The impact of Community Health Centers on inappropriate use of emergency services

Abstract

Community Health Centres (CHCs) offer coordinated and comprehensive responses to primary care needs. Our study aims at assessing whether the introduction of such organisational model improved health outcomes measured by inappropriate emergency visits among diabetics in the Emilia-Romagna region of Italy. Using difference-in-differences methods within a staggered treatment setting, we estimate the effect of CHC participation on inappropriate hospital emergency visits between year 2010 and year 2016. We distinguish between ED admissions for varying time spans, occurring at daytime during working days, at night-time, as well as during weekends. We show that, the causal effect of the adoption of the community care model leads to a reduction in the probability of inappropriate admissions by an amount ranging between 1.6-1.7 percentage points during working days at daytime, with large facilities responsible for most gains. Conversely, we detect no difference at night-time and weekends. Our results point out that the coordinated care model increases appropriateness among vulnerable patients, and that extending opening hours and the range of services can further enhance such benefits.

1. Introduction

The European health policy framework identifies community and primary care as strategic areas and underlines that strengthening territorial care through a coordinated approach can improve efficiency, effectiveness, and responsiveness of health care systems [WHO, 2013]. Although a common definition for integrated community care centres remains elusive - they are called Medical Homes, Patient-centered Medical Homes, Community Health Centres etc.-, they share distinctive features that include team-based care, integrated information technology systems, tools supporting clinical decisions such as population-based registries and elements of the Chronic Care Model [Klein et al. 2013]. Developed first in the United States as Medical Homes to manage complex patients [Bojadziewski and Gabbay, 2011], similar initiatives are now implemented worldwide. Examples can be found in Canada [The College of Family Physicians of Canada, 2019] and in several European countries, including England where the primary care home program introduced in 2015 counts more than 240 sites across the country serving 17% of the population [Kumpunen et al, 2017].

In the initial phase, the launch of the Community Health Centres (CHCs) initiative in North America was mainly motivated by growing concerns with the accelerating rate of healthcare spending. They were part of a strategy to curb down raising health care costs by shifting demand away from acute hospital care in favor of less complex settings. Later, the focus has moved from containment of resource use to quality improvement in territorial healthcare through integrated approaches to patient treatment.

In Italy, the National Healthcare System (NHS) has been traditionally based on a hospital-centric model, characterized by limited investments in prevention, heavy pressure on the EDs, limited development of primary care and weak integration between hospital and territorial care. Given this background, the underlying motivations of the initiative are mainly supply related. The Italian policy debate promoting the large-scale establishment of CHCs ("Case della Salute") starts in the years 2000s', and CHCs are seen as an investment to improve the organization and the appropriateness of care, the quality of the processes and, ultimately, the effectiveness of treatments, especially for patients suffering from chronic conditions. Increased supply and accessibility to community care is also meant to relieve congestion at the EDs.

The Ministry of Health identified the deployment of CHCs as a national priority. They are defined as "public physical places and – at the same time – active and dynamic centres for health and well-being for local community that collect citizens' demand for healthcare and supply it in the most

appropriate way in time and space" [Italian Ministry of Health, 2007]. In the light of that, Regional Governments have launched large-scale programs, but the dissemination of the model has been gradual, depending on local infrastructures and funding. A key step in such process is the involvement of General Practitioners (GPs), who lie at the forefront of chronic care management. Rather than referring patients systematically from generalist to specialist physicians, CHCs develop a team-based approach and coordinate care through partnerships encompassing primary, mental health, community, and social care as well as the voluntary sector. However, similarly to the traditional gatekeeping model, the GP remains in charge of the patient and is responsible for the coordination of the whole set of treatments.

Despite the prominence achieved in many institutional contexts, the evidence of the impact of such programs on patients' outcomes is surprisingly scant outside the US and, in most cases, it remains descriptive in nature. The aim of this paper is to contribute to fill this gap by assessing the causal impact of GPs' participation into CHCs in the Italian NHS: we test whether patients whose GP operates in a CHC display better outcomes compared to patients whose GP does not. We focus on Diabetes Mellitus type II and measure outcomes as inappropriate emergency admissions consisting of minor conditions that should be treated in primary care. As diabetes treatment requires multiple processes and resources, these patients are well suited for studying the consequences of the new coordinated care system because assistance relies strongly on GP's monitoring, regular screening, and routine controls. At the same time, CHCs multidisciplinary teams providing support, education and care are deemed to improve diabetes management by enhancing regular specialist advice.

We consider an unbalanced panel of individual-level data covering the diabetic population of the Italian region Emilia-Romagna over seven years (2010–16) and employ a series of difference-in-differences (DID) estimators. At each point in time, patients whose GP operates in a CHC are assigned to the treatment group, while the control group consists of the patients registered with physicians working in traditional practices. The identification of the effect of the CHC model is challenged by the fact that GPs' participation occurs on a voluntary basis, and unobserved factors may affect GP propensity to join a CHC and patients' outcomes at the same time. We employ multiple strategies that allow to tackle such potential endogeneity bias. As baseline estimation procedure, we use two-way fixed effects (TWFE) linear regression models, controlling for fixed differences between GPs that entered a CHC and those that did not via the GP fixed effects. We evaluate the robustness of our estimates by using alternative DID methods that allow to compare

outcomes in treatment and control groups so that the estimated effect can have a causal interpretation.

Very recently, a number of methodological contributions have highlighted potential pitfalls associated with the TWFE estimator within settings with multiple time periods and variation of treatment timing (e.g., Callaway and Sant'Anna, 2021, Goodman-Bacon, 2021, Sun and Abraham, 2021 and de Chaisemartin and D'Haultfœuille, 2020).[1] As these studies point out, in staggered treatment setup, TWFE estimators may lead to biased estimates in cases with heterogeneous treatment effects. In this paper, we test the robustness of our analysis to treatment effects heterogeneity by using the Callaway and Sant'Anna difference-in-differences (CS-DID) methodology. We exploit the flexibility of the CS-DID approach to examine treatment effect dynamics both with respect to length of exposure to the treatment and with respect to calendar year.

Our estimates of the effect of CHC participation on inappropriate hospital emergency visits are robust across different estimation methods. We find that, other things equal, being registered with GPs operating in a CHC reduces the probability of inappropriate admissions to emergency departments compared to traditional practices. The impact of the new organizational model ranges between 1.5-1.7 percentage points and is driven by the decrease in week-days admissions occurring at daytime, while we find no significant effects either at night-time or on weekends when CHCs are closed. We find no evidence of significant heterogeneity effects with respect to elapsed treatment time neither the average treatment effect significantly evolves over calendar years. Finally, we document that CHC size and the scope of services matter, with larger and better equipped centres bearing the merit of the reduction in inappropriate admissions.

2. Related literature

Our analysis relates in the first place to the studies on the diffusion of Medical Homes and Community Health Centres. Such literature is mainly US-based and covers areas such as patient experience, quality improvement, cost containment and utilization of hospital services. While there is some evidence of positive effects on clinical and economic outcomes [NCQA, 2019; Shippee, Finch and Wholey, 2018; David et al., 2015, 2018a; Hearld and Alexander, 2012; Weinick et al., 2010], other studies find limited or no effect on quality, utilization, or expenditure [Friedberg et al., 2014; Jackson et al., 2013; Peikes et al., 2012]. As for Europe, the literature on the impact of such a model

[1] Roth et al (2022) provide an excellent review of the studies conducted in this area.

is limited and largely based on surveys, qualitative interviews, practice observations and focus groups [Faber et al. 2013; Lionis and Petelos, 2015].

Whereas the first wave of works has mainly explored correlations between the new organizational arrangements and the outcomes of interest, a bunch of more recent studies assesses the causal impact of such initiatives. In particular, David et al (2018a) study the influence of Patient-Centered Medical Homes (PCMHs) in the US and emphasize the importance of accounting for the heterogeneity in the internal organization of these centres. When PCMHs are taken as a homogenous setting, the authors find little evidence of a link between PCMH status and patient outcomes, whereas this emerges when three distinct clusters, based on PCMH characteristics, are considered. Strumpf et al. (2017) study differences in healthcare utilization and costs between patients participating and non-participating in Family Medicine Groups in Quebec (Canada). They use propensity score matching (PSM) to account for voluntary participation into the program and pinpoint a decrease in the use of outpatient services, but no effect on secondary care (hospitalizations and ED admissions). Similarly, in our study GPs joined the scheme at different times and they do it on a voluntary basis, hence we also use propensity scoring to address the potential bias arising from self-selection. Yet, we perform the analysis in a different institutional setting, and we widen the scope of the analysis by separately estimating the treatment effect for different types of CHCs to capture diversity in size, internal organization and services provided. Last, from a methodological perspective, we take advantage of the most recent advances in DID methods by applying the estimator proposed by Callaway and Sant'Anna (2021), which is robust to treatment effect heterogeneity in settings with staggered treatment timing.

