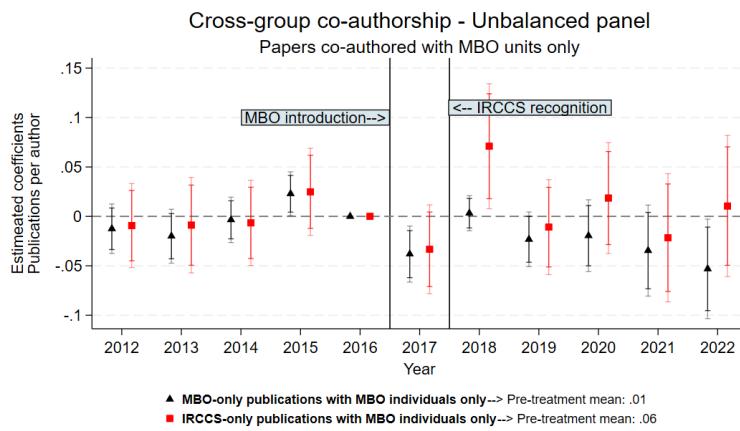


c) Removing top 15% researchers.

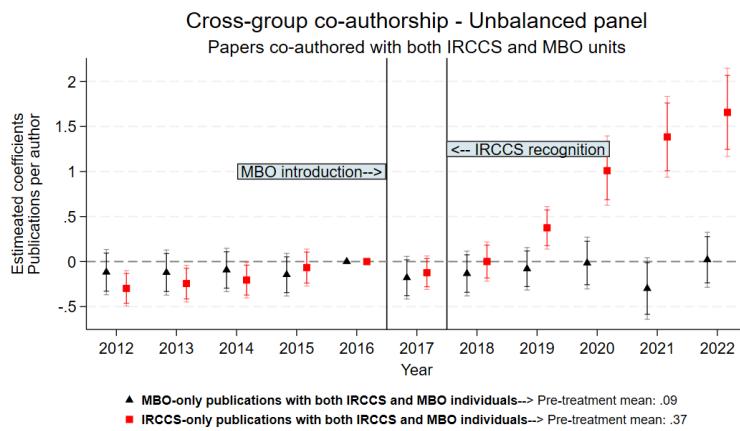
Figure A15: Unbalanced panel - Dynamic impact of MBO-implementation and IRCCS recognition on annual publications in the sub-samples obtained by excluding most prolific authors.



a) Outcome: publications co-authored with IRCCS units only.



b) Outcome: publications co-authored with MBO-eligible units only.



c) Outcome: publications co-authored with both MBO-eligible units and IRCCS researchers.

Figure A16: Dynamic impact of the policy implementation on annual cross-group publications in the balanced panel.

	(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 <sup>2</sup>	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 <sup>2</sup>	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 <sup>2</sup>	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 <sup>2</sup>	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 A8: Regressions between enrollment into IRCCS perimeter and several factors.