esade

EsadeEcPol - Center
for Economic Policy

Attracting Talent to General Practice through Financial Incentives

AUTHORS

Lucía Cobreros
UAM, EsadeEcPol

Carlos Sunyer
UC3M

Judit Vall Castelló
UB, IEB, CRES-UPF

June, 2026

Attracting Talent to General Practice through Financial Incentives

Lucía Cobreros — UAM
Carlos Sunyer — UC3M
Judit Vall Castelló — UB, IEB, CRES-UPF

June 18, 2026

Abstract

Shortages of general practitioners (GPs) strain healthcare systems worldwide. We test whether wage incentives can attract higher-ability physicians using a 2022 reform in Catalonia, Spain, that granted incoming GP residents unconditional salary supplements of 26% to 47% of base pay across three tiers. Using administrative data on medical residency choices nationwide and a synthetic difference-in-differences design, we document two effects. First, the reform raised entrants' standardized exam scores by 0.067 standard deviations on average (0.174 in the highest subsidy tier) and increased the international resident share by nearly 30 percentage points. Second, GP vacancies emerged nationwide after 2022 but rose 4.4 percentage points less in Catalonia, drawing lower-scoring candidates to the newly filled slots and mechanically diluting the average score gain. Lee bounds correcting for this compositional shift raise the estimated effects to 0.088 and 0.191 standard deviations in a fully cleared market. Our findings show that salary supplements can improve the allocation of medical talent to shortage occupations, though their effectiveness depends on local labor-market tightness and physician supply.

JEL Codes: I11, J24, H75

Key words: Monetary incentives, GP recruitment, healthcare workforce quality

1 Introduction

Primary health care (PHC) is a central input into health production. With general practitioners (GPs) at its core, PHC improves population health, reduces avoidable hospital use, and expands access to care.[1] Yet this core function is increasingly strained by shortages in GP across high-income countries (OECD 2024). Policy responses have largely focused on one side of this problem, the *extensive margin*, seeking to fill vacant GP positions. Shortages, however, may also arise along the *intensive margin* if GP fails to attract physicians with higher human capital. This second margin has received less attention, even though growing evidence shows that patient outcomes depend not only on provider availability, but also on who provides care.[2]

In GP, both margins are empirically relevant. Vacancies remain common, particularly in rural and underserved areas (Khoury et al. 2025), while general practice disproportionately attracts graduates from the lower tail of the ability distribution, a pattern reflecting working conditions, workload, prestige and income differences relative to other specialties (Nicholson 2002; Sivey et al. 2012; Russo et al. 2023; OECD 2024). A central question is therefore whether financial incentives can improve the human capital that GP attracts and, if so, which candidates respond. We answer this question by exploiting a 2022 reform in Catalonia, Spain, that introduced salary supplements for incoming GP residents, ranging from 26% to 47% of base pay and varying across three geographic intensity tiers.

The Spanish setting is well suited to address this question. Medical graduates sit a standardized national residency entrance exam, the MIR, which is common to all candidates and closely related to the medical knowledge used during residency. Since candidates are ranked almost entirely on the basis of this score and then choose a specialty–hospital pair sequentially in rank order, the system provides both a transparent proxy for candidate ability and revealed preferences over specialty and location (Machado et al. 2012; Martinez-Miera and Sunyer 2025).[3] Spain is also not an exception to the broader challenge of attracting

1. See, among others, Starfield et al. (2005), Dusheiko et al. (2011), Bailey and Goodman-Bacon (2015), Gong et al. (2019), and Mora-García et al. (2024).
2. Evidence on physician quality and medical human capital includes Aakvik and Holmås (2006), Doyle Jr et al. (2010), Chen (2021), Kunz and Propper (2023), Ginja et al. (2025), and Posso et al. (2025).
3. Standardized medical exam scores have also been used to proxy for physician talent in other matching markets, such as USMLE scores in the United States (Gottlieb et al. 2025). The Spanish MIR offers a cleaner setting for studying preferences because allocation is one-sided rather than two-sided. Conditional on exam rank, hospitals do not select residents. Instead, candidates sequentially choose among the remaining specialty–hospital pairs, making observed choices a clean measure of revealed preferences over specialties and locations.

talent into general practice. GP positions attract lower-scoring candidates than other specialties, and Catalonia exhibits one of the country's largest gaps between the ability of the graduates it trains and that of the residents who enter its GP positions, as shown in Figure 1. This makes Catalonia both a policy priority and an informative case for studying whether financial incentives can improve the human capital of physicians entering general practice.

Our analysis draws on administrative records covering the universe of medical residents in Spain between 2019 and 2024. We estimate the policy's causal effect using a synthetic difference-in-differences (SDID) design (Arkhangelsky et al. 2021), comparing GP positions in Catalonia by intensity tier with the remaining choices in the MIR system. We then trace the dynamics of the response using event-study specifications.

We obtain two main sets of results. First, the reform raised the average MIR score of GP entrants in Catalonia by 0.067 standard deviations, although this average masks sharp heterogeneity across intensity tiers. The effect is concentrated among positions receiving the largest wage supplements, where scores rise by 0.174 standard deviations, while effects in low- and medium-intensity areas are close to zero. The reform also reshaped the geographic composition of entrants. The share of internationally trained residents increased by 29.7 percentage points, while the share of residents trained in other Spanish regions fell by a similar amount and the share of Catalan-trained residents remained unchanged. This pattern is consistent with evidence that international professionals are more geographically mobile and more responsive to economic incentives (Kleven et al. 2013; Schündeln 2014; Adovor et al. 2021; Braga et al. 2024).

Second, although the reform did not explicitly target vacancies, it reduced the share of GP positions left unfilled by 4.4 percentage points relative to comparable GP positions elsewhere. This matters for interpreting the quality effect. By filling positions that would otherwise have remained vacant, the policy may bring in marginal candidates from the lower tail of the score distribution, which mechanically attenuates the average score gain among entrants (Agarwal 2015, 2017). We therefore use Lee bounds to ask how large the score effect would be after equalizing fill rates between treated and control positions (Lee 2009). This adjustment yields slightly larger estimates, increasing the pooled effect from 0.067 to 0.088 standard deviations and the high-intensity effect from 0.174 to 0.191.

This paper contributes to the literature on financial incentives and human capital allocation in public-sector occupations (Clotfelter et al. 2008; Dal Bó et al. 2013; Biasi 2021). First, we shift the focus of the physician-attraction literature from *quantity* and *location* to *quality*. Existing work finds limited effects of location-choice incentives (Costa et al. 2024), supply-side expansions (McNamara and Pineda-Torres 2025), and loan-forgiveness programs (Kulka and McWeeny 2019), consistent with the low short-run wage elasticity of physician supply (Agarwal 2017; Khoury et al. 2025). We instead study whether an *unconditional* salary supplement changes the academic ability of physicians entering general practice, and how this response varies with incentive size. Second, by exploiting a regional supplement within a single national matching market, we provide causal evidence that internationally trained professionals respond strongly to purely financial incentives. This adds to the literature on tax and wage differentials and high-skilled mobility (Kleven et al. 2013, 2014; Akcigit et al. 2016; Adovor et al. 2021; Braga et al. 2024). Third, we observe both the intensive and extensive margins of a matching market within the same policy. This allows us to connect changes in entrant quality to changes in vacancies, providing direct empirical evidence on the mechanism highlighted by Agarwal (2015, 2017).

Beyond these contributions, our findings inform the design of wage policies for hard-to-staff public roles. Unconditional incentives can markedly improve the academic profile of entrants, but their effectiveness is sensitive to local cost-of-living conditions and must be calibrated accordingly.

2 Institutional Setting

2.1 The MIR System

The Spanish medical residency system, known as the MIR (*Médico Interno Residente*), allocates new physicians to residency positions through a centralized national procedure. All medical graduates who wish to train as specialists in the Spanish National Health System must complete a four- or five-year residency. Positions are defined as specialty–hospital pairs and cover approximately 46 specialties across around 850 accredited centers, with about 8,500 positions offered in the last year of our sample.

The allocation order is determined by candidates' total score, computed as a weighted average of two merit-based components: a standardized nationwide exam, the MIR examination, which accounts for 90% of the score, and the candidate's undergraduate GPA, which accounts for the remaining 10%. To participate in the position selection, candidates must clear a minimum threshold on the exam component, set at 35% of the average score of the top-ten candidates throughout most of our study period. Candidates are then ranked by their total score and select positions in rank order, with the highest-ranked candidate choosing first, until all positions are filled. The exam is held in late January or early February, scores are released in March or April, and residents begin training in early summer.[4]

This institutional design is useful for two reasons. First, the MIR score provides a common, high-stakes, and nationally comparable measure of academic medical human capital. The exam produces stable rankings that discriminate consistently across cohorts, and previous work using the MIR has shown that residency choices and exam-based allocation contain information about physician and hospital quality (Machado et al. 2012; Martinez-Miera and Sunyer 2025). Second, because hospitals do not choose among candidates, observed allocations directly reflect candidates' choices conditional on rank. The MIR therefore allows us to study not only which candidates enter general practice, but also how financial incentives reshape revealed preferences over specialties and locations.

2.2 The Policy Intervention

In March 2022, after the MIR exam had already taken place but resident placements had not yet started, the regional government of Catalonia introduced a program of unconditional salary supplements targeting incoming GP residents.[5]

The supplement is geographically tiered into three intensity levels. Residents assigned to the Barcelona Metropolitan Area receive the lowest annual supplement, €5,125 (*Low intensity*); those in the rest of the province of Barcelona receive €7,175 (*Medium intensity*); and those in the remaining provinces of Catalonia receive €9,225 (*High intensity*).[6] Figure A1 displays the spatial distribution. Given a 2022 base salary of approximately €19,500, these supplements represent annual increases of 26%, 37%, and 47% over base pay. Paid throughout the four-year residency, the cumulative transfer ranges from €20,500 to €36,900, comparable in magnitude to the loan-forgiveness amounts evaluated in the U.S. literature (Falcettoni 2019; Kulka and McWeeny 2019).

The geographic structure is consistent with the government's stated objectives of *territorial equity* and *talent attraction*. Before the reform, GP entrants' MIR scores declined mono-

4. See Díez-Rituerto et al. (forthcoming) for further details on the MIR examination.
5. Comprehensive details are publicly available at the Catalan Health Department's official portal: `https://canalsalut.gencat.cat/ca/professionals/formacio/especialitzada/mir-medicina-familiar-comunitaria/`.
6. Hospitals in the Catalunya Central healthcare region, although geographically within the province of Barcelona, are administratively classified as high-intensity.

tonically across treatment tiers, from the 20th percentile in low-intensity areas to the 8th and 5th percentiles in medium- and high-intensity areas, respectively. The largest supplements therefore targeted the areas with the lowest pre-policy academic profile among incoming GP residents (Figure A2).

3 Data

Our analysis draws on administrative data from the Spanish Ministry of Health, covering the universe of candidates who registered for the MIR exam between 2019 and 2024. The dataset contains 47,380 individuals who started their medical residency during this period, with detailed information on each candidate's hospital–specialty choice, MIR exam score, undergraduate GPA, university of origin, and nationality. Of these, 28% chose GP and 15.6% started their training in Catalonia. The treated population, candidates choosing GP in Catalonia, comprises 2,052 entrants, of which 42.3%, 20.7%, and 37.0% were placed in low-, medium-, and high-intensity tiers, respectively.

For estimation, we aggregate to the hospital–specialty level, building a balanced panel of 16,146 hospital–specialty–year observations covering all hospitals offering each specialty in every year between 2019 and 2024.[7] For each cell, we compute the median MIR score and the mean of the compositional outcomes.[8]

Summary statistics are reported in Table A2. MIR scores are systematically lower among GP entrants than among entrants into other specialties. Before the reform, the median GP entrant nationally was at the 19th percentile of the MIR distribution, compared with the 66th percentile for other specialties. The gap was even larger in Catalonia, where the corresponding figures were the 10th and 65th percentiles, with substantial heterogeneity across treatment tiers (Figure A3). GP entrants in Catalonia also differed in their training origin. Relative to GP positions elsewhere in Spain, Catalonia had a lower share of locally trained residents and a higher share of internationally trained residents, with this imbalance particularly pronounced in higher-intensity areas.

7. Constructing the balanced panel requires excluding hospital–specialty pairs not recorded in every year of the sample (8.1% of cells). Only one excluded pair is treated, corresponding to a hospital that did not offer GP before the policy, so its omission is unlikely to introduce bias. Results are unchanged when aggregating at the province–specialty level with and without these cells.

8. We use standardized MIR scores rather than percentile ranks as our main outcome. A fixed difference in standardized scores corresponds to a fixed difference in the underlying score distribution, whereas a fixed percentile difference depends on the density of candidates around each point of the distribution. Percentile-based estimates are reported as a robustness check in Table A1.

