

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Does the extension of primary care practice opening hours reduce the use of emergency services?

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

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, 144-155 [10.1016/j.jhealeco.2016.09.011].

Availability:

This version is available at: <https://hdl.handle.net/11585/570046> since: 2022-02-28

Published:

DOI: <http://doi.org/10.1016/j.jhealeco.2016.09.011>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

Bruni, M. L., 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, 144-155.

The final published version is available online at:

<https://doi.org/10.1016/j.jhealeco.2016.09.011>

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)

When citing, please refer to the published version.

Does the extension of primary care practice opening hours reduce the use of emergency services?

Matteo Lippi Bruni ^{a1}, Irene Mammi ^a, Cristina Ugolini ^a

^aDepartment of Economics, University of Bologna

Piazza Scaravilli, 2 40126, Bologna (Italy)

Highlights JHE

We examine patients' attendances at the emergency department in Italy's Region Emilia-Romagna.

We study the impact of an extension in the daily coverage of primary care services.

We estimate count data models and test for endogenous participation into the program.

The extension of primary care availability significantly reduces inappropriate admissions.

The estimated reduction in inappropriate admissions ranges between 10-15%.

Abstract

Overcrowding in emergency departments generates potential inefficiencies. Using regional administrative data, we investigate the impact that an increase in the accessibility of primary care has on emergency visits in Italy. We consider two measures of avoidable emergency visits recorded at list level for each General Practitioner. We test whether extending practices' opening hours to up to 12 hours/day reduces the inappropriate utilisation of emergency services. Since subscribing to the extension program is voluntary, we account for the potential endogeneity of participation in a count model for emergency admissions in two ways: first, we use a two-stage residual inclusion approach. Then we exploit panel methods on data covering a three-year period, thus accounting directly for individual heterogeneity. Our

¹ Corresponding Author: Matteo Lippi Bruni (Tel. +39 051 2098145).

E-mail addresses: matteo.lippibruni2@unibo.it (M. Lippi Bruni); irene.mammi@unibo.it (I. Mammi); cristina.ugolini@unibo.it (C. Ugolini).

results show that increasing primary care accessibility acts as a restraint on the inappropriate use of emergency departments. The estimated effect is in the range of a 10-15% reduction in inappropriate admissions.

Key words: count data; two-stage residual inclusion; panel data models; emergency services; primary care.

JEL classification: C2; H5; I1.

Acknowledgments:

The paper is part of the research project “*The role of economic incentives in the governance of primary care*” financed by the Health Department of Emilia-Romagna, Italy, that also provided the data. The paper greatly benefited from suggestions and comments by Roberto Golinelli, Maria Elena Bontempi, Rossella Verzulli, Stephen G. Hall, Marcello Montefiori and participants to the LAGV12 (Aix en Provence), SIE (Italian Economists Association), iHEA (Dublin) conferences and to the Health Econometrics Workshop 2014 (University of Padua). Two anonymous referees of the journal greatly contributed to improve the work. We are also grateful to Andrea Donatini of the Health Department of Emilia-Romagna for useful suggestions and valuable support throughout the research. The work reflects exclusively authors’ opinions and does not involve the above mentioned institution. Other usual disclaimers apply.

1. Introduction

Overcrowding in Emergency Departments (EDs) raises concerns, since it results in the inefficient allocation of resources and reduces the capacity to provide critically-ill patients with timely responses (Pines *et al.*, 2011; Flores-Mateo *et al.*, 2012). It also increases pressure on hospital resources, and the disruption in the continuity of care may adversely affect health outcomes. Such issues are especially serious in those cases where outpatient care could be equally as effective as emergency services.

The share of avoidable ED visits is estimated at around 25% in Canada (Afilalo *et al.*, 2004), 30% in France (Lang, 1996), Sweden (Hansagi *et al.*, 1987) and Spain (Sempere-Selva *et al.*, 2001), and up to 50% in Taiwan (Chan *et al.*, 2013). More recently, Thompson *et al.* (2013) found for the UK that 43% of patients who went to the ED could have been treated in primary care instead. Similar results were also found in New Zealand (Elley *et al.*, 2007). In Italy, Bianco *et al.* (2003) estimate that non-urgent visits to EDs in the Calabria Region amount to 20% of total visits, while in the USA they range from between 5% and 13% (GAO, 2009). Figures vary depending on the institutional framework and on the criteria used to identify patients who could be treated in a primary care setting (Durand *et al.*, 2011). In particular, non-urgent attendances may overestimate the avoidable episodes by relying on primary care services only, since a fraction of non-urgent visits may still require hospital care. An alternative approach focuses on the classification of ED admissions using a triage coding system based on clinical assessment, which suggests that the share of inappropriate ED visits ranges between 20% and 80% of all non-urgent cases (Afilalo *et al.*, 2004). Despite these differences, researchers agree that a relevant fraction of emergency visits could be treated effectively in less intensive settings, and that ensuring improved access to family physicians may help contain the excessive use of EDs.

We investigate the determinants of inappropriate ED attendances in Italy's Emilia-Romagna Region, and study to what extent practice organisation influences the use of emergency services. Our main goal is to assess whether differences in daytime accessibility to primary care are linked to ED utilisation rates. We focus on a policy that extends the availability of GPs' primary care services to between 10 and 12 hours/day, and we test whether it reduces the number of ED visits that are not followed by hospitalisation. We estimate count models for ED admissions from data for 2008-2010. Since participation in the program is voluntary, we first adopt an instrumental variable strategy: using two-stage residual inclusion (2SRI), we test for the potential endogeneity of the policy variable, and account for the GPs' non-random participation in the program. Thanks to the availability of data over three years, we further exploit variations in health outcomes and program participation, and explicitly account for individual unobserved time-invariant heterogeneity using panel data models.

Our paper contributes to the existing literature in several ways: first, we investigate the link between primary care and emergency services on a large scale, whereas most studies rely on survey data or case studies; second, we expand existing analyses of whether the greater role of primary care policies and GPs fosters the more appropriate use of hospital services (e.g. Dusheiko *et al.*, 2011); third, we partly overcome the limitations associated with the use of imprecise measures of inappropriate ED utilisation, since we identify avoidable ED visits on the basis of a comprehensive ex-post clinical assessment which considers the overall complexity of a patient's condition. Each episode is coded using colour labels, whereby those episodes identified as inappropriate are given a *white code*. In this context, we can adopt alternative definitions of avoidable ED visits. A narrow definition includes only those episodes classified as inappropriate according to the on-site clinical assessment (*white codes*). A less restrictive definition uses information about the intensity of treatment received at EDs, and pools together white codes with those attendances but which are given minor attention at the ED (*potentially inappropriate visits*). This occurs when patients only undergo a general check-up with no diagnostic or specialist follow-up.

Our findings indicate that ED attendances are affected by the increased availability of practice-based care. An extension in practices' daytime opening hours results in a reduced use of ED wards. The results of the panel data analysis indicate that the estimated reduction in avoidable ED admissions is around 14% for white codes, and around 7% for potentially inappropriate visits.

2. Background Literature

Patients requiring timely medical care may turn to an ED or to a primary care practice: while this choice reflects patients' assessment of their own medical needs, it is also influenced by the relative accessibility of alternative solutions. Financial and geographical barriers may affect access to ED and primary care differently. As financial costs of access vary depending on institutional conditions, the type of treatment required or the patient's insurance status, their impact on ED service use is highly context- and patient-specific. In some countries, primary care services are free at the point of use, whereas in others GPs charge fees for visits. As with the ED, the copayment is usually of a limited amount, and may be limited to inappropriate attendances (Sabik and Gandhi, 2015). There is evidence that cost sharing significantly reduces admissions, although the size of the copayment has little effect on ED utilisation (Roberts and Mays, 1998). Similarly, geographical barriers depend on the density and location of GP practices compared to EDs. GP practices are usually located closer to patients' homes, thus rendering the use of EDs relatively less attractive. However, individuals in remote areas may be more subject to complications if they lack regular access to community care (Booth *et al.* 2005), and may thus use emergency services more frequently.

The propensity to use emergency care also depends on a patient's preferences, and may be influenced by characteristics such as age, cultural background and health literacy. For instance, poverty and minority status increase the likelihood of people attending EDs for non-urgent care (Lang *et al.*, 1996), although some studies show that ED attenders are often middle or upper class (Shah *et al.*, 1996; Lee *et al.*, 2000).

Furthermore, the different types of care may be important. While ED visits might be occasional, GP care entails a personal relationship, which is especially valuable for chronic patients for whom continuity of care may reduce hospital utilisation in the long run (Ansari *et al.* 2002). On the other hand, EDs may appeal to patients who seek more direct access to hospital diagnostic and specialist services, and thus use the ED to “jump the queue” (Agarwal *et al.*, 2012). Differences in expected waiting time at the ward is another potentially relevant factor, as substitution between ED and GP visits may occur in response to lengthy waiting times to see local GPs (Puig-Junoy *et al.*, 1998).