Our analysis delivers novel insights on several issues that are relevant for policy. First, we study the utilization of downstream hospital treatments focusing on inappropriate episodes, rather than on the overall use of services. Assessing whether integrated, team-based care centres improve appropriateness of treatments is a key piece of information to establish whether the new model can achieve one of its main objectives, that is channeling demand into the proper care settings relatively better than the traditional GP gatekeeping model. Second, such effect is investigated for a large, vulnerable population of chronic patients, that represents a focal target for primary care policies. Third, we consider a highly planned public system, where all patients benefit from comprehensive universal coverage and are registered with a primary care physician. In similar institutional contexts, the causal impact of collaborative care models has not yet been investigated.

The paper also contributes to the streams of literature that address the role of primary care on diabetes outcomes and on the congestion of emergency departments (EDs), respectively. While the first body of works focuses mainly on economic incentives paid to GPs [e.g., Scott et al. 2009; Dusheiko et al. 2011; Kantarevich and Kralj 2013; Iezzi et al. 2014], we consider a policy wider in scope which results in a major organizational change of the practice. The studies that link expanded access to primary care with outpatient and emergency services have considered policies ranging from the mere extension of opening hours [Dolton and Pathania 2016; Lippi Bruni et al 2016], to the opening of Walk-in-clinics (WiCs) where the primary care physician is bypassed. Pinchbeck (2019) provides an insightful analysis of the deployment of WiCs in the UK that has several analogies with our contribution, including staggered program implementation and extended opening hours of community-based centers. Differently from us, the UK policy initiative is aimed also at increasing the points of delivery for treating emergency conditions outside the ED, especially in underserved areas. The contribution uses area-level data and studies the role of improved spatial accessibility to services, as well as of easier admission rules, with patients no longer required to be registered in advance with the community service. The Italian CHC initiative falls in-between the mere extension of opening hours and a richer availability of supply centers, as it is based on the reorganization of the management of existing clinics, preserving the GPs' pivotal role and continuity of care, while taking advantage from a multidisciplinary care model. Differently from the UK experience, the organizational changes occur in a context where the role of spatial and institutional barriers in the access to services do not vary. Hence, the evaluation of its effects is important also to design viable strategies that reduce congestion at the ED when policymakers do not engage in large programs for establishing new centers, but they reorganize the existing supply. We also take advantage of the availability of individual-level data that allows us to control for heterogeneity across units of observation at the level individual patients. Moreover, we have the possibility to assess the robustness of baseline findings to the Callaway and Sant'Anna (2021) DID framework that was not still settled by the time Pinchbeck's (2019) paper was published.

3. Community Health Centres in primary care

In Italy, GPs are self-employed physicians contracted with the NHS paid mainly through capitation. They operate on a list-based system and can freely decide where to establish their premises, although they are encouraged to aggregate in group practices. Physicians operating in groups locate

their clinic in the same premises. They may also decide to share the costs of nursing services and to substitute each other's in case of illness, but each GP fully preserves its professional autonomy.

The establishment of Community Health Centres is a leading, more ambitious initiative whose aim is to fosters integration between community and primary care. When a CHC opens in the area where a GP operates, the latter has the opportunity to join the new team-based organizational model. The CHC model preserves the pivotal role of the GP in guiding patients in their contacts with the NHS, but it promotes closer collaboration between GPs and the territorial network of providers. Because of that, GPs lose part of the autonomy of the traditional gatekeeping model and face added organizational complexity, as teamwork may impose coordination costs. The choice to adhere to the program is an individual one and is taken by each GP autonomously from the group he/she may belong to. From the financial viewpoint, physicians receive no direct cash subsidy for joining a CHC, but the running costs of the center is borne by the NHS through the local LHA. Participation is incentivized as the costs of renting the practice premises and of hiring personnel are waived for participating physicians.

Faced with counteracting incentives, the diffusion of the CHC model grounds on the propensity of GPs to adhere to the initiative. The policymaker's view is that, operating in a context with strong governance, leadership, and advanced infrastructures, such as innovative equipment, skilled health workforce and digital supports, improves quality of care and facilitates the connections with other layers of the system. This is expected to yield social benefits, in terms of a more efficient and effective care environment, high value treatments and improvements in population health.

In 2010, guidelines were issued to support Local Health Authorities (LHAs) in the development of the CHC initiative [Emilia-Romagna Region, 2010]. CHCs are managed by the LHA Primary Care Department that establishes them as reference centres to improve population health management through a better integration between hospital and outpatient services and between health and social care. As such, they are expected to facilitate the management of chronic conditions at the territorial level, including prevention of hospital referrals for minor conditions. CHCs provide citizens with a unique access point for outpatient care; organize and coordinate care and health communication to patients; strengthen the integration between hospitals and community care, also providing outpatient emergency treatments; develop diagnostic and integrated care pathways together with prevention programs targeting individuals, specific subgroups, and the general population; manage chronic conditions through primary and specialist care integration.

Regional guidelines set standards for CHCs' premises to ensure that the infrastructures are uniform and adequate to the functions and services delivered. Two types of CHCs can be identified: small and large complexity centres which differ not only by size, but also by range of social and health services delivered. Small-size CHCs provide ambulatory care nursing, medical primary care, pediatric and obstetric care, specialized outpatient clinics, social assistance, and primary prevention services (including vaccines), and guarantee access to care 12 hours a day. Large CHCs supplement services offered in small CHCs, with outpatient clinics for the integrated management of chronic illnesses and conference rooms meeting to host health education programs for the general population, X-ray diagnostic services, blood samples service, specialist outpatient care, homecare service coordination and rehabilitation care, family counselling, mental and addiction care, as well as secondary prevention services, including screening [Odone et al. 2016]. With the initiative beginning in 2010, the number of running CHCs amounts to 42 in 2011, raising up to 55 in 2013, and to 84 in 2016 (63 small CHCs and 21 large CHCs) [Emilia-Romagna Region, 2016].

For improving the quality of diabetes care, the Regional Department of Health issued the "Clinical Guidelines for Management of Diabetes Mellitus" introducing a Disease Management Program called "Integrated Management" (IM) at LHA-level, which is based on an integrated proactive approach managed by the GP for the assumption of responsibility of type-2 diabetes patients and supported by monetary incentives [Emilia-Romagna Region, 2003]. Patients can be enrolled in the IM program if they do not exceed a given severity threshold (i.e., they are without complications or with minor complications, otherwise they have to be managed by the Diabetic Centre) [Ugolini et al., 2019]. By joining the CHC, GPs also have the opportunity to improve the quality of care provided to their diabetes patients thanks to the advantages that the new model has over traditional practice organization. First, CHCs provide up-to-date advanced diagnostics that allow accurate on-site checks of patients' health status. Second, the "Outpatient Clinic for integrated management of chronicity" run by multi-professional teams within CHCs gives patients the opportunity of a more direct access to specialist consultations, and specialized nurses ensure a more structured follow-up and the implementation of personalized educational programs to improve patients' compliance. Third, CHCs guarantee access to primary care services 12 hours a day, giving patients the opportunity of consulting a doctor in case of a potential emergency even when the GP the individual is enrolled with is not working.

From the patient's perspective, under the Italian list-based system, when a GP takes the decision to join a CHC, all rostered patients follow the physician. In general, patients may change their GPs at

any time, including if they are unsatisfied with (or lack of) the enrolment into a CHC. However, in practice we observe very limited patients' movements between GPs in any circumstance, thus confirming that adherence to a CHC is a physician-driven choice. [2] As GPs move their clinics inside the CHCs, visits and consultations take place in the new setting, and, even though patients remain registered with a single GP, they gain access to all services provided within the CHC.

4. Data

The data is drawn from the Regional Healthcare Information System, which provides information on GPs operating in the Region and on patients' utilisation of regional health services. The diabetic patient is our unit of observation. The main advantage of using patient-level instead of GP-level data relies on the possibility to include individual controls for patients' health status which contribute to reducing unobserved heterogeneity across observations.[3] The study population consists of all type-2 diabetic patients living in the region between 2010-2016. Patients' identification is based on an algorithm developed by the Regional Department of Health that selects individuals over the age of 18 years who have received at least two diabetes drug prescriptions (oral agents or insulin) over the previous three years. As this criterion may fail to identify cases treated through diet and exercise only, the dataset is augmented by outpatients who, over the previous three years, attended a Diabetes Centre and inpatients diagnosed with diabetes. Given the nature of the dataset, based on anonymized administrative records, no clinical information such as blood glucose levels or HbA1c levels is available.

To retain only observations that are viable for our identification strategy, we omit from our estimating sample the GPs that have already adhered to the CHC model in year 2010, for whom untreated potential outcomes are never observed as they are subject to treatment over the whole estimation period (*always adopters*). We apply additional exclusion criteria in order to prevent uncertain assignment of patients to a specific GP, and to leave out from the analysis physicians only marginally involved in primary care activities, for whom CHC membership is a non-viable option. For each year we include in the estimating sample only patients with 12 months of continuous

[2] In the data section, we will show that also in our estimating sample there is no evidence of any change in the attractiveness of GP practices following a change in their affiliation status with a CHC.