4 Empirical Identification

To assess the causal impact of the policy, we compare outcomes before and after 2022 for treated units, defined as GP positions in Catalonia, separately by intensity tier, relative to untreated units in the MIR system. A purely regional comparison, using GP positions elsewhere in Spain, is insufficient because the policy may reallocate GP candidates across regions. A purely specialty-based comparison, using other specialties within Catalonia, is also insufficient because the policy may change the relative attractiveness of GP within the regional specialty menu. Since the MIR is a centralized national market with a common pool of candidates, both comparison groups may be indirectly affected by the reform. This motivates a triple-difference specification (DDD) that exploits variation across both geography and specialty.[9]

As the DDD event studies reveal differential pre-policy trends in scores and compositional outcomes, especially across intensity tiers (Figure A4), we rely on a synthetic difference-in-differences (SDID) design (Arkhangelsky et al. 2021). SDID combines synthetic-control reweighting with the fixed-effects structure of difference-in-differences by choosing unit and time weights that improve pre-treatment fit between treated and comparison units. This approach is particularly useful in our setting, where treated cells account for only 4.3% of the market and the relevant comparison group is not obvious ex ante. Moreover, the weighting procedure reduces reliance on an arbitrary choice of controls and improves balance in pre-policy dynamics (Ferman and Pinto 2021).

For estimation, we treat each hospital (h) specialty (j) pair as a single unit, indexed by i, with α_i denoting unit fixed effects and W_{it} denoting the corresponding treatment indicator. The first step of the procedure recovers two sets of weights from pre-treatment data: unit weights $\hat{\omega}_i^{sdid}$, which match the trajectory of control units to that of the treated ones, and time weights $\hat{\lambda}_t^{sdid}$, which balance the contribution of pre- and post- treatment periods. The treatment effect is then recovered from a weighted two-way fixed-effects regression that uses these weights:

9. Formally, the DDD specification would be

$$Y_{hjt} = \beta\, GP_j \times CAT_{r(h)} \times Post_t + \gamma\, GP_j \times CAT_{r(h)} + \delta\, GP_j \times Post_t + \zeta\, CAT_{r(h)} \times Post_t + \mu_{r(h)jt} + \varepsilon_{hjt},$$

where h, j, and t denote hospital, specialty, and year, $r(h)$ is the region of hospital h, and GP_j, $CAT_{r(h)}$, $Post_t$ are indicators for the GP specialty, hospitals in Catalonia, and years from 2022 onward. The vector $\mu_{r(h)jt}$ collects region, specialty, and time fixed effects in their various saturated combinations.

$$\left(\hat{\tau}^{sdid}, \hat{\mu}, \hat{\alpha}, \hat{\beta}\right) = \arg\min_{\tau,\mu,\alpha,\beta} \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{it} - \mu - \alpha_i - \beta_t - W_{it}\,\tau)^2\, \hat{\omega}_i^{sdid}\, \hat{\lambda}_t^{sdid}, \tag{1}$$

where Y_{it} denotes the outcome of interest (score or compositional variables), β_t are year fixed effects, and W_{it} takes value one for units offering GP in Catalonia from 2022 onward and zero otherwise. The coefficient τ identifies the average effect of the policy on the treated cells. We compute standard errors via bootstrap.[10] We complement the pooled estimate with event-study specifications to trace dynamic effects.[11]

Crucially, identification requires the Stable Unit Treatment Value Assumption (SUTVA) (Rubin 1980). This is a relevant concern in our setting because the MIR is a centralized national allocation system, implying that if the policy attracts a candidate into a treated GP position, that candidate is no longer available to fill another position elsewhere, creating potential spillovers across regions or specialties (Rubin 1980; Sobel 2006). In practice, however, the scope for such spillovers is limited by the scale of the intervention, as treated positions account for only 4.3% of the resident pool. To assess whether remaining spillovers affect our estimates, we implement a ring approach (Kline and Moretti 2014; Butts 2023) that progressively excludes from the donor pool the units most exposed to re-sorting, both geographically and across specialties. Ressuringly, the estimates remain quantitatively stable across these exclusions (Table A3), suggesting that SUTVA violations do not drive our results.

5 Results

5.1 Intensive margin: human capital attraction

Table 1 reports SDID estimates of the policy's effect on standardized MIR scores. The pooled effect is modest but statistically significant, at 0.067 standard deviations ($p = 0.036$). This average, however, masks substantial heterogeneity across treatment intensity. Effects in low- and medium-intensity areas are close to zero and statistically insignificant. The response is concentrated in high-intensity areas, where scores increase by 0.174 standard deviations ($p < 0.01$). This is a sizable effect given that these areas had the lowest pre-policy GP scores

10. The two alternative inference procedures proposed by Arkhangelsky et al. (2021), placebo and jackknife, rely on assumptions about the variance structure of the errors that need not hold in our setting. Bootstrap inference imposes no such restrictions, at the cost of being more computationally demanding.

11. We implement the estimator using the Stata `sdid_event` package (Ciccia 2024), an event-study extension of the parent `sdid` command (Clarke et al. 2024).

in the country.

The dynamics reinforce this interpretation. Figure 2 plots median MIR scores for GP entrants by treatment intensity and for entrants in other specialties, indexed to 2021. After the reform, high-intensity GP positions display a sharp upward shift relative to their pre-policy trajectory, while low- and medium-intensity GP positions, as well as non-GP positions, show no comparable movement. The SDID event-study estimates in Figure A6 align closely with this raw pattern. The high-intensity response appears immediately after the reform, reaching 0.37 standard deviations in 2022. It declines in 2023 and rises again to 0.30 in 2024, while low- and medium-intensity areas remain close to zero throughout. We return to the 2023 decline in Section 6.

The magnitude of the high-intensity response is economically meaningful. To our knowledge, there are no directly comparable causal estimates for wage incentives targeting the human-capital composition of physicians entering shortage specialties. We therefore benchmark our estimates against the broader literature on educational interventions, where Kraft (2020) provides a widely used reference point. He reports a median effect size of 0.10 standard deviations for interventions on standardized scores, with effects closer to 0.05 when implemented at scale and effects above 0.20 typically considered *large*. Against this benchmark, the average high-intensity effect of 0.174 is large, and the first-year effect of 0.37 is particularly substantial.

5.1.1 The non-linear structure of the effect

The score effect is concentrated in high-intensity units, even though supplements in low- and medium-intensity units also exceeded 26% of base salary. Why did residents respond so differently across tiers? We explore two non-mutually exclusive explanations.

The first is within-region re-sorting: residents who would otherwise have chosen GP in low- or medium-intensity hospitals may have been drawn instead to high-intensity ones. The evidence does not support this interpretation. Score effects in low- and medium-intensity units are essentially zero rather than negative (Table 1), as would be expected if high-intensity hospitals drained higher-scoring residents from the rest of Catalonia. Moreover, the SDID effect on the share of Barcelona-trained residents placed in high-intensity hospitals is –9.6 percentage points ($p = 0.053$), pointing, if anything, to the opposite pattern. These results are inconsistent with within-region drainage of talent as the main driver of the heterogeneous score response.

The second explanation is that the higher cost of living in low- and medium-intensity areas may have offset part of the financial incentive. To examine this channel, we estimate the policy's effect separately for each treated hospital and regress the resulting estimates on the rental price index of the hospital's municipality.[12] The relationship is negative and statistically significant ($p = 0.01$). It remains significant after controlling for working conditions, workload, and broad geography ($\hat{\beta} = -0.017$, $p = 0.076$), suggesting that the cost-of-living channel operates beyond non-pecuniary job characteristics or location (Figure A8). While suggestive, this evidence is consistent with higher living costs muting the response in low- and medium-intensity areas.

This geography of the effect has a direct distributional implication. Because the response is concentrated in high-intensity units, where pre-policy scores were lowest, the reform reduced human-capital disparities across treated areas, one of its stated objectives (Section 2.2). We document this convergence by regressing hospital-level effects on pre-policy median MIR scores. Figure A9 and Table A5 show a negative relationship that survives the inclusion of intensity fixed effects, indicating that convergence operates not only across tiers, but also within them.

5.2 Compositional shifts in residency preferences

Table 2 reports SDID estimates for the composition of GP entrants by training origin. The reform raises the share of internationally trained residents by 29.7 percentage points on average, with effects increasing from 13.1 percentage points in low-intensity areas to 30.8 and 39.7 percentage points in medium- and high-intensity areas, respectively. This increase is matched primarily by a decline in residents trained in other Spanish regions, whose share falls by 17.1, 13.7, and 35.1 percentage points across the three tiers. By contrast, the share of locally trained residents remains largely unchanged, except in medium-intensity areas, where it declines by 15.6 percentage points. The raw trends in Figure 3 illustrate the dynamics behind these estimates. They show a sharp post-reform increase in internationally trained entrants and a corresponding decline among entrants trained in other Spanish regions, both of which persist throughout the post-reform period. The SDID event-study estimates in Figure A10 align closely with this raw pattern.

12. For each treated hospital we re-run the SDID design treating only its GP cells from 2022 onward, and rebuild the synthetic control from untreated cells, namely GP cells outside Catalonia and all non-GP cells, while excluding the other treated hospitals from the donor pool. The hospital-level effect is the post-policy SDID coefficient averaged over 2022–2024.

The strong international response is consistent with evidence that internationally trained professionals face fewer location-specific constraints and are more responsive to economic incentives (Kleven et al. 2013, 2014; Schündeln 2014). This margin is particularly relevant in our setting, where roughly 91.4% of international entrants come from Latin American countries with substantially lower income levels. Decomposing the score effect by origin group (Table A6) suggests that the new international entrants have a slightly lower academic profile (–0.112 standard deviations, $p = 0.12$), while nationally trained residents show positive and marginally significant score gains. The estimates are 0.170 standard deviations for locally trained residents ($p = 0.08$) and 0.287 for residents trained in other Spanish regions ($p = 0.08$), although these estimates should be interpreted cautiously given the small within-group samples.

5.3 Reconciling the intensive and extensive margins

Although the policy was not designed to reduce vacancies, and no GP positions had gone unfilled before 2022, vacancies emerged precisely in the first year of the policy and concentrated in GP. Figure A11 shows that vacancy rates rose in both Catalonia and the rest of Spain, but remained lower in Catalonia. A DiD specification restricted to GP confirms this divergence: vacancy rates in Catalonia are approximately 4.4 percentage points lower than in control regions after 2022 (Table A7).[13]

This divergence affects the interpretation of our intensive-margin estimates. Catalonia's higher post-policy fill rate implies that 2.87 additional percentage points of positions are filled in treated units but remain vacant in controls. Since vacancies weaken competition for the remaining positions (Agarwal 2015, 2017), these marginal entrants are likely drawn from the lower tail of the score distribution. Their inclusion depresses the observed average score in treated units, so the baseline SDID estimates likely understate the policy's effect on academic quality in a fully cleared market.

We formalize this attenuation using Lee bounds (Lee 2009). We trim the lowest-scoring 2.87% of post-policy GP observations in Catalonia and re-estimate the SDID. The pooled upper bound is 0.088 standard deviations, compared with the baseline estimate of 0.067.

13. Since vacancy rates were zero across all regions before 2022, the specification effectively reduces to a post-policy comparison between treated and control regions with hospital fixed effects, rather than a conventional DiD exploiting pre-trends. Specifically, we estimate $Vacancies_{h(r)t} = \eta_h + \alpha_t + \beta, Treat_r \times Post_t + \varepsilon_{h(r)t}$, where $Vacancies_{h(r)t}$ is the vacancy rate in hospital h in year t, $Treat_r$ equals one for hospitals in Catalonia, and $Post_t$ equals one from 2022 onward. The specification includes hospital and year fixed effects, with standard errors clustered at the province level.

In high-intensity units, the upper bound is 0.191, compared with 0.174 at baseline. In low-intensity units, where all positions are filled, the bound coincides with the baseline, while in medium-intensity units the reverse fill-rate gap requires trimming the control distribution and yields an estimate slightly below the baseline. Full results are reported in Table A8. Overall, the bounds are slightly above the baseline estimates wherever vacancies are present, confirming that our main estimates are conservative.

6 Robustness and Sensitivity

The dip in 2023. The dynamic estimates in Section 5.1 show a temporary decline in 2023 that, although not statistically significant, warrants comment. This year coincided with three national MIR reforms. The qualifying threshold was lowered from 35% to 25% of the top-ten candidates' average score, reducing exam failures by approximately 58%. A fast-track homologation procedure for foreign medical degrees also came into effect under Real Decreto 889/2022, and the non-EU quota of MIR positions increased from 4% to 6%. Although national in scope, these reforms are particularly relevant for our setting because Catalonia hosts a structurally larger share of internationally trained residents and GP positions in Catalonia historically attract candidates from the lower tail of the score distribution. Any expansion of the MIR pool at the lower end is therefore likely to affect treated cells disproportionately. This interpretation is consistent with the leftward shift in internationally trained candidates' scores observed across both treated and control regions in 2023. The dip likely reflects a combination of the policy's effect and these contemporaneous national reforms, which we cannot separately identify. Re-estimating the SDID specification excluding 2023 yields a pooled effect of 0.353 standard deviations and a high-intensity effect of 0.615 standard deviations (Table A9), consistent with the baseline estimates being conservative.