As regards physicians' organisations, in the US Lowe *et al.* (2005) estimate that overall ED use by Medicaid patients would fall by 13% if all GP practices opened 12 or more evening hours a week, and by a further 5% if they also opened at weekends. Moreover, the integration of primary care with hospital-based services may reduce utilization of ED services (Feachem *et al.*, 2002). In the UK, practices' organisation seems to have a limited impact, as patients' characteristics account for most of the variation in the use of EDs (Saxena *et al.*, 2006; Calderón-Larranaga *et al.*, 2011). Harris *et al.* (2011) test whether ED attendance can be explained by varying accessibility to GP practices proxied by total opening hours per week. This proxy turns out to be not significant, suggesting that ED use is mainly driven by patients' characteristics. Focusing on the introduction of out-of-hours primary care services, Thompson *et al.* (2010) find little change in ED attendances, whereas the introduction of Pay-for-Performance programs reduces emergency admissions, but only for the incentivised conditions (Harrison *et al.* 2014). The Netherlands' reorganisation of primary care in favour of large-scale GP cooperatives providing out-of-hours assistance appears more successful. Moll van Charante *et al.* (2007) show that the cooperatives treat the vast majority of out-of-hours requests, whereas only a small group of patients self-refer to the ED, mainly for appropriate reasons.

Overall, the literature shows that the reasons for high ED attendances vary and often reflect perceived problems with primary care: they include frustration with scheduling appointments and lengthy waiting times, the perception of long waits before gaining access to secondary services, and greater trust in hospital care. Our work is intended to provide new insights into the potential reduction in inappropriate ED utilisation through the increased accessibility to primary care services represented by longer opening hours.

3. Primary care in the Italian National Health Service

3.1 The institutional framework and GPs' payment scheme

The Italian National Health Service (NHS) established in 1978, introduced a Beveridgian system based on universalism, comprehensiveness and equity. In the 1990s, a series of reforms led to the progressive regionalisation of the NHS, giving regions political, administrative and financial responsibility for the organisation and delivery of health care; public health expenditure in 2013 amounts to 6.8% of GDP, whereas total health expenditure accounts for 8.8% of GDP (just below the OECD average - OECD, 2015).

Primary care is organised on a list system, and family physicians are independent contractors with the NHS acting as gatekeepers. Access is free of charge and registration with family physicians is compulsory². Capitation is the main income source for GPs, and is negotiated at national level. This is then topped up by a variable and an additional part. The variable part consists of fee-for-service payments for specific types of treatment (minor surgery, post-surgery follow-up, etc...). The additional part rewards high quality, appropriate care or participation in organisational innovations (Lo Scalzo *et al.*, 2009; Fiorentini *et al.*, 2011), but GPs are not directly rewarded for containing ED admissions.

Cooperation among GPs has been encouraged since the mid-1990s, designed to improve the quality and comprehensiveness of care offered by professional partnerships (Shaw and Meads, 2012). Indeed, individual practices are believed to be less effective in providing high quality treatment, due to their limited ability to offer continuity of care (Fantini *et al.*, 2012), and of acquiring diagnostic equipment. Groups are composed of between 3 and 10 physicians. However, even if their GP is part of a group, patients continue to be registered with a specific physician, who is responsible for them. GPs can hire receptionists, nurses or administrative personnel. While in individual practices, the presence of a nurse is infrequent, groups more often acquire skilled supporting staff for their activity, since the GPs can share the corresponding costs.³

3.2 Primary care and inappropriate ED utilisation

Emergency services in Italy are characterised by a triage coding system for those ED attendances not followed by hospitalisation. This four-level triage system is based on an acuity scale, with *white codes* representing the lowest priority and classified as inappropriate (Levaggi and Montefiori, 2013). The assignment of the triage code is strictly regulated: at their arrival, patients are initially assessed based on an interview regarding the main symptoms and on a visual inspection aimed at identifying need for

² Each GP has a maximum list size of 1,500 patients. Citizens can freely change their GP, although the turnover rate is very low and is mostly due to changes in residence rather than any dissatisfaction with the GP (Italian Ministry of Health, 2010).

³ Specialists do not participate in these teams, and there are no nurse-led practices either. Therefore, the presence of nursing staff signals a fairly advanced organisational model.

immediate treatment; then, following a more detailed physical examination and measurement of vital parameters (airway, breathing, circulation and disability), the initial triage code is assigned. However, the patient is constantly monitored and his/her code can be changed if new information or developments emerge. Finally, at the time of the patient's discharge, the entire episode is subject to a final evaluation under medical supervision. The final triage code is reported in the ED's discharge records⁴.

The reliability, validity and robustness of the Italian Triage Scale (I-4L) have been recently investigated: Parenti *et al.* (2010) assess its reliability and predictive validity in an urban hospital in Emilia-Romagna⁵. The study uses paper triage scenarios submitted to different raters, whose evaluation is then repeated six months later. The same scenarios are also evaluated by qualified experts, and their assessment is taken as a reference standard. The findings indicate that the I-4L displays a good inter-rater and intra-rater reliability. Moreover, it validly predicts the triage code taken as the reference standard, and the risk of hospitalisation. Cremonesi *et al.* (2015) investigate whether the triage system offers a timely response to more urgent conditions regardless of ED congestion. They find that average waiting times between arrival and initial consultation are independent of congestion in the case of patients in a severe condition, but that they increase significantly for white codes. Despite this, the Italian NHS is still faced with the intense, often inappropriate, use of its EDs. As regards Emilia-Romagna, an attempted solution to the problem is based on the view that improved accessibility to primary care could meet part of the demand from patients who would otherwise self-refer to the ED. Since the national contract establishes limited requirements in terms of practice opening hours, the policy focuses on extending weekday hours during the daytime. However, there are still no empirical findings regarding the effectiveness of these measures, and we contribute to fill this gap.

The policy was launched in 2007 and involved the coordination of opening hours to ensure a daily coverage by the group of up to 12 hours. This measure targeted GPs who were already members of a group, rather than individual practices. Since the coordination of opening hours entails close cooperation, established formal connections among physicians were recommended. In the light of this, our study evaluates the impact of the policy among GPs belonging to groups. However, since patients are registered with a specific physician and not with a group, patients' use of hospital services can be univocally associated to a specific GP. The main advantage of being registered with a GP participating in the extension program is the increased possibility of obtaining initial care from another member of the group if one's own GP is not available, should the patient's conditions require a prompt consultation. Top-up

⁴ Interestingly, in order to promote uniform standards the guidelines recommend that the staff in charge of the triage rotate, thus avoiding situations where staff members act as the only raters.

⁵ Reliability refers to consistency in evaluation across raters and over time, and it measures the degree of agreement beyond what would be observed by chance. Predictive validity refers to the triage system's capacity to correctly predict reference standard measures or related outcomes such as hospitalisation, mortality, length of stay and resource use.

payments for the extension are established by the collective contract for primary care: participating GPs receive an additional fee of 4 € per patient if the group ensures 10 hours coverage, and an additional 1 € per patient for each additional hour up to a maximum of 12 hours coverage.⁶

4. Data

4.1 Data sources

The GP is our unit of observation. The data cover all primary care physicians working in the Emilia-Romagna Region during the period 2008-2010⁷, and offer information on patients' attendances of regional EDs not followed by hospitalisation. ED records contain information about the treatment received, and this can be used to establish whether the episodes consisted of a general visit only, or whether they also entailed subsequent diagnostic or specialist examinations. We consider two measures of ED attendance: inappropriate visits according to the triage system (*white codes*), and the sum of white codes and those episodes which, although not coded as such, consisted in a general visit with no diagnostic or specialist follow-up (*potentially inappropriate visits*)⁸.

Our study population consists of GPs working in groups, as these are the ones eligible for the program. The variable of main policy interest is a dummy taking a value of 1 if the GP works in a group subscribing to the program for the coordination of practice opening hours, or of 0 otherwise. Other controls include the GP's gender and seniority, a dummy for nursing staff, a dummy for practices in (totally or partially) mountainous municipalities, and the distance between the premises and the closest ED. Since citizens usually enrol with a physician located close to their home, distance can be taken as a proxy for geographical accessibility to the ED. We specify list size as an offset variable to account for heterogeneous exposure to the probability of ED admissions (Chauhan *et al.*, 2012). We also consider the share of male patients, the patients' average age, and the share of foreign patients in the list. As we lack information other than age and gender with which to adjust for patients' case-mix, we include hospital admission rates as a proxy for list severity⁹. Finally, we control for territorial effects by including dummies for the regional sub-areas responsible for coordinating health care policies (Iezzi *et al.*, 2014)¹⁰.

⁶ A GP with an average size list (1,272 patients) receives an extra compensation amounting to 5,088 € for a 10-hour extension, and this rises to 7,632 € for a 12-hour extension. Groups of five physicians (median group size) thus receive an extra payment of 25,440 € in the first case, and of 38,160 € in the second case.

⁷ The total number of GPs was 3,280 in 2008, 3,271 in 2009, and 3,282 in 2010.

⁸ This indicator is also used, together with white codes, by health authorities to monitor the regional emergency system. Levaggi and Montefiori (2013) analyse cost data regarding 1011 ED patients, and show that laboratory tests and examinations are the key drivers of cost. Given this, they pool white coded episodes and non-white episodes which received no lab tests, in a homogenous category in terms of the risk of inappropriateness and resource use.

⁹ Hospitalisations do not overlap with our dependent variables, as the latter consider those ED visits that are not followed by hospitalisation.

