# Physician workload and treatment choice: the case of primary care*

Ity Shurtz†1, Alon Eizenberg‡2, Adi Alkalay[3], and Amnon Lahad[3,4]

[1]Department of Economics, The Hebrew University of Jerusalem
[2]Department of Economics, The Hebrew University of Jerusalem, and CEPR
[3]Clalit Health Services
[4]School of public health, The Hebrew University of Jerusalem

July 2019

**Abstract**

Primary care is a notable example of a service industry where capacity-constrained suppliers face fluctuating demand levels. Unable to adjust prices, such providers may degrade service quality when faced with high demand levels. Little is known, however, on the nature of such adjustments in the primary care context. We study how physicians trade off one key input — their time with patients — with other inputs, such as prescriptions, lab tests and referrals. Employing detailed administrative data from eleven clinics of a large Israeli HMO, we use the absence of colleagues as a source of exogenous variation in physician workload. We find no evidence that physicians' workload affects the intensity with which they prescribe painkillers, or refer patients to the Emergency Room, and very little evidence for an effect on the prescription of antibiotics. We do find, however, that physician time and the use of diagnostic inputs are complements: a one minute decrease in the (daily) average visit length causes a 9 percent decrease in referrals to specialists, and a 3.8 percent decrease in referrals to lab tests. Following recent literature, we complement the traditional use of an exclusion restriction within a linear model by estimating non-parametric bounds on Average Treatment Effects using alternative assumptions. Such alternative estimators rule out the possibility that physician time and the use of diagnostic tools are substitutes in an economically-meaningful fashion, while still leaving a broad scope for the possibility that those are complements. Taken together, the results indicate that the shadow cost of physician capacity is not reflected in poor treatment decisions, but may instead be manifested in the underprovision of tests and referrals to specialists — fundamental aspects of long-term preventive care.

*First public version: February 2018. We are grateful for helpful conversations and comments from Tobias Kline, Jianjing Lin, Charles Manski, Fernanda Marquez-Padilla, and John Pepper. All errors are our own. We have also benefitted from comments by seminar participants at Ben Gurion, Carnegie Mellon, Mannheim, Tel-Aviv, and Tilburg, and conference participants at the Israeli IO Day 2017, iHEA Boston 2017 Congress, Barcelona GSE Summer Forum, Policy Evaluation in Health 2017, the 7th Annual Conference of the American Society of Health Economists, and the IIOC 2018. Nadav Perlov provided excellent research assistance. Financial support from the Maurice Falk Institute for Economic Research in Israel and from the Israel National Institute for Health Policy Research is gratefully acknowledged.

†ity.shurtz@huji.ac.il

‡alon.eizenberg@mail.huji.ac.il

# 1 Introduction

In many industries, firms' capacity is fixed or modifiable at a high cost while the demand they face fluctuates between high and low states. Airlines, hotels and car rentals commonly adjust prices to reflect the varying shadow cost of the capacity constraint, i.e., the cost of serving the marginal consumer. In professional service industries such as banking, legal services and healthcare, firms often refrain from price adjustments and may instead adjust other margins such as service quality (Chatain and Eizenberg (2015)). This paper studies the nature of such adjustments in the healthcare sector, with a specific focus on primary care.

Primary care is a particularly interesting setup because better primary care provision is thought to be associated with improved population health and lower health care costs (Starfield et al. (2005)).[1] Moreover, the capacity-quality tradeoff has been identified as a crucial policy issue in the primary care context.[2] Nonetheless, very little is known regarding the manner with which primary care physicians trade off one key input — namely, their time with patients — with other inputs such as the prescription of medication, or referrals to specialists.

In this paper we study the components of the shadow cost of physician capacity. First, we ask whether physicians respond to a tightening time constraint by raising the intensity of treatment administered during the visit: namely, do they prescribe more painkillers or antibiotics, or issue more referrals to the Emergency Room? If this is the case, then the shadow cost of physician capacity includes the cost of those intensified treatments.

Second, we ask whether diagnostic tools like referrals to specialists or to tests serve as substitutes to the physician's time with patients, or, rather, as complements. Substitution may imply that the shadow cost of interest manifests itself (at least in part) in more intense use of such tools. Complementarity, in contrast, would imply that a tightening time constraint leads to a lower provision of referrals, potentially involving an adverse effect on long-term preventive care. Importantly, the map from our findings to statements about the efficiency of the system

[1] Scott (2000) notes that "GPs make many different types of decisions that influence the amount, type and location of care received by patients. These include decisions to refer to a specialist or other health professionals, prescribe medication, arrange follow-up, and order tests."

[2] Anand et al. (2011) note that "(a) major difficulty in improving productivity in such customer-intensive services is the sensitivity of...service quality... to the speed of service: as the service speed increases, the quality of service inevitably declines...*(p)rimary health-care practice in the United States epitomizes this problem*" (*italics* added by the authors).

is quite complex, and we discuss this issue in more detail below when interpreting our findings.

In the primary care setting, high demand states interact with physician capacity in a manner that increases the number of patients a physician sees within a given amount of time (or, equivalently, the average visit length). This number is commonly used as the definition of physician workload. We examine how workload affects physicians' decisions using detailed administrative data from eleven clinics of a large Israeli HMO during 2011-2014. In addition to visit length, the data record visit-level information regarding the physician's actions, including, in particular, prescriptions and referrals.

Primary care physicians play an important role in the healthcare system. They serve as the HMO's gatekeepers, regulating treatment and referrals to avoid unnecessary and inefficient treatment. They also administer preventive care by referring patients to tests and taking other action to detect and solve health issues before they deteriorate. Measuring the quality of primary care from data is therefore quite complex, as the map from physicians' decisions to patient outcomes such as well being or mortality rates is far from obvious. For this reason, we study the impact of workload on a number of clearly identifiable visit-level outcomes, pertaining to the employment of a variety of elements in the physician's toolcase. Those can be broadly categorized into three types: the use of diagnostics, the choice of treatment, and the utilization of non face-to-face interaction with patients.

Identifying a causal effect of workload on these outcomes is frustrated by endogeneity, which may arise from measurement error in the workload variable, or from the presence of unobserved factors that simultaneously affect both the outcome, and the physician's workload. We therefore propose an instrumental variable approach: the absence of fellow physicians at the clinic is used as a source of an exogenous variation in the physician's workload. In practice, our instrument is the fraction of the total count of patients that visit a physician on a given day that is attributed to an absent colleague's patients.

While the presence of the absent colleague's patients shifts the physician's workload, we only analyze her decisions with respect to her regular patients. We therefore avoid confounding the effect of workload with the effect of treating unfamiliar patients. We employ the instrument in standard Two Stage Least Squares (2SLS) linear models as well as in nonparametric models that place bounds on the Average Treatment Effect (ATE) following Manski and Pepper (2000).

The partial identification approach supplements the linear models by allowing us to exploit the data in additional fashions, employing assumptions that are naturally motivated by the economic setup. A Monotone Instrumental Variable (MIV) assumption allows higher levels of the instrument to be associated with higher levels of the response function at any workload level, thus weakening the exclusion restriction. This accommodates the possibility that patients with non-acute conditions give up their slot when their physician is busy on account of a colleague's absence. Such "deterrence" effects would violate a classic exclusion restriction, but are allowed by the MIV assumption. Just the same, we provide robust evidence against the possibility that the instrument generates such deterrence effects, supporting the standard exclusion restriction embedded in the 2SLS specifications.

We combine this MIV assumption with a Monotone Treatment Selection (MTS) assumption that allows the response function to take higher values at any workload level conditional on a high level of the *realized workload*. The assumption is consistent with the notion that higher physician workload may be correlated with a sicker pool of patients. This correlation — which is the source of the endogeneity concern to begin with — may stem from unobserved factors (e.g., a flu epidemic) affecting both physician workload and the distribution of medical issues presented by patients. It may also stem from unobserved choices by patients with minor issues to reschedule (or avoid making an appointment on a given day) on account of longer-than-usual wait times on days when their physician experiences high workload levels.

These complementary strategies, along with the rich data structure, help us address some additional potential threats to our identification strategy. One such threat is that seasonal effects, such as a flu epidemic, may affect both the absence of colleagues and patients' health. The use of time fixed effects helps us mitigate this concern. Another concern is that the allocation of an absent colleague's patients among the non-absent colleagues may not be random, e.g., that those "extra patients" are referred by the clinic's manager to physicians who have less bargaining power within the clinic. We address this issue via the inclusion in the model of physician fixed effects.

**Results.** One may hypothesize that physicians would be more conservative when workload is higher and increase treatment intensity. We find very limited evidence for this in practice. Our 2SLS regressions imply that, in the context of face-to-face visits, physician workload has no significant impact on referrals to the Emergency Room. Nor do we find evidence that workload affects the prescription of painkillers. Some specifications do provide (mixed) evidence that higher workload is associated with an increase in prescription of antibiotics: a one minute decrease in average visit length increases the prescription of antibiotics by about 5%.

It does not appear, therefore, that the shadow cost of physician capacity involves a substantial degree of over-provision of medication, or of referrals to the ER. Our 2SLS analysis does imply, however, that when physicians experience higher workload they substantially decrease the amount of non face-to-face encounters with patients: a one minute reduction in average visit length decreases the amount of response to patient online queries by 8 percent, and a 10 percent decrease in the number of phone calls with patients.

Finally, we examine the impact of workload on the use of diagnostic inputs. Our 2SLS regressions suggest that diagnostic inputs are complements rather than substitutes to physician time. A one minute decrease in average visit length causes a 9 percent decrease in referrals to specialists, a 3.8 percent decrease in referrals to lab tests, and a statistically insignificant decrease in referrals to imaging. These results suggest that when physicians experience high workload they tend to limit the scope of issues they address during a single visit. Examining heterogeneity, we find that the effects on patients of age 60 and above are stronger than those on younger patients. The broader scope of medical issues presented by older patients may suggest that a larger number of issues are left unattended given intense physician workload.

The significance of these findings would likely be qualified if, having not been referred to a specialist or to a lab test on account of the physician being busy, patients return to the clinic to obtain such referrals on a later date. We find, however, no evidence that workload increases the likelihood of subsequent visits: such effects are small and statistically insignificant.