[3] The use of data grouped at the GP-level would allow to control for list characteristics only. In such a setting, the inference on the actual drivers of ED would be relatively less informative since the impact of the covariates would refer to the list characteristics and could not be directly associated to those of the individuals responsible of ED admissions.

enrolment with the same GP and patients followed by GPs with more than 426 registered patients.[4] The number of GPs excluded according to the latter criterion is around 30 per year, corresponding roughly to 4,200 patient-year observations.[5] The number of patients registered with multiple GPs during the same calendar year is negligible (around 550 cases per year). Moreover, such switches are well-balanced between GPs that have adhered to a CHC and those who have not, with yearly outflows and inflows ranging between 0.20% and 0.30% of the list size, respectively. These patterns point to a high loyalty of patients towards their GP, with no evidence of strategic enrolment related to the CHC membership of the physician. Hence, for the purposes of our analysis the composition of the list can be assumed as predetermined.

The final dataset consists of a total of 1,621,592 patient-year observations and 2,733 GPs located in 8 LHAs between year 2010 and year 2016. We record 359 GPs that have switched to CHC status by June 2016, corresponding to 17% of the total. The timing of transitions is reported in Figure 1a and Figure 1b for GPs and for the patients registered with them, respectively. These figures document that switches to the CHC status were staggered over the period considered, supporting our identification strategy. In particular, the data shows that the number of CHC admissions increases between year 2011 and year 2015, while decreases in the last year of analysis.

The institutional context supports the conjecture that the timing of adoption of the CHC model by GPs is independent of physician's underlying characteristics. When a new CHC opens, a very high fraction of *adopters* joins the CHC during the first year. Such evidence suggests that, for GPs willing to participate, the staggered adoption of the model is due to supply side constraints and not to physician's own choice. Even more so considering that local delays in openings are mostly due to the need of completing the complex bureaucratic procedures and of establishing the infrastructures required for running a CHC. Consequently, while entry decision is likely to be endogenous to GP characteristics, the timing of such entry can be attributed mainly to the delayed availability of local infrastructures, a feature that is plausibly exogenous to the outcome of interest and supports identification through the staggered roll out of the program.

INSERT FIGURES 1a and 1b

Table 1a provides the descriptive statistics for our diabetic population. Patients' characteristics include gender, age, foreign citizenship, insulin use and the presence of at least one chronic disease

[4] The threshold of 426 patients corresponds to the first percentile of patient distribution in the GPs' list.
[5] Our findings are robust to the inclusion of the full set of GPs and the results are available upon request.

other than diabetes (asthma, hypertension, coronary artery disease, chronic obstructive pulmonary disease, congestive heart failure). Observations can be divided in two subgroups: we define patients whose GP adhered to a CHC between year 2011 and year 2016 as *switchers*, whereas those patients whose GP did not join a CHC over the period of interest are labelled *non-adopters*.

Approximately half of patients (47%) are females, the average age is 69, and about 60% of them suffer from at least one additional chronic disease. Foreigners account for about 5% of the sample, while 21% of registered patients received at least one insulin prescription. To assess the balance of patients' characteristics between *switchers* and *non-adopters*, we compute the standardized difference for each covariate, distinguishing between continuous and dichotomous variables. All standardized differences for patient-year observations were lower than 0.10, showing a balance in observable characteristics between the two groups.

INSERT TABLE 1a

GPs' characteristics include age, gender, practice type (individual vs. group practice), list size, degree of urbanization of practice location (low-medium vs high), the share of diabetic patients in the list, and a dummy for GP's participation to the local IM program for diabetes care. To control for GP's involvement in care-improvement activities and for their compliance with regional and local guidelines, we also include two variables accounting for financial incentives obtained from (1) any incentivised program the GP adheres to (excluding diabetes) and (2) any diabetes program the GP adheres to. Both indicators are defined as share of the annual GP income paid by the LHA.

INSERT TABLE 1b

About two thirds of GPs are males, their average age is 60 years, 85% of them work in group practices and 57% adhere to the local IM program, with an average 6% of their enrolled patients with diabetes. Incentivised programs amount to 10% of total income, whereas incentives for diabetes programs represent 2.5% of total income. The bottom panel of Table 1b reports in the first column standardized differences between GPs who switch to the CHC model (*GP-switchers*) and those who do not (*GP-non-adopters*) and highlights significant differences in observed characteristics across groups. Physicians that join a CHC are more likely to work in group-practices, to adhere to the local IM program, to operate in low-medium urbanized areas and to have a larger list size (and a higher share of diabetic patients in their list).

Our outcome measure is given by inappropriate admissions to Emergency Departments (EDs), identified from to the Italian four-level triage system (I-4L). Upon admission to the ED, all patients receive a triage assessment, whereas no comparable classification system for severity/urgency condition exists for patients visiting ambulatory care facilities, including CHCs. The lowest priority category in the acuity scale (*white codes*) corresponds to minor conditions that can be effectively treated in outpatient settings. The triage criteria for classifying emergency admissions are established by the Ministry of Health. The classification adjusts to changes in patient's health status that may occur during the period spent at the ED and the system meets satisfactory reliability and validity standards [Parenti et al. 2010]. The steps of the process are the following: upon arrival at the ED, patients' conditions are subject to a quick, preliminary assessment for identifying cases in need of immediate treatment. A successive, more accurate evaluation leads to the assignment of the triage code by a specialized nurse. Finally, at the time of discharge the physician in charge reassesses the entire episode in the light of the information acquired and assigns the final triage. In line with the criteria set by the regional health authorities, we classify as inappropriate the episode receiving the lowest grade in the urgency/severity ranking at both the time of admission and of discharge.

Table 2 displays the frequency of inappropriate ED visits, distinguishing patients whose GP belongs to a local CHC from those patients whose GP does not. While hospital emergency services are available on 24h/7days basis, access to CHCs is limited to daytime on weekdays. Still, compared to traditional practices, CHCs are accessible over a longer time span on weekdays offering an alternative to the ED especially valuable for patients experiencing minor problems. This reactionary channel can be expected to relieve the ED from inappropriate visits especially during CHC opening times. Alongside, high quality of preventive service delivered at the CHC (on-site diagnostic and specialist care, personalized educational programs, etc.) may contribute to reduce overall ED utilization irrespective of the timing of admission. Hence, separate analysis by timing of admission may be important to establish the empirical relevance of the different channels on the use of hospital emergency treatments. To get further insights, we separately consider admissions occurring during daytime on weekdays by those that occurred at nigh-time or during weekends to analyze separately the time spans when the CHCs are open from those when they are not.

Inappropriate ED admissions involve 5% of our sample of patient-year observations (3.6% for patients whose GP adheres to a CHC and 5.2% for those patients whose GP does not enter a CHC).

Most visits are registered during the day on weekdays covering 3.5% in the total sample (2.5% for the CHCs sample vs 3.8% for the control units).

INSERT TABLE 2

5. Econometric strategy

Our empirical strategy draws on the comparison of inappropriate ED admissions between patients followed by GPs that operate in a CHC between January 2011 and June 2016 (treatment group) and patients whose GPs do not (control group). Treatment is an absorbing state, so that patients are considered treated since year *t* if their GP entered a CHC during the first semester of that year. In contrast, the control group consists of patients assisted by GPs that have never entered a CHC until June 2016. We exploit the staggered deployment of the program, with *switchers* entering a CHC at different points in time, as source of variation for estimating the causal effect of CHC participation.

While in the study treatment is assigned at the patient level, the decision to adhere to a CHC is taken by the GP. Compliance to treatment might be threatened if patients move out of the practice in response to their GP's joining a CHC. However, the institutional features of the system ensure full patients' compliance to treatment. We should bear in mind that in Italy rostering of patients in the list is binding and patients cannot attend a practice different from the one they are registered with. Although in principle patients can change practice at any time and with no transaction costs, our data shows that switches to a different list are extremely rare, with no evidence that patients strategic enroll (signing in/out of the list) following GP admission into a CHC.[6] These considerations strongly support the conjecture that, when a GP joins a CHC, the new setting is shared by all listed patients.

5.1 Difference-in-differences approach

Our baseline estimation strategy is given by a two-way fixed effects (TWFE) difference-in-differences design. More precisely, we run a TWFE specification as described by the following equation:

$$E(y_{ijt}/X_{ijt}) = \alpha_0 + \alpha_1 Post\_X\_Treatment_{jt} + \alpha_2 Z_{it} + \alpha_3 W_{jt} + \gamma_i + \delta_t + \varepsilon_{ijt} \quad (1)$$

where y_{ijt} is a dummy variable taking value 1 if patient *i* in the list of GP *j* had at least 1 inappropriate ED visit at time *t*, and 0 otherwise; $Post\_X\_Treatment_{jt}$ is a dummy variable equal to 1 if GP *j* is

[6] On this point see also our discussion at the beginning of section 4.

part of a CHC by the first semester of year *t*, and 0 otherwise; Z_{it} and W_{jt} are vectors of patient and GP covariates, respectively; γ_j and δ_t are practice and year fixed effects, with 2010 set as the baseline year; and ε_{ijt} is a white noise error term.[7] Treatment adoption at different points in time implies that, in each sample year, our control group is represented by all practices that have not yet entered a CHC in the first semester of that year.

One issue that arises in identifying the causal effect of CHC participation is the non-random assignment of GPs, and consequently of their patients, to the different groups. Descriptive analysis reported in Table 1b) suggests that GPs joining a CHC differ in observable characteristics from those who do not, thus pointing to potential endogeneity bias.