Donor pool. The ring-based exclusion strategy described in Section 4 yields stable results.[14] The high-intensity score effect ranges from 0.151 to 0.180 standard deviations across the three exclusion specifications, close to the baseline estimate of 0.174 and statistically significant throughout (Table A3). The SDID unit weights are also dispersed, with no single donor unit accounting for more than 1.5% of the synthetic control (Figure A12), indicating that

14. The excluded regions are Valencia, Madrid, Andalusia, and the Canary Islands. The excluded specialties are Preventive Medicine and Public Health, Clinical Analysis, Allergology, Occupational Medicine, Microbiology and Parasitology, Pathological Anatomy, Clinical Neurophysiology, Geriatrics, Clinical Pharmacology, and Nuclear Medicine.

inference is not driven by any particular comparison unit.

Empirical specification. Three additional checks are reported in the appendix. First, the results are stable across levels of geographic aggregation (Figure A7). Second, including the annual share of GP positions offered as a supply-side covariate leaves the high-intensity effect essentially unchanged (Table A4). Third, replacing the median MIR score with the mean yields a high-intensity effect of 0.128 standard deviations, somewhat below the baseline estimate of 0.174. This attenuation is consistent with the right-skew of the within-cell score distribution pulling the mean toward the upper tail (Table A10).

7 Conclusion

This paper provides new causal evidence on the impact of unconditional financial incentives on the human capital of physicians entering GP. We study a large salary supplement introduced in Catalonia and show that it raised the average MIR score of GP entrants by 0.067 standard deviations. This average masks substantial heterogeneity. The effect is concentrated in high-intensity areas, where scores increase by 0.174 standard deviations, precisely where pre-policy human capital was lowest, while effects in low- and medium-intensity areas are close to zero. The reform also changed the composition of entrants, increasing the share of internationally trained residents by nearly 30 percentage points and reducing the share of residents trained in other Spanish regions by a similar amount, with little change among locally trained residents. Although the reform did not explicitly target vacancies, it reduced the share of GP positions left unfilled by 4.4 percentage points relative to control regions. Lee bounds suggest that accounting for this differential selection into filled positions yields somewhat larger quality effects.

Taken together, the results show that unconditional wage incentives can improve the allocation of medical talent into shortage occupations, but that their effectiveness is highly uneven across local labor markets. The reform was most effective where need was greatest, suggesting that targeted financial incentives can reduce disparities in the academic profile of GP entrants. At the same time, the absence of detectable score gains in tiers covering close to 60% of treated GP positions, despite supplements of up to 35% of base pay, points to limits of nominal wage policies when local amenities, cost of living, or working conditions offset the value of the transfer. This raises questions about the efficiency of uniform or insufficiently calibrated incentive schemes.

Our findings point to two directions for future research. First, the strongly non-linear response to incentive size calls for a better understanding of the local conditions that mediate the effectiveness of wage policies, especially cost of living, labor-market tightness, and non-wage job attributes. Second, given the evidence that physician quality affects patient outcomes, an important open question is whether improvements in the academic composition of GP entrants translate into measurable health gains, particularly in underserved areas where the social returns to provider quality may be highest.

References

Aakvik, A., Holmås, T.H., 2006. Access to primary health care and health outcomes: the relationships between gp characteristics and mortality rates. Journal of Health Economics 25, 1139–1153.

Adovor, E., Czaika, M., Docquier, F., Moullan, Y., 2021. Medical brain drain: How many, where and why? Journal of Health Economics 76.

Agarwal, N., 2015. An empirical model of the medical match. American Economic Review 105, 1939–1978.

Agarwal, N., 2017. Policy analysis in matching markets. American Economic Review 107, 246–250.

Akcigit, U., Baslandze, S., Stantcheva, S., 2016. Taxation and the international mobility of inventors. American Economic Review 106, 2930–2981.

Arkhangelsky, D., Athey, S., Hirshberg, D.A., Imbens, G.W., Wager, S., 2021. Synthetic difference-in-differences. American Economic Review 111, 4088–4118.

Bailey, M.J., Goodman-Bacon, A., 2015. The war on poverty's experiment in public medicine: Community health centers and the mortality of older americans. American Economic Review 105, 1067–1104.

Biasi, B., 2021. The labor market for teachers under different pay schemes. American Economic Journal: Economic Policy 13, 63–102.

Braga, B., Khanna, G., Turner, S., 2024. Migration policy and the supply of foreign physicians: evidence from the conrad 30 waiver program. Journal of Economic Behavior & Organization 226, 106682.

Butts, K., 2023. Jue insight: Difference-in-differences with geocoded microdata. Journal of Urban Economics 133, 103493.

Chen, Y., 2021. Team-specific human capital and team performance: Evidence from doctors. American Economic Review 111, 3923–3962. doi:10.1257/aer.20201238.

Ciccia, D., 2024. A Short Note on Event-Study Synthetic Difference-in-Differences Estimators. Working Paper arXiv:2407.09565.

Clarke, D., Pailanir, D., Athey, S., Imbens, G., 2024. On synthetic difference-in-differences and related estimation methods in stata. The Stata Journal 24, 557–598.

Clotfelter, C., Glennie, E., Ladd, H., Vigdor, J., 2008. Would higher salaries keep teachers in high-poverty schools? evidence from a policy intervention in north carolina. Journal of Public Economics 92, 1352–1370.

Costa, F., Nunes, L., Sanchez, F., 2024. How to attract physicians to underserved areas? policy recommendations from a structural model. Review of Economics and Statistics

106, 36–52.

Dal Bó, E., Finan, F., Rossi, M.A., 2013. Strengthening state capabilities: The role of financial incentives in the call to public service. The Quarterly Journal of Economics 128, 1169–1218.

Díez-Rituerto, M., Gardeazabal, J., Iriberri, N., Rey-Biel, P., forthcoming. Gender gaps in access to medical intern positions: The role of competition. Journal of Labor Economics doi:10.1086/736551.

Doyle Jr, J.J., Ewer, S.M., Wagner, T.H., 2010. Returns to physician human capital: Evidence from patients randomized to physician teams. Journal of Health Economics 29, 866–882.

Dusheiko, M., Gravelle, H., Martin, S., Rice, N., Smith, P.C., 2011. Does better disease management in primary care reduce hospital costs? evidence from english primary care. Journal of Health Economics 30, 919–932.

Falcettoni, E., 2019. The determinants of physicians' location choice: Understanding the rural shortage. Unpublished manuscript.

Ferman, B., Pinto, C., 2021. Synthetic controls with imperfect pretreatment fit. Quantitative Economics 12, 1197–1221.

Ginja, R., Riise, J., Willage, B., Willén, A., 2025. Does your doctor matter? Journal of Political Economy Microeconomics 3, 497–538. doi:10.1086/732838.

Gong, G., Phillips, S.G., Hudson, C., Curti, D., Philips, B.U., 2019. Higher us rural mortality rates linked to socioeconomic status, physician shortages, and lack of health insurance. Health Affairs 38, 2003–2010.

Gottlieb, J.D., Polyakova, M., Rinz, K., Shiplett, H., Udalova, V., 2025. The earnings and labor supply of us physicians. The Quarterly Journal of Economics , qjaf001.

Khoury, S., Leganza, J.M., Masucci, A., 2025. Health professional shortage areas and physician location decisions. American Journal of Health Economics 11, 487–527.

Kleven, H.J., Landais, C., Saez, E., 2013. Taxation and international migration of superstars: Evidence from the european football market. American Economic Review 103, 1892–1924.

Kleven, H.J., Landais, C., Saez, E., Schultz, E., 2014. Migration and wage effects of taxing top earners: Evidence from the foreigners' tax scheme in denmark. The Quarterly Journal of Economics 129, 333–378.

Kline, P., Moretti, E., 2014. Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority. The Quarterly Journal of Economics 129, 275–331.

Kraft, M.A., 2020. Interpreting effect sizes of education interventions. Educational researcher 49, 241–253.

Kulka, A., McWeeny, D., 2019. Rural physician shortages and policy intervention. Unpublished manuscript.

Kunz, J.S., Propper, C., 2023. Jue insight: Is hospital quality predictive of pandemic deaths? evidence from us counties. Journal of Urban Economics 133, 103472.

Lee, D.S., 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. The Review of Economic Studies 76, 1071–1102.

Machado, M.P., Mora, R., Romero-Medina, A., 2012. Can we infer hospital quality from medical graduates' residency choices? Journal of the European Economic Association 10, 1400–1424.

Martinez-Miera, D., Sunyer, C., 2025. Political Polarization and Human Capital Allocation. CEPR Discussion Paper 20682. Centre for Economic Policy Research.

McNamara, C., Pineda-Torres, M., 2025. Medical residency subsidies and physician shortages. Journal of Public Economics 251, 105494. doi:10.1016/j.jpubeco.2025.105494.

Mora-García, C.A., Pesec, M., Prado, A.M., 2024. The effect of primary healthcare on mortality: Evidence from costa rica. Journal of Health Economics 93, 102833.

Nicholson, S., 2002. Physician specialty choice under uncertainty. Journal of labor Economics 20, 816–847.

OECD, 2024. Health at a Glance: Europe 2024: State of Health in the EU Cycle. OECD Publishing, Paris. URL: https://doi.org/10.1787/b3704e14-en, doi:10.1787/b3704e14-en.

Posso, C., Tamayo, J., Guarin, A., Saravia, E., 2025. Luck of the draw: The causal effect of physicians on birth outcomes. Harvard Business School Working Paper 22-015, accepted at *American Economic Journal: Applied Economics*.

Rubin, D.B., 1980. Randomization analysis of experimental data: The fisher randomization test comment. Journal of the American Statistical Association 75, 591–593.

Russo, G., Perelman, J., Zapata, T., Šantrić-Milićević, M., 2023. The layered crisis of the primary care medical workforce in the european region: what evidence do we need to identify causes and solutions? Human Resources for Health 21, 55.

Schündeln, M., 2014. Are immigrants more mobile than natives? evidence from germany. Journal of Regional Science 54, 70–95.

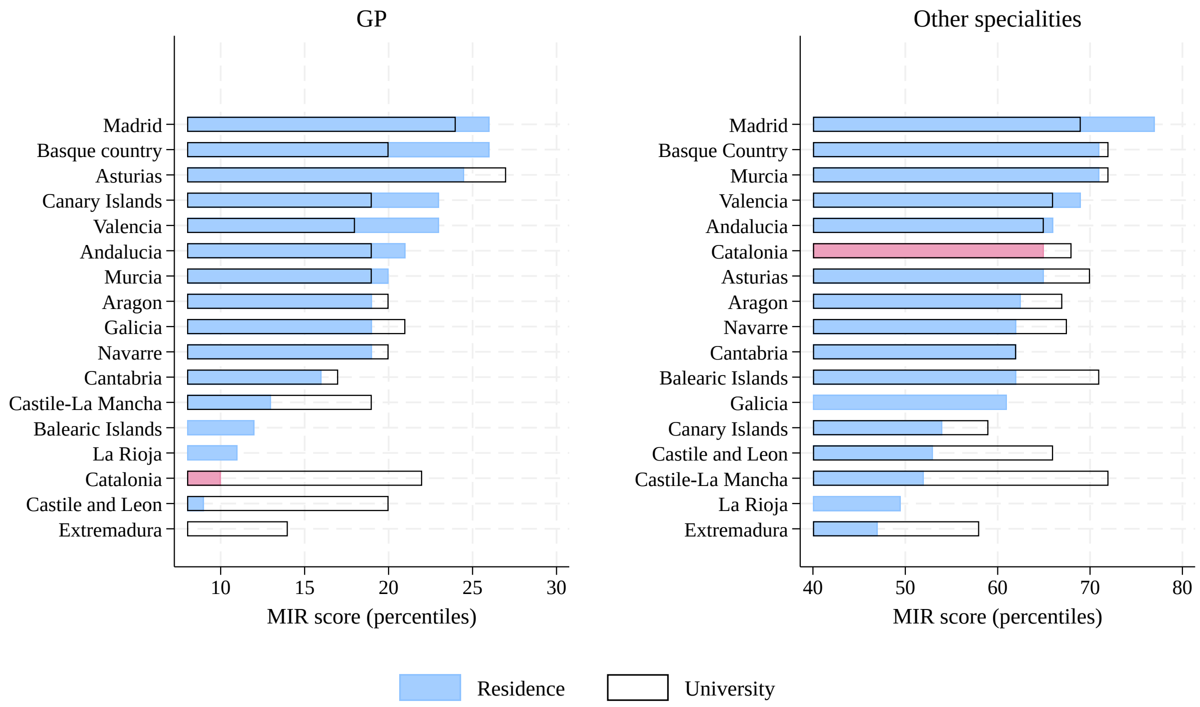
Sivey, P., Scott, A., Witt, J., Joyce, C., Humphreys, J., 2012. Junior doctors' preferences for specialty choice. Journal of Health Economics 31, 813–823.