To further investigate the important relationship between workload and the use of diagnostics, we turn to nonparametric bounds analysis. Here we relax the linear structure and replace the exclusion restriction by the weaker MIV assumption, yet obtain additional identifying power via the Monotone Treatment Selection assumption. We estimate the ATE of a switch from low workload to high workload (where low and high workload refer to workload below and above the 75th percentile of the physician-specific workload distribution). The MIV-MTS estimated set for this ATE is $[-.432, -.025]$, with a 95 percent confidence interval of $[-.438, .013]$.

While the confidence interval contains positive values, they suggest that higher workload can increase the probability of using diagnostics by at most 1.3 percent. This effectively rules out that time with patients and the use of diagnostics are substitutes in an economically-meaningful sense. Consistent with the baseline linear analysis, the confidence interval does leave a broad scope for complementarity, suggesting that higher workload (i.e., less time with patients) may reduce the probability of using diagnostics by as much as 43.8 percent.

**Policy implications.** A better understanding of the shadow cost of physician capacity should inform policymakers' optimal decisions regarding the provided amount of physician hours. Our analysis is indeed related to the so-called "primary-care crunch" — the shortage in primary care physicians in the United States. It is often argued that this shortage results in increased workload, and lower quality healthcare. Our analysis reveals the channels via which such effects are realized, noting that physicians are able to offset some of the harm to patients by making optimal treatment choices under the workload constraint (at least when it comes to the prescription of painkillers or ER referrals).

The impact on referrals to specialists and to other tests is of particular importance: an early and accurate diagnosis of medical problems is considered one of the main benefits to managed care systems such as HMOs, and such referrals are very important in that context. Our analysis rules out that physicians increase the use of such diagnostics as they become busier, while suggesting the strong possibility that they substantially *decrease* it. We therefore identify the underprovision of such referrals and tests as the channel via which the shadow cost of physician capacity is likely realized.

One could wonder whether physicians, to begin with, provide excessive referrals and tests — in which case the shadow cost may actually be negative. Determining the efficiency of use of diagnostics is beyond the scope of our paper. Nonetheless, our findings are informative to policy makers, given any beliefs they may hold regarding this matter. A policy maker who believes that such referrals are excessively provided in general would be reassured by our findings that tight physician capacity does not exacerbate this problem, whereas a policy maker who believes that such tools are used efficiently (or are under-provided) in general would learn that tight capacity may result in under provision of long-term preventive care services. And, both hypothetical policy makers would learn that tight capacity does not significantly affect the course of treatment (e.g., prescriptions of medications or referrals to the ER).

The validity of our analysis does not depend, therefore, on an implicit assumption regarding the efficiency of primary care. Just the same, experts have expressed grave concerns that tight capacity limits physicians' ability to provide quality care, and our interpretation of the results is that this limitation mostly has to do with underprovision of referrals to specialists and tests.

Our findings would still be of limited value to policy makers if providers are aware of their physicians' response to the capacity constraint, *and* optimally trade off their investment in expanded capacity (i.e., the hiring of additional physicians) versus their investment along other margins (i.e., referrals and diagnostics). Neither one of these conditions is likely to be met. The complex set of tools at the physicians' disposal suggests that providers are not likely to have a clear sense of the channels via which physicians respond to workload. Our analysis is therefore useful in identifying these channels. Second, even if providers obtained perfect knowledge regarding these mechanisms, and optimally solved their own cost-minimizing problem, they may fail to fully internalize the social costs associated with physician workload. As taxpayers end up shouldering some of the costs associated with poorer long-run health outcomes, a misalignment of incentives is to be expected.

Our analysis also leaves some open questions: while workload affects physician behavior, the ultimate impact on patient well-being and the efficiency of the system has yet to be completely understood. Finally, it is worth noting that our analysis examines temporary increases in workload which allow intertemporal substitution of tasks. The effect of permanent increases in workload however may have a stronger impact on patient well-being.

**Literature.** This paper contributes to a growing literature on the role played by supply side factors in driving the well documented variation in health care utilization (Finkelstein et al., 2016; Currie et al., 2016). Recent work in this area has focused on the role of financial incentives (Clemens and Gottlieb, 2014; Ho and Pakes, 2014), liability concerns (Currie and

MacLeod, 2008; Frakes, 2013), group practice size (Gaynor et al., 2004) and the degree of team work among physicians (Chan, 2016).

The supply-side mechanism we focus on involves input substitution: we study how time with patients is traded off versus the use of diagnostic inputs, treatment choices and non face-to-face interactions. This connects our study to a literature that has examined a workload-quality tradeoff in various work environments. Tan and Netessine (2014) use data from restaurants to examine the effect of workload, defined as the number of tables handled, on performance measured by sales and meal duration. Surprisingly, they find that higher workload is associated with higher effort in a manner that may lead to higher sales and to lower labor costs. Perdikaki et al. (2012) study the relationship between store traffic, labor, and sales performance. They find that the conversion of incoming traffic into sales declines with shopper's traffic. Chatain and Eizenberg (2015) study a legal service provider and find that service quality is an increasing function of the amount of available resources.

In healthcare, the workload issue has received attention in the context of hospitals. Using operational data from a hospital emergency department, Batt and Terwiesch (2012) find that workload induced service slowdown and that care providers adjust their clinical behavior to accelerate the service. Kc and Terwiesch (2009) show that the system load increases the service rate and results in a reduction in the quality of care. Kim et al. (2016) study admission to intensive care units (ICU) and find that ICU congestion can have a significant impact on ICU admission decisions and patient outcomes. Powell et al. (2012) find that physician workload reduces that share of "severe" patients and consequently hospital reimbursement.

Our study addresses these issues in the context of primary care. While the primary care crunch is considered a major policy issue (see Anand et al. (2011)), surprisingly little empirical work has attempted to identify a causal effect of primary care physicians' capacity on medical decisions. One possible reason is that, unlike the ICU or ER contexts, one cannot use a simple indicator, such as the mortality rate, to measure service quality.[3] Given this difficulty, we instead attempt to identify the specific channels via which workload affects primary care

[3]The health economics literature has attempted to measure quality in a variety of fashions to overcome this issue. Some examples include Fleitas (2018) who uses test scores from graduate school medical specialty admissions to measure physician quality, and Brunt et al. (2018) who measure the quality of treatment using data from the Medicare's Physician Compare Quality Reporting System.

physician's choices.

Some recent working papers that are related to our analysis are Neprash (2016) and Freedman et al. (2018) who examine the impact of schedule disruptions at the individual visit level on physician actions. Our approach differs from that pursued in these papers in three ways. First, we explicitly study the impact of the overall daily level of workload on treatment choices at the individual visit level — as opposed to studying the impact of visit-level schedule disruptions on visit-level outcomes. Our approach is complementary to that of Neprash (2016) and Freedman et al. (2018) in that it speaks more directly to the broad question of how service providers trade off the use of inputs when facing capacity constraints.

Second, our identification strategy is different: we rely on absence of colleagues as a shifter of the physician's daily workload level, as opposed to the use of unexpected schedule disruptions as shifters of the visit-level time pressure. Finally, we employ both point- and partial-identification strategies to identify the effects of interest. This, in particular, allows us to account for possible violations of our exclusion restriction.

Despite these differences, and the fact that neither the outcomes nor the treatments considered by these papers coincide exactly with our definitions, it is interesting to note that important aspects of the findings in these papers are broadly in line with our findings. In particular, Freedman et al. (2018) report that increased time pressure causes physicians to limit the number of topics with which they deal during a visit and to use fewer referrals. Neprash (2016) shows that in response to schedule disruptions, physicians tend to shorten the visit's duration, perform fewer procedures, increase referrals of a new patient to a specialist and prescribe more antibiotics and painkillers. Some differences also obtain: for example, we do not find that workload causes a significant increase in the likelihood of a return visit.

Our paper joins a growing empirical literature that places nonparametric bounds on the Average Treatment Effect under various sets of assumptions following Manski and Pepper (2000). Examples of such applications include health economics (Gerfin and Schellhorn, 2006; Bhattacharya et al., 2012), the economics of education (Gonzalez, 2005; De Haan, 2011; De Haan and Leuven, 2016), public economics (Kreider et al., 2012), and online network effects within social media (Shriver et al., 2013). An example of an alternative approach for placing bounds on treatment effects is provided in Mogstad et al. (2018). Interest in partial identification strategies in applied work has been on the rise, as surveyed by Ho and Rosen (2015).

The remainder of the paper is structured as follows. Section 2 describes the data, section 3 describes our baseline empirical strategy (i.e., the linear models), and section 4 reports baseline findings. Section 5 presents the nonparametric bounds analysis. Section 6 concludes.

## 2 Data and Environment

We use a detailed administrative database that covers all primary care visits in eleven clinics in the Jerusalem area of Clalit Health Services — the largest of four HMOs that provide the vast majority of health insurance in the country and deliver most of its primary care — in the period 2011-2014.[4]

In Clalit Health Services, primary care physicians are (for the most part) directly employed by the HMO and patients are enrolled with a regular primary care physician. They may schedule ten minute appointments, online or by calling the clinic's offices, at the fixed price of zero. Appointments are scheduled until the schedule is full, and sometimes beyond that. Additionally, patients with urgent medical issues would drop-in, asking to be seen by a physician even if the schedule is full. The clinic's office would then try to fit them in, so long as it is feasible. As a consequence, the number of patients that are seen by physicians at the clinic fluctuates between high and low demand states.

Physician work times at the clinic are fixed, and there is very little wiggle room around them. Discussions with clinicians suggest that they tend to stick to their regular hours, and do not stay at the clinic beyond them. One reason is that often the clinic closes down at the end of the shift. As a consequence of these institutional features, a fixed capacity of physician time interacts with fluctuating levels of demand, absent an ability to regulate those demand levels via price adjustments. This gives rise to our research question: do physicians alter the quality of service as a function of the degree to which their time constraint binds, and if so, how?

[4]The Israeli primary care system is largely publicly-funded, with much of the HMOs' budget being derived from health taxes. Israeli residents may freely choose an HMO, and HMOs compete over enrolees, in part, by striving to improve the quality of care.

Normally, a primary care visit is scheduled with the patient's regular physician. There are exceptions to this rule: if patients need urgent care, outside of their physician's office hours or when their physician is absent, they are typically referred to another physician at the clinic. We restrict our sample to visits that occur during weekdays, in which physicians encounter their regular patients, on days when they see at least twelve of their regular patients. This allows us to capture physicians' actions with respect to their regular patients in a typical day.[5]

The data include information about visit characteristics. Specifically, the patient's regular physician identity, the identity of the attending physician and the visit actual start time (as opposed to the visit's scheduled start time). The visit length is calculated as the time between the visit's start time, and the start time of the subsequent visit.