The use of practice fixed effects in equation (1) crucially allows to control for fixed differences between practices that entered a CHC and those that did not. So, if treated patients always had less inappropriate ED visits than the control group, then this effect will be absorbed into the GP fixed effects. Changes in inappropriate ED visits that are common to all GPs are captured by the year FEs. Our key coefficient of interest is α_1, which estimates the effect of CHC as changes in inappropriate ED visits specific to practices after they switched to CHC status.

5.2 Multilevel analysis and propensity score

Given the hierarchical structure of our data, where patients are clustered within GPs, we test the robustness of our estimates by using a two-level generalized linear probability DID model [Imbens and Wooldridge, 2009; Lencher, 2010; Arpino and Mealli, 2011], where a random intercept is introduced to deal with the fact that the same patient can be observed across years:

$$E(y_{iit}/X_{iit}) = \alpha_0 + \alpha_1 Post\_X\_Treatment_{it} + \alpha_2 Z_{it} + \alpha_3 W_{it} + \alpha_4 CHC_i + \delta_t + v_i + \omega_i + \varepsilon_{iii} \qquad (2)$$

CHC_j is a dummy taking value 1 if the *j-th* GP ever adopts the CHC model by June 2016 (*GP-switchers*), and 0 otherwise (*GP-non-adopters*); $v_j \sim N(0,\sigma_v^2)$ is a Gaussian-distributed random-effect term specific for the *j-th* GP; $\omega_i \sim N(0,\sigma_v^2)$ is a Gaussian-distributed random-effect term specific for

[7] Even if patients may visit the ED more than once during the same year, we adopt a binary rather than a count indicator for inappropriate ED visits, because multiple inappropriate ED visits associated to the same patient are rare events. In our sample, the number of inappropriate ED visits associated to the same patient in each year is up to 1 in almost 98% of the sample. It is equal to 2 for only 1.8% of cases, 3 for about 0.4% of patients, and more than 3 for only about 0.2% of the sample.

the *i-th patient;* and $\varepsilon_{ijt} \sim N(0,\sigma_v^2)$ is a Gaussian-distributed error term specific for the *i-th* patient and the *j-th* GP at time *t*.

The CHC_j indicator captures whether the patient's GP ever joins a CHC or not. It can be interpreted as a proxy for "GP-type", accounting for time-invariant factors that distinguish *switchers* from *non-adopters*. It is expected to absorb (at least part of) the unobserved heterogeneity across physicians' groups, thereby attenuating the consequences of GP self-selection into the program. The purpose is to control for professional attitudes and practice style characteristics associated with the propensity to adopt the CHC model that may also affect outcomes via patient-physician interactions.

As a more comprehensive way to tackle the possible selection bias due to GPs voluntary adherence to the CHC model, we further adjust our analysis by using one of the propensity score-based methods developed by the statistical literature for observational studies [Rosenbaum and Rubin, 1983; Austin and Mamdani, 2006; Austin, 2008]. The aim is to ensure that, conditional on the propensity score, *GP-switchers* and *GP-non-adopters* have similar distributions of covariates in the pre-treatment period. The propensity score is obtained by estimating a logistic regression model, in which the probability of adhering to the local CHS is regressed on the GP's characteristics listed in Table 1b.[8] All these variables are assumed to be related to both treatment and outcomes, and are measured at the baseline pre-treatment year 2010.[9] In the bottom panel of Table 1b, we report effects size computed before and after PS estimation. The balance of characteristics between the group of *GP-switchers* and *GP-non-adopters* was reassessed calculating conditional weighed standardized differences [Austin, 2008]. Our findings confirm that, after balancing with the propensity score, physician characteristics are comparable between the two groups.[10]

While matched paired analysis based on the propensity score is most used in the economic literature [Strumpf et al., 2017], given the characteristics of our data where only 13% of GPs switched to CHC status by June 2016, such approach would have largely reduced our sample size and negatively affected the precision of the estimated treatment effect. For this reason, we used

[8] The variable that captures participation into local management programs (GP_IM) is dropped due to collinearity with financial incentives received for diabetes-related care improvement activities.

[9] Following Austin and Mamdani (2006), *non-adopters* who had an estimated propensity score lower than any GPs in CHCs were excluded from the analysis (4 physicians in total), as they emerge as outliers compared to the remaining GP population. The coefficients of the propensity model were then re-estimated on the restricted dataset.

[10] We use the Stata program developed by Becker and Ichino (2002) to estimate a propensity score that satisfies the so-called "common support" assumption, which restricts the set of data points over which the test of the balancing property is satisfied to those belonging to the intersection of the supports of the propensity score of treated and control units.

the estimated propensity score in a covariate adjustment procedure [D'Agostino,1998; Austin and Mamdani, 2006], where the outcome is regressed on an indicator variable denoting treatment status together with the propensity score estimated for each GP and linked to patient-level data. To accommodate non-linear relationships between outcome (inappropriate ED visits) and propensity score, the latter was included in the model via a cubic spline function [Austin and Mamdani, 2006; Franklin et al.,2017; Tian, Baro and Zhang, 2019]. This strategy leads to the following equation (2):

$$E(y_{ijt}/X_{ijt}) = \alpha_0 + \alpha_1 Post\_X\_Treatment_{jt} + \alpha_2 Z_{it} + \alpha_3 s(PS_{jt}) + \alpha_4 CHC_j + \delta_t + v_j + \omega_i + \varepsilon_{ijt} \quad (3)$$

where s denotes the cubic spline function and PS_{jt} is the propensity score estimated for *j-th* GP at time *t* [Strumpf et al., 2017].

5.3 Staggered difference-in-differences estimation

As highlighted in the most recent DID literature, in settings here there are multiple time periods and units can become treated at different points in time, TWFE regression models suffer from potentially severe weaknesses that threatened identification of the causal effect (Goodman-Bacon, 2021). The TWFE approach provides estimates of a weighted average of treatment effects based on all possible pairs of treated and untreated units at different points in time. However, some of these comparisons are newly treated units relative to already treated units, and this can generate biased estimates in contexts with treatment effects heterogeneity.

To deal with this issue, we use the Callaway and Sant'Anna difference-in-differences (CS-DID) methodology (Callaway and Sant'Anna, 2021). This approach bypasses the weaknesses associated with TWFE by explicitly identifying group-time average treatment effects for different groups and different times, where "groups" are defined based on the time period when units become treated. This method is robust to arbitrary treatment effect heterogeneity and allows to examine treatment effect dynamics. We exploit this flexibility to test whether average treatment effects (ATTs) are heterogeneous with respect to the length of exposure to the treatment, and how cumulative average treatment effects evolve over calendar year.

Within this framework, we use the "never treated" units as our counterfactual, and we apply the doubly-robust method proposed by Sant'Anna and Zhao (2020) as estimation procedure. We identify the causal effects parameters by relying on the conditional parallel trends assumption. This assumption can be considered more plausible than an unconditional parallel trends assumption, as

it holds after conditioning on observed pre-treatment covariates. More precisely, it assumes that, conditional on GP pre-treatment covariates, the average outcomes for the treated group and for the "never treated" group would have followed parallel paths in the absence of treatment. To assess the validity of this assumption, we conduct a pre-trend test for the null that all pre-treatments are equal to zero. The results of this test, which are available upon request, suggest that we cannot reject the null of parallel trends, therefore supporting the conditional parallel trends assumption.

6. Results

Table 3 shows our baseline results for the TWFE DID model as specified in equation (1). Column (1) presents our findings for all ED admissions, whereas columns (2)-(4) show the results for ED admissions over different time-horizons, i.e., daily visits occurring in working days, night visits and weekend visits, respectively.

INSERT TABLE 3

The estimates for the treatment effect ($Post\_X\_Treatment_{jt}$) pin down a significant reduction in inappropriate ED admissions by diabetic patients associated to GP's participation to a CHC. On average, we find a statistically significant decline in the probability of inappropriate admissions by 1.45% for all ED visits and by 1.7% for daily ED visits occurring during working days. Conversely, we find no evidence of significant effects of CHC participation when ED admissions are confined to night and weekend ED visits when CHC are closed. The estimated drop in admissions during CHC opening hours is consistent with the view that, compared to enrolment with traditional primary care practices, the opportunity to directly access a CHC at the point of need generates a protective effect against the risk of using the ED inappropriately. While patient-based controls affect ED admissions, in particular those capturing frailty conditions, time-varying personal characteristics of the GP have a small influence on the outcome of interest.

We next exploit the hierarchical structure of the data by running the two-level generalized linear probability DID analysis. We present our findings for all ED visits in Table 4. Column (1) shows the results by using the model as specified in equation (2), where self-selection is controlled for only by means of the dichotomous indicator that identifies *adopters*, as proxy for physician type.

INSERT TABLE 4

The results are very similar to our baseline findings, both in qualitative and quantitative terms. As for our key variable of interest, the impact of treatment points to a significant reduction in inappropriate ED admissions by diabetic patients associated to GP's participation to a CHC. In terms of size, the drop in the probability of inappropriate ED visits is of about 1.4 percentage points for total admissions.

The indicator intended to capture GP's propensity to adopt the CHC model (CHC_j variable) absorbs part of the otherwise unobserved heterogeneity between *adopters* and *non-adopters*. The estimated coefficient suggests that *adopters* record a significantly lower propensity to visit the ED inappropriately. The finding is consistent with the conjecture that GP-type may affect both program participation and health outcomes.