¹⁰ The Emilia-Romagna health care system is divided into three areas representing the main institutional units in charge of planning and coordinating hospital activity, including ED services. Each area is designed to ensure self-sufficiency of hospital production capacity and to minimise patient outflows.

Those physicians who are part of a group, and those who are not, are fairly similar in terms of observables, although those working in groups have a slightly larger list size. Comparing ED admissions per 1000 patients, we find that the average frequency of white codes is highest for GPs not working in groups (70), followed by GPs in groups that do not take part in the extension program (55), and lowest by those participating in the program (44). Similar rankings emerge for potentially inappropriate admissions (151; 126; 96 per 1000 patients respectively).

4.2 Estimating sample and descriptive statistics

The estimating sample is composed of those GPs working in groups who had more than 300 registered patients each during the period 2008-2010¹¹. The panel covers 1,069, 1,075 and 1,075 GPs over the three years respectively. As for the extension, less than a quarter of GPs participated in the program (23 %) in 2008. This increased to 30% in 2009 and remained stable in 2010 (31 %).

Our pooled sample comprises 1,182,168 ED admissions, 221,010 of which are white codes (19% of the total), while 458,968 fall into the category of potentially inappropriate visits according to the aforementioned definition (39% of the total).

TABLE 1a

Table 1a displays the yearly frequency of ED visits per GP. The average number of attendances per GP for physicians in group practices, amounts to 367, while there are 69 white-coded cases and 143 potentially inappropriate attendances. When we consider those GPs who extended their opening hours compared to those who did not separately, the following differences emerge: 336 vs. 379 for total attendances, 53 vs. 75 for white codes, and 116 vs. 153 for potentially inappropriate visits, respectively. The last column in Table 1a shows the standard *t*-test for mean comparison: all differences are statistically significant.

TABLE 1b

Table 1b shows the descriptive statistics for the controls averaged over the period 2008-2010. About 73% of GPs are male, list size is around 1,271 patients and the share of foreign patients is in line with census data. Around a quarter of GPs employ nursing staff, while nearly 30% of the practices are in municipalities located in totally or partially mountainous areas.

5. Econometric Strategy

¹¹ The primary care agreement sets the threshold at which the Department of Health reserves the right to terminate a GP's contract, at 300 patients.

5.1 Latent heterogeneity and potential endogeneity of the policy variable

An empirical challenge stems from the unobserved heterogeneity not captured by the predictors included. Even after controlling for observed GP and list characteristics, there is still room for latent heterogeneity, which is problematic whenever correlated with the covariates. In fact, maximum likelihood estimators only provide consistent estimates under the exogeneity of the regressors. This correlation is a concern here, in particular for the participation dummy. As the groups of GPs can choose whether to extend opening hours, unobserved factors may affect, at one and the same time, GPs' propensity to participate in the extension program, and patients' frequency of attendance of EDs, thus leading to endogeneity of participation. This may depend on unobserved differences in practice style and list composition: for instance, GPs more committed to providing preventive care and effective counselling, may also be keener to take part in the extension program; GPs' propensity to participate may also reflect patients' needs. Since the nature of patient-physician interaction and the case-mix of the list are stable over time, the sources of latent heterogeneity can be considered almost time-invariant in the short-run.

Our strategy consists of two steps, both tackling latent heterogeneity and potential endogeneity. We first address the potential endogeneity of the policy in a count model for ED admissions, by adopting an instrumental variable approach. We then estimate panel models on data for the years 2008 to 2010, which explicitly account for time-invariant latent heterogeneity through the inclusion of GP-specific effects.

5.2 Instrumental variable estimation

5.2.1 Exclusion restrictions

For the Instrumental Variables (IVs) estimation to be viable, the IVs have to be correlated with the potentially endogenous variable, but they must not exert a direct effect on the outcome, thus being safely excludable from the outcome equation. The extension program relies on consolidated cooperative agreements among GPs, and two group characteristics emerge as good candidates for IVs: the number of physicians participating in the group, and the time (in years) since the GP joined the group.

The features of the Regional Healthcare System support this choice. When patients are affected by a non-deferrable condition, and their physician is unavailable, the chance of initial treatment in a primary care setting may act as a restraint to ED admissions. The extension of opening hours is a demanding task for a group's members, in terms of the coordination needed to ensure the collective daytime availability called for. This effort decreases in larger groups, since the burden is shared among a greater number of GPs. Thus we would expect the number of physicians in a group to positively affect participation in the extension policy. From the patient's perspective, the important factor is the availability of primary care in the event of non-deferrable need: such availability does not depend on the size of the group but on the

requirements all groups have to meet. Therefore, provided that the group as a whole guarantees a continuous service for the contracted 10-12 hours, the fact that a group consists of a larger or smaller number of physicians is not expected to directly affect patients' use of the ED.

Concerns about the validity of the instrument may derive from potential economies of scale in the quality of primary care, which may lead to a positive association between group size and patients' health, this correlation potentially affecting the use of hospital services. However, our outcomes do not measure patients' health status, but only represent patients' choice of (inappropriate) care settings. In fact, we focus on minor conditions not requiring hospital treatment and for which primary care would have been effective, if the patient had chosen to attend the practice. Therefore, the crucial feature deemed to affect inappropriate ED visits is the availability of a fellow doctor to substitute one's own. All groups participating in the program, regardless of the group's size, guarantee such availability over the 10-12 hours.

Similarly, physicians who have been working as a group for several years may find it easier to participate in the program, since their cooperative routines have been in place for longer than those of physicians who have recently set up a group. The link between patients' ED utilisation and the number of years their GP has spent in a group is assumed not to be a direct one, but only one that is correlated to the increased probability of participation in the program. Indeed, while the length of group cooperation may affect the interaction among GPs, thus favouring the establishment of closer professional links, the argument that the number of years spent in a group directly influences the rate of patients' ED attendance is not as strong.

The institutional arrangements of the Italian NHS establish that the relationship between GPs and patients is direct and personal even for those physicians in group practices. This implies that patients are required to go to the physician they are registered with unless the health problem is non-deferrable. Consequently, a patient sees a different member of the group only under a limited set of circumstances.

Moreover, the fellow doctor standing in for his/her colleague, varies according to the time of the day and the day of week. Therefore, repeated interaction with a specific member of the group other than one's own GP is rare, thus limiting the possibility of a GP establishing any familiarity with patients registered with a colleague, regardless of the moment when he/she first joined the group. This implies that group arrangements can be expected to directly affect interaction among GPs, but not the patient-physician relationship.

Looking at matters from a different perspective, we acknowledge that the IVs only capture features of a group's internal organisation and of a GP's professional history, but do not control for unobserved patient

heterogeneity. The inclusion of additional IVs better able to capture patients' characteristics would strengthen the analysis. However, data limitations prevent the inclusion of good proxies for such latent confounders.

The descriptive statistics for the instruments show that on average, the groups participating in the program are larger than those that do not extend their opening hours (5.58 vs. 4.77 members), thus supporting the hypothesis that larger groups are keener to take up the extension option. Furthermore, those GPs extending opening hours have been working in a group for 8.5 years on average, compared with 7.6 years in the case of those not subscribing to the program. These differences are statistically significant according to the *t-test* for mean comparison.

5.2.2 IV methods

Let y_i be the count variable of interest measuring ED attendances. We first assume y_i to be Poisson distributed with exponential conditional mean:

$$E(y_i|\mathbf{x}_i) = \mu_i = \mu_i(\mathbf{x}_i) = \exp(\mathbf{x}_i'\boldsymbol{\beta}) \quad (1)$$

with μ_i being also the conditional variance of y_i . Under standard assumptions, the parameters in (1) can be consistently and efficiently estimated by Maximum Likelihood (ML). To account for overdispersion and latent heterogeneity, a multiplicative random effect v_i can be included in the conditional mean, leading to the Negative Binomial (NB) model:

$$E(y_i|\mathbf{x}_i, v_i) = \mu_i(\mathbf{x}_i)v_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}) v_i. \quad (2)$$

The estimating model for the year 2009 is as follows:

$$y_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}) + \varepsilon_i = \exp(\beta_0 + \beta_1 extop_i + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \ln(listsize_i)) + \varepsilon_i \quad (3)$$

where the vector $\mathbf{x}_i = (x_1, x_2, \dots, x_p)$ includes the predictors presented in Table 1b, *extop* is the dummy for the extension of practice opening hours, ε_i is an idiosyncratic error and the list size in logarithms acts as offset variable, accounting for heterogeneous exposure, with the coefficient constrained to 1¹². The Poisson and NB ML estimators are only consistent if the regressors are strictly exogenous.

To accommodate the potential endogeneity of participation, we allow for the presence of unobservable factors by introducing a latent confounder x_{iu} in equation (3). The model thus becomes:

$$y_i = \exp(\beta_0 + \beta_1 extop_i + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \rho x_{iu} + \ln(listsize_i)) + \varepsilon_i \quad (4)$$

where *extop* is potentially correlated with x_{iu} .

¹² The number of patients represents, for each GP, the population at risk of attending the ED, and as such represents the *exposure* term. See Cowling et al. (2013), Gagnon et al. (2008).