Also observed are patient characteristics such as gender, age, country of origin and chronic conditions. The latter are recorded in a very detailed fashion, allowing us to control for 113 different chronic conditions. Finally, a detailed description of the visit's outcomes is recorded including diagnoses, prescriptions, and referrals to specialists, laboratory tests, and imaging.

In this final sample there are 823,349 office visits made by 78,959 patients to 93 physicians. Table 1 provides descriptive statistics regarding these face-to-face visits. The mean patient age is 47.6, and 58 percent of the patient visits are by women. Thirty percent of the visiting patients are smokers and 26 percent are obese. Hypertension characterizes 34 percent of the visiting patients, almost 45 percent of them have hyperlipidemia, while 15 percent suffer from ischemic heart disease. A substantial mass of office visits is, therefore, generated by a population exhibiting chronic medical conditions.

Office visits last 11.56 minutes on average. Fourteen percent of visits result in a referral to a specialist, 8 percent result in referrals to imaging and 20 percent result in a referral to lab tests. The employment of such diagnostic tools is, therefore, an important part of the normal activity of primary care physicians. Patients are referred to the emergency room in one percent of the visits. Antibiotics are prescribed in ten percent of the visits, and pain killers are prescribed in five percent of the visits.

[5]In Israel, weekdays are Sunday through Thursday, thus we drop Friday and Saturday visits. The data include about 1,090,000 weekday office visits.

# 3 Empirical strategy

## 3.1 Workload definition and the identification challenge

Our goal is to estimate the relationship between a visit-level outcome variable (e.g., an indicator for the prescription of antibiotics) and an explanatory variable: a measure of the physician's daily level of workload. A common measure of workload in the primary care setting is the number of patients a physician sees per-hour or, equivalently, the average visit length (see e.g. Hobbs et al. (2016)). We therefore define the main explanatory variable *workload* as the daily average of a physician's visit lengths.[6] Consider, for example, a physician who had seen patients during a total of two hours and, within that time, saw ten patients. For this physician, on that day, our workload measure takes the value of twelve minutes per-patient.

Figure 1 depicts the distribution of the workload measure, the daily average of a physician's visit lengths.[7] The mean of the workload measure is 11.5 minutes per patient. Workload exhibits substantial variation. The daily average of visit lengths is 7.6 minutes at the $10^{th}$ percentile of the workload distribution and it more than doubles to 16 minutes at the $90^{th}$ percentile. The primary variable of interest, the workload level, therefore exhibits substantial variation over physician-day data cells.

Consider the following model for the relationship between workload and physician behavior

$$y = \alpha + \beta \cdot workload + x \cdot \gamma + \epsilon \tag{1}$$

where $y$ is an outcome and $x$ is a rich set of controls.

Analyzing this model using OLS may result in biased estimates for a couple of reasons. First, our workload measure is subject to measurement error: true workload is determined not only by the number of patients seen per hour, which is what we measure, but also by random shocks that arise from patient characteristics such as age or chronic conditions. Second, while the data are highly detailed and contain much of the relevant information that underlies the realization of

[6] Note that the length of the last visit of the day, which we are unable to calculate, is not used to generate the *workload* variable.

[7] As the $99th$ percentile of the distribution is 21.5, about 2000 observations with a visit length in excess of 25 minutes were excluded from this figure for illustrative purposes.

the outcomes we study, omitted variables may still be correlated with our measure of workload. For example, a local infection may increase both the number of patients the physician sees per hour, and the probability of prescribing antibiotics.

Third, a simultaneity mechanism may present itself: a high workload level of the physician on a given day may induce some of her patients to give up their slot on account of longer-than usual wait times. Such patients may also refrain from scheduling the appointment in the first place, if, upon calling the clinic and asking to be seen by a physician on that day, they are advised by the office that their physician's schedule is already quite tight, and they would likely need to endure a substantial wait time. We do not observe schedules (and definitely not visits that were never scheduled) but rather actual visits, suggesting that such behavior is unobserved to us. If the patients who give up on seeing the physician on account of her high workload level tend to present non-acute medical issues, as one would expect, then high workload may affect the distribution of medical conditions the physician observes, and, therefore, the outcomes (e.g., the incidence of antibiotics prescriptions). As we discuss next, we address these various endogeneity concerns via an instrumental variable approach.

## 3.2 An exclusion restriction

To identify the causal effect of workload on physician choices we use the absence of colleagues at the clinic as source of exogenous variation in a physician's workload. When a colleague is absent, her patients are referred to other physicians at the clinic, affecting their workload.

Physicians have fixed shifts at the clinic. We exploit this regularity and define a physician as absent on a given day if two conditions are satisfied. First, on this day, a physician treats zero of her (positive number of) patients, namely, the physician is not present at the clinic on that day, yet some of her patients do arrive to seek treatment. Second, the physician has worked (and has seen at least 5 of her patients) in the two weeks before and after the relevant day on the same weekday. The observation therefore pertains to a weekday in which the physician normally works at the clinic.

Having defined physicians' days of absence from the clinic, we calculate, for each physician who is present at the clinic on a given day, a proxy for the added workload brought about by absent colleagues. Our proxy is the share of the absent colleague's patients out of the total patients seen by the physician on that day. This total includes both the physician's regular patients, and patients of the absent colleague.[8] We refer to this proxy as the share of the missing physician's patients hereinafter, and use it as an instrumental variable for physician workload.

We also define an extensive margin instrumental variable: an indicator taking the value 1 for physicians who see any absent colleagues' patients on the given day, and zero otherwise. Finally, in Appendix section A.2 we also report results using a clinic-level version of this instrument, namely, an indicator taking the value of one given any absence at the clinic, and zero otherwise. As shown below, these variations result in qualitatively similar results as in our baseline analysis.

We provide below evidence that our instrument is strongly correlated with the endogenous workload variable, ruling out weak IV concerns. The validity of the instrument also hinges on an assumed exclusion restriction: the instrument should affect the outcome only via its effect of the workload variable, i.e., it should not be correlated with the error term in (1). In our nonparametric bounds approach we shall consider a weaker version of this assumption. In this section, however, we discuss the merit of the exclusion restriction, referring to a number of concerns that arise with respect to it — and to the manner with which we address them.

A first concern is that, in periods that are prone to disease (e.g., the winter when the flu is prevalent), the disease may affect the distribution of patient conditions presented at the clinic while *also* leading to higher rates of physician absence. To minimize the potential biases from such issues we include a rich set of time-period controls that should ameliorate seasonal effects.

Second, there may be another channel via which the distribution of the physician's own patients is correlated with the absence of a colleague. Imagine that, when a physician's colleague is absent, her patients observe (or are told on the phone prior to arriving) that wait times are longer than usual. Such patients who do not display acute medical conditions may then give up their appointment. As we show below, the data do not give much credence to such concerns. In particular, we find that the number of *own-patients* seen by a physician per-hour is not affected by the absence of a colleague. Furthermore, including patient characteristics (notably,

[8]We exclude patients who are not regular patients of either the physician or her missing colleague.

age, gender and chronic conditions) does not affect our first-stage regression results. In light of this evidence, we do not believe that this mechanism is important (also noting that our MIV assumption would account for it).

Finally, a third issue pertains to a non-random assignment of a missing colleague's patients among the non-missing physicians. An intricate set of incentives affects the decision of the clinic manager to divert such additional patients to a given physician, and the physician's willingness to accept them. Seeing a missing colleague's patients increases the physician's workload, which is likely undesirable.[9] Some physicians may be more reluctant to see their colleague's patients than others, or may have more "bargaining power" that allows them to stir the additional workload towards others. It is also possible that the clinic manager uses her familiarity with the physicians to divert extra patients, as much as possible, towards those physicians who are likely to perform well under pressure. Our inclusion of physician fixed effects largely diminishes concerns for bias arising from this issue. Fixed effects imply that we rely on within physician variation so that differences in practice style across physicians do not drive the results.

In the next section we employ the exclusion restriction in 2SLS analysis and report results. In Section 5 below, we relax the exclusion restriction to an MIV assumption.

# 4 Baseline results: 2SLS analysis

## 4.1 The first stage

We first illustrate graphically the source of variation we use in our instrumental variable approach. Figure 2 depicts the relationship between our instrument (the share of absent physician patients out of the physician's total daily patient count) and our measure of the physician's daily workload. To do so, we classify the instrument into bins of 0.025 percentage points. For each bin, we calculate the average of our measure of workload — the average visit length. We also regress the workload measure on the instrument, and use the solid line to display the relationship predicted by this regression. As the figure shows, an increase of ten percentage points in the share of the absent physician patients seen by a physician is associated with a

[9] To potentially address this, physicians receive a monetary compensation of about 7 dollars per such visit.

decrease of about one minute in that physician's average visit length. Given that the average visit length in the data is about 11.6 minutes, this is a sizable effect.

To further illustrate the effect of seeing an absent colleague's patients on the physician's daily workload at the clinic, we aggregate the data at the physician-day level and analyze an event study model. We let $D_{st}$ be an indicator that takes the value one when at least one physician is absent from clinic $s$ at time $t$, and zero otherwise. Suppose for example that a physician was absent on January 5th, 2013 at clinic 5, then $D_{5,1/5/2013} = 1$. Next, define $\tau_{st}$, the *event relative time*, as the number of days that elapsed since the absence. Thus, in our example $\tau_{5,1/5/2013} = 0$, $\tau_{5,1/4/2013} = -1$ and $\tau_{5,1/7/2013} = 2$. Indexing physicians by $j$ we estimate a model of the form:

$$
\begin{aligned}
\text{(2)} \qquad workload_{jst} =& \alpha + \nu_j + \nu_t + \\
& \gamma_1 \cdot \tau_{-k} + ... + \gamma_{k+1} \cdot \tau_0 + \gamma_{k+2} \cdot \tau_1 + ... + \gamma_{2k+1} \cdot \tau_{k-1} + \epsilon_{jst},
\end{aligned}
$$

where $\nu_j$ are physician fixed effects, and $\nu_t$ contains year-month and day of the week fixed effects. The objects of interest are the coefficients on the $\tau$ variables. Our hypothesis is that the coefficient on $\tau_0$ should be negative (recalling that seeing a higher number of patients implies a *lower* average visit length), while the coefficients on days before and after the absence should not be different than zero.