Although insightful, the findings above should be cautiously interpreted as causal effects. A drawback being that the empirical approach relies on the dummy indicator for *adopters* to address the underlying differences between *adopters* and *non-adopters* possibly induced by a non-random assignment of GPs between treatment and control group. To tackle this issue, we estimate a two-level generalized linear probability DID model as defined in equation (3), where we allow for nonlinear effects by modeling a PS spline function calculated at GP level. Such a strategy deals with potential selection bias by comparing similar distributions of baseline covariates across GPs.

Column (2) of Table 4 shows the results for this model specification, with the dependent variable being based on all ED visits. As for the coefficient of main policy interest, the estimated parameter for the treatment effect is very stable and points to a significant reduction in the probability of inappropriate ED access, with a magnitude of the CHC effect being about 1.5 percentage points for all ED admissions. Patients' characteristics exert a significant, albeit generally modest in magnitude, influence on the probability of inappropriate ED admissions.

Finally, we test the sensitivity of our results to treatment effects heterogeneity by using the CS-DID methodology. Table 5 presents the results derived by using the CS-DID approach for staggered treatment assignment. We show the overall estimated average treatment effect for all inappropriate ED visits and by distinguishing between alternative time span (i.e., daily visits occurring in working days, night visits and weekend visits). The estimates are very similar to our baseline findings. The overall estimated average treatment effect points to a statistically significant decline in the probability of inappropriate ED visits during CHC opening hours, while we continue to find no evidence of a significant effect of CHC participation when the outcome is restricted to night

and weekend ED visits. The estimated coefficients suggest a drop in the probability of inappropriate admissions by 1.5 percentage points for all ED visits, and by 1.6 percentage points for daily visits occurring during working days.

INSERT TABLE 5

We exploit the flexibility of the CS-DID approach to examine the treatment effect dynamics with respect to length of treatment exposure and with respect to calendar year. Figure 2 plots the average treatment effects of CHC participation by length of exposure to the treatment. For the periods before GPs' participation into CHCs, the plotted values provide a "pre-test" for the conditional parallel trends assumption. As this graph shows, the only statistically significant effects refer to the post-treatment period: we find a significant drop in the probability of inappropriate ED visits since the second year of treatment exposure, with the effect being fairly stable afterwards. Moreover, the average treatment effects in the pre-treatment period are insignificant, providing further support to the validity of the conditional parallel trends assumption.

INSERT FIGURE 2

Table 6 estimates the average treatment effects by post-treatment year. The estimated parameters suggest that the effect is statistically significant in all periods of study from year 2012 onwards, with the size of the effect ranging between 1.3 and 2.4 percentage points.

INSERT TABLE 6

7. Extensions

In this section, we provide a twofold extension of the main analysis. First, we perform a falsification test for the impact of participation to CHCs. Second, we explore the role of the internal organization of CHC as potential determinant of the observed outcome. We conduct these additional analyses to provide further validity to our estimation results, and to gain further insights on the institutional drivers of our main findings.

7.1 Falsification test

For the falsification test we take as dependent variable episodes that a priori should not be affected by physician's CHC membership. Good candidates are ED visits for highly severe and urgent

conditions (s.c. *red codes*). They typically involve ill-health and trauma episodes that, due to their complexity and urgency, must receive a timely and highly specialised treatment in a hospital-based setting. As such they are not deemed to be sensitive to the organisation of community and primary care. Any evidence of a significant "CHC-effect" for red codes would point to the presence of unobserved factors that challenge identification for inappropriate admissions as well.

The results of the falsification test are presented in Tables 7-8, where we apply our baseline analysis (equation 1) and the CS-DID approach, respectively, but include *red* instead of *white* codes as dependent variable. We detect no significant impact of the adoption of the CHC model on the probability of urgent and highly severe episodes leading to inappropriate ED admissions. Overall, these findings are consistent with the conjecture that the integrated care model helps physicians retain patients suffering from minor conditions that may otherwise attend the ED, but it has no influence on admissions for patients indisputably requiring secondary or tertiary care. The lack of significance for the treatment indicator for red codes supports the identification of a genuine rather than a spurious "CHC effect" in the main analysis.

INSERT TABLES 7-8

7.2 The role of CHC size

We further extend our analysis to investigate possible channels that may drive our results. In particular, we assess potential heterogenous effects across different types of CHCs. In our context, they can be identified according to the structural characteristics established in the Regional Guidelines, which define operating and infrastructural standards for large and small size CHCs. Our approach follows recent evidence investigating the heterogenous impact of CHCs across types of facility [David et al 2018b], whereas most previous studies considered them as black boxes. Lack of consideration of the differences in CHCs' internal organisation may fail to grasp valuable indications on whether particular types of CHCs influence patients' outcomes the most. As discussed in section 3, in Emilia-Romagna large size CHCs are endowed with more advanced technological equipment, offer a wider array of preventive and diagnostic activities, and require larger financial investments.

Tables 9-10 examine the role of CHC size by using TWFE (equation 1) as our baseline estimation model, and the CS-DID estimation strategy allowing for arbitrary treatment effect heterogeneity. We employ two estimating sub-samples separately, by excluding large and small CHCs in turn.

INSERT TABLES 9-10

Our findings are robust across specification models. While confirming previous results for all the other covariates, the empirical evidence suggests that CHC size and internal organization are key determinants in affecting the capacity of the new team-based organizational model to reduce inappropriate use of hospital services. In particular, the estimated parameters indicate that the significant drop in white code ED admissions can be primarily ascribed to the role exerted by large CHCs. The contribution of the latter group of facilities points to a statistically significant reduction in the probability of inappropriate ED visits ranging between 3 and 4 percentage points approximately. Conversely, the effect tends to fade out when the analysis is confined to small CHCs.

8. Conclusions

We investigated whether the introduction of innovative team-based organisational models in primary care based on the establishment of Community Health Centres contributed to curb down inappropriate utilisation of emergency care services. Using data from the Italian Region Emilia-Romagna, we fitted a series of DID models to estimate the impact of GPs' participation into CHCs on the probability of inappropriate ED visits of diabetic patients between year 2010 and year 2016. Our results suggest that CHC status had the effect to significantly reduce the probability of inappropriate ED admissions compared with the traditional primary care model, although the magnitude of the effect is modest. Most importantly, the result holds after accounting for GPs' self-selection into the program, controlling for patients' and GPs' characteristics, and allowing for treatment effect heterogeneity. We document that the effect is driven by daytime visits registered during working days, whereas the finding is not confirmed for night and weekend admissions, when CHCs are closed. This reinforces the credibility of the existence of a causal link between CHC accessibility and reduction in inappropriate utilisation of hospital emergency services.

Our findings are robust across alternative DID estimation procedures. The estimates for the treatment effect point to a reduction in the probability of inappropriate ED admissions by about 1.5 percentage points for all visits. Using a TWFE estimator, we find that the effect of CHC status increases to 1.6 percentage points for daytime visits occurring during working days, and to about 3 percentage points for large CHCs. By contrast, the effect turns to be insignificant when the analysis is confined to night admissions and weekend visits, and the sample is restricted to small CHC facilities. Such findings suggest that when chronic patients need treatment, the opportunity to access community care centres that deliver integrated care instead of traditional primary care

practices limits inappropriate utilisation of emergency departments. The results are by and large confirmed when we employ alternative DID estimation strategies, including the inference procedure for staggered DID setups proposed by Callaway and Sant'Anna (2021). The analysis on treatment effect dynamics by using the CS-DID approach suggests that the effect of participating into a CHC does not significantly vary with respect to the length of exposure to treatment or with respect to calendar year.

We also show that the impact of CHCs on the use of emergency services is heterogenous across patients' conditions: while the probability of ED visits falls for minor cases, admissions due to traumas and highly severe conditions are unaffected by the policy. Whereas most of the existing literature on community care models considers hospital services as homogeneous sets of treatments, this new insight allows to draw more clear-cut indications on the variation in demand composition that hospitals are expected to face in response to the opening of CHCs in nearby areas. Overall, our findings show that the adoption of collaborative integrated organisational models for the treatment of chronic patients leads to a more appropriate utilization of emergency departments, thus contributing to improve efficiency in resource allocation within the system.

The policy implications of the analysis suggest that investing in CHCs – especially those providing a wide array of integrated services - yields benefits in terms of higher appropriateness in the utilisation of different care settings. Nevertheless, it should also be remarked that, even if the direction is in line with policymaker's expectations, the magnitude of the effect remains fairly small, thus suggesting a limited -albeit positive- relief for Emergency Department overcrowding for diabetics. The development of primary care-based emergency services during nights and weekends could strengthen the effect further, provided that the prevailing mechanism in reducing inappropriate ED visits for diabetic patients is expanded access to high quality community care services, which at the moment is available through CHCs only during daytime in working days. With the CHC-model development process in its infancy, these achievements can be further enhanced in the future.