Count data specifications constitute a particular challenge to the implementation of IV methods (e.g. Windmeijer and Santos Silva, 1997). The linear IV estimator (2SLS) suffers from severe drawbacks in this framework. Terza, Bradford and Dismuke (2008) show that conventional linear IV methods can lead to biased estimates in non-linear models. Terza, Basu and Rathouz (2008) suggest a two-stage residual inclusion (2SRI) strategy to tackle endogeneity in count/exponential models. We adopt this approach to estimate the model in (4) where the (binary) regressor *extop* is allowed to be endogenous. The 2SRI estimator yields consistent estimates in the presence of endogeneity, whilst allowing unobservable confounders to be correlated with the regressors, and requiring only weak structural assumptions regarding the data-generating process.

2SRI is a version of the control function approach, first developed for count data by Wooldridge (1997, 2010), and can be seen as an extension of the Hausman (1978) endogeneity test for nonlinear models. Indeed, exogeneity can be tested through an asymptotically-efficient Wald test on the coefficients of the first-stage residuals included in the second stage.

The main econometric issue to be addressed here is the potential correlation between *extop* and the unobservable factors captured by x_{iu} in (4). As the latent factors in x_{iu} are not observable, x_{iu} in (4) has to be replaced by a consistent estimate from the first stage. At the first stage, we estimate the following reduced form for the probability of extending opening hours:

$$\Pr(extop_i = 1) = g(\gamma_0 + \gamma_2 x_{i2} + \dots + \gamma_p x_{ip} + \delta_1 sizeprac + \delta_2 yearsgroup) + x_{iu} \quad (5)$$

where *sizeprac* and *yearsgroup* are the IVs, x_{iu} represents the unmeasured latent factors affecting both *extop* and y , and x_{i2}, \dots, x_{ip} are the additional covariates. For dichotomous indicators, the function g usually adopted is the probit or the logit. The residuals from equation (5), defined as \hat{x}_{iu} , provide a consistent estimate for x_{iu} and are included in the outcome equation at the second stage. There is no agreement in the literature on the appropriate specification of the residuals (Basu and Manning (2009), Garrido et al. (2012)) and, therefore, we consider alternative specifications of the residuals \hat{x}_{iu} . We first compute the *response* or raw residuals as:

$$\hat{x}_{iu} = extop_i - \Pr(extop_i = 1). \quad (6)$$

We also calculate the generalised residuals (Gourieroux et al., 1987) and the *deviance* residuals for the probit link, and the *Pearson* and *standardised Pearson* residuals for the logit. We then test for the joint relevance of the IVs by means of a χ^2 test. We also estimate equation (5) on the three-year panel 2008-2010, taking g as a linear function: in this case, the linear probability model (LPM) is estimated by OLS and conventional residuals are included in the second-stage equation. Table 2 reports the first-stage estimates.

In the second stage, the unobserved confounder x_{iu} is replaced by the first-stage residual \hat{x}_{iu} , as follows:

$$y_i = \exp(b_0 + b_1 extop_i + b_2 x_{iz} + \dots + b_p x_{ip} + \rho \hat{x}_{iu} + \ln(listsize_i)) + \epsilon_i \quad (7)$$

with *extop* among the regressors. We estimate equation (7) by NB MLE. The inclusion of the residuals from the reduced form in the second stage both provides a straightforward Wald exogeneity test within a nonlinear framework, and controls for the endogeneity of participation. A statistically significant ρ coefficient provides evidence of underlying unobserved factors affecting both opening and outcome variables. To account for potential nonlinearities, we also include the square of the residuals (Garrido et al. 2012). Finally, given that we include the residuals from the first stage, standard errors are bootstrapped.

In both steps, along with robust standard errors for the cross section, and clustered standard errors at the GP level for the panel, we also get standard errors clustered at the group practice level to account for potential cluster effects and for the participation dummy being a grouped regressor (Moulton, 1986, 1990; Bertrand et al., 2004). Tables 3a and 3b present the second-stage estimates.

5.3 Panel data analysis

Count panel data models allow latent individual heterogeneity to be straightforwardly controlled for by means of a multiplicative individual-specific time-invariant effect that enters the conditional mean and that can be assumed to be fixed or random. The dependent variable has the following exponential conditional mean:

$$E(y_{it} | \mathbf{x}_{it}, c_i) = \mu_{it} = c_i \exp(\mathbf{x}'_{it} \boldsymbol{\beta}) \quad (8)$$

where c_i is the heterogeneity term, and \mathbf{x}_{it} is a set of covariates possibly including time dummies (Cameron and Trivedi, 2015).

Different assumptions regarding the term c_i lead to the specification of alternative Poisson and NB panel data models, the most common being the random effects (RE) and fixed effects (FE) models.

RE models assume that the individual effects c_i are unobservable i.i.d. random terms uncorrelated with the included predictors \mathbf{x}_{it} . In the Poisson RE, c_i can be assumed Gamma or log-Normal distributed. The model is estimated by conditional ML after integrating out c_i from the conditional mean. The Poisson RE estimator, proposed by Hausman et al. (1984), is consistent under the assumption that c_i is not correlated with the covariates. The NB RE model adds a latent heterogeneity term to the conditional mean to explicitly account for overdispersion; this can be seen as introducing heterogeneity a second time (Greene, 2007). In the NB RE model, both heterogeneity terms are assumed to be Gamma-distributed with parameters allowed to vary across individuals. The coefficients are estimated by the conditional NB RE ML estimator, also proposed by Hausman et al. (1984).

FE models allow c_i to be correlated with the predictors. In order to consistently estimate the parameters in β , the regressors have to be assumed strictly exogenous, conditional on c_i . Both the Poisson and the NB model can be specified as FE, the latter including a second latent heterogeneity term to account for overdispersion. Several estimation strategies can be adopted for the Poisson FE¹³. We jointly estimate β and c_i by unconditional ML, including a full set of individual dummy variables. Although this strategy is computationally intensive, the Poisson FE does not suffer from the incidental parameter problem in short panels (Lancaster, 2000). The extension of the FE specification to the NB model is more problematic. Latent heterogeneity enters the model twice, and the conditional ML estimator of Hausman et al. (1984) requires restrictive assumptions on the relationship between c_i and the overdispersion parameter (Guimarães, 2008). Moreover, the conditional NB FE estimator results as not being a true fixed effect (Allison and Waterman, 2002), and allows the coefficients of time-invariant regressors to be identified. The unconditional estimator is not suitable here, as it suffers from an incidental parameter problem in short panels (Cameron and Trivedi, 2015).

On account of the drawbacks of the NB FE, an alternative is the conditionally correlated random effect (CCRE) model, which generalises the RE model and combines features of both the FE and RE formulations. Extending the approach first developed by Mundlak (1978) and Chamberlain (1982) for linear panels, the CCRE model relaxes the assumption of non-correlation between individual heterogeneity and the regressors, and instead assumes c_i to be conditionally correlated to the covariates. This formulation specifies c_i as an exponential function of the time-averages of the regressors, as follows:

$$c_i = \exp(\bar{x}_i' \gamma + \varepsilon_i) \quad (9)$$

where \bar{x}_i is the vector of time-averaged covariates and ε_i is the component of individual heterogeneity uncorrelated with the regressors. The resulting CCRE model with time-averages as additional regressors

$$E(y_{it} | x_{it}, \bar{x}_i, c_i) = \exp(x_{it}' \beta + \bar{x}_i' \gamma + \varepsilon_i) \quad (10)$$

can be estimated as a RE model by ML (Trivedi, 2014).

The choice between FE and RE models relies on specific assumptions about unobserved heterogeneity, with CCRE being a compromise between the two. RE estimators are only consistent if c_i is uncorrelated with the regressors. If such a restriction holds, the RE is more efficient. On the other hand, the FE estimator is more likely to be consistent and is a valid alternative to IV approaches to addressing latent heterogeneity whenever the regressors are correlated with time-invariant unobservable factors only. On

¹³ β and c_i can be jointly estimated by ML. Alternatively, a conditional ML estimator can be used to estimate β conditional on sufficient statistics for c_i ; likewise, one can exploit a moment-based estimator that eliminates the fixed effects from the conditional mean through a quasi-differencing transformation. Lancaster (2000) shows that for the Poisson FE model, the conditional and unconditional ML estimators give numerically equivalent estimates. See Cameron and Trivedi (2013).

the downside, the coefficients of predictors with little time variability might be identified less precisely and display larger standard errors than those of the RE specification.

As for the choice between Poisson and NB, the Poisson model is reasonably robust to overdispersion, even more so as cluster-robust estimation of the variance-covariance matrix is feasible for all the specifications considered. The NB model, instead, relies on specific distributional assumptions for the two sources of heterogeneity.

Tables 4 presents estimates for the panel model comprising the regressors in Table 1b, year dummies and $\ln(listsize_{it})$ as the offset variable. We exclude from the Poisson FE specification the time-invariant regressors and also GP's seniority because of its collinearity with time dummies. We estimate Poisson FE and RE, NB RE and CCRE models, the latter also including time averages among the predictors. We obtain panel-robust standard errors to account for serial correlation; in addition, we get standard errors clustered at the group practice level, to account for the policy variable being a grouped regressor (Moulton, 1986, 1990; Cameron and Miller, 2015).