Sobel, M.E., 2006. What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. Journal of the American Statistical Association 101, 1398–1407.

Starfield, B., Shi, L., Macinko, J., 2005. Contribution of primary care to health systems and health. The Milbank Quarterly 83, 457–502.

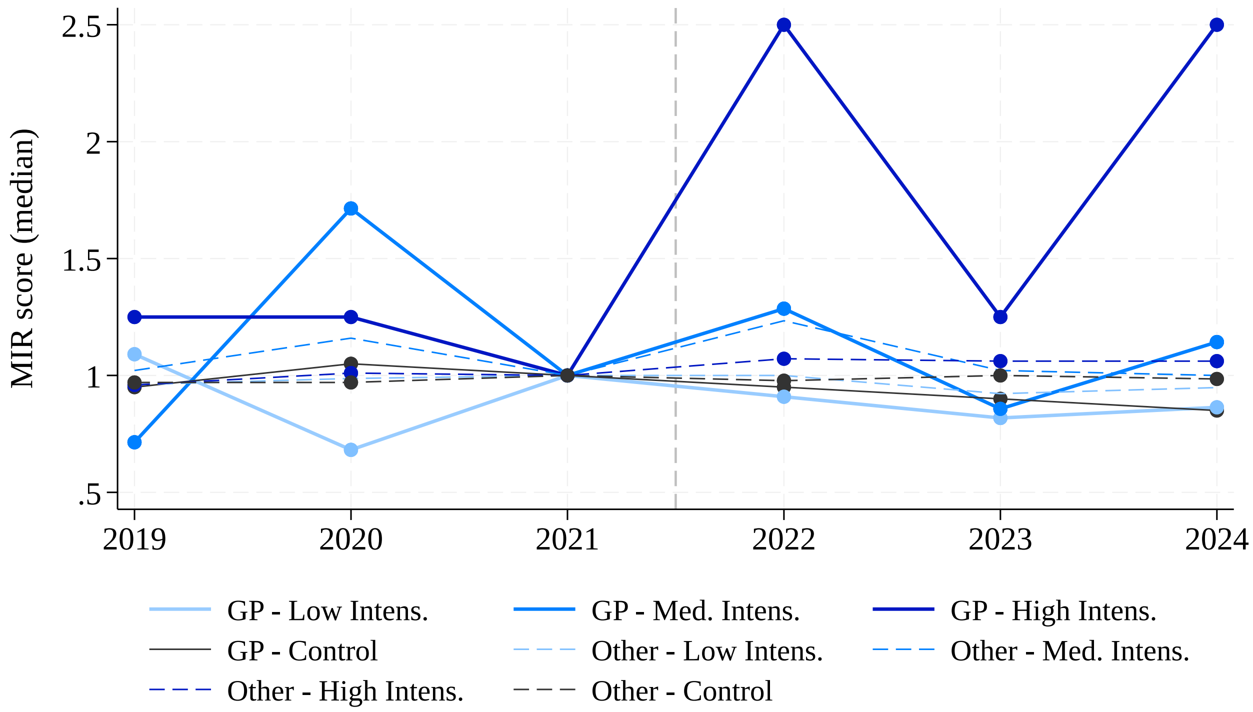
Figures and Tables

Figure 1. MIR percentile of medical residents by region of residency and by region of training, pre-policy



Notes: The figure plots the median percentile score in the MIR exam of medical residents, by region of residency (light blue bars) and by region of undergraduate medical training (outlined bars), pooled over the pre-policy period (2019–2021). The left panel reports the distribution for General Practice; the right panel reports the distribution for all other specialties. Bars shaded in red correspond to Catalonia. The difference between filled and outlined bars indicates the academic gap between incoming residents and locally trained graduates: a positive gap suggests a net inflow of higher-scoring candidates from other regions, whereas a negative gap suggests a relative outflow of academic talent.

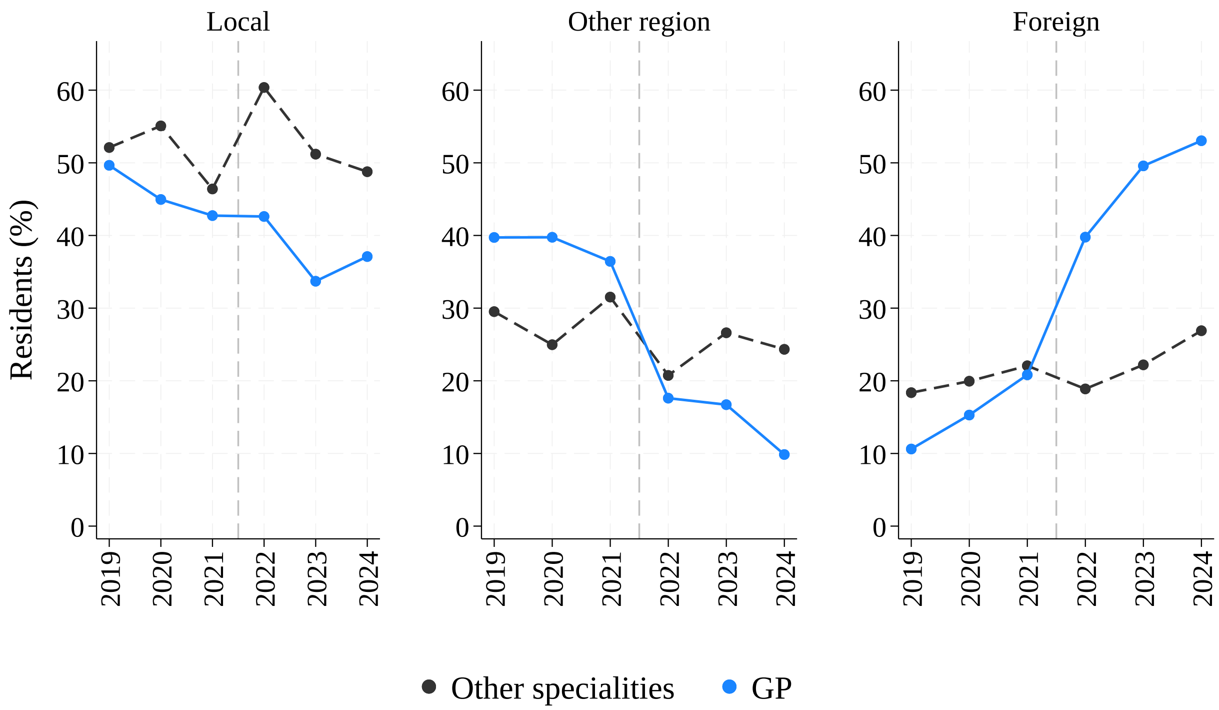
Figure 2. Median MIR score in General Practice and other specialties by treatment status



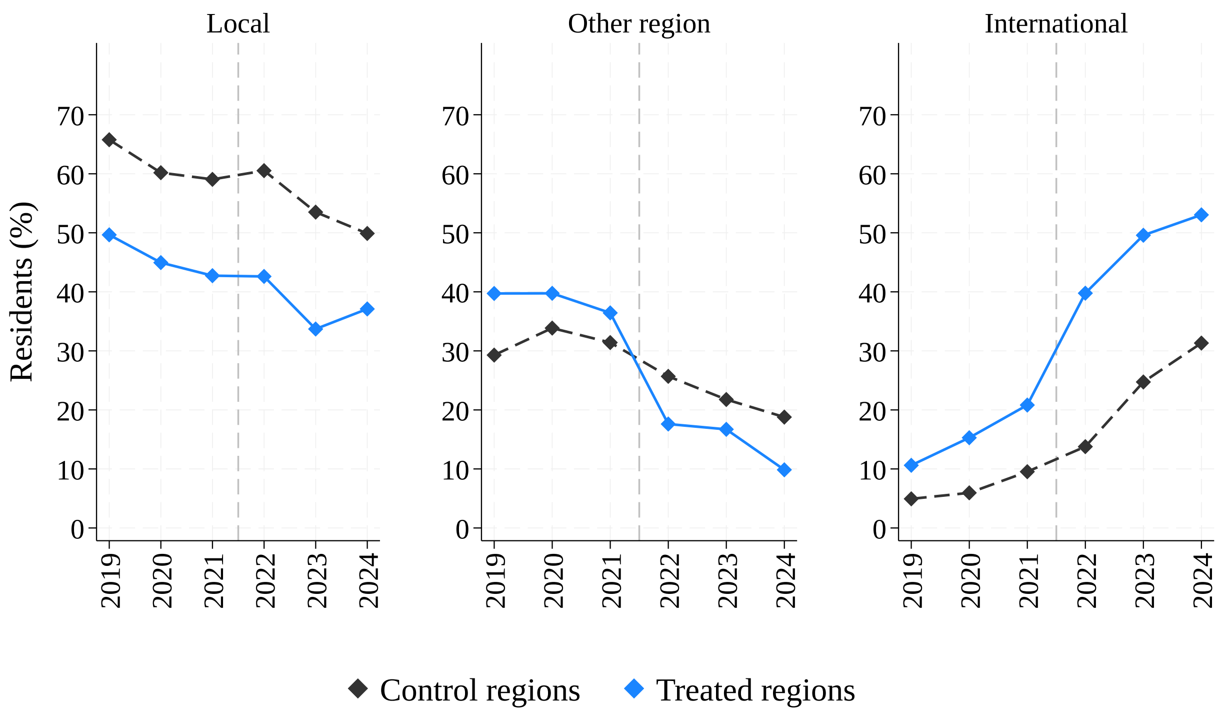
Notes: This graph represents the median MIR score in percentiles, standardized by year and indexed to 2021 (the year previous to the policy), by treatment status and specialty. Solid lines represent General Practice (GP), while dashed lines represent other specialties. Colors indicate treatment intensity groups in Catalonia (low, medium, and high), and black lines correspond to the control regions (rest of Spain).

Figure 3. Composition of GP entrants by origin

(a) In Catalonia, by specialty



(b) In General Practice, by treatment status



Notes: The figure plots the share of MIR entrants by origin—trained locally (left), in another Spanish region (center), or abroad (right). Panel (a) compares General Practice with the rest of medical specialties *within* the treated region (Catalonia). Panel (b) compares the treated region with the rest of Spain *within* General Practice. The vertical dashed line marks the introduction of the financial incentive policy in 2022.

Table 1. Impact of the financial incentives policy on score

	(1) All	(2) Low	(3) Medium	(4) High
Treat × Post	0.067*	-0.005	-0.008	0.174***
	(0.036)	(0.075)	(0.059)	(0.051)
Observations	16218	16146	16140	16164

Notes: This table presents the impact of the financial incentives policy on MIR standardized scores (a proxy of human capital attraction) by treatment intensity. Estimates are obtained using the synthetic difference-in-differences (Arkhangelsky et al. 2021). The empirical analysis is based on annual panel data at the hospital-specialty level. The dependent variable in all columns is *Score*. Column 1 presents the estimated effect for the entire treated region, while columns 2 through 4 report the effects for low-, medium-, and high-intensity areas, respectively. Standard errors are displayed in parentheses and obtained using bootstrapping procedures.