Figure 3 displays the estimates of these indicators, ranging from $\tau_{-7}$ to $\tau_6$, along with 95 percent confidence intervals. The pattern confirms our hypothesis: in the days before the event, and in the days after it, the effect of the absence is not different than zero in a statistical significance sense. On the day of the event, the average visit length at the clinic drops by about a third of a minute, and this effect is statistically significant.

We next turn to formally estimating the relationship between absences and physician workload. Indexing visits by $i$, the first stage regression is:

$$
\text{(3)} \qquad workload_{jsti} = \alpha + \nu_j + \nu_t + sa_{jst} \cdot \beta_1 + x_{jsti} \cdot \beta_2 + \epsilon_{jsti}
$$

where again $\nu_j$ and $\nu_t$ capture fixed effects for physician, year-month, and day of the week. The variable $sa_{jst}$ is our instrument: the share of an absent physician's patients out of physician $j$'s total count of patients at clinic $s$ on day $t$. Below we denote this instrument by $IV1$. We also define a second instrument, IV2, taking the value 1 if $sa_{jst} > 0$, and zero otherwise. The vector $x$ is a set of visit-level characteristics including patient characteristics.[10]

Table 2 displays first stage results. Column 1 of panel (a) shows that our first instrument, $IV1$, has a negative, statistically-significant effect on workload with a point estimate of about -4.8, implying that an increase of ten percentage points in the share of an absent physician's patients is associated with a decrease of 0.48 minutes in average visit length.[11]

One of the threats to our instrumental variable approach, discussed above, is a potential interaction between the absence of colleagues and the composition of the physician's own patient pool. This could arise if patients with non-acute conditions are deterred by the physician's workload and decide to return on a different day. To assess this issue, we present in column (2) of Panel (a) of the table results that control for patient characteristics. We also control for visits of an administrative nature. If the instrument affects the type of visits or the patient pool, the first stage estimates in this specification would be different from those presented in column (1). The results demonstrate that adding those patient and visit level controls leaves the first stage results virtually unchanged.

As yet another test of the "deterrence" mechanism, we examine if the number of the physicians' own patients is affected by a colleague's absence (in the spirit of McCrary (2008)). Concretely, we examine whether the number of own-patients seen by a physician is lower on a day when a colleague is absent. We run an event study analysis, similar in nature to the analysis described in Equation (2) (and that was displayed in Figure 3). The dependent variable here is a daily measure of the number of the physician's own patients per hour.[12] Namely, the number of the physician's regular patients encountered during a day, divided by the number of

[10]The patient level characteristics we use are: age, age squared, a gender dummy, and 113 indicators for chronic conditions. We additionally include dummy variables for visits for which the main reason is: issue a medical certificate, prescription renewal, filling out forms, and an administrative visit.

[11]The hypothesis that one may exclude the instrument from the model is rejected at 99.9% significance, alleviating weak instruments concerns.

[12]We use patients per-hour, rather than simply counting patients per-day, to account for the fact physicians' shift-lengths may vary across physicians and over time.

hours the physician spent at the clinic in that day. Figure 4 displays the results of this analysis. There appears to be no change in the number of patients a physician sees per hour on days on which a colleague is absent. Namely, the number of visits of the regular patients of the present physician, on a day of a colleague's absence, is not statistically different from this number in other days. This result further alleviates the concerns that absence affects the outcomes we measure in channels other than through its effect on workload.

Finally, in panel (b) we turn to the first stage performance of our alternative instrument, IV2. This results in exactly the same patterns as we saw for IV1. Column (1) shows a point estimate of -0.63, implying that seeing any of the absent physician's patients results in a decrease of 0.63 minutes in average visit length. Adding patient characteristics and visit level controls in column (2) does not change the estimates.

## 4.2 The second stage: the effect of workload on physician behavior

We now address our research question: what is the impact of workload on physician behavior? To that end, we estimate the following version of the model from equation (1):

$$y_{jsti} = \alpha + \nu_j + \nu_t + \beta_1 \cdot workload_{jst} + x_{jsti} \cdot \beta_2 + \epsilon_{jsti}, \tag{4}$$

where, again, $j, s, t, i$ index physicians, clinics, time and visits, respectively. The dependent variable $y_{jsti}$ is an indicator capturing an outcome of interest, e.g. referring a patient to a specialist. Our main explanatory variable is $workload_{jst}$, measuring physician $j$'s average visit length at clinic $s$, time $t$. We use IV1 and IV2 to instrument for this endogenous variable in 2SLS regressions.

## 4.3 The effect of workload: face-to-face visits

We first analyze the effect of workload on physician behavior in face-to-face encounters with patients, focusing on two sets of outcomes. The first involves diagnostic outcomes: the dependent variables are indicator variables for the following visit-level outcomes: referrals to specialists, referrals to imaging (such as x-ray, ultrasound, CT or MRI) and referrals to lab tests (e.g. blood or urine tests). The second set involves treatment outcomes: indicators for referrals to the emergency room, prescriptions of antibiotics, and prescriptions of pain killers. We estimate linear probability models and multiply estimates by one hundred.

#### 4.3.1 The effect of workload on diagnostic outcomes

It is not *a-priori* clear whether workload and diagnostic outcomes are substitutes or complements. On the one hand, when workload is higher, physicians may be able to substitute face-to-face time and physical examination with diagnostic procedures. On the other hand, under a tightening time constraint, physicians may limit the scope of the medical issues they address during the visit, and therefore may use fewer diagnostic procedures. This motivates empirical work of the sign and magnitude of this effect. Such estimates provide insights into the tradeoffs made by physicians under a tightening time constraint.

Table 3 displays the results of the diagnostic outcomes analysis. All specifications include year-month, day of the week and physician fixed effects. The time fixed effects are helpful in addressing the possibility that an omitted factor such as weather conditions affects both the absence of colleagues at the clinic, and the composition of patient medical conditions on a given day. The physician fixed effects, for their part, imply that we rely on within physician variation for identification. We therefore avoid basing our causal inference on differences in practice style among non-absent physicians.

In Panel (A), we analyze the overall utilization of diagnostic inputs by considering an indicator dependent variable, taking the value 1 if any of the diagnostic inputs considered (i.e., referrals to specialists, lab tests or imaging) are used, and zero otherwise. The OLS estimate of the effect of the daily average visit length, our measure of workload, is reported in column (1) to equal 0.48, indicating that time with patients and the use of diagnostic inputs are complements. Adding patient characteristics (including fixed effects for specific patient chronic conditions) in column (2) decreases the estimates slightly to 0.45.

Columns (3)-(6) address endogeneity concerns by using IV1 and IV2 as instruments for the workload variable. The sign of the coefficient on workload continues to be positive in all these specifications, reconfirming the OLS findings that the utilization of diagnostics is an increasing function of the time with patients. Using IV1, the estimated coefficient of 1.64 on the average visit length is positive and considerably larger than the OLS coefficients. This estimate is barely affected by adding patient characteristics in column (4). Using IV2 generates even larger coefficients of 1.86 and 1.84 in columns (5) and (6), respectively, showing again that the effect is not sensitive to the inclusion of patient characteristics.

To assess the quantitative implications, recall that diagnostic inputs are used in 35 percent of the visits. The results using IV1 therefore imply that a 1 minute decrease in average visit length causes a 4.6 percent decrease in the probability of utilization of diagnostic inputs.

Additional panels of Table 3 examine the impact of workload on the utilization rate of each diagnostic input separately. Panel (B) displays results concerning referrals to specialists. The results are qualitatively similar as those in Panel (A). Both OLS and IV coefficients on the average visit length are positive, suggesting that additional time with patients results in more referrals to specialists. Once again, controlling for patient characteristics has a very minimal effect on these estimates. Given the mean of the dependent variable, 0.14, the results using IV1 imply that a 1 minute decrease in average visit length causes a 9 percent decrease in the probability of a referral to a specialist.

Panel (C) repeats the analysis with the dependent variable being an indicator for a referral to lab test results. The exact same pattern of results obtains. Here too, the OLS estimates in columns (1) and (2) are positive. The IV estimates using the first instrument (column (3)-(4)) remain positive and they are much larger with point estimates of 0.76. The estimates that use IV2 are yet larger with point estimates of about 0.9. Given that 20 percent of visits result in a referral to a lab test, the results indicate that a 1 minute decrease in average visit length causes a 3.8 percent decrease in the probability of a referral to a lab test.

Finally, panel (D) reports results concerning referrals to imaging. The OLS estimates in columns (1) and (2) are positive with point estimate of 0.22 and 0.2 respectively. The IV estimates using IV1 (columns (3)-(4)) are positive, but are not statistically significant. The estimates that use IV2 are also statistically insignificant.

Taken together, the results in Table 3 indicate that, under a tightening time constraint, physicians reduce their utilization of diagnostic tools. The results are not driven by any particular diagnostic tool but rather hold for each type separately (except that the results for imaging are statistically insignificant). To interpret these results, we note two specific issues. First, issuing a referral entails an administrative and professional burden: the physician needs to come to the conclusion that the referral is medically justified, and to type into her computer a detailed note explaining the reasons and background for the referral. A tightening time constraint interferes with the ability to perform this task.

Perhaps more importantly, more time with patients may allow the physician to explore a wider scope of issues. She can take an overall look at the patient's health, suggesting routine checkups (e.g., blood work). She may also ask the patient about issues beyond the specific medical issue that brought the patient into the clinic. Consider, for example, an elderly patient who scheduled the appointment on account of a sore elbow. Given a tight schedule, the physician may quickly attend to the elbow issue and avoid discussing any other matters. Given more time, the physician can ask about additional symptoms such as fatigue, appetite, memory loss and so on, and suggest tests and examination by specialists on additional issues. This more comprehensive approach allows physicians to effectively administer preventive care, and our findings suggest that this element of primary care is restricted as the time constraint tightens.

**Heterogeneity.** While the results in Table 3 showed that controlling for patient characteristics did not affect the estimated impact of workload on the utilization of diagnostic inputs, it is still of interest to explore how the impact differs across patient populations. We explore this in Table 4, focusing on two margins of heterogeneity. The first is the patient's age: we split the sample to subsamples with patients above and below the age of 60. For each subsample we analyze the same regression model as in Table 3. A second dimension of heterogeneity we explore involves the extent to which the patients use the primary care system.