5.4 Discussion of the empirical strategy

The two strategies adopted to tackle latent heterogeneity come out as being complementary. On the one hand, the control function approach may entail several inconveniences. The sample is relatively small and includes grouped regressors, possibly implying an overfitting of endogenous variables at the second stage (Cameron and Trivedi, 2013). Moreover, if the IVs are weak, or not strictly exogenous, or only capture part of the heterogeneity, the estimates might result as inconsistent. On the other hand, panel data models allow us to explicitly account for latent time-invariant heterogeneity. By permitting the correlation between latent factors and the regressors, the fixed effects specification is a valid alternative for consistently estimating causal effects, provided that the predictors are correlated only with the effects c_i . The sources of latent heterogeneity possibly affecting the participation in the program include physicians' attitudes and patients' characteristics: these factors are persistent over time, at least in the short run. Since physician-specific effects allow us to control for time-invariant heterogeneity, and given that the main sources of potential correlation with predictors are stable over time, the FE specification is expected to account for most, if not all, of the latent heterogeneity. On the down side, many of the included predictors show little time variation, and this may impact the precision of the FE estimates on a short panel.

6. Results

6.1 IV estimates

To implement the 2SRI strategy discussed previously, we first estimate the reduced form in equation (5) for the probability of extending opening hours. The two IVs are *group size*, that is the number of

physicians participating in each group, and the *number of years* each GP has been working in the group. At the second stage, we include the residuals from the first stage in the outcome equation, together with the full set of controls.¹⁴ To allow for potential nonlinearities we also include the squared residuals.

Table 2 displays the Probit and Logit estimates of equation (5) for 2009, and the LPM estimates for the panel.

TABLE 2

The first-stage results are consistent across specifications. The two IVs are significant predictors of the participation in the program, and their sign is in line with expectations. The larger the number of GPs in the group, the greater the probability the group opts for the extension of opening hours. Similarly, having worked in association for a longer period, increases the probability of GPs extending their practice's opening hours. The χ^2 -test and the F-test on the IV relevance confirm the instruments as jointly relevant and good predictors of the extension.

Tables 3a and 3b show the NB estimates¹⁵ of equation (7) which includes the first-stage residuals (Response, Generalized, Deviance for the Probit model; Response, Pearson and Standardised Pearson for the Logit model, OLS residuals for the LPM). The second-stage standard errors are bootstrapped, as we include a generated regressor from the first stage.

TABLES 3a and 3b

For white codes, the second-stage results suggest the possible endogeneity of program participation, whereas the estimates of the coefficient ρ for potentially inappropriate admissions are not significant. With respect to the square of the residuals, we find some evidence of nonlinear effects. Not surprisingly, the estimates for ρ are sensitive to the type of residuals considered. The dummy for extended opening hours displays a negative and significant coefficient, in most cases at 5%, for both dependent variables. When clustering the standard errors at the group level, in the first stage the coefficients of the IVs generally remain highly significant, proving them to be strongly relevant. In the second stage, the standard errors are clustered to deal with within-group correlation, and are bootstrapped to account for the inclusion of generated regressors. The coefficient of the policy variable is no longer significant, nor are those of first-stage residuals. In general, the estimated effect of participation in this framework comes out as weakly significant and further impaired, once the standard error have been clustered.

¹⁴ For the sake of conciseness, we have omitted the estimates for the covariates from the tables, and have postponed any discussion of their role to the panel analysis. The complete set of results is available upon request.

¹⁵ We also estimate the Poisson model and find strong evidence against equidispersion: the deviance and Pearson χ^2 tests reject the null of good fit; the high deviance and the Pearson dispersion statistic (well above 1) point to overdispersion. In the presence of overdispersion, the NB estimator is more efficient and it better accommodates the data (Hilbe and Greene, 2007), making it our preferred specification. The LM test for Poisson against NB rejects equidispersion (Hilbe (2011)). The deviance and the Pearson dispersion statistics for the NB model approximate to 1, confirming the improved fit.

6.2. Panel data estimates

This section discusses the Poisson FE, Poisson RE and NB RE estimates of the panel data model in equation (8), and those of the NB CCRE model in equation (10). As long as the included predictors are only correlated with latent confounders that are fairly stable over time, the FE model comes out as a valid alternative to 2SRI for the purpose of addressing the potential endogeneity of participation, and is therefore our benchmark specification. The standard errors are clustered first at GP level, in order to account for serial correlation, then at group level to control for within-group correlation and for the inclusion of grouped regressors. The results are shown in Table 4.

TABLE 4

The previous findings are confirmed and further corroborated: extended opening hours reduce inappropriate ED attendances. The estimated effects follow the same direction seen in the 2SRI, as the participation dummy displays a negative and significant coefficient for both dependent variables.

The estimated effects of opening-hour extension are generally greater for white codes than for potentially inappropriate admissions. Unlike in the IV case, despite the fact that practice-clustered standard errors are higher than GP-clustered ones, the coefficients of the extension program are significant at the 5% level, the only exception being the NB CCRE for potentially inappropriate visits, for which the coefficient is significant at the 10% level.

As for the covariates, the controls affect ED utilisation in a similar way across specifications. Physicians with longer professional experience record a lower number of ED visits, and male GPs are associated with lower expected counts. As for nursing staff, practices employing nurses perform significantly better in preventing white-coded attendances. Accessibility of the ED also plays an important role: the coefficient of distance is always negative and significant. The coefficient of practices located in mountainous municipalities, when statistically significant, is positive. This seems to indicate that, as patients in such areas face higher travelling costs for seeing their GP than urban residents do, once we control for the distance between the practice and the hospital, the frequency of ED utilisation increases. Finally, statistically significant effects emerge for list characteristics. ED visits are in most cases negatively correlated with average age, and positively correlated with the proportion of males. As the analysis is based on GP-level information, drawing conclusions about the use of the ED by specific groups of patients may lead to an ecological fallacy (Greenland, 2001). Although caution is required, our findings are nevertheless consistent with the conjecture that a high proportion of patients who may enjoy

more regular access to their GPs because of the lower opportunity cost of time, such as the elderly and females, reduces ED use.

To assess the economic relevance of the impact of the program, we present the Incidence Rate Ratios (IRRs), the average marginal effects (AMEs) and the marginal effects at the means (MEMs) for the policy variable (Table 6). All the estimates are significant at the 5% level.

TABLE 6

The magnitude of the drop in ED visits ranges from between 9.6% and 15.2% for white codes, and from between 6.5% and 10.1% for potentially inappropriate visits, thus constituting sizeable reductions in inappropriate ED utilisation associated with the extension program. The fact that we record a relatively larger reduction in ED admissions for white codes than for potentially inappropriate visits, is consistent with the less restrictive definition of inappropriateness of the latter outcome indicator. The AMEs calculated for *white codes* signal that participation in the program leads to an estimated average annual reduction of between 6.7 and 11.3 episodes per GP. For potentially inappropriate visits, the extension results in an average estimated reduction of 10-14 episodes. Similar patterns are confirmed by the marginal effects at means.

Finally, it is interesting to assess the potential savings from estimated reductions in inappropriate ED attendances. A full cost-benefit analysis is beyond the scope of the present work, as the benefits associated with improved primary care accessibility extend beyond ED utilisation. Moreover, this would require detailed data on the costs of ED services and on the workload created by patients receiving different triage codes. Despite these limitations, we can use a study by the Italian Ministry of Health (2007), that estimated an average cost of 226 € per white code in six hospitals within the Lazio region. In our case, such a figure would imply savings per GP ranging from around 1515 € in the most conservative estimate, up to 2560 € in the most favourable scenario. Overall, these figures are lower than the financial incentives received by a participating GP with average sample characteristics (see Section 3.2). This would suggest that the reduction in inappropriate ED utilisation alone does not seem to generate sufficient savings to fully compensate for the cost of the program. However, this in itself does not point to a negative conclusion regarding the cost-effectiveness of the policy. Indeed, the improvements in care provision achieved through the extension program may involve dimensions not accounted for by our outcomes. Moreover, a proper evaluation would require information on the actual costs of the ED services in Emilia-Romagna, which is not available here.

6.3 Robustness checks

To assess the robustness of the results of both IV and panel models, we also estimate a linear panel specification with the ratio of ED admissions to the list size as an outcome variable. The model is estimated in the linear pooled, FE, FE IV, RE and RE IV formulations, exploiting the same IVs as in 2SRI. We refer the reader to the online Appendix for the full set of results. Overall, the findings from the count data analysis are also confirmed for the linear model.

To gain further insights into the correlation between practice opening hours and ED attendances, we run a number of additional checks. Firstly, we consider, as dependent variables, conditions that should not be influenced by the accessibility of primary care (placebo test). Secondly, we examine ED visits on weekdays and at weekends, separately. Finally, we evaluate the sensitivity of our findings to the presence of controls for case-mix and to the characteristics of the other members of the group.

The placebo tests support a genuine identification of the impact of the policy, while we find no evidence of a “weekend effect”, and the coefficient for the extension program is not substantially affected by the changes in the set of covariates. These results confirm the robustness of our main findings. For the sake of brevity, these robustness checks are examined in the online Appendix.