Despite our efforts to provide a thorough assessment of the impact of the CHC initiative, the paper is not free from limitations. The internal validity of our findings would have come out stronger, had we had the possibility to extend the dataset both backward and forward. Non only the analysis would have benefitted from a longer coverage of the pre-implementation phase, but also from the possibility to assess the long-term effects of the policy, especially relevant when targeting

preventive activities. Moreover, richer information for patients' socio-economic status than it is currently available in administrative data, would have improved the capacity of the analysis to account for underlying patients' heterogeneity.

As for the external validity, our data only covers diabetic patients, and the findings cannot be straightforwardly extended to other population groups. To broaden the scope of the analysis, future work should include groups of patients affected by other chronic conditions, and if possible, the entire patient population. Finally, due to lack of data availability, we could not evaluate the impact on clinical outcomes (e.g., blood glucose levels). With the availability of rich data sources, future studies may examine a wider spectrum of the expected benefits of the policy to be weighed with the costs of the program.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

The study, based on routine administrative information, was carried out in compliance with Emilia-Romagna Regional Authority data processing regulations and the Italian Data Protection Act that harmonized with the European GDPR 2016/679 by means of the Legislative Decree 101/2018.

Administrative data were anonymized prior to the analysis at the regional statistical office, where each patient is assigned a unique identifier. This identifier does not allow to trace the patient's identity and other sensitive information. Anonymized regional administrative data may be used for retrospective studies, with no specific written patient consent, when the aim is health-care quality evaluation and improvement, which was the primary objective of this analysis. Given the characteristics of the study, no ethical approval was required.

References

Arpino B, Mealli F (2011) The specification of the propensity score in multilevel observational studies, Computational Statistics & Data Analysis, 55 (4): 1770-1780.

Austin PC (2008) Goodness-of-fit diagnostics for the propensity score model when estimating treatment effects using covariate adjustment with the propensity score. Pharmacoepidemiology and Drug Safety, 17 (2): 1202-17.

Austin PC, Mamdani MM (2006) A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use. Statistics in Medicine, 25(12):2084-106.

Becker SO, Ichino, A (2002) Estimation of average treatment effects based on propensity scores. The Stata journal. 2(4): 358-377.

Bojadzievski T, Gabbay RA (2011) Patient-centered Medical Home and diabetes. Diabetes Care (34):1047-53.

Callaway B, Sant'Anna PH (2021) Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2):200-230.

Cohen J (1988) Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

David G, Gunnarsson C, Saynisch C et al. (2015) Do patient-centered medical homes reduce emergency department visits? Health Services Research. 50 (2): 418–439.

David G, Saynisch PA, Luster S, Smith-McLallen A, Chawla R. (2018a) The impact of patient-centered medical homes on medication adherence? Health Economics, 27: 1805-1820.

David G, Saynisch PA, Smith-McLallen A. (2018b) The economics of patient-centered care, Journal of Health Economics, 59: 60-77.

D'Agostino RB Jr (1998) Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Statistics in Medicine, 17(19): 2265-81.

de Chaisemartin C, d'Haultfoeuille X. (2020) Two-way fixed effects estimators with heterogeneous treatment effects. American Economic Review, 110(9): 2964-96.

Dolton, P, Pathania V (2016) Can increased primary care access reduce demand for emergency care? Evidence from England's 7-day GP opening. Journal of Health Economics, 49: 193-208.

Dusheiko M Doran T, Gravelle H, Fullwood C, Roland M (2011) Does higher quality of diabetes management in family practice reduce unplanned hospital admissions? Health Services Research, 46: 27–46.

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

Emilia-Romagna Region (2003) Linee guida clinico-organizzative per il management del diabete mellito, (in Italian) http://salute.regione.emilia-romagna.it/@@search?Subject%3Alist=Diabete

Emilia-Romagna Region (2010) Resolution n. 291/2010. Casa Della Salute: Indicazioni Regionali per la realizzazione e L'Organizzazione Funzionale, (in Italian). Available from: http://bur.regione.emilia-

romagna.it/bur/area-bollettini/bollettino-ufficiale-18-marzo-2010-1/casa-della-salute-indicazioni-regionali-per-la-realizzazione-e-lorganizzazione-funzionale/casa-della-salute-indicazioni-regionali

Emilia-Romagna Region (2016). Le "Case della Salute" in Emilia-Romagna il monitoraggio regionale. Report 2016, (in Italian). Available from: http://salute.regione.emilia-romagna.it/cure-primarie/case-della-salute

Faber M, Voerman G, Erler A, et al. (2013) Survey of 5 European countries suggests that more elements of Patient-Centered Medical Homes could improve primary care. Health Affairs, 32 (4): 797–806.

Franklin JM, Eddings W, Austin PC, Stuart EA, and Schneeweiss S (2017) Comparing the performance of propensity score methods in healthcare database studies with rare outcomes. Statistics in Medicine 36 (12): 1946–1963.

Friedberg MW, Schneider EC, Rosenthal MB, Volpp KG, Werner RM (2014) Association between participation in a multipayer medical home intervention and changes in quality, utilization, and costs of care. Journal of the American Medical Association 311 (8): 815–825.

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

Hearld LR, Alexander JA (2012) Patient-centered care and emergency department utilization: a path analysis of the mediating effects of care coordination and delays in care. Medical Care Research and Review, 69 (5): 560–580.

Heckman JJ, LaLonde RJ, Smith JA (1999) The economics and econometrics of active labor market programs. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, 3. Elsevier, Amsterdam: 1865–2097.

Imbens GW, Wooldridge JM (2009), Recent developments in the econometrics of program evaluation. Journal of Economic Literature 47: 5–86.

Italian Ministry of Health (2007) Medical homes: description and objectives. Available from https://www.salute.gov.it/resources/static/primopiano/390/Casa_salute_Intervento_Ministro.pdf (in Italian)

Iezzi E, Lippi Bruni M, Ugolini C (2014), The role of GP compensation schemes in diabetes care, Journal of Health Economics, 50, 144-156.

Jackson GL, Powers BJ, Chatterjee R, Bettger JP, Kemper AR et al. (2013) The patient-centered medical home: a systematic review. Annals of Internal Medicine. 158(3): 169–178.

Kantarevic J, Kralj B (2013), Link between pay for performance incentives and physician payment mechanisms: evidence from the diabetes management incentive in Ontario, Health Economics, 22: 1417-39.

Klein D B, Laugesen MJ, Liu N (2013) The Patient-Centered Medical Home: A future standard for American health care? Public Administration Review, 73: S82-S92.

Kumpunen S, Rosen R, Kossarova L, Sherlaw-Johnson C. (2017) Primary Care Home: evaluating a new model of primary care, Nuffield Trust, London.

Lencher M (2010) The estimation of causal effects by difference-in-difference methods. Foundations and Trends in Econometrics (4), 3: 165–224.

Lionis C, Petelos E (2015) The Patient-Centered Medical Home model within an Integrated primary care system: Towards a European model? In: O'Donohue W., Maragakis A. (eds) Integrated Primary and Behavioral Care. Springer, Cham.

Lippi Bruni M, Mammi I, Ugolini C (2016) Does the extension of primary care practice opening hours reduce the use of emergency services? Journal of Health Economics, 50: 144-55.

NCQA (2019) Benefits of NCQA Patient-Centered Medical Home Recognition, downloadable at https://www.ncqa.org/wp-content/uploads/2019/09/20190926_PCMH_Evidence_Report.pdf

Odone A, Saccani E, Chiesa V, Brambilla A, Brianti E, et al. (2016) The implementation of a Community Health Centre-based primary care model in Italy. The experience of the Case della Salute in the Emilia-Romagna Region. Annali Istituto Superiore di Sanità. 52(1):70-7.

Parenti, N., Manfredi, R., Bacchi Reggiani, M.L., Sangiorgi, D., Lenzi, T. (2010), Reliability and validity of an Italian four-level emergency triage system. Emergency Medicine Journal, 27(7):495-8.

Peikes D, Zutshi A, Genevro JL, Parchman ML, Meyers DS (2012) Early evaluations of the medical home: building on a promising start. American Journal of Management Care 18 (2): 105–116.

Pinchbeck (2019), Convenient primary care and emergency hospital utilization, Journal of Health Economics, 68, 102242.

Rios-Avila F, Sant'Anna P, Callaway B (2021) CSDID: Stata module for the estimation of Difference-in-Difference models with multiple time periods.

Rosenbaum PR, Rubin DB. (1983) The central role of the propensity score in observational studies for causal effects. Biometrika; 70:41–55.

Roth J, Sant'Anna PH, Bilinski A, Poe J (2022) What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. arXiv preprint arXiv:2201.01194.

Sant'Anna PH, Zhao J (2020) Doubly robust difference-in-differences estimators. Journal of Econometrics, 219(1): 101-122.

Scott, A., Schurer, S., Jensen, P.H., Sivey, P. (2009) The effects of an incentive program on quality of care in diabetes management. Health Economics, 18: 1091–1108.

Shippee ND, Finch M, Wholey D (2018) Using statewide data on health care quality to assess the effect of a Patient-Centered Medical Home initiative on quality of care. Population Health Management. 21(2): 148-154.

Strumpf E, Ammi M, Diop M, Tousignant P (2017) The impact of team-based primary care on health care services utilization and costs: Quebec's family medicine groups. Journal of Health Economics, 55:76-94.