The reasons to explore heterogeneity are twofold. First, we wish to know whether the results reported above are driven by some particular patient segment, or, alternatively, characterize patients of all types. Second, this analysis can inform an important policy debate. According to a current report, young adults use the healthcare system less than groups older or younger than them, and are admitted to the emergency room at higher rates (Bonnie et al., 2015). Young adults may, therefore, fail to make sufficient use of the preventive care aspects of primary care, compared to older groups, despite the fact that, as stated by the report, "(y)oung adulthood provides an important opportunity for prevention" of serious illnesses and disorders.

It is, therefore, of interest to know whether physicians, possibly being aware of this issue, prioritize the provision of preventive care to younger patients (in the sense of referring them to specialists and tests). In our analysis, this could reflect itself in a smaller impact of workload on the use of diagnostics when it comes to younger patients. Put differently, if physicians are particularly aware of the importance of preventive care at young ages, they may not allow time constraints to deny those patients from referrals to specialists and tests as much as they do in the case of older patients.

We therefore begin by analyzing the potentially differential impact of workload on the utilization of diagnostic tools across patients above and below the age of 60. We report the results of this analysis in columns (1)-(6) of Table 4. Panel (A) of the table shows the results for the overall utilization of diagnostic inputs. The point estimates for IV1, in columns (2) and (5) are 2.04 and 1.39 for patients over 60 and patients aged 60 or less, respectively. These estimates indicate a 6% and a 3.6% decrease in utilization for the older and younger patients, respectively, in response to a 1 minute decrease in average visit length. This difference across subpopulations disappears, however, when we use IV2 rather than IV1.

Panel (B) shows a similar and even more pronounced pattern in referrals to specialists. Namely, the effect is stronger among older patients, using either IV1 or IV2. The referral to lab tests analysis in Panel (C) is also quite similar, but again the difference between the two age groups is only apparent with IV1. The referrals to imaging results, shown in Panel (D) remain insignificant.

To recap, there appears to be some evidence that the effect of workload on the use of diagnostic tools is smaller for younger patients. This is consistent with the hypothesis stated above: that physicians are aware of the fact that younger patients, to begin with, suffer from under-utilization of preventive care. Of course, there could also be other interpretations for this result. It could be, for example, that physicians communicate more effectively with younger patients and can therefore accomplish tasks quicker, including the referral of such patients to specialists or to lab tests. Testing among such competing hypotheses is beyond the scope of our study. Nonetheless, demonstrating the existence of this heterogeneity is informative regarding the channels via which workload affects the provision of preventive care.

The second dimension of heterogeneity we explore involves the patient's condition. Here, we take advantage of the fact that we observe detailed patient-level characteristics as well as the number of visits made over the sample period. We use this information to attribute a "utilization score" to each patient, capturing the extent to which the patient is likely to use the primary care system, based on her observed characteristics.

To this end, we analyze the data at the patient-year level and regress the number of visits per year against patient characteristics. We then use the regression to predict the number of visits for each patient in the data in every year. This prediction is defined as the patient's utilization score. Using the predicted, rather than the actual utilization rate helps us avoid a potential simultaneity bias.[13]

We then split our sample to high (low) utilization visits by patients with score above (below) the median utilization score, and repeat our analysis of the relationship between workload and the employment of diagnostic tools. The results of this analysis are displayed in columns (7)-(12) of Table 4. Panel (A) again addresses the impact of workload on the overall use of diagnostic inputs. The point estimates for IV1, reported in columns (8) and (11) are 1.86 and 1.45 indicating a 5.4% and a 3.9% decrease for high utilization and low utilization patients, respectively, in response to a 1 minute decrease in average visit length. This difference arises also with IV2. Panel (B), referrals to specialists, shows a similar pattern. Namely, that the effect is stronger among high utilization patients. Panel (C) also shows a stronger decrease in referrals to lab tests among high utilization patients. The referrals to imaging results, shown in Panel (D) are again insignificant.

It therefore appears that not only younger patients, but also low-utilization patients more generally, experience a smaller impact from their physician workload when it comes to the provision of diagnostic tools. It is plausible that physicians are able to identify high-utilization patients, and, since they predict that they will have more frequent opportunities to see those

[13]We describe the process of creating the utilization score in detail in Appendix B.

patients, they associate a smaller sense of urgency to addressing all their medical issues during the current visit. Such possibilities once again demonstrate that the manner with which physicians manage their workload is far from obvious, motivating empirical investigation.

#### 4.3.2 The effect of workload on the choice of treatment

We next analyze the relationship between workload and the choice of treatment. Here, one may hypothesize that under higher workload, physicians would tend to be more conservative and provide more treatment, namely, substitute office time and examination with treatment such as the prescription of antibiotics or painkillers, or referrals to the emergency room. As we shall see next, however, we find little support for this hypothesis.

We examine this issue in Table 5. We first analyze the overall provision of treatment. Analogously to the previous section, we use an indicator that takes the value 1 if any of the treatments we consider were used, and zero otherwise. Panel (A) shows the results of this specification, with columns (1) and (2) showing the OLS estimates without, and with accounting for patients characteristics, respectively. The estimate in column (1) is positive but when we add patient characteristics it becomes small and statistically insignificant. The instrumental variable estimates in columns (3)-(6) are all negative. The estimates with IV1 are insignificant with point estimates of about -0.5 percentage points. With IV2 the estimates are larger and statistically significant with point estimates of about -0.8 percentage points, that reflects an increase of about 5 percent in the probability of receiving treatment.

Unpacking this aggregate effect to consider specific treatment outcomes, Panel (B) reports estimates of the relationship between workload and referral to the emergency room. The OLS estimates in column (1) and (2) are positive and significant. However, the instrumental variable estimates in columns (3)-(6) are all very small and statistically insignificant. Estimates for specifications in which an indicator for the prescription of pain killers serve as the dependent variable are reported in Panel (C). Again, the OLS estimates are positive and significant, while the IV estimates are negative and statistically insignificant.

Finally, Panel (D) displays the results regarding the prescription of antibiotics. The OLS estimate in column (1) is -0.03 percentage points, and is statistically insignificant. Controlling for patient characteristics, the estimate, shown in column (2), is statistically significant at -0.04 percentage points. The estimates using IV1, shown in column (3)-(4), are -0.32 yet they are statistically insignificant. Using the second instrument, IV2 (recall this is an indicator for whether the physician sees any patients of an absent colleague) the estimate in column (5) is -0.51 and controlling for patient characteristics in column (6) the result becomes statistically significant with a point estimate of -0.55. As the probability of receiving antibiotics is on average 10 percent, This result implies that a 1 minute decrease in the daily mean visit length increases the visit-level probability of receiving antibiotics by 5 percent.

Overall, these results provide little evidence for a tendency to increase the intensity of treatment in response to higher physician workload. There appears to be no effect on the incidence of referrals to the emergency room or on the prescription of painkillers, yet there is some (mixed) evidence that increased workload tends to increase the use of antibiotics.

**Heterogeneity.** In Table 6, we explore heterogeneity in the effect of workload on treatment choices in much the same fashion as we did above in the case of the diagnostic tools analysis. Again, we allow for heterogeneity along two dimensions: age, and the patient's utilization score.

Here, our main motivation is to determine whether our conclusion from the analysis that used the entire sample — that there appears to be little effect of physician workload on the course of treatment — holds within each subsample as well. In principle, it could be that the "non-result" in the full sample masks important heterogeneity across groups. Table 6 shows, however, that this is not the case: little evidence for a workload effect is obtained within each subsample, just like in the full sample.

More specifically, we report the results from the analysis by age in columns (1)-(6) of Table 6. Panel (A) of the table shows the results for the overall provision of treatment. The point estimates for IV1, in columns (2) and (5) are -0.89 and -0.3 for patients over 60 and patients aged 60 or less, respectively. These estimates are statistically indistinguishable from zero. With IV2, the corresponding point estimates are -1.25 and -0.57, respectively, for patients above and below the age of 60. The negative effect for those above the age of 60 is statistically significant. This suggests that older patients do obtain more intense treatment when physicians experience higher workload, but, again, this finding obtains only with IV2.

The referral to the emergency room results in Panel (B) are statistically insignificant in all IV specifications. In Panel (C) we report the results for the prescription of pain killers as the dependent variable. The estimates using IV1 and IV2 are -0.65 and (a significant) -0.72 for older patients and 0.06 and 0 for the younger patients. In other words, there appears to be an increase in prescription of pain killer only in older patients — but, as in the case of the overall treatment intensity results from Panel A, this only obtains under IV2. The results regarding prescription of antibiotics, reported in Panel (D) are all statistically insignificant.

To summarize the analysis of the impact of age, while there is some evidence that the probability of being prescribed painkillers increases with physician workload, this evidence does not appear strong as it is only observed when using one of our instrumental variables. With respect to other forms of treatment, we continue to see no effect of workload, consistent with the full-sample analysis from Table 5.

The results of the analysis by patient condition are displayed in columns (7)-(12) of Table 6. By and large, these regressions deliver insignificant estimates of the effect of workload — again reconfirming the "non-result" obtained in the full sample analysis.

## 4.4 The effect of workload on subsequent face-to-face encounters

Our results indicate that high workload diminishes the probability with which physicians refers patients to specialists and to lab tests. We next examine whether physicians and patients are able to make up for this by simply meeting again, at a later date, to complete such tasks under more suitable conditions. Specifically, a physician may ask the patient to return to the clinic in a couple of weeks to complete the exploration of a non urgent matter. Or, the patient may decide to come back and see the physician again, if she feels that some medical issue has not received proper attention on account of the physician's workload. If such behaviors are prevalent, the harm to the patient's long-run health conditions caused by her primary care physician's temporary workload could be minimal, and cause no (or little) concern.

To investigate such possibilities, we examine the impact of workload at a given visit on the probability of there being subsequent face-to-face visits within a reasonable time period. We create four indicator variables, taking the value one if, after the visit, the patient arrives at the clinic again within 15, 30, 60 or 90 days, respectively. These are then used as dependent variables in regressions of the same nature as those reported above.