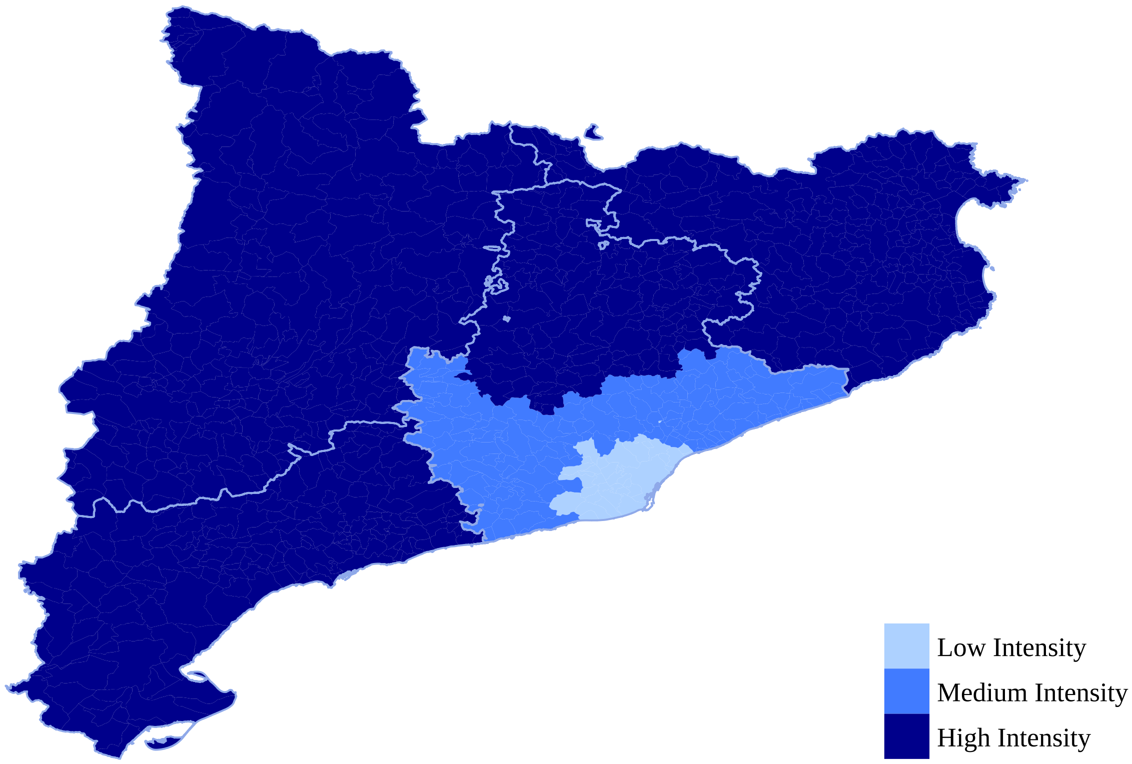
Table 2. Compositional effects of the financial incentives policy

	(1) All	(2) Low	(3) Medium	(4) High
International	0.297***	0.131*	0.308***	0.397***
	(0.035)	(0.068)	(0.025)	(0.056)
Local	-0.030	0.049	-0.156**	-0.016
	(0.038)	(0.090)	(0.080)	(0.040)
Other region	-0.248***	-0.171***	-0.137*	-0.351***
	(0.037)	(0.051)	(0.081)	(0.031)
Observations	16218	16146	16140	16164

Notes: This table presents the impact of the financial incentives policy on individuals characteristics by treatment intensity. Estimates are obtained using the synthetic difference-in-differences (Arkhangelsky et al. 2021). The empirical analysis is based on annual panel data at the hospital-specialty level. Each coefficient corresponds to a separate regression, with outcome variables listed in the rows. Column 1 presents the estimated effect for the entire treated region, while columns 2 through 4 report the effects for low-, medium-, and high-intensity areas, respectively. Standard errors are displayed in parentheses and obtained using bootstrapping procedures.

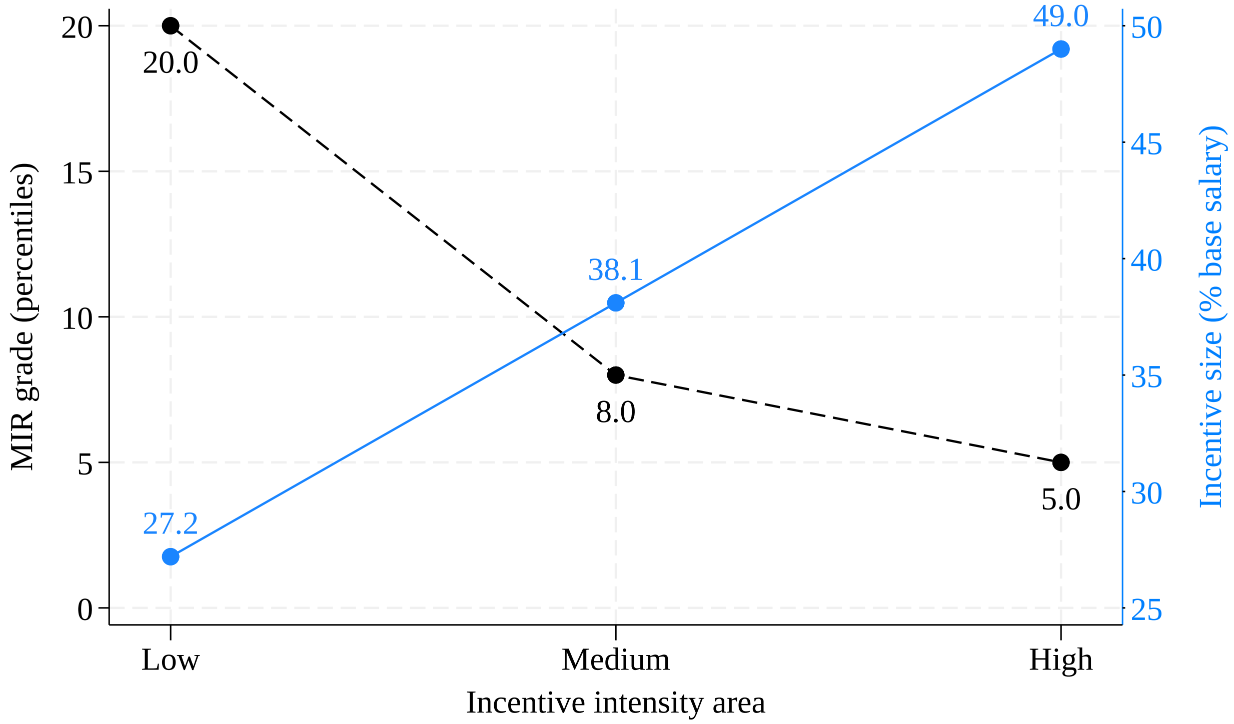
A Online Appendix

Figure A1. Spatial variation in treatment intensity



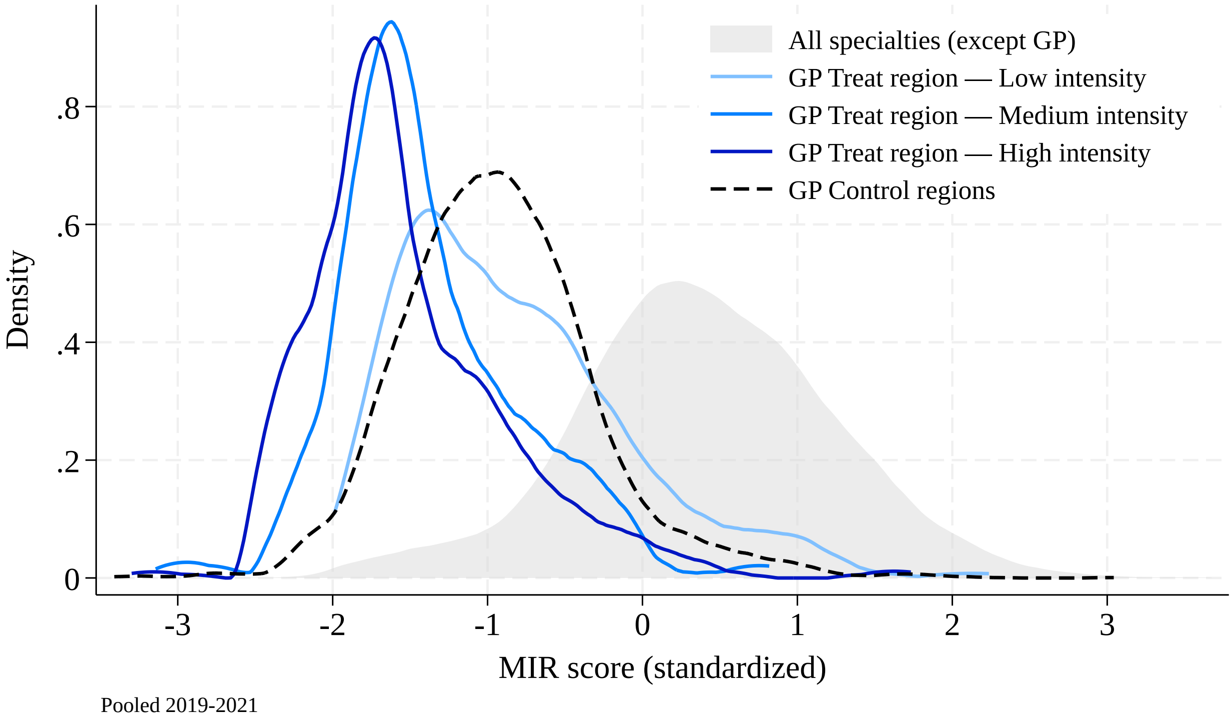
Notes: Map of municipalities in Catalonia by treatment intensity. Municipalities shaded in light blue represent the Barcelona metropolitan area (low-intensity treatment). Those in blue correspond to the rest of the province of Barcelona (medium-intensity treatment), excluding the Catalunya Central healthcare region. Municipalities shaded in dark blue represent the provinces of Girona, Lleida, and Tarragona, together with the Catalunya Central healthcare region, which form the high-intensity tier.

Figure A2. Treatment intensity, salary supplement, and pre-policy MIR scores



Notes: The figure plots, for each treatment intensity tier, the annual salary supplement (right axis, blue) as a percentage of the pre-policy gross base salary of a first-year MIR resident, and the median MIR score percentile of GP entrants placed in that tier during the pre-policy period 2019–2021 (left axis, black). The supplement amounts (€5,125, €7,175, and €9,225 per year, in 2023) are expressed as a fraction of the 2022 base salary (approximately €19,500).

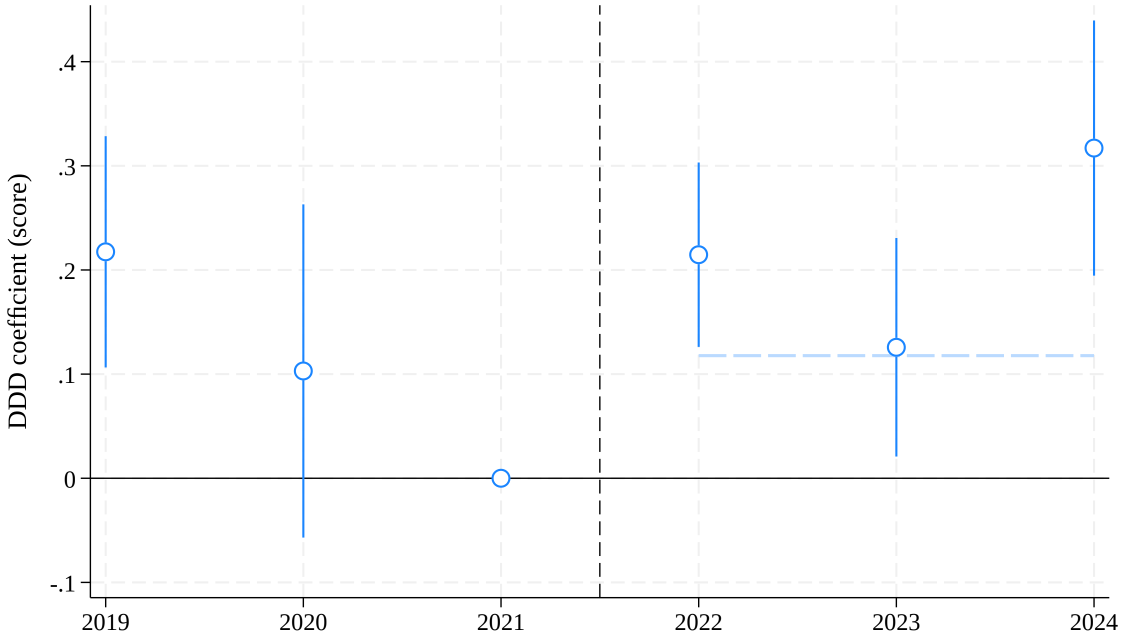
Figure A3. Pre-policy distribution of standardized MIR scores by group



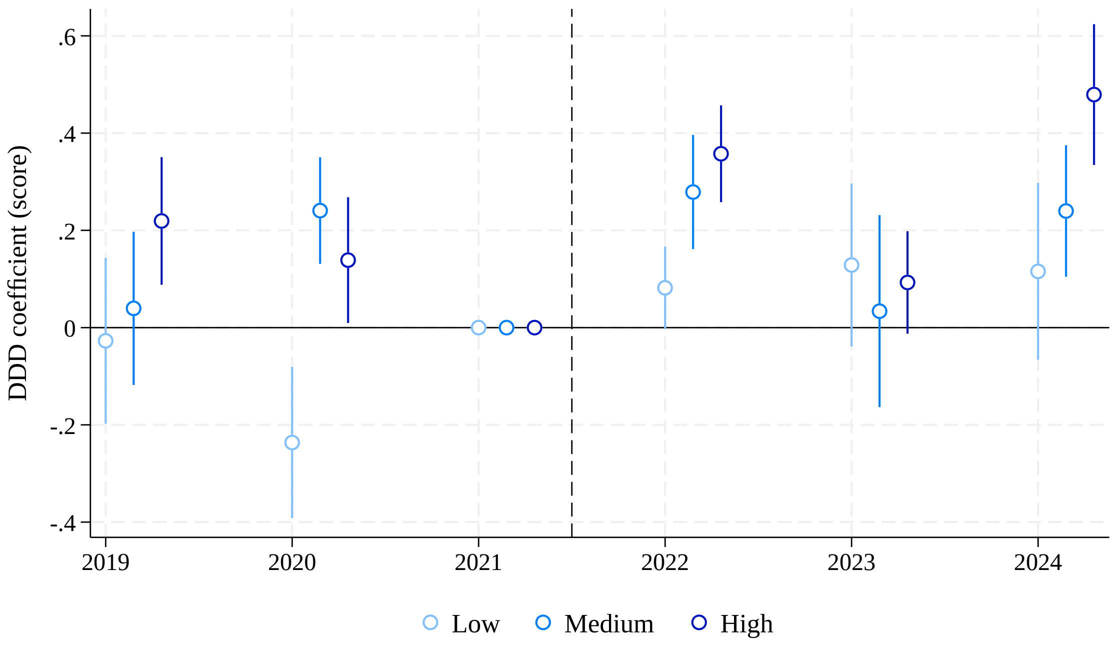
Notes: The figure plots the kernel density distribution of standardized MIR scores by group, pooled over the pre-policy period (2019–2021). The grey shaded distribution corresponds to entrants in all specialties other than General Practice. The dashed black line corresponds to GP entrants outside Catalonia. The three blue lines correspond to GP entrants in Catalonia, disaggregated by intensity tier (low, medium, and high).