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

The College of Family Physicians of Canada (2019) A new vision for Canada: Family Practice - The Patient's Medical Home. Mississauga, ON: College of Family Physicians of Canada. Available from: www.cfpc.ca/A_Vision_for_Canada

Tian Y, Baro E, Zhang R (2019) Performance evaluation of regression splines for propensity score adjustment in post-market safety analysis with multiple treatments. J Biopharm Stat. 29(5):810-821.

Ugolini C., Lippi Bruni M., Leucci AC et al. (2019) Disease management in diabetes care: when involving GPs improves patient compliance and health outcomes. Health Policy 123 (10): 995-962.

Weinick RM, Burns RM, Mehrotra A (2010) Many emergency department visits could be managed at urgent care centers and retail clinics. Health Affairs 29 (9):1630–1636.

World Health Organization – Regional Office for Europe (2013). Health 2020: a European policy framework supporting action across government and society for health and well-being. Geneva, available from: https://www.euro.who.int/__data/assets/pdf_file/0006/199536/Health2020-Short.pdf

Table 1a) Descriptive statistics: patient-level control variables

	Switchers (17%)	**Non-adopters (83%)**	**Total (N = 1,621,592)**	**Standardized difference**
Female patients	0.468	0.468	0.468	0.000
	(0.499)	(0.499)	(0.499)	
Insulin user patients	0.207	0.210	0.209	-0.007
	(0.405)	(0.407)	(0.407)	
Patients with at least one chronic disease	0.589	0.598	0.596	-0.017
	(0.492)	(0.490)	(0.491)	
Foreign patients	0.042	0.048	0.047	-0.007
	(0.199)	(0.214)	(0.212)	
Patient age	69.060	68.909	68.935	0.011
	(13.548)	(13.659)	(13.640)	

Notes.; Patients' characteristics, pooled data for the years 2010-16. All variables are dummies 0-1, except "Patient age" in years. Standard Deviations in parentheses. Patients assigned to multiple GPs in a given year are excluded for that year. Patients followed by GPs with less than 540 enrolled patients or that were always adopters are omitted from the sample of study.

Table 1b) Descriptive statistics: GP-level control variables

	GP-switchers (13%)	**GP-non-adopters (87%)**	**Total (N = 2733)**
GP associated practices	0.952	0.830	0.846
	(0.214)	(0.375)	(0.361)
GP_IM program	0.627	0.563	0.571
	(0.484)	(0.496)	(0.495)
GP female	0.260	0.306	0.299
	(0.439)	(0.461)	(0.458)
GP_low_medium_urbanization	0.528	0.254	0.289
	(0.439)	(0.436)	(0.454)
GP list size	1,426.952	1,370.14	1,377.45
	(315.960)	(327.29)	(326.35)
GP_%diabetic patients in list	6.193	6.011	6.035
	(1.373)	(1.561)	(1.539)
GP_incentivised programs on total income	0.102	0.102	0.102
	(0.006)	(0.058)	(0.059)
GP_%diabetes programs on total income	2.467	2.441	2.444
	(2.081)	(2.358)	(2.322)
GP_age	60.385	60.236	60.255
	(4.414)	(4.784)	(4.737)

Notes. GPs' characteristics, pooled data for the years 2010-16. Standard Deviations in parentheses.

Table 1b) Descriptive statistics: GP-level control variables (*continued*)

	Pre-Propensity Score	**Post-Propensity Score**
	Standardised Difference	**Conditional Weighted Standardised Difference**
GP associated practices	0.399***	0.010
GP_IM program	0.130*	-0.067
GP female	-0.103	0.033
GP_low_medium_urbanization	0.584***	-0.006
GP list size	0.177**	0.023
GP_% diabetic patients in list	0.124*	-0.058
GP_incentivised programs on total income	0.003	-0.019
GP_% diabetes programs on total income	0.015	-0.091
GP age	0.032	0.069

Notes. GPs' characteristics, pooled data for the years 2010-16. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Table 2. Descriptive statistics: frequency distribution of the dependent variables

Dependent Variable	Patient-obs in the treatment group						Patient-obs in the control group		Total patient-obs	
	Large CHC		Small CHC		Total CHC					
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Inappr. ED visits-total	0.031	0.174	0.038	0.192	0.036	0.186	0.052	0.222	0.049	0.216
Inappr. ED visits-workdays-day shift	0.022	0.147	0.027	0.162	0.025	0.157	0.038	0.190	0.035	0.185
Inappr. ED visits-night shift	0.007	0.081	0.008	0.089	0.008	0.086	0.011	0.105	0.011	0.102
Inappr. ED visits-weekends	0.011	0.104	0.013	0.115	0.013	0.111	0.017	0.130	0.016	0.127

Notes: Yearly averages over the period 2010-16. Patients followed by GPs with less than 540 enrolled patients or that were always adopters are omitted from the sample of study. "Patient-obs in the treatment group" denotes observations referred to patients enrolled with GPs operating in a CHC at time *t*. "Patient-obs in the control group" denotes observations referred to patients enrolled with GPs not operating in a CHC at time *t*. "Inappr. ED visits-total" includes all inappropriate ED admissions; "Inappr. ED visits-workdays-day shift" includes inappropriate ED admissions occurring Monday to Fridays between 8 am-8 pm; "Inappr. ED visits-night shift" includes inappropriate ED admissions occurring between 8 pm-8 am; "Inappr. ED visits-weekends" includes inappropriate ED admissions occurring on Saturdays and Sundays.

Figure 1. GPs and patients entering a Community Health Centre

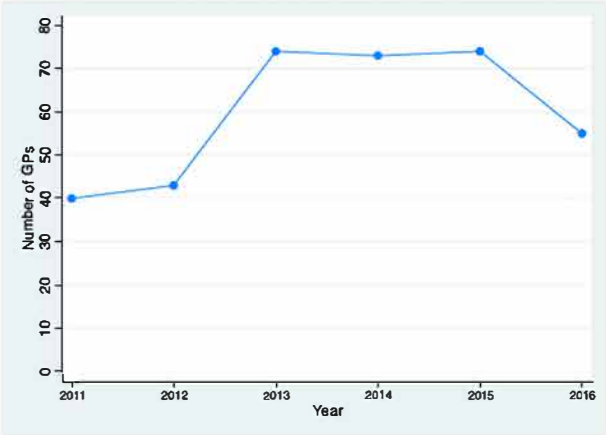


Figure 1a
Number of GPs entering a CHC each year. Years 2011-16.

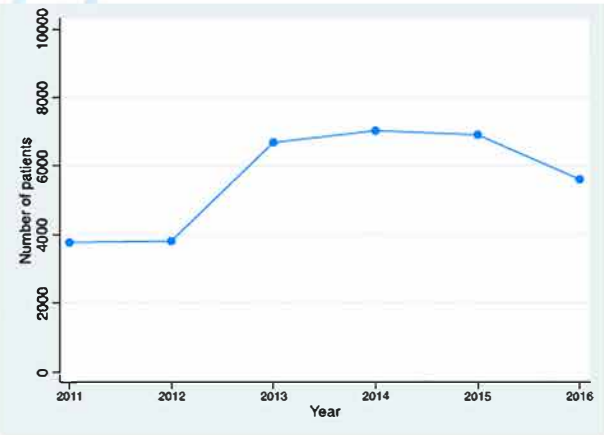


Figure 1b
Number of patients assisted by GPs entering a CHC each year. Years 2011-16.

Table 3. Two-way fixed effects

	INAPPROPR ED VISITS - TOTAL	INAPPROPR ED VISITS - WORKDAYS-DAY TIME	INAPPROPR ED VISITS - NIGHT SHIFT	INAPPROPR ED VISITS - WEEKEND
POST X TREAT	-0.0145***	-0.0174***	-0.0123	-0.00919
	(-3.72)	(-3.79)	(-1.94)	(-1.43)
GP list size	0.0000330***	0.0000252*	0.0000651***	0.0000286
	(3.64)	(2.38)	(4.07)	(1.95)
GP_%diabetic patients in list	0.0951	0.279	0.308	-0.413
	(0.60)	(1.47)	(1.15)	(-1.58)
GP_IM program	0.0138***	0.0128***	0.0139**	0.0156***
	(4.76)	(3.72)	(2.78)	(3.37)
GP_% incentivised programs on total income	-0.00110	-0.00268	-0.00497	0.00185
	(-0.11)	(-0.22)	(-0.27)	(0.11)
Patient_female	0.00524***	0.00589***	-0.000981	0.00479*
	(4.71)	(4.21)	(-0.45)	(2.28)
Patient_age	0.00209***	0.00247***	-0.000883	0.00244***
	(6.94)	(6.37)	(-1.54)	(4.39)
Patient_age2	-0.0000398***	-0.0000425***	-0.0000157***	-0.0000441***
	(-17.93)	(-14.94)	(-3.75)	(-10.92)
Patient_foreign	0.0462***	0.0457***	0.0400***	0.0520***
	(13.56)	(10.39)	(6.01)	(8.14)
Patient ins user	-0.0230***	-0.0240***	-0.0195***	-0.0211***
	(-19.26)	(-15.28)	(-8.40)	(-9.48)
Chronic disease	0.0185***	0.0203***	0.0157***	0.0156***
	(16.16)	(13.67)	(6.70)	(7.12)
Constant	0.126***	0.120***	0.123***	0.164***
	(6.04)	(4.80)	(3.31)	(4.62)
Year FEs	Y	Y	Y	Y
GP FEs	Y	Y	Y	Y