Table 7 reports the results. It is immediately apparent that we find no statistically significant impact of the physician's workload on a given visit on the probability of such subsequent patient visits. This holds true whether or not we use our instruments to address the endogeneity concerns with respect to the workload variable, and whether or not we control for patient characteristics. Table 8 reveals that the same non-result obtains for subsamples defined by age and utilization scores.

To conclude, we find no evidence that the system "corrects itself" by generating additional opportunities for physicians to provide adequate consideration of patients' medical issues, if, on account of intense workload, such issues did not receive proper attention originally.

## 4.5 The effect of workload on non face-to-face encounters

We complete our baseline results by examining how workload affects the quantity of non face-to-face physician-patient encounters. Particularly, we look at two outcomes: responses to patients' online queries, and phone calls made with patients.

Patients can make an online query to their physician using the Clalit Health Services website. The vast majority of those queries are prescription renewals and administrative requests. Response time is four business days and the Clalit Health Services website explicitly notes that such queries are not intended to be used for urgent issues. Physicians respond to those queries during the work day. This is an interesting outcome to explore since, unlike face-to-face office visits, physicians have a substantial degree of freedom to choose the timing in which they respond to those queries. Thus, this is a "cheap" channel, in terms of patient well being, that is available to physicians to manage their time. Our hypothesis is, therefore, that higher workload would result in a decreased number of replies to queries that physicians make during the day.

Another mode of communication between physicians and patients is phone calls. The extent to which a physician returns phone calls is another channel via which she may manage her time under the workload constraint. Therefore, we expect a reduced number of phone calls between physicians and their patients when workload is higher.

We analyze the response of the daily amount of online queries and phone calls to the daily level of workload, measured, as before, by the physician's average visit length (in face-to-face visits). The analysis is therefore conducted at the physician-day level and is summarized in Table 9. Panel (A) displays the online queries results. The OLS estimate in column (1) is negative, but the IV estimates in columns (2) and (3) are positive with (statistically significant) magnitudes of 0.07 and 0.08, respectively, using IV1 and IV2. These estimates imply that a 1 minute decrease in average visit length decreases responses to queries by 4 percent. Since the mean visit length is 11.5, the estimates reflect an elasticity of about 0.5.

Panel (B) of the table summarizes the results for phone calls, displaying a similar pattern: the OLS estimate in column (1) is, again, negative, while the IV estimates in columns (2) and (3) are again positive. The estimates imply that a 1 minute decrease in average visit length decreases responses to phone calls by 10 percent, reflecting an elasticity of 1.1.

Consistent with our expectation, therefore, higher workload is associated with fewer responses to online queries and fewer phone calls with patients.

**Baseline analysis takeaways.** Overall, our baseline analysis, presented in this section, indicated that the intensity of treatment is not significantly affected by workload. Instead, physicians adjust their choices via other channels to cope with higher workload: they respond less to online queries and perform fewer phone call with patients, and, importantly, make a lesser use of diagnostic tests or referrals to specialists. We also note that Appendix A provides reduced-form regressions (i.e., where the explanatory variable is the instrument) that further support these findings.

These results inform us regarding the nature of the shadow cost of physician capacity: it does not involve the oversubscription of medication, but rather a poorer long-term management of patient health via a reduced amount of diagnostic tests and thorough examination of medical problems by specialists.

# 5 Nonparametric bounds on the effect of physician workload on the utilization of diagnostic inputs

The baseline analysis in the previous section entailed several important features: an exclusion restriction, a linear functional form, and concrete interpretations of the estimated coefficient of interest. Namely, the 2SLS estimate of the coefficient on the workload measure may be interpreted either as a homogeneous treatment effect across units, or as a Local Average Treatment Effect, associated with a "complier" population.[14]

In this section we complement the baseline analysis by exploring additional sets of identifying assumptions. Specifically, we explore nonparametric bounds on the average treatment effect (ATE) of workload on physician behavior following Manski and Pepper (2000). We first explain how such bounds are derived under an exclusion restriction akin to the one used above. Then, we explore the possibility of imposing a weaker assumption, Monotone Instrumental Variable (MIV), on the relationship between our instrument and the response function. Finally, we combine the MIV assumption with a Monotone Treatment Selection (MTS) assumption that is intuitively linked to the nature of the endogeneity concern.

We focus on a particular set of outcomes — the use of diagnostic inputs such as referrals to specialists and tests — as these are the most important outcomes for which a workload effect was prominently documented in the baseline analysis above. We emphasize that we view the baseline linear models, and the nonparametric bounds approach developed next, as complementary strategies. Each entails costs and benefits, and, taken together, they provide a broad picture of the information that can be gleaned from the data under alternative assumptions. We return to this below when we summarize the main findings from this exercise.

## 5.1 Notation and assumptions

Adopting the framework and notation from Manski and Pepper (2000), the population of interest contains a set $j \in \mathcal{J}$ of individual units: in our case, patient visits. Each individual

[14] This latter interpretation rests on an additional assumption that there are no "defier" units, i.e., units that would be characterized by lower workload given the absence of a colleague, and by higher workload if such absence does not occur. See Imbens and Angrist (1994).

unit is characterized by a *response function* $y_j(\cdot): T \to Y$, where $t \in T$ are discrete treatments, and $y \in Y$ are discrete outcomes.

In our application the set $\mathcal{J}$ contains all visits (triplets of patient, physician and day). Our outcome space is binary, i.e., $Y = \{0, 1\}$: a patient is either referred to some diagnostic procedure (e.g., an examination by a specialist or a lab test), or not. In the baseline linear model, the treatment $t$ (the level of physician workload on the relevant day, measured by the average visit length) was a continuous measure. In the nonparametric analysis, to keep the framework as transparent as possible, we define this treatment as a binary variable, so $T = \{0, 1\}$. A value of $t = 1$ implies that the physician experiences workload above a certain threshold (we set the threshold at the 75th percentile of the physician-specific workload distribution), while $t = 0$ implies values below or at that threshold.

The observed variables in this framework are $(x, z, y)$, where $x_j$ is a covariate vector for visit $j$. The variable $z_j \in T$ is the *realized treatment*. That is, it takes the value 1 for visits that take place on a day when the physician is *observed* to experience higher-than-normal workload, and zero otherwise. Finally, $y_j = y_j(z_j)$ is the observed outcome.

The covariate vector can be written as $x = (w, \nu) \in \mathcal{X} = W \times V$, where $\nu \in V$ is our instrument: the daily share of patients seen by the physician that are attributed to the absence of a colleague. In the baseline linear model, the instrument was treated as a continuous variable. In the nonparametric analysis, we treat the space of instrument values $V$ as discrete, dividing it into 10 bins ranging from 0 to 0.4, with a fixed width of 0.04. We do not consider days in which more than 40 percent of the patients seen by the physician are due to an absent colleague. Those events are rare and likely not representative of the phenomenon of interest.

Our object of interest is the distribution $\mathcal{P}[y(\cdot)]$ of response functions, or, its conditional version $\mathcal{P}[y(\cdot)|w]$. Specifically, as we focus on a binary outcome the ATE is defined by:

$$ATE(1, 0|w) = P[y(1)|w] - P[y(0)|w] \tag{5}$$

Namely, we are interested in the treatment effect of physician workload on the probability with which the physician employs diagnostic outcomes. A negative ATE would correspond to the findings from the baseline linear model, where physician time and the use of diagnostic tools were found to be complements.

We next provide two alternative identifying assumptions that characterize the relationship between the instrument, the treatment (i.e., the workload level), and the outcome.

**Assumption 1** ***IV:***

$$E\big[y(t)|w, \nu = u'\big] = E\big[y(t)|w, \nu = u\big] \;\; \forall t \in T, w \in W, (u, u') \in V \times V$$

In words, the IV assumption does not allow the instrument to affect the response function. It can, therefore, only affect the outcome via its effect on the treatment $t$. This corresponds to the exclusion restriction employed in the baseline linear analysis.

Alternatively, a weaker Monotone Instrumental Variable assumption allows the response function to depend on the instrument, but in a pre-specified direction:

**Assumption 2** ***MIV:***

$$E\big[y(t)|w, \nu = u_2\big] \geq E\big[y(t)|w, \nu = u_1\big] \;\; \forall t \in T, w \in W, (u_1, u_2) \in V \times V \;\; \textit{such that } u_2 \geq u_1$$

In the context of the outcome we examine, the MIV assumption allows the absence of a colleague to be associated with higher use of diagnostic inputs *conditional on the physician's workload level.* This may hold if unobserved factors such as an especially cold winter affect both patient and physician health, causing the absence of a colleague to be associated with a sicker population of patients. Such scenarios would violate the IV assumption, but are allowed under MIV. In the baseline linear analysis we addressed this concern via year-month fixed effects. The MIV assumption presents an alternative way of addressing the same issue.[15]

We combine the MIV and IV assumptions with yet another assumption presented in Manski and Pepper (2000), Monotone Treatment Selection (MTS). The MTS essentially allows the

[15]Recall also another concern discussed above: that an absent colleague creates longer wait times that cause patients with non-acute medical complaints to give up their appointment. The absence may then affect not only the workload, but increase the tendency to utilize diagnostic inputs conditional on the workload level. As shown in previous sections, we find little evidence for such mechanisms in the data. Nonetheless, it is interesting that the MIV assumption allows for consistent estimation even if such a mechanism were, in fact, present.

realized treatment $z$ to serve as an additional monotone instrumental variable:

**Assumption 3** ***MTS:***

$$E\big[y(t)|w, \nu = u, z = 1\big] \geq E\big[y(t)|w, \nu = u, z = 0\big] \ \forall t \in T, w \in W, u \in V$$

The MTS assumption implies that the response function shifts upward, at any treatment level (either $t = 1$ or $t = 0$), if the *realized* treatment is that of high workload, i.e., if $z = 1$. This would hold, for example, if congestion in the waiting room given the realization of a high-workload day deters patients with non-acute conditions from keeping their appointment, thereby inducing a sicker population of examined patients. This sicker population has a higher probability of requiring the use of diagnostics (e.g., tests) regardless of the level $t$ of their physician's workload, which is precisely what the MTS assumption asserts.

This assumption is, therefore, consistent with the fundamental endogeneity concern: that the realized treatment may be associated with the outcome via confounding variables, in this case, the patient pool that ends up being examined. Incorporating it into the analysis allows us to obtain additional identifying power, on top of that provided by the IV or MIV assumptions.