Figure A4. Triple-difference event study: pre-trend tests

(a) Pooled

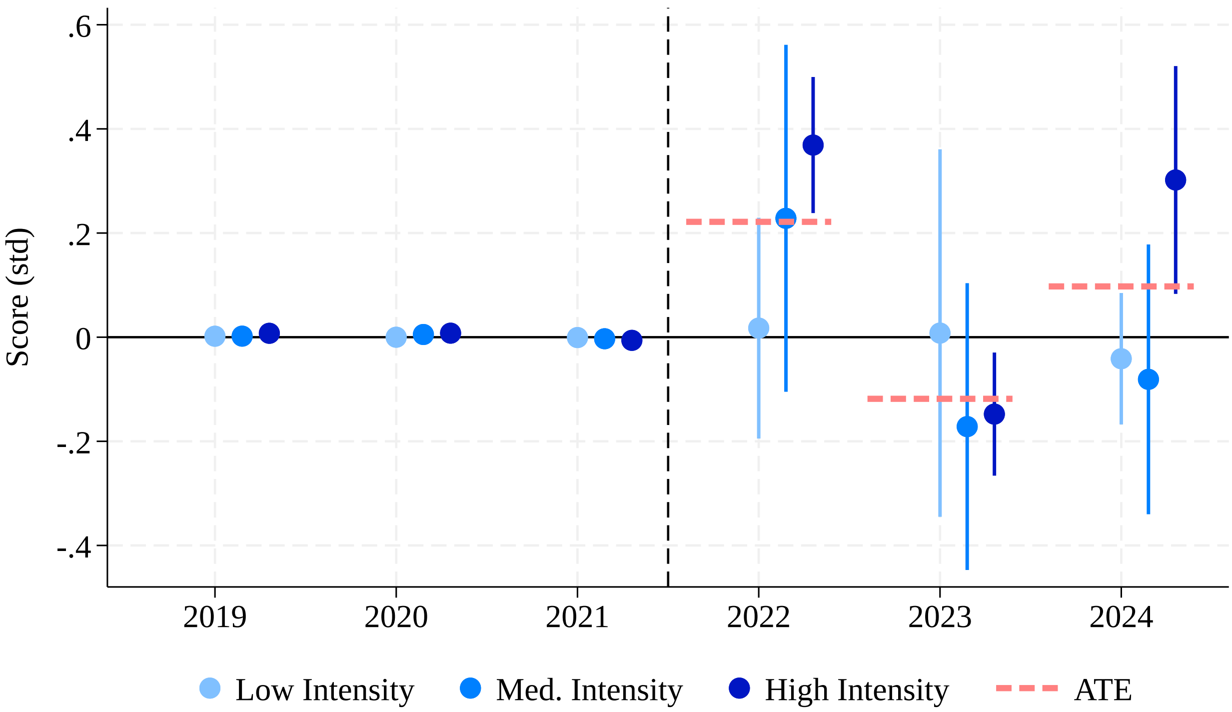


(b) By intensity tier



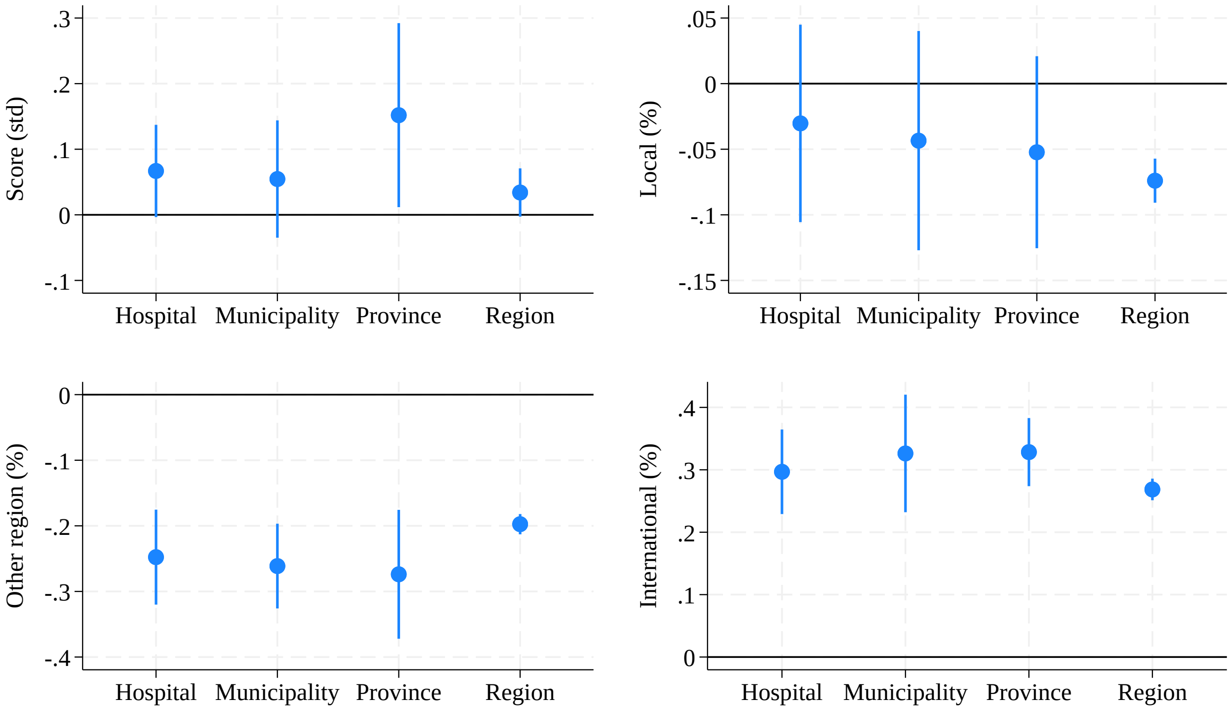
Notes: The figure plots event-study coefficients from the triple-difference (DDD) specification described in Section 4 on standardized MIR scores. Bars denote 95% confidence intervals; standard errors are clustered at the hospital–specialty level. The horizontal dashed line in panel (a) marks the ATE estimated. F-tests of joint significance of the pre-period coefficients (2019, 2020): Pooled F = 7.38 ($p = 0.001$); Low F = 4.54 ($p = 0.011$); Medium F = 9.31 ($p < 0.001$); High F = 5.38 ($p = 0.005$).

Figure A6. Dynamic effects of financial incentives on score by treatment intensity



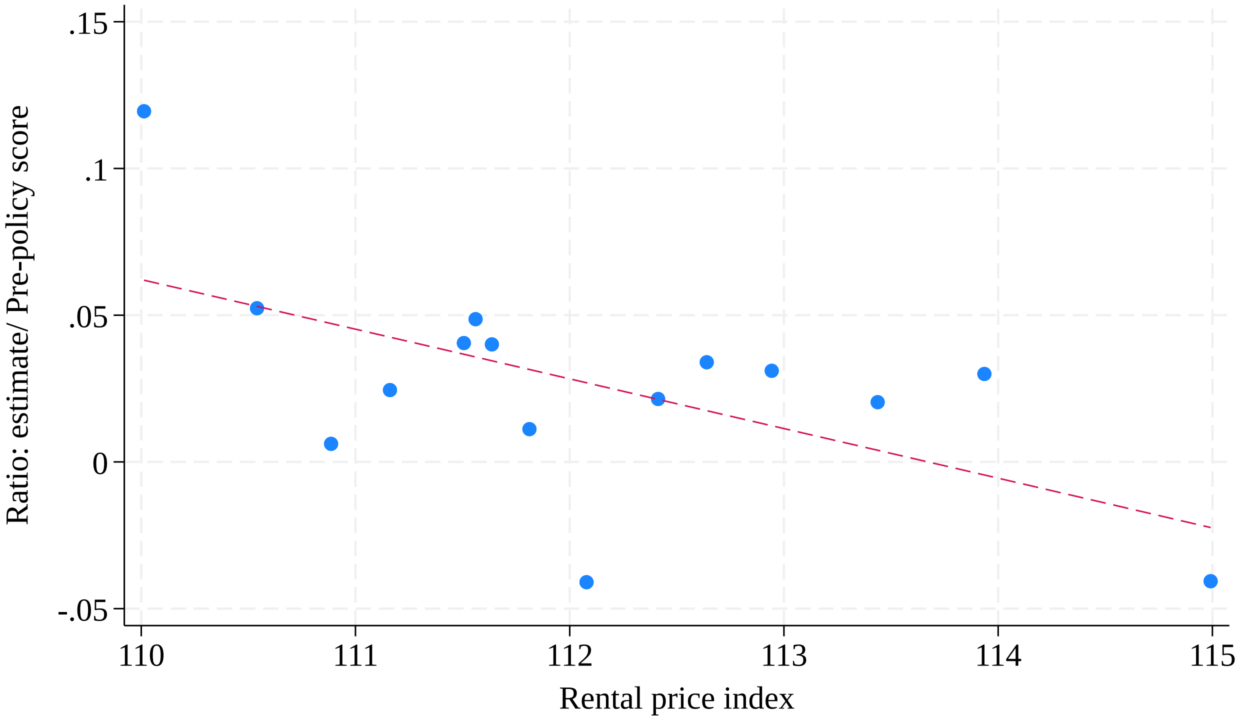
Notes: The figure reports event-study estimates of the effect of the financial incentive policy on the standardized MIR score (a proxy for *human capital*), separately by treatment intensity. Light, medium, and dark blue markers correspond to low-, medium-, and high-intensity units, respectively. Dashed red lines indicate the pooled average treatment effect (ATE) for each intensity tier. Estimates are obtained from the synthetic difference-in-differences procedure of Arkhangelsky et al. (2021) applied to the balanced panel of hospital–specialty units. Vertical bars denote 95% confidence intervals based on bootstrap standard errors.

Figure A7. Robustness of treatment effects to the level of geographic aggregation



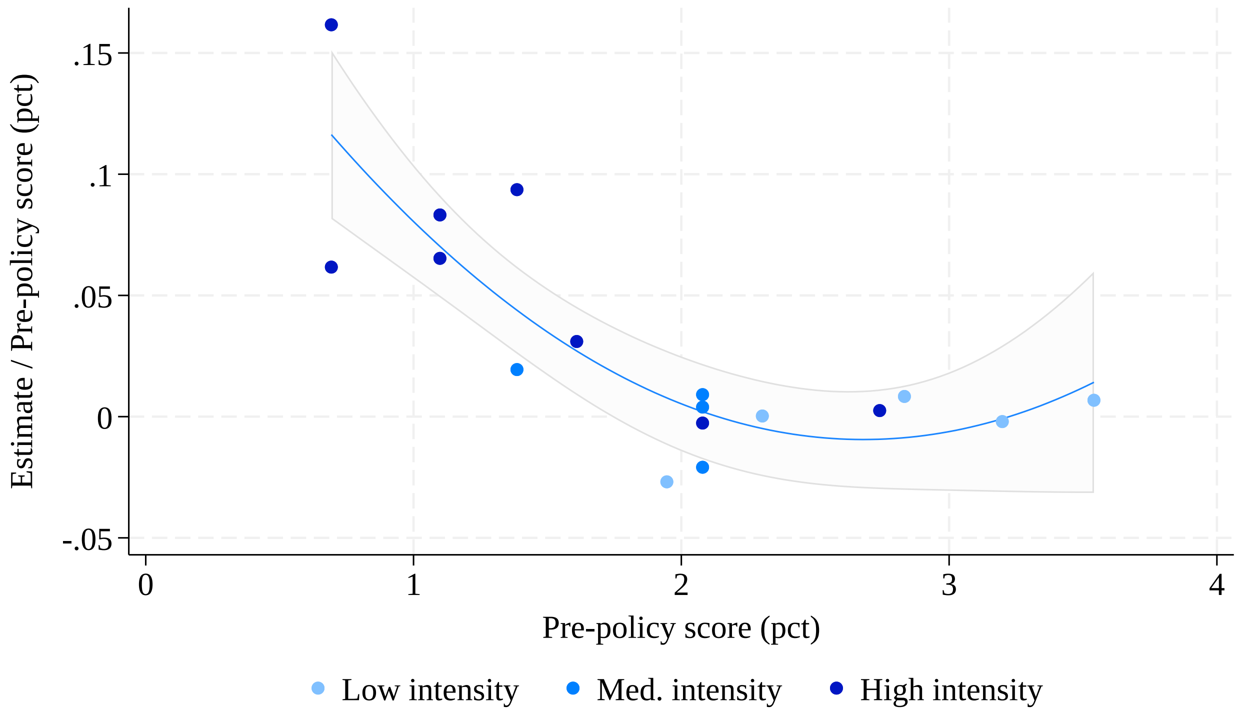
Notes: This figure assesses the robustness of the estimated treatment effects to the level of geographic aggregation at which the synthetic control is constructed. Each panel corresponds to a different outcome: standardized scores, the share of local entrants, the share of entrants from other regions, and the share of international entrants. Within each panel, every point reports the average treatment effect when units are aggregated at the hospital, municipality, province, and region level, respectively. Vertical bars denote 95% confidence intervals. The horizontal line at zero marks the absence of an effect.