7. Concluding remarks

Improving accessibility to primary care has been a recurring headline topic on the health policy agenda. This is an important target for the Emilia-Romagna Region of Italy as well, where policymakers have incentivised the extension of the opening hours of GPs working in groups, in order to ensure a daily coverage of up to 12 hours. The present work’s main goal is to assess the impact of this policy in terms of the improved appropriateness of the use of emergency services. To this end, we consider two measures of those ED visits not followed by hospitalisation, aggregated at the GP level, and focus on cases that could effectively be treated in a primary care setting.

Our empirical approach consists of two steps. Since subscribing to the extension program is voluntary, we estimate a count model for ED admissions by adopting an instrumental variable strategy. Using two relevant instruments for GPs’ participation in the program and using a two-stage residual inclusion (2SRI) approach, we test for the potential endogeneity of the policy variable and we account for the possible bias due to GPs’ non-random participation. The 2SRI estimates suggest the possible endogeneity of program participation when dealing with white code visits.

To further assess the effectiveness of the policy, we extend the analysis to a three-year panel. This allows us to exploit variation over time in health outcomes and program participation, and to account for individual time-invariant heterogeneity that could be the source of endogeneity. In the panel analysis, we

take the Poisson fixed-effect as the benchmark specification, since it requires mild assumptions and allows for a correlation between individual heterogeneity and the regressors.

The evidence is consistent across different estimation strategies and for both measures of avoidable ED utilisation, in terms of the sign and, in most cases, the significance of the estimated coefficients. The estimated effects are stronger in the panel data analysis, once we have controlled for latent heterogeneity. Overall, our findings support the hypothesis that increasing practices' daytime opening hours permits a reduction in ED utilisation. Based on the results of the panel data analysis, the expected reduction in white codes is estimated at between 10 and 15%.

References

- Afilalo, J., Marinovich, A., Afilalo M., et al. (2004) Nonurgent emergency department patient characteristics and barriers to primary care. *Academy of Emergency Medicine*, **11**, 1302-10.
- Agarwal, S., Banerjee, J., Baker, R. et al. (2012) Potentially avoidable emergency department attendance: interview study of patients' reasons for attendance. *Emergency Medicine Journal*, **29**, e3.
- Allison, P. and Waterman, R. (2002) Fixed effects negative binomial regression models. *Sociological Methodology*, **32**, 247-65.
- Ansari, Z., Carson, N., Serraglio, A., Barbetti, T., and Cicuttini, F. (2002) The Victorian Ambulatory Care Sensitive Conditions study: reducing demand on hospital services in Victoria. *Australian Health Review*, **25**, 71-77.
- Basu, A. and Manning, W.G. (2009) Issues for the Next Generation of Health Care Cost Analysis. *Medical Care*, **47**, s109-14.
- Bertrand, M., Duflo, E. and Mullainathan, S. (2004) How Much Should we Trust Differences-in-Differences Estimates. *Quarterly Journal of Economics*, **119**, 249-275.
- Bianco, A., Pileggi, C. and Angelillo, I.F. (2003) Non-urgent visits to a hospital emergency department in Italy. *Public Health*, **117**, 250-55.
- Booth, G. L., Hux, J.E., Fang, J. and Chan, B.T. (2005) Time trends and geographic disparities in acute complications of diabetes in Ontario, Canada. *Diabetes Care*, **28**, 1045-1050.
- Calderón-Larranaga, A., Carney, L., Soljak, M. et al. (2011) Association of population and primary healthcare factors with hospital admissions rates for chronic obstructive pulmonary disease in England. *Thorax*, **66**, 191-96.

- Cameron, A.C. and Miller, D. L. (2015) A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, **50**, 317-73.
- Cameron, A.C. and Trivedi, P.K. (2013) *Regression Analysis of Count Data* (2nd ed). Econometric Society Monograph No.53, Cambridge University Press.
- Cameron, A.C. and Trivedi, P.K. (2015) Count Panel Data. In B. Baltagi (ed) *Oxford Handbook of Panel Data*, Oxford University Press, 233-56.
- Chamberlain, G. (1982) Multivariate Regression Model for Panel Data, *Journal of Econometrics*, **18**, 5–46.
- Chan, C.L., Lin, W., Yang, N.P. and Huang, H.T. (2013) The association between the availability of ambulatory care and non-emergency treatment in emergency medicine departments: A comprehensive and nationwide validation. *Health Policy*, **110**, 271-279.
- Chauhan, M., Bankart, M. J., Labeit, A. and Baker, R. (2012) Characteristics of general practices associated with numbers of elective admissions. *Journal of Public Health*, **34**, 584-590.
- Cowling, T.E., Cecil, E.V., Soljak, M.A., Tayu Lee, J., Millett, C., Majeed, A., Wachter, R.M. and Harris, M.J. (2013) Access to primary care and visits to Emergency Departments in England: A cross-sectional, population based study, *Plos One*, **8**, 1-6.
- Cremonesi, P., di Bella, E., Montefiori, M. and Persico, L. (2015) The robustness and effectiveness of the triage system at times of overcrowding and the extra costs due to inappropriate use of emergency departments. *Applied Health Economics and Health Policy*, **13**, 507-514.
- Durand, A.C., Gentile, S., Devictor, B. et al. (2011) ED patients: how nonurgent are they? Systematic review of the emergency medicine literature. *American Journal of Emergency Medicine*, **29**, 333-345.
- Dusheiko, M., Gravelle, H., Martin, S., Rice, N. and Smith, P. (2011) Does better disease management in primary care reduce hospital costs? Evidence from English primary care. *Journal of Health Economics*, **30**, 919-932.
- Elley, C.R., Randall, P.J., Bratt, D. et al. (2007) Can primary care patients be identified within an emergency department workload? *New Zealand Medical Journal*, **120**, U2583.
- Fantini, M.P., Compagni, A., Rucci, P., Mimmi, S. and Longo, F. (2012) General practitioners' adherence to evidence-based guidelines: A multilevel analysis. *Health Care Managing Review*, **37**, 67–76.
- Feachem, R.G., Sekhri, N.K. and White, K.L. (2002) Getting more for their dollar: a comparison of the NHS with California's Kaiser Permanente. *British Medical Journal*, **324**, 135-41.

- Fiorentini, G., Iezzi, E., Lippi Bruni, M. and Ugolini, C. (2011) Incentives in primary care and their impact on potentially avoidable hospital admissions. *European Journal of Health Economics*, **12**, 297-309.
- Flores-Mateo, G., Violan-Fors, C., Carrillo-Santistevé, P., Peiro, S. and Argimon, J.M. (2012) Effectiveness of organizational interventions to reduce emergency department utilization: a systematic review. *PLoS One*; 7(5):e35903.
- Gagnon, D.R., Doron-LaMarca, S., Bell, M., O'Farrell, T.J., Taft, C.T. (2008) Poisson regression for modeling count and frequency outcomes in trauma research. *Journal of Traumatic Stress*, **21**, 448-454.
- GAO, General Accounting Office (2009) *Hospital emergency departments*. Publication 09-347, Washington DC.
- Garrido, M.M., Deb, P., Burgess, J.F. and Penrod, J.D. (2012) Choosing Models for Health Care Cost Analysis: Issues of Nonlinearity and Endogeneity. *Health Services Research*, **47**, 2377-97.
- Gourieroux, C., Monfort, A., Renault, E. and Trognon, A. (1987) Generalized residuals. *Journal of Econometrics*, **34**, 5-32.
- Greene, W. (2007) Functional Form and Heterogeneity in Models for Count Data. Foundations and Trends® in *Econometrics*, **1**, 113-218.
- Greenland, S. (2001) Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *International Journal of Epidemiology*, **30**, 1341-1350.
- Guimarães, P. (2008) The fixed effects negative binomial model revisited. *Economics Letters*, **99**, 63-6.
- Hansagi, H., Carlsson, B., Olsson, M. and Edhag, O. (1987) Trial of a method of reducing inappropriate demands on a hospital emergency department. *Public Health*, **10**, 99-105.
- Harris, M.J., Patel, B. and Bowen, S. (2011) Primary care access and its relationship with emergency department utilisation. *British Journal of General Practice*, **61**, e787-93.
- Harrison, M., Dusheiko M. Sutton, M. and Gravelle, H. (2014) Effect of a national primary care pay for performance scheme on emergency hospital admissions for ambulatory care sensitive conditions: controlled longitudinal study, *British Medical Journal*, **349**, g6423.
- Hausman, J.A. (1978) Specification Tests in Econometrics. *Econometrica*, **46**, 1251-71.
- Hausman, J.A., Hall, B. and Griliches, Z. (1984) Econometric Models for Count Data with an Application to the Patents-R&D Relationship. *Econometrica*, **52**, 909-38.
- Hilbe, J.M. (2011) Negative Binomial Regression. Cambridge University Press, 2nd edition.