Note that the version of the MTS assumption employed here conditions on the instrument $\nu$ (i.e., on the share of patients seen by the physician that are accounted for by a missing colleague), in addition to the covariate vector $w$. In Manski and Pepper (2000), the MTS assumption is not used in conjunction with an additional instrumental variable, and so the conditioning on $\nu$ does not arise there. De Haan and Leuven (2016) combine the MTS assumption along with an additional MIV assumption, as we do here. As a consequence, the MTS assumption in that paper also conditions on the value of the additional instrumental variable.

### 5.1.1 Bounds under IV assumption

Following Manski and Pepper (2000) exactly, to derive bounds on the ATE, the Law of Iterated Expectations can first be used as follows:

$$(6) \qquad E\big[y(t)|w,\nu=u\big] = E\big[y|w,\nu=u,z=t\big]\cdot Pr(z=t|w,\nu=u) + E\big[y(t)|w,\nu=u,z\neq t\big]\cdot Pr(z\neq t|w,\nu=u)$$

Recalling that our observed variables are $(y, w, \nu, z)$, the only unidentified object on the RHS of this expression is the "counterfactual" outcome under the treatment that was not realized, $E\big[y(t)|w,\nu = u, z \neq t\big]$. As the outcome $y$ only takes the values 0 or 1, we can bound this unknown term by 0 from below, and by 1 from above, resulting in the following bounds:

$$(7) \qquad \underline{b}(w,u,t) \leq E\big[y(t)|w,\nu=u\big] \leq \overline{b}(w,u,t), \text{ with}$$

$$\underline{b}(w,u,t) \equiv E\big[y|w,\nu=u,z=t\big]\cdot Pr(z=t|w,\nu=u)$$
$$\overline{b}(w,u,t) \equiv E\big[y|w,\nu=u,z=t\big]\cdot Pr(z=t|w,\nu=u) + P(z\neq t|w,\nu=u)$$

These are often referred to as the "no assumption bounds." We next incorporate the IV assumption to derive bounds on the objects of interest. The IV assumption (along with the Law of Iterated Expectations) implies that $E\big[y(t)|w,\nu=u\big] = E\big[y(t)|w\big]\ \forall u \in V$.As a consequence, each value of the instrument generates bounds on the same quantity of interest $E\big[y(t)|w\big]$, and we can obtain the tightest upper (lower) bounds on it by sweeping over the instrument values to obtain the smallest (largest) values. We can therefore bound $E\big[y(t)|w\big]$:

$$(8) \qquad \max_{u\in\mathcal{V}} \underline{b}(w,u,t) \leq E\big[y(t)|w\big] \leq \min_{u\in\mathcal{V}} \overline{b}(w,u,t)$$

Condition (8) defines a lower bound, $LB^t \equiv \max_{u\in\mathcal{V}} \underline{b}(w,u,t)$, and an upper bound, $UB^t \equiv \min_{u\in\mathcal{V}} \overline{b}(w,u,t)$, on the average value of the response function, evaluated at a specific treatment level $t$. In our case, this average value is the probability of the binary outcome (the use of diagnostics) taking place given a specific workload level. Recalling that $t = 0, 1$ values correspond to low and high workload levels, respectively, we obtain bounds on the ATE:

$$[LB^1 - UB^0, UB^1 - LB^0] \tag{9}$$

The quantities $LB^1, UB^0, UB^1$, and $LB^0$ are easily estimated using nonparametric methods.

### 5.1.2 Bounds under MIV

Continuing to follow Manski and Pepper (2000), by the weaker MIV assumption, for any $(u, u_1, u_2) \in V \times V \times V$ such that $u_1 \leq u \leq u_2$, we have:

$$E\big[y(t)|w, \nu = u_1\big] \leq E\big[y(t)|w, \nu = u\big] \leq E\big[y(t)|w, \nu = u_2\big]$$

Combining with the no-assumptions bounds in (7), we obtain for any $u_1 \leq u \leq u_2$:

$$\underline{b}(w, u_1, t) \leq E\big[y(t)|w, \nu = u\big] \leq \overline{b}(w, u_2, t)$$

Fixing $u$, and sweeping over all values $(u_1, u_2)$ that satisfy the condition $u_1 \leq u \leq u_2$, we obtain bounds on $E\big[y(t)|w, \nu = u\big]$:

$$\max_{u_1 \leq u} \underline{b}(w, u_1, t) \leq E\big[y(t)|w, \nu = u\big] \leq \min_{u_2 \geq u} \overline{b}(w, u_2, t) \tag{10}$$

But we are interested in bounds on $E\big[y(t)|w\big]$. To eliminate the conditioning on the instrument $\nu$, we use the Law of Iterated Expectations to obtain:

$$E\big[y(t)|w\big] = \sum_{u \in \mathcal{V}} E\big[y(t)|w, \nu = u\big] \cdot Pr(\nu = u)$$

By substituting the bounds from (10), we obtain the following:

$$(11) \qquad \sum_{u\in\mathcal{V}}\Big[Pr(\nu=u)\max_{u_1\le u}\underline{b}(w,u_1,t)\Big] \le E\big[y(t)|w\big] \le \sum_{u\in\mathcal{V}}\Big[Pr(\nu=u)\min_{u_2\ge u}\overline{b}(w,u_2,t)\Big]$$

Equation (11) therefore provides bounds on $E\big[y(t)|w\big]$ under the MIV assumption. Bounding the ATE follows in the same fashion as described above for the IV case.

#### 5.1.3 Tightening the bounds using the MTS assumption

We can combine the MTS assumption together with either the IV or the MIV assumptions above to tighten the upper bounds on $E\big[y(t)|w\big]$ that appear in (8) or (11), respectively, for the high workload treatment value $t = 1$.

Recalling the no-assumptions bounds in (6), for the high workload case ($t = 1$) we obtain:

$$\begin{aligned} E\big[y(1)|w,\nu=u\big] &= E\big[y|w,\nu=u,z=1\big]\cdot P(z=1|w,\nu=u) \\ &\quad + E\big[y(1)|w,\nu=u,z=0\big]\cdot P(z=0|w,\nu=u) \end{aligned}$$

Our next step was to bound the "counterfactual" expression $E\big[y(1)|w,\nu=u,z=0\big]$ from above by 1, yielding the upper bound $\overline{b}(w,u,1)$ as defined in (7). The MTS assumption, however, asserts that $E\big[y(1)|w,\nu=u,z=0\big] \le E\big[y(1)|w,\nu=u,z=1\big]$, allowing us to bound the counterfactual expression from above by an expression that is potentially smaller than 1.

We can now, therefore, define a tighter upper bound on $E\big[y(1)|w,\nu=u\big]$ than the one that was defined in (7). This tighter upper bound, denoted $\overline{b}_{MTS}(w,u,1)$, is given by:

$$\begin{aligned} \overline{b}_{MTS}(w,u,1) &\equiv E\big[y|w,\nu=u,z=1\big]\cdot P(z=1|w,\nu=u) \\ &\quad + E\big[y|w,\nu=u,z=1\big]\cdot P(z=0|w,\nu=u) \\ &= E\big[y|w,\nu=u,z=1\big] \end{aligned}$$

To combine the MTS assumption with the IV assumption, we only need to substitute

$\bar{b}_{MTS}(w, u, 1)$ for $\bar{b}(w, u, 1)$ in (8). This results in what we refer to below as the IV-MTS bounds. Compared to the IV bounds, the IV-MTS obtains a tighter upper bound on $E[y(t)|w]$ for $t = 1$. Similarly, to combine MTS with MIV, we make the same substitution into (11), resulting in the MIV-MTS bounds.

### 5.2 Nonparametric bounds: estimation results

In this section we report the estimated sets on the Average Treatment Effects of interest. We focus attention on the ATE of physician workload on the probability with which physicians use diagnostic inputs, which was found to be negative in the baseline linear analysis.

We begin with the IV bounds obtained in (8). Figure 5 provides a graphic illustration of these bounds, while Panel (A) of Table 10 provides the estimates.

Panel (a) of Figure 5 displays bounds on the probability of using any diagnostic input. The vertical red and blue lines show the bounds under low-workload and high-workload, respectively. The figure provides these bounds within the full sample, and within particular subsamples such as patients above the age of 60. Formally, the subsamples correspond to different conditioning covariates, denoted by $w$ in (8), and they match the heterogeneity specification in the baseline linear analysis. Panel (b) displays the bounds on the ATE — the average effect of a switch from low to high workload on the probability of using diagnostic inputs, shown in (9).

In all subsamples, the upper bound on the estimated probability given low workload lies above the upper bound given high workload, and the same holds for the respective lower bounds. Nonetheless, it is not possible to sign the treatment effect. This can be seen in Panel (a) because the intervals on the probability of using diagnostics given high and low workload always overlap, or in Panel (b), as the bounds on the ATE contain the value of zero. The same information is displayed numerically in panel (A) of Table 10. For example, the estimated set for the ATE within the full sample is $[-.321, 0.138]$. The 95 percent confidence interval is, of course, wider, at $[-.345, .179]$.[16] The confidence intervals in the IV case are displayed graphically in Panel

[16] We follow the common practice in the relevant literature (e.g., de Haan and Leuven 2016, Kreider et al. 2012) of relying on Imbens and Manski (2004) to derive the 95% confidence intervals. Those are obtained using 100 bootstrap replications. As our sample size is much larger than that commonly used in such applications, we do not perform the correction for finite sample bias prescribed in these papers.

(b) of Figure 5 and reported numerically in Columns (7)-(8) of Table 10.

We next present results, in a similar fashion, for the bounds generated under the weaker MIV assumption, given in (11). Figure 6 provides the graphic illustration of the bounds under MIV assumption and Panel (B) of Table 10 provides the corresponding estimates. The same pattern is observed: the intervals on the probability of using diagnostics given low workload are higher than those given under high workload — but they overlap. Therefore, once again, it is not possible to sign the ATE. This is hardly surprising, given that the MIV assumption is weaker than the IV assumption employed above.

Comparing the IV to the MIV results by examining, again, Panels (A) and (B) of Table 10, we note that the MIV lower bound is always smaller than the IV lower bound. This is natural, given that the MIV assumption is weaker than the IV assumption. At the same time, the two assumptions deliver the exact same upper bounds. This happens in our case because the smallest "no-assumptions" upper bound $\bar{b}(w, u, t)$ (defined in (7)) is obtained at the largest value of $u$ (recall that this is a value of our instrument $\nu$, the fraction of patients of a missing colleague seen by the physician, which values have been divided into discrete bins for the purpose of the bounds analysis).

As a consequence of this, the upper IV bound $min_{u \in V}\bar{b}(w, u, t)$ in (8) coincides exactly with the MIV upper bound which is the solution to $min_{u_2 \geq u}\bar{b}(w, u_2, t)$ for all $u$ in (11). Interestingly, Manski and Pepper (2000) mention that the MIV and IV assumptions would yield exactly the same identifying information (i.e., generating the same upper *and lower* bounds) if the no-assumptions bounds were to weakly decrease with $u$. This does not hold in our case and, again, only the upper bounds delivered by the two assumptions coincide in our analysis.

We next use the MTS assumption in conjunction with the IV and the MIV assumptions to obtain a tighter upper bound on the probability of using diagnostics given high workload ($t = 1$). The IV-MTS and MIV-MTS bounds are presented in Panels (A) and (B) of Table 11, respectively. Corresponding graphic illustrations are available in Figures 7 and 8.

Columns (5) and (6) in Table 11 show that the estimated intervals for the ATE of workload on the probability of using diagnostic inputs now contain only negative values, in both the IV-MTS and the MIV-MTS cases. For example, under MIV-MTS, within the full sample, this interval is $[-.432, -.025]$. The same pattern holds for all other subsamples.

The estimated intervals under MIV-MTS therefore imply — consistent with our baseline linear analysis — that physician time and the employment of diagnostic tools are complements. As workload intensifies, and less time can be spent with patients, the probability of using diagnostics declines by 2.5 to 43.2 percent.

The 95 percent confidence intervals for this ATE, however, do contain the value of zero: for example, under MIV-MTS, within the full sample, the confidence interval is $[-.438, .013]$. One cannot, therefore, reject the hypothesis that the average effect of workload on the probability of using diagnostic tools is zero. These patterns are demonstrated graphically in Figures 7 and 8: the estimated intervals for the relevant probabilities under low workload lie above those that correspond to high workload. The confidence intervals, however, very mildly overlap.

Despite this overlap, the MIV-MTS analysis is informative regarding the key question of whether physician time and the use of diagnostic tools are complements or substitutes. Whether within the full sample or in various subsamples, the ATE is bounded between a substantial negative value, and a very small positive value. Full-sample results imply that diagnostics could be very strong *complements* to physician time: switching from low to high workload may reduce the probability of using diagnostics by as much as 43.8 percent. At the same time, diagnostics cannot be strong *substitutes* for time: at the upper bound, a switch to high workload could increase the probability of using diagnostics by at most 1.3 percent.

The MIV-MTS analysis therefore provides strong evidence against the possibility that physician time and the use of diagnostic tools are substitutes, while leaving a substantial scope for complementarity. We next summarize the overall message that emerges from this nonparametric bounds analysis, and from our baseline linear analysis of the same question.

**The effect of workload on the use of diagnostic tools: takeaways.** Our baseline analysis, relying on a linear functional form, and on a strict exclusion restriction, provided a negative, statistically-significant point estimate for the effect of workload on the probability of using diagnostic tools, suggesting that physician time and the employment of such tools are complements. This estimated effect could be interpreted either as a homogeneous treatment effect pertaining to all units, or as a Local Average Treatment Effect.

The nonparametric bounds analysis employing the combined MIV-MTS assumptions complements the baseline analysis by pursuing different assumptions, and by estimating a different object. It does not impose a linear functional form, and relies on an MIV assumption which is weaker than the strict exclusion restriction used in the baseline analysis. It also adds the MTS assumption which we view as a natural reflection of the very endogeneity issue we tackle. In terms of the estimated object, rather than providing a point estimate of a LATE, the MIV-MTS delivers bounds on the ATE.

The estimated MIV-MTS bounds reinforce the results from our baseline linear analysis: they effectively rule out the possibility that diagnostic tools serve as substitutes to physician time in a quantitatively important fashion, while leaving a substantial scope for complementarity. The partial identification techniques employed here complement the classic use of instrumental variables in the familiar linear model. By presenting both the linear IV results and the nonparametric MIV-MTS results, we effectively offer a sequence of identifying assumptions that result in a sequence of conclusions. This allows us to transparently demonstrate what the data tell us regarding the effect in question under different identifying assumptions.

Across these analyses, we obtain the robust finding that diagnostic tools do not serve as a substitute for physician time. The baseline linear results go further and unequivocally establish that those tools are complements to physician time. The sections above provided very detailed arguments in favor of the exclusion restriction employed in our baseline analysis, and so *we view the complementarity result as robust*. But even if one were to discredit this exclusion restriction and place more trust in a weaker assumption such as MIV, that person would still learn the valuable information that the ATE of workload on the probability of using diagnostics is not likely to be positive at an economically meaningful level, and that the data do not rule out the possibility that this effect is strongly negative.

From an economic perspective, the linear IV analysis implies that the shadow cost of physician capacity includes an important component: a reduction in the utilization of diagnostic tools, and, therefore, in the ability to provide effective preventive care. Our nonparametric MIV-MTS analysis does not contradict this conclusion, and, at the very least, inform us that the contrary is not true: that is, we learn that the shadow cost of capacity does not include intensified referrals to tests and examination by specialists. We, therefore, obtain a multi-faceted picture of the intricate role played by physician workload in determining clinical courses of action, and, ultimately, in the administration of primary care.

# 6 Conclusions

In this study we examine the effect of workload on physician behavior. We find that when workload is higher physicians respond less to online patient queries and make fewer phone calls with patients. Physicians' utilization of diagnostic inputs decreases with workload (or, at least, does not increase - i.e., physicians do not use these diagnostic tools as a substitute to time with patients). Finally, increased workload does not affect the intensity with which physicians refer patients to the emergency room or prescribe pain killers. There is some very weak evidence that the prescription of antibiotics may increase with workload.

Our identification strategy relies on the absence of colleagues as a source of exogenous variation in workload. This setting corresponds conceptually to a very simple "comparative statics" analysis, in which we are able to examine the effect of a temporary increase in workload on physicians' behavior holding other parameters of the environment constant.

From a policy perspective, the responses we find should be accounted for in determining the optimal size of the workforce, i.e., the capacity of primary care physicians hours with patients. Our results indicate what components are included in the shadow cost of physicians' time constraint — and what components are not included. For example, while one may have expected physicians to prescribe more medication and refer more intensively to the emergency room as they become busier, the data does not provide support for such mechanism.

We paid particular attention to the relationship between workload and the utilization of diagnostic tools such as referrals to specialists or to lab tests. A-priori, such diagnostic tools could be used as substitutes to physician time (i.e., physicians could use them more intensely as they become busier to compensate for the shorter time they have with patients), or they could be complements to physician time.

Our baseline results suggested that, as the time constraint becomes more stringent, physicians substantially reduce their use of diagnostic tools. This would imply that the shadow cost of this constraint may include a very important reduction in the administration of preventive care. We then reexamine this possibility via a nonparametric bounds analysis (MIV-MTS). This additional analysis is largely consistent the conclusions of the baseline linear analysis: at the very least, it tells us that diagnostic tools are not used as substitutes to physicians' time in a quantitatively important fashion, while leaving a very broad scope open for the possibility that they are, in fact, complements — just as our baseline analysis asserted.

Additional research, using varied data sources and identification strategies, may further enhance our understanding of the important relationship between primary care physicians' time with patients and their ability to provide preventive care.

# References


Krishnan S Anand, M Fazil Pac, and Senthil Veeraraghavan. Quality-speed conundrum: Trade-offs in customer-intensive services. *Management Science*, 57(1):40–56, 2011.

Robert J Batt and Christian Terwiesch. Doctors under load: An empirical study of state-dependent service times in emergency care. *Working Paper, The Wharton School*, 1, 2012.

Jay Bhattacharya, Azeem M Shaikh, and Edward Vytlacil. Treatment effect bounds: An application to swan–ganz catheterization. *Journal of Econometrics*, 168(2):223–243, 2012.

RJ Bonnie, C Stroud, H Breiner, et al. Investing in the health and well-being of young adults. 2015.

Christopher S Brunt, John Bowblis, and Joshua R Hendrickson. Physician competition and quality of care: Empirical evidence from medicare's physician quality reporting system. Technical report, Mimeo, 2018.

David C Chan. Teamwork and moral hazard: evidence from the emergency department. *Journal of Political Economy*, 124(3):734–770, 2016.

Olivier Chatain and Alon Eizenberg. Demand fluctuations, capacity constraints and repeated interaction: An empirical analysis of service quality adjustments. Technical report, CEPR, DP10545, 2015.

Jeffrey Clemens and Joshua D Gottlieb. Do physicians' financial incentives affect medical treatment and patient health? *American Economic Review*, 104(4):1320–49, 2014.

Janet Currie and W Bentley MacLeod. First do no harm? tort reform and birth outcomes. *The Quarterly Journal of Economics*, 123(2):795–830, 2008.

Janet Currie, W Bentley MacLeod, and Jessica Van Parys. Provider practice style and patient health outcomes: the case of heart attacks. *Journal of health economics*, 47:64–80, 2016.

Monique De Haan. The effect of parents' schooling on child's schooling: a nonparametric bounds analysis. *Journal of Labor Economics*, 29(4):859–892, 2011.

Monique De Haan and Edwin Leuven. Head start and the distribution of long term education and labor market outcomes. *CESifo Working Paper No. 5870*, 2016.

Amy Finkelstein, Matthew Gentzkow, and Heidi Williams. Sources of geographic variation in health care: Evidence from patient migration. *The quarterly journal of economics*, 131(4): 1681–1726, 2016.

Sebastián Fleitas. Who benefits when inertia is reduced? competition, quality and returns to skill in health care markets. Technical report, Mimeo, 2018.

Michael Frakes. The impact of medical liability standards on regional variations in physician behavior: evidence from the adoption of national-standard rules. *American Economic Review*, 103(1):257–76, 2013.

Seth Freedman, Ezra Golberstein, Tsan-Yao Huang, David Satin, and Laura Barrie Smith. Docs with their eyes on the clock? the effect of time pressures on primary care productivity. *mimeo*, 2018.

Martin Gaynor, James B Rebitzer, and Lowell J Taylor. Physician incentives