Figure A8. Rental prices and financial incentives effects on score



Notes: This figure shows the relationship between the relative effect of the financial incentives policy on *Score* and the rental price index, after controlling for intensity-area fixed effects, an index of working conditions, and a proxy for workload. Each dot represents a hospital. The red line depicts the linear fit between the two variables.

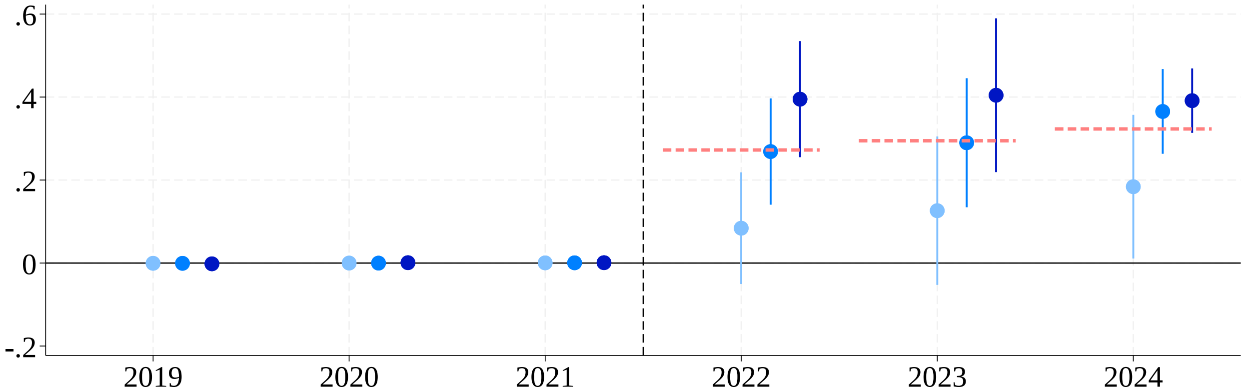
Figure A9. Convergence



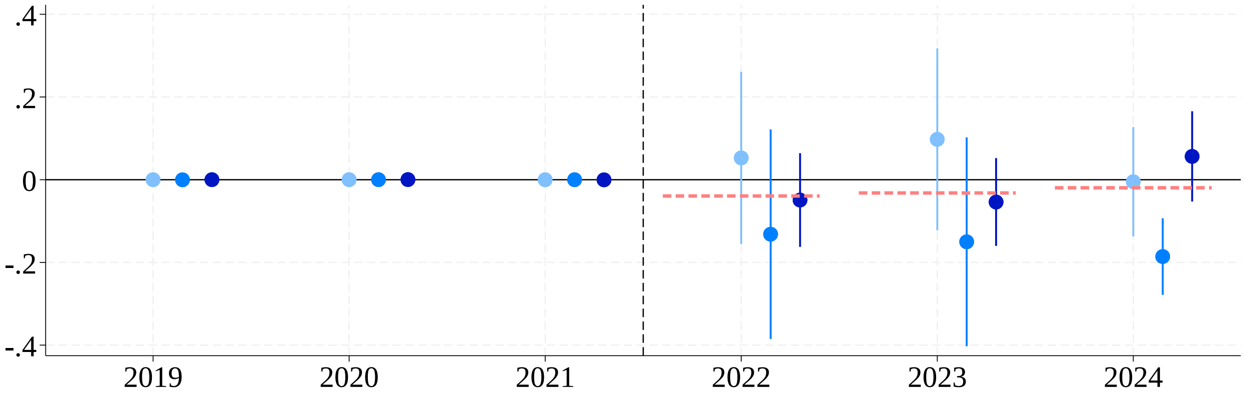
Notes: This figure depicts the relationship between the percentage change in standardized scores by hospital and the pre-policy median score (calculated prior to 2022). Each dot represents a treated hospital. Dots in light blue correspond to hospitals in low-intensity treatment areas, medium blue to medium-intensity areas, and dark blue to high-intensity areas. The red line represents the fitted quadratic relationship, with the shaded area indicating the 95% confidence interval.

Figure A10. Event-study estimates of compositional effects by treatment intensity.

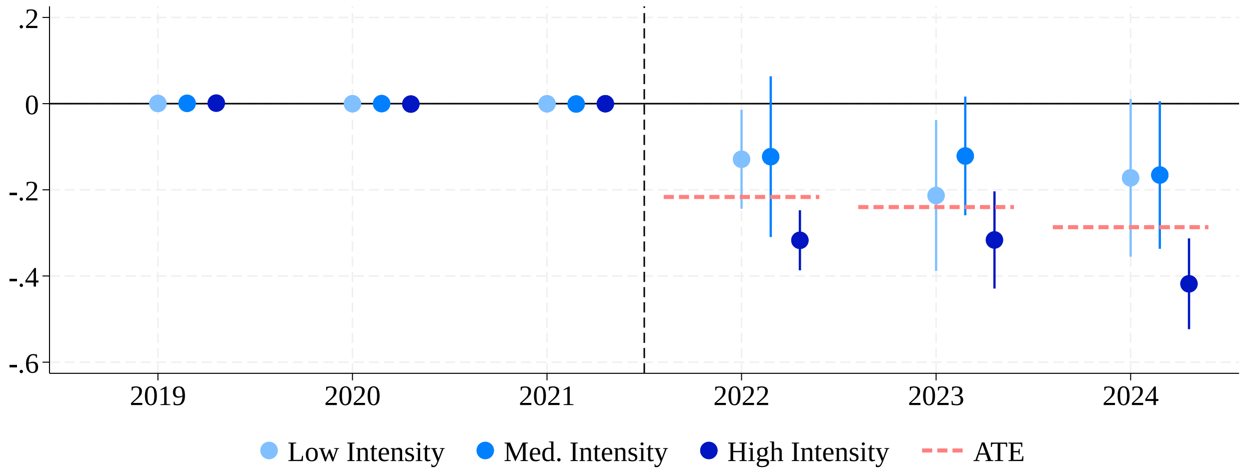
(a) International (%)



(b) Local (%)

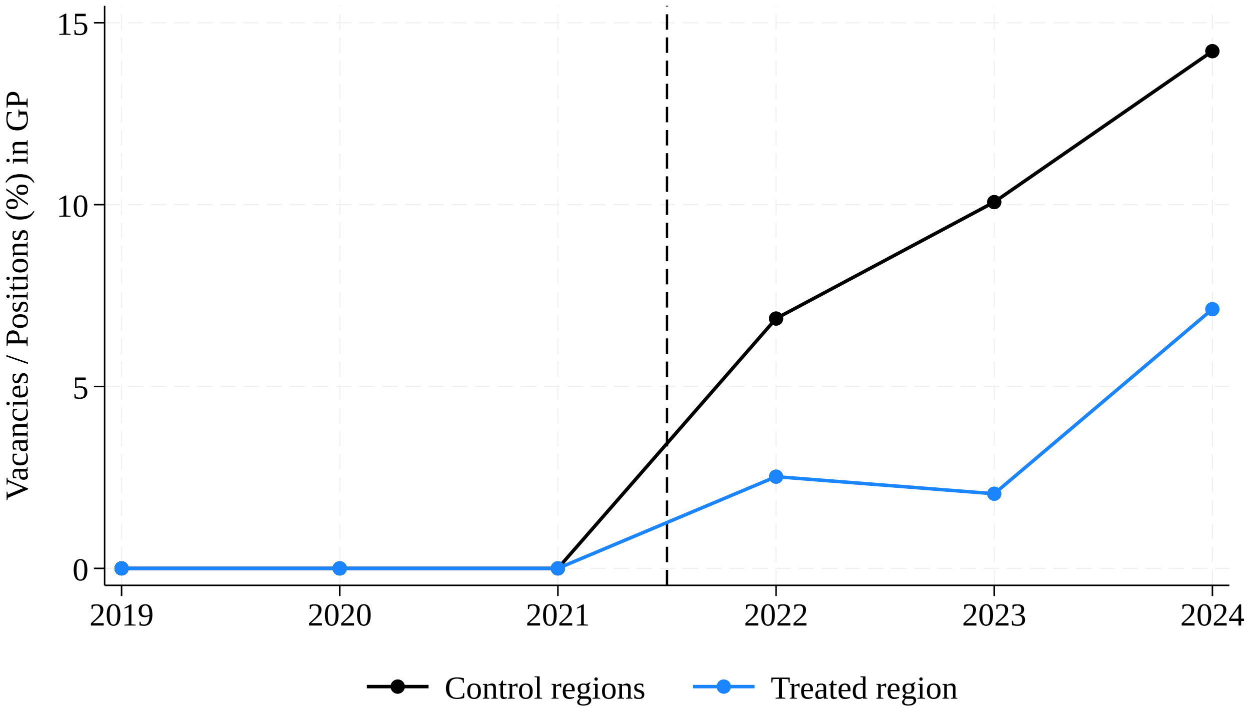


(c) Other region (%)



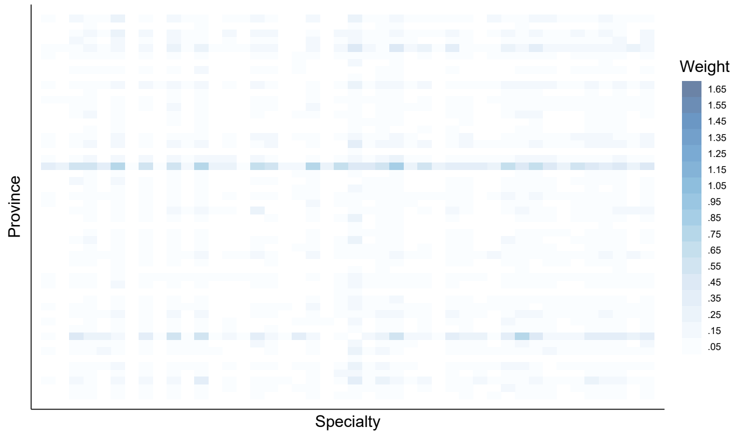
Notes: The figure reports event-study estimates of the effect of the financial incentive policy on the composition of GP entrants by origin, separately by treatment intensity. Panel (a) reports the share of internationally trained residents, panel (b) the share of locally trained residents, and panel (c) the share of residents trained in other Spanish regions. Light, medium, and dark blue markers correspond to low-, medium-, and high-intensity units, respectively. Dashed red lines indicate the pooled ATE for each intensity tier. Estimates are obtained from the SDID procedure of Arkhangelsky et al. (2021) applied to the balanced panel of hospital–specialty units. Vertical bars denote 95% confidence intervals based on bootstrap standard errors.

Figure A11. Share of vacancies in General Practice

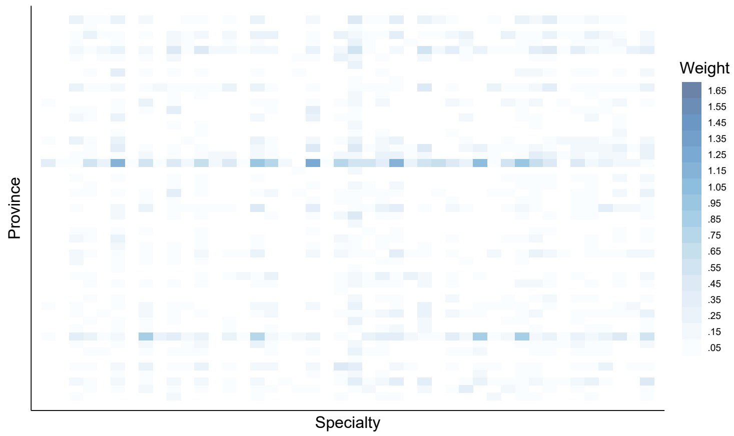


Notes: The figure plots the share of unfilled GP positions over total GP positions offered each year, separately for the treated region (Catalonia, blue) and the rest of Spain (black). Both series are essentially zero between 2019 and 2021, reflecting that all GP positions were filled nationally before 2022. The vertical dashed line marks the introduction of the financial incentive policy in 2022.