- Hilbe, J.M. and Greene, W.H. (2007) Count Response Regression Models. In A. Chaudhury, T.C. Christofides, C.R. Rao *Handbook of Statistics*, Elsevier, **7**, 210-52.
- Iezzi, E., Lippi Bruni, M. and Ugolini, C. (2014) The role of GP's compensation schemes in diabetes care: Evidence from panel data. *Journal of Health Economics*, **34**, 104–120
- Italian Ministry of Health (2007) Pronto Soccorso e sistema 118. Proposta metodologica per la valutazione dei costi dell'emergenza (in Italian). Progetto Mattoni SSN.
- Italian Ministry of Health (2010) *Cittadini e salute: La soddisfazione degli Italiani per la sanità* (in Italian). Quaderni del Ministero della Salute, n.5.
- Lancaster, T. (2000) The incidental parameters problem since 1948. *Journal of Econometrics*, **95**, 391–414.
- Lang, T., David, A., Diakité, B., Agay, E., et al. (1996) Non-urgent care in the hospital medical emergency department in France: how much and which health needs does it reflect? *Journal of Epidemiology and Community Health*, **50**, 45-62.
- Lee, A., Lau, F.L., Hazlett, C.B. et al. (2000) Factors associated with non-urgent utilization of Accident and Emergency services: a case-control study in Hong Kong. *Social Science and Medicine*, **51**, 1075-85.
- Levaggi, R. and Montefiori, M. (2013) Definition of a prospective payment system to reimburse emergency department. *BMC Health Services Research*, **13**, 409.
- Lo Scalzo, A., Donatini, A., Orzella, L. et al. (2009) *Health System in Transition-Italy*. European Observatory on Health Systems and Policies, **11** (6).
- Lowe, R.A., Localio, A.R., Schwarx, D.F. et al. (2005) Association between primary care practice characteristics and emergency department use in a Medicaid managed care organization. *Medical Care*, **3**, 792-800.
- Moll van Charante, E.P., van Steewijk-Opdam, P. and Bindels, P. (2007) Out-of-hours demand for GP care and emergency services: patients' choice and referrals by general practitioners and ambulance services. *BMC Family Practice*, **8**, 46.
- Moulton, B.R. (1986) Random Group Effects and the Precision of Regression Estimates. *Journal of Econometrics*, **32**, 385-97.
- Moulton, B.R. (1990) An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units. *Review of Economics and Statistics*, **72**, 334-38.
- Mundlak Y. (1978) On the pooling of time series and cross sectional data, *Econometrica*, **46**, 69-85.

OECD (2015) *OECD Health Statistics 2015*. Permanent URL: <http://www.oecd.org/health/healthdata>.

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

Pines, J.M. et al. (22 authors) (2011) International perspectives on emergency department crowding. *Academy of Emergency Medicine*, **18**, 1358-70.

Puig-Junoy, J., Saez, M. and Martínez-García, E. (1998) Why do patients prefer hospital emergency visits? A nested multinomial logit analysis for patient-initiated contacts. *Health Care Management Science*, **1**, 39-52.

Roberts, E. and Mays, N. (1998) Can primary care and community-based models of emergency care substitute for the hospital A&E department? *Health Policy*, **44**, 191-214.

Sabik, L. M., and Gandhi, S.O. (2015) Copayments and Emergency Department Use Among Adult Medicaid Enrollees. *Health Economics*., doi: [10.1002/hec.3164](https://doi.org/10.1002/hec.3164).

Saxena, S., George, J., Barber, J. et al. (2006) Association of population and practice factors with potentially avoidable admission rates for chronic diseases in London. *Journal of the Royal Society of Medicine*, **99**, 81-89.

Sempere-Selva, T., Peirò, S., Sendra-Pina, P., Martinez-Espin, C. and Lòpez-Aguilera, I. (2001) Inappropriate use of an accident and emergency department: magnitude, associate factors and reasons. *Annals of Emergency Medicine*, **37**, 568-79.

Shah, N.M., Shah, M.A. and Behbehani, J. (1996) Predictors of non-urgent utilization of hospital emergency services in Kuwait. *Social Science and Medicine*, **42**, 1313-23.

Shaw, S.E. and Meads, G. (2012) Extending primary care: potential learning from Italy. *Primary Health Care Research & Development*, **13**, 289-93.

Terza, J.V., Basu, A. and Rathouz, P.J. (2008) Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modelling. *Journal of Health Economics*, **27**, 531-43.

Terza, J.V., Bradford, W.D. and Dismuke, C.E. (2008) The use of linear instrumental variables methods in health services research and health economics: a cautionary note. *Health Services Research*, **43**, 1102-20.

Thompson, C., Hayhurst, C. and Boyle, A. (2010) How have changes to out-of-hours primary care services since 2004 affected emergency department attendances at a UK District General Hospital? A longitudinal study. *Emergency Medicine*, **27**, 22-25.

Thompson, M.I., Lasserson, D., McCann, L., Thompson, M., Heneghan, C. and Lambert E. (2013) Suitability of emergency department attenders to be assessed in primary care: survey of general practitioner agreement in a random sample of triage records analysed in a service evaluation project. *British Medical Journal Open*, 3(12): e003612.

Trivedi, P.K. (2014) Models of count data. In A.J. Culyer (ed) *Encyclopedia of Health Economics*, Elsevier, 2: 306-11.

Windmeijer, F.A.G. and Santos Silva, J.M.C. (1997) Endogeneity in count data models: an application to demand for health care. *Journal of Applied Econometrics*, **12**, 281-94.

Wooldridge, J.M. (1997) *Quasi-likelihood methods for count data*, in Pesaran H, Schmidt P, (eds) *Handbook of Applied Econometrics. Volume II: Microeconometrics*, Blackwell Publishers Ltd, Malden, MA.

Wooldridge, J.M. (2010) *Econometric analysis of cross section and panel data* (2nd ed), MIT Press, Cambridge, MA.

Table 1a. Summary statistics for yearly ED visits per GP

	WHOLE SAMPLE		NO EXTENSION		EXTENSION		t-test
<i>ED admissions</i>	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	p-value
Year 2008							
White codes	63.39	(43.53)	66.34	(46.31)	53.56	(30.73)	0.000
Potentially inapprop. visits	143.88	(73.97)	151.98	(74.61)	116.95	(65.00)	0.000
Total ED visits	368.87	(132.96)	380.41	(134.45)	330.49	(120.40)	0.000
Observations	1069		822		247		
Year 2009							
White codes	71.68	(59.31)	79.83	(64.91)	52.79	(37.44)	0.000
Potentially inapprop. visits	151.48	(82.22)	164.77	(84.72)	120.67	(66.69)	0.000
Total ED visits	379.02	(128.87)	392.38	(131.39)	348.06	(117.30)	0.000
Observations	1075		751		324		
Year 2010							
White codes	70.87	(52.04)	79.65	(54.97)	51.57	(38.51)	0.000
Potentially inapprop. visits	132.39	(70.12)	142.40	(72.82)	110.35	(58.07)	0.000
Total ED visits	353.86	(123.80)	365.11	(128.74)	329.09	108.30)	0.000
Observations	1075		739		336		
Years 2008-2010							
White codes	68.66	(52.16)	74.98	(55.99)	52.55	(36.13)	0.000
Potentially inapprop. visits	142.58	(75.99)	153.07	(77.99)	115.83	(63.24)	0.000
Total ED visits	367.25	(128.96)	379.41	(132.05)	336.25	(115.14)	0.000
Observations	3219		2312		907		
Years 2008-2010							
Instrumental variables							
GPs in the group	4.783	(1.79)	4.470	(1.49)	5.582	(2.20)	0.000
Years in group	7.865	(4.28)	7.608	(4.20)	8.520	(4.43)	0.000

Table 1b. Descriptive statistics for the covariates, years 2008-2010

	WHOLE SAMPLE (3219 Obs)		NO EXTENSION (2312 Obs)		EXTENSION (907 Obs)	
<i>Control variables</i>	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Extension 10-12	0.282	0.450	0	0	1	0
Male GP	0.735	0.441	0.721	0.448	0.768	0.422
GP seniority	21.369	8.225	21.160	8.247	21.902	8.147
Mountainous area	0.283	0.451	0.309	0.462	0.217	0.412
Nursing staff	0.237	0.425	0.175	0.380	0.395	0.489
List avg age	51.011	3.327	51.198	3.324	50.534	3.289
Foreign patients	0.069	0.060	0.067	0.056	0.071	0.070
Male patients	0.478	0.038	0.475	0.038	0.485	0.038
Previous hospitalisation	237.633	68.250	238.951	68.282	234.273	68.092
Distance	5.974	6.064	6.059	6.340	5.758	5.291

Previous hospitalisation	237.633	68.250	238.951	68.282	234.273	68.092
List size	1,271.716	315.988	1,267.167	311.148	1,283.313	327.893

Accepted Manuscript

Table 2 – 2SRI: FIRST STAGE Probit and Logit ML estimates for programme participation, year 2009*

FIRST STAGE	PROBIT	LOGIT	LPM (2008-2010)
Dep. Variable:	Extension 10-12	Extension 10-12	Extension 10-12
GPs in the group	0.2594	0.4459	0.0734
<i>GP clustered standard errors</i>	(0.027)	(0.047)	(0.007)
<i>p-value</i>	0.000	0.000	0.000
<i>Group clustered standard errors</i>	(0.063)	(0.109)	(0.017)
<i>p-value</i>	0.000	0.000	0.000
Years in group	0.0415	0.0729	0.0075
<i>GP clustered standard errors</i>	(0.012)	(0.021)	(0.003)
<i>p-value</i>	0.001	0.000	0.008
<i>Group clustered standard errors</i>	(0.018)	(0.032)	(0.005)
<i>p-value</i>	0.024	0.024	0.105
Full set of controls (see Table 1b)	Yes	Yes	Yes
Area dummies	Yes	Yes	Yes
Time dummies	No	No	Yes
Observations	1,075	1,075	3,219
Test on the joint relevance of IVs	$\chi^2=100.51$ ($p=0.000$)	$\chi^2=96.26$ ($p=0.000$)	$F=64.60$ ($p=0.000$)