Notes. Years 2010-16. Linear Probability Model specification; *t*-statistics in parentheses. Dependent variables: dichotomous indicators taking value 1 if the patient has at least one inappropriate ED admission during the year, 0 otherwise. "Inappr. ED visits-total" includes all inappropriate ED admissions; "Inappr. ED visits-workdays-day-time" includes inappropriate ED admissions occurring from Mondays to Friday between 8 am-8 pm; "Inappr. ED visits-night shift" includes inappropriate ED admissions occurring between 8 pm-8 am; "Inappr. ED visits-weekends" includes inappropriate ED admissions occurring on Saturdays and Sundays. $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

Table 4 Two-levels generalized DID models

	Multilevel	**Multilevel with PS covariate adjustment**
POST X TREAT	-0.0144***	-0.0152***
	(-5.02)	(-5.22)
CHC	-0.0257***	-0.0241***
	(-4.64)	(-4.22)
GP female	0.00901*	
	(2.22)	
GP_age	0.000623	
	(1.60)	
GP list size	0.0000202***	
	(4.69)	
GP_associated practice	-0.00651	
	(-1.29)	
GP_low_medium urbanization	0.000216	
	(0.05)	
GP_%diabetic patients in list	-0.268**	
	(-3.01)	
GP_% diabetes program on total income	0.00344***	
	(4.34)	
GP_IM program	0.00721***	
	(3.57)	
GP_% incentivised programs on total income	-0.00179	
	(-0.21)	
Patient female	0.00524***	0.00513***
	(4.86)	(4.74)
Patient_age	0.00206***	0.00215***
	(7.84)	(8.12)
Patient_age2	-0.0000396***	-0.0000403***
	(-20.08)	(-20.32)
Patient_foreign	0.0463***	0.0472***
	(17.08)	(17.31)
Patient ins user	-0.0234***	-0.0231***
	(-20.03)	(-19.67)
Chronic disease	0.0184***	0.0181***
	(16.83)	(16.49)
Constant	0.130***	0.198***
	(4.86)	(16.42)
GP-level variance	0.008***	0.008***
	(32.88)	(32.82)
Individual-level variance	0.013***	0.013***
	65.14	65.07

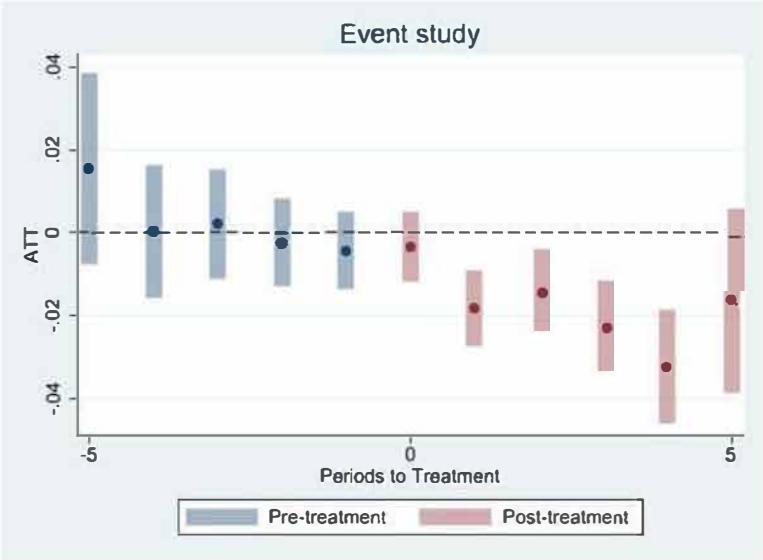
Notes. Cols (1)-(2) reports the coefficient estimates derived from the two-level generalized difference-in-differences regressions as described in Equations (2)-(3), respectively. Years 2010-16. Dependent variables: dichotomous indicators taking value 1 if the patient has at least one inappropriate ED admission during the year, 0 otherwise. "Inappr. ED visits-total" includes all inappropriate ED admission. *t*-statistics in parentheses. $*p < 0.05$; $**p<0.01$; $***p<0.001$.

Table 5. Callaway and Sant'Anna (2020) results

	INAPPROPR ED VISITS - TOTAL	INAPPROPR ED VISITS - WORKDAYS-DAY TIME	INAPPROPR ED VISITS - NIGHT SHIFT	INAPPROPR ED VISITS - WEEKEND
ATT	-0.015***	-0.016**	-0.015	-0.012
	(-3.93)	(-3.30)	(-1.88)	(-1.63)

Notes. Models are computed using the `csdid Stata` command described by Rios-Avila et al (2021), using the improved doubly robust DID estimator proposed by Sant'Anna and Zhao (2020). *t*-statistics in parentheses. Abbreviation: ATT, Average Treatment Effect on the Treated. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 2. Average treatment effects by periods before and after treatment (inappropriate ED visits - total)



Notes. Plots of average treatment effects by length of exposure to treatment and 95% confidence intervals derived with the `Stata` command `csdid` described by Rios-Avila et al (2021), using the improved doubly robust DID estimator proposed by Sant'Anna and Zhao (2020).

Table 6. Calendar time effects: average treatment effects by calendar year (inappropriate ED visits - total)

	INAPPROPR ED VISITS - TOTAL
Year 2011	-0.012
	(-1.05)
Year 2012	-0.024**
	(-3.38)
Year 2013	-0.017**
	(-2.91)
Year 2014	-0.017**
	(-3.05)
Year 2015	-0.011*
	(-2.05)
Year 2016	-0.013**
	-2.81

Notes: Results derived with the `Stata` command `csdid` described by Rios-Avila et al (2021), using the improved doubly robust DID estimator proposed by Sant'Anna and Zhao (2020). *t*-statistics in parentheses. ***$p < 0.01$; **$p < 0.05$, *$p < 0.1$.

Table 7. Red codes (inappr ED visits - total)

	TWFE
POST X TREAT	-0.00116
	(-0.44)
GP list size	-0.00000816
	(-1.74)
GP_%diabetic patients in list	0.396***
	(4.37)
GP_IM program	-0.00474**
	(-2.59)
GP_% incentivised programs on total income	0.00214
	(0.34)
Patient female	-0.0118***
	(-14.49)
Patient_age	-0.00142***
	(-8.08)
Patient_age2	0.0000295***
	(20.60)
Patient_foreign	-0.00782***
	(-5.16)
Patient ins user	0.0380***
	(36.73)
Chronic disease	-0.0377***
	(-38.40)
Constant	0.0259*
	(2.38)
Year FEs	Y
GP FEs	Y

Notes. Years 2010-16. Dependent variables: dichotomous indicators taking value 1 if the patient has at least one inappropriate ED admission during the year, 0 otherwise. *t*-statistics in parentheses. *** $p < 0.01$; ** $p < 0.05$, * $p < 0.1$.

Table 8. Callaway and Sant'Anna estimation - Red codes

	INAPPROPR ED VISITS - TOTAL
ATT	0.004
	(1.44)

Notes: Model estimated with the `Stata` command `csdid` as described in Callaway and Sant'Anna (2021), using the improved doubly robust DID estimator based on inverse probability of tilting and weighted least squares.). *t*-statistics in parentheses. *** $p < 0.01$; ** $p < 0.05$, * $p < 0.1$.

Table 9. CHC size effect (inappropriate ED visits - total)

	TWFE	
	Large CHCs	**Small CHCs**
POST X TREAT	-0.0299***	-0.00831
	(-4.67)	(-1.79)
GP list size	0.0000298**	0.0000302**
	(3.15)	(3.22)
GP_%diabetic patients in list	0.0202	0.101
	(0.12)	(0.62)
GP_IM program	0.0171***	0.0136***
	(5.40)	(4.66)
GP_% incentivised programs on total income	-0.00937	0.00358
	(-0.87)	(0.36)
Patient female	0.00435***	0.00595***
	(3.64)	(5.22)
Patient_age	0.00209***	0.00214***
	(6.48)	(6.86)
Patient_age2	-0.0000402***	-0.0000405***
	(-16.91)	(-17.65)
Patient_foreign	0.0460***	0.0458***
	(12.86)	(13.26)
Patient ins user	-0.0226***	-0.0232***
	(-18.04)	(-18.98)
Chronic disease	0.0191***	0.0185***
	(15.71)	(15.73)
Constant	0.140***	0.131***
	(6.37)	(6.09)
Year FEs	Y	Y
GP FEs	Y	Y

Notes. Years 2010-16. Linear Probability Model specification; *t* statistics in parentheses. Dependent variables: dichotomous indicators taking value 1 if the patient has at least one inappropriate ED admission during the year, 0 otherwise. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 10. Callaway and Sant'Anna estimation - CHC size effect (inappropriate ED visits - total)

	Large CHCs	**Small CHCs**
ATT	-0.037***	-0.009*
	(-4.32)	(-1.97)

Notes: Model estimated with the Stata command csdid as described in Callaway and Sant'Anna (2021), by using the improved doubly robust DID estimator based on inverse probability of tilting and weighted least squares. *t* statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.