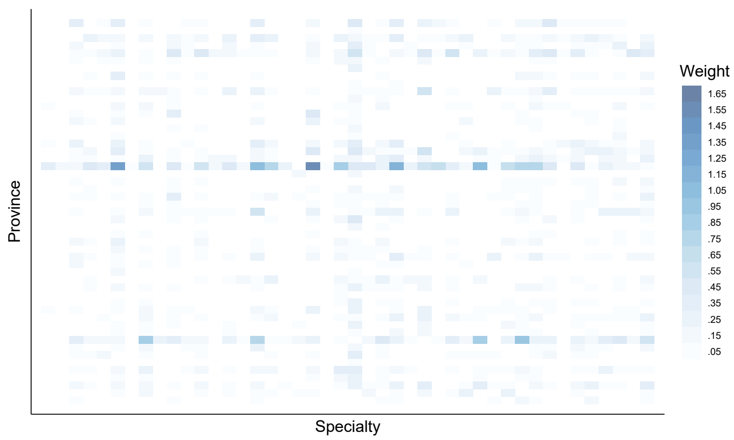
Figure A12. Unit weights by intensity



(a) Low intensity area



(b) Medium intensity area



(c) High intensity area

Notes: This figure plots the unit weights from the synthetic difference-in-differences estimator, using *Score* as the outcome variable, by treatment intensity.

Table A1. Impact of the financial incentives policy on score (percentiles)

	(1) All	(2) Low	(3) Medium	(4) High
Treat × Post	0.819	-1.470***	-0.454	3.146***
	(0.536)	(0.566)	(0.969)	(0.490)
Baseline median (pre-policy)	15.25	17.00	8.00	3.50
Observations	16218	16146	16140	16164

Notes: This table presents the impact of the financial incentives policy on human capital attraction by treatment intensity. Estimates are obtained using the synthetic difference-in-differences (Arkhangelsky et al. 2021). The empirical analysis is based on annual panel data at the hospital-specialty level. The dependent variable in all columns is the *percentile rank* of the MIR score. Column 1 presents the estimated effect for the entire treated region, while columns 2 through 4 report the effects for low-, medium-, and high-intensity areas, respectively. Standard errors are displayed in parentheses and obtained using bootstrapping procedures.

Table A2. Summary statistics of the MIR exam by treatment status before and after 2022

	Pre-policy				Post-policy			
	Low	**Medium**	**High**	**Control**	**Low**	**Medium**	**High**	**Control**
Treat: General Practitioner								
Score								
Score (std)	-1.1 (0.4)	-1.5 (0.1)	-1.7 (0.2)	-1.1 (0.4)	-1.3 (0.6)	-1.7 (0.1)	-1.8 (0.3)	-1.3 (0.6)
Score percentiles	17.0 (11.2)	8.0 (2.0)	3.5 (4.6)	17.0 (11.1)	13.0 (10.5)	8.2 (0.9)	7.5 (3.7)	13.5 (9.7)
Individual characteristics								
International (%)	22.4	22.7	15.6	7.7	38.7	55.6	56.1	27.8
Interregional (%)	37.5	24.7	53.4	30.7	17.6	8.2	15.7	21.1
Local (%)	40.1	52.7	31.0	61.6	43.8	36.2	28.2	51.1
# MIR residents	417	203	364	5317	439	222	395	5524
Control: Rest specialities								
Score								
Score (std)	0.6 (0.8)	0.1 (0.6)	-0.0 (0.6)	0.3 (0.7)	0.6 (0.6)	0.2 (0.5)	0.2 (0.6)	0.4 (0.7)
Score percentiles	72.8 (22.3)	52.0 (18.9)	49.0 (18.2)	59.0 (22.9)	71.0 (22.3)	52.5 (18.2)	52.0 (18.7)	60.0 (22.7)
Individual characteristics								
International (%)	22.7	21.9	19.1	12.0	27.2	26.7	19.1	14.4
Interregional (%)	24.0	23.8	40.6	37.1	19.9	14.9	40.4	35.6
Local (%)	53.4	54.3	40.4	50.9	53.0	58.4	40.4	49.9
# MIR residents	1526	343	436	12401	1632	364	506	13443

Notes: This table provides summary statistics by treatment status (specialty and geographical intensity) before and after the 2022 policy. The three intensities of the monetary policy are represented by the columns labelled *Low*, *Medium* and *High*, while *Control* represents the regions where the policy did not take place. On the other hand, the two horizontal panels represent the treatment status by specialty. The *Score (standardised)* is the total score obtained in the MIR exam, standardised by year. *Score percentiles* denotes the percentile rank of the previous variable. In both cases, the median is displayed to avoid the effect of extreme values, with the standard deviation shown in brackets. For individual characteristics, the mean value is presented. The *International*, *Interregional* and *Local* variables represent the percentage of students who studied abroad, at a Spanish university but in a different region to their placement hospital, and at a spanish university located in the same region as their placement hospital, respectively. The variable *MIR residents* represents the total number of residents in each group. Note that the descriptives are calculated at hospital-specialty level using the balanced panel, i.e. considering only hospitals offering specialties every year from 2019 to 2022.

Table A3. Effects of the financial incentives policy: ring approach

	(1) All	(2) Low	(3) Medium	(4) High
Panel A: Excluding high-mobility regions				
Treat × Post	0.042	-0.030	-0.035	0.151**
	(0.045)	(0.082)	(0.059)	(0.064)
Observations	8,400	8,328	8,322	8,346
Panel B: Excluding specialties closest to GP in pre-policy scores				
Treat × Post	0.069*	-0.005	-0.009	0.180***
	(0.037)	(0.065)	(0.060)	(0.045)
Observations	13,404	13,332	13,326	13,350
Panel C: Excluding both regions and specialties				
Treat × Post	0.046	-0.026	-0.031	0.160***
	(0.041)	(0.072)	(0.059)	(0.061)
Observations	7,080	7,008	7,002	7,026

Notes: The table reports the effect of the financial incentives policy on the standardized MIR score under three alternative ring-based exclusion strategies. Estimates are obtained using the synthetic difference-in-differences (Arkhangelsky et al. 2021), applied to the balanced panel of hospital–specialty units. Panel A excludes the four Spanish regions with the highest pre-policy inflow of medical students to Catalonia (Valencia, Madrid, Andalusia, and the Canary Islands) from the donor pool. Panel B excludes the specialties whose pre-policy entry scores were closest to those of GP (Preventive Medicine and Public Health, Clinical Analysis, Allergology, Occupational Medicine, Microbiology and Parasitology, Pathological Anatomy, Clinical Neurophysiology, Geriatrics, Clinical Pharmacology, and Nuclear Medicine). Panel C combines both exclusions simultaneously. Column 1 reports the estimated effect for the entire treated region; Columns 2–4 report the effects separately for low-, medium-, and high-intensity units. Standard errors are displayed in parentheses and obtained using bootstrapping procedures.

Table A4. Effects of the financial incentives policy: including covariates

	(1) All	(2) Low	(3) Medium	(4) High
Treat × Post	0.067	-0.005	-0.008	0.175***
	(0.045)	(0.070)	(0.062)	(0.043)
Observations	16218	16146	16140	16164

Notes: This table presents the impact of the financial incentives policy on human capital attraction by treatment intensity. Estimates are obtained using the synthetic difference-in-differences (Arkhangelsky et al. 2021). The empirical analysis is based on annual panel data at the hospital-specialty level. The dependent variable in all columns is *Score*. Covariates include the annual share of posts offered by specialty at the hospital and province levels. Column 1 presents the estimated effect for the entire treated region, while columns 2 through 4 report the effects for low-, medium-, and high-intensity areas, respectively. Standard errors are displayed in parentheses and obtained using bootstrapping procedures.

Table A5. The effects of the financial incentives policy on territorial equity

	(1)	(2)	(3)
Score	-0.041***	-0.171**	-0.150**
	(0.014)	(0.059)	(0.063)
$Score^2$		0.032**	0.029*
		(0.012)	(0.013)
Adjusted R^2	0.459	0.682	0.677
Observations	17	17	17
Akaike IC	-62.947	-71.168	-69.505

Notes: This table presents the impact of the financial incentives policy on territorial equity. Estimates are obtained using OLS regressions, where each observation corresponds to a treated hospital. The dependent variable in all columns is the percentage change in *Score*. For this specification, *Score* corresponds to the median pre-policy *Score*, while *Score2* denotes its squared term. Robust standard errors (HC2) in parentheses.

Table A6. Effects of the financial incentives policy by individuals characteristics

	(1)	(2)	(3)	(4)
Treat × Post × International	-0.112			
	(0.072)			
Treat × Post × Local		0.170*		
		(0.097)		
Treat × Post × Other region			0.287*	
			(0.165)	
Observations	11754	8472	8862	

Notes: This table presents the heterogeneous impact of the financial incentives policy on score by individuals characteristics for the entire treated region. Estimates are obtained using the synthetic difference-in-differences (Arkhangelsky et al. 2021). The dependent variable in all columns is the standardized MIR *Score*. Columns 1 and 3 report the estimated effect of the triple interaction using the full sample, while Column 2 restricts the analysis to nationally trained medical students. Standard errors are displayed in parentheses and obtained using bootstrapping procedures.

Table A7. Effect of the financial incentives policy on GP vacancy rates

	(1) Vacancy rate (%)
Treat $\times$ Post	-4.185*
	(2.218)
Hospital fixed effects	Yes
Year fixed effects	Yes
Observations	1,078

Notes: The table reports the effect of the policy on the share of unfilled GP positions, estimated via a difference-in-differences specification restricted to General Practice. The dependent variable is the vacancy rate at the hospital-year level, expressed in percentage points. The specification includes hospital and year fixed effects, with standard errors clustered at the province level. Because vacancy rates were zero across all regions before 2022, the specification reduces effectively to a post-policy comparison between treated and control regions with hospital fixed effects.

Table A8. Lee bounds on the effect of the financial incentives policy on score

	(1) All	(2) Low	(3) Medium	(4) High
Treat × Post	0.088**	0.051	-0.012	0.191***
	(0.036)	(0.068)	(0.058)	(0.047)
Observations	16218	15168	14388	14610

Notes: The table reports upper Lee (2009) bounds on the average treatment effect of the financial incentives policy on the standardized MIR score, both for the pooled treated region (Column 1) and separately by intensity tier (Columns 2–4). The bounds are constructed by trimming the treated sample to equalize the share of filled positions across treated and control groups: the post-policy fill rate in Catalonia (96.0%) exceeds the rate in the rest of Spain (93.1%), so the bottom of the post-policy score distribution in treated units is trimmed. This corresponds to the most pessimistic assumption about the marginal entrants attracted by the policy, namely that they are of below-average academic profile; the resulting estimate is therefore an upper bound on the score effect. The trimming proportion is 3.1% in the pooled specification, and 6.9%, 4.1%, and 4.5% in the low-, medium-, and high-intensity tiers, respectively. In medium-intensity units, the post-policy fill rate in treated units is below that of control units, so the top of the control post-policy score distribution is trimmed instead. Standard errors are displayed in parentheses and obtained using bootstrapping procedures.

Table A9. Impact of the financial incentives policy on score, without 2023

	(1) All	(2) Low	(3) Medium	(4) High
Treat × Post	0.353***	0.091	0.203	0.615***
	(0.074)	(0.094)	(0.125)	(0.108)
Observations	9575	9515	9510	9530

Notes: This table presents the impact of the financial incentives policy on human capital attraction by treatment intensity, excluding year 2023. Estimates are obtained using the synthetic difference-in-differences (Arkhangelsky et al. 2021). The empirical analysis is based on annual panel data (2022 and 2024) at the hospital-specialty level. The dependent variable in all columns is *Score*. Column 1 presents the estimated effect for the entire treated region, while columns 2 through 4 report the effects for low-, medium-, and high-intensity areas, respectively. Standard errors are displayed in parentheses and obtained using bootstrapping procedures.

Table A10. Effects of the financial incentives policy: effect on mean score

	(1) All	(2) Low	(3) Medium	(4) High
Treat × Post	-0.003	-0.106	-0.085*	0.128***
	(0.039)	(0.069)	(0.044)	(0.049)
Observations	16218	16146	16140	16164

Notes: This table presents the impact of the financial incentives policy on human capital attraction by treatment intensity. Estimates are obtained using the synthetic difference-in-differences (Arkhangelsky et al. 2021). The empirical analysis is based on annual panel data at the hospital-specialty level. The dependent variable in all columns is the *mean* standardized MIR score, in contrast to the median used in the baseline specification. Covariates include the annual share of posts offered by specialty at the hospital and province levels. Column 1 presents the estimated effect for the entire treated region, while columns 2 through 4 report the effects for low-, medium-, and high-intensity areas, respectively. Standard errors are displayed in parentheses and obtained using bootstrapping procedures.