*Standard errors in parentheses, *p*-values in italics

Table 3a – 2RSI: SECOND STAGE NB ML estimates, White code visits per GP, year 2009*

SECOND STAGE NEGBIN	PROBIT Res	PROBIT Gen Res	PROBIT Dev Res	LOGIT Res	LOGIT Pearson	LOGIT Std Pearson	LPM 2008-2010
Extension 10-12	-0.1955	-0.2722	-0.3267	-0.1886	-0.2607	-0.2614	-0.2110
<i>GP clustered standard errors</i>	(0.100)	(0.099)	(0.121)	(0.101)	(0.098)	(0.094)	(0.090)
<i>p-value</i>	0.051	0.006	0.007	0.063	0.008	0.006	0.019
<i>Group clustered standard errors</i>	(0.220)	(0.188)	(0.247)	(0.202)	(0.190)	(0.184)	(0.195)
<i>p-value</i>	0.376	0.148	0.187	0.352	0.171	0.149	0.278
First-stage residuals	-0.2387	0.1798	0.1491	-0.2312	0.1039	0.1035	0.1677
	(0.107)	(0.063)	(0.053)	(0.103)	(0.047)	(0.045)	(0.092)
	0.025	0.004	0.005	0.024	0.026	0.021	0.068
(First-stage residuals)²	-0.3476	-0.0824	-0.0579	-0.3422	-0.0128	-0.0128	-0.1302
	(0.090)	(0.027)	(0.017)	(0.083)	(0.012)	(0.010)	(0.092)
	0.000	0.003	0.001	0.000	0.289	0.206	0.157
Full set of controls (see Table 1b)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ln(List size)	offset	offset	offset	offset	offset	offset	offset
Area dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	No	No	No	No	No	No	Yes
Alpha	0.2044	0.2054	0.2050	0.2045	0.2068	0.2068	0.1182
	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.007)
Observations	1,075	1,075	1,075	1,075	1,075	1,075	3,219

*Standard errors in parentheses, *p*-values in italics

Table 3b –2RSI: SECOND STAGE NB ML estimates, Potentially Inappropriate visits per GP, year 2009*

SECOND STAGE NEGBIN	PROBIT Res	PROBIT Gen Res	PROBIT Dev Res	LOGIT Res	LOGIT Pearson	LOGIT Std Pearson	LPM 2008-2010
Extension 10-12	-0.1260	-0.1555	-0.1610	-0.1235	-0.1512	-0.1515	-0.0985
<i>GP clustered standard errors</i>	(0.076)	(0.072)	(0.092)	(0.075)	(0.068)	(0.068)	(0.073)
<i>p-value</i>	<i>0.099</i>	<i>0.030</i>	<i>0.081</i>	<i>0.098</i>	<i>0.026</i>	<i>0.027</i>	<i>0.180</i>
<i>Group clustered standard errors</i>	(0.164)	(0.160)	(0.206)	(0.160)	(0.138)	(0.135)	(0.160)
<i>p-value</i>	<i>0.443</i>	<i>0.333</i>	<i>0.434</i>	<i>0.441</i>	<i>0.274</i>	<i>0.255</i>	<i>0.539</i>
First-stage residuals	-0.0752	0.0578	0.0430	-0.0724	0.0313	0.0313	0.0079
	(0.081)	(0.047)	(0.042)	(0.082)	(0.035)	(0.035)	(0.080)
	<i>0.355</i>	<i>0.221</i>	<i>0.303</i>	<i>0.378</i>	<i>0.365</i>	<i>0.372</i>	<i>0.921</i>
(First-stage residuals)²	-0.1295	-0.0283	-0.0205	-0.1277	-0.0032	-0.0032	-0.0426
	(0.072)	(0.020)	(0.013)	(0.065)	(0.006)	(0.006)	(0.062)
	<i>0.073</i>	<i>0.157</i>	<i>0.128</i>	<i>0.049</i>	<i>0.594</i>	<i>0.597</i>	<i>0.494</i>
Full set of controls (see Table 1b)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ln(List size)	offset	offset	offset	offset	offset	offset	offset
Area dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	No	No	No	No	No	No	Yes
Alpha	0.1071	0.1073	0.1072	0.1071	0.1074	0.1074	0.1034
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.046)	(0.005)
Observations	1,075	1,075	1,075	1,075	1,075	1,075	3,219

*Standard errors in parentheses, *p*-values in italics

Table 4 – Panel analysis for White code visits and Potentially inappropriate visits, years 2008-2010*

Dependent Variable	White code visits				Potentially inappropriate visits			
	Poisson FE	Poisson RE	NB RE	NB CCRE	Poisson FE	Poisson RE	NB RE	NB CCRE
Extension 10-12	-0.1650	-0.1433	-0.1010	-0.1205	-0.0671	-0.0751	-0.1063	-0.0466
<i>GP clustered standard errors</i>	(0.031)	(0.024)	(0.025)	(0.033)	(0.021)	(0.017)	(0.023)	(0.020)
<i>p-value</i>	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.022
<i>Group clustered standard errors</i>	(0.048)	(0.038)	(0.041)	(0.051)	(0.031)	(0.027)	(0.034)	(0.028)
<i>p-value</i>	0.001	0.000	0.014	0.018	0.031	0.005	0.002	0.091
Male GP		-0.0990 (0.046) 0.032	-0.1490 (0.041) 0.000	-0.1529 (0.039) 0.000		-0.0829 (0.036) 0.020	-0.1117 (0.046) 0.014	-0.1219 (0.036) 0.001
GP seniority		-0.0049 (0.002) 0.004	-0.0047 (0.001) 0.003	-0.2278 (0.139) 0.102		-0.0050 (0.001) 0.000	-0.0038 (0.002) 0.035	-0.2463 (0.111) 0.026
Mountainous area		0.2347 (0.031) 0.000	0.2393 (0.034) 0.000	0.2417 (0.034) 0.000		0.0745 (0.026) 0.004	-0.0042 (0.048) 0.931	-0.0064 (0.029) 0.824
Nursing staff	-0.0751 (0.042) 0.076	-0.0978 (0.035) 0.006	-0.1361 (0.033) 0.000	-0.0973 (0.038) 0.010	-0.0036 (0.019) 0.852	-0.0293 (0.017) 0.093	-0.0632 (0.018) 0.001	-0.0142 (0.019) 0.461
List avg age	-0.0129 (0.019) 0.490	-0.0245 (0.006) 0.000	-0.0244 (0.045) 0.000	-0.0122 (0.018) 0.497	-0.0335 (0.015) 0.024	-0.0231 (0.005) 0.000	-0.0153 (0.005) 0.001	-0.0338 (0.012) 0.007
Foreign patients (% list)	0.3039 (0.881) 0.730	0.4431 (0.389) 0.254	0.5110 (0.252) 0.043	0.1288 (0.662) 0.846	-0.8663 (0.647) 0.181	0.0245 (0.291) 0.933	0.4883 (0.231) 0.035	-1.0376 (0.540) 0.055
Male patients (% list)	1.0612 (1.468) 0.470	1.5715 (0.638) 0.014	1.952 (0.408) 0.000	1.0987 (1.279) 0.390	0.6737 (1.101) 0.540	1.2352 (0.510) 0.015	1.3434 (0.417) 0.001	0.6435 (1.147) 0.575
Distance		-0.0249 (0.002) 0.000	-0.0239 (0.002) 0.000	-0.0241 (0.002) 0.000		-0.0227 (0.002) 0.000	-0.0211 (0.003) 0.000	-0.0215 (0.002) 0.000
Previous hospitalisation	-0.0002 (0.000) 0.471	-0.0002 (0.000) 0.476	-0.0002 (0.000) 0.244	-0.0001 (0.000) 0.725	-0.0004 (0.000) 0.041	-0.0004 (0.000) 0.009	-0.0004 (0.000) 0.002	-0.0003 (0.000) 0.020
Constant		-2.6245 (0.3891) 0.000	-3.7880 (0.312) 0.000	-4.0870 (0.285) 0.000		--1.3186 (0.312) 0.000	-2.6462 (0.394) 0.000	-3.1372 (0.380) 0.000
Area dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ln(List size)	offset	offset	offset	offset	offset	offset	offset	offset
Time averages of the regressors	No	No	No	Yes	No	No	No	Yes
Observations	3,219	3,219	3,219	3,219	3,219	3,219	3,219	3,219

*Standard errors in parentheses, *p*-values in italics

Table 5 – Marginal effects*

White code visits	IRR	AME	MEM
FE Poisson	0.848	-11.33	-9.33
RE Poisson	0.866	-8.61	-7.88
RE Negbin	0.904	-6.71	-6.13
Potentially inappropriate visits	IRR	AME	MEM
FE Poisson	0.935	-9.56	-8.65
RE Poisson	0.928	-10.00	-9.53
RE Negbin	0.899	-14.75	-14.09

* all p-values < 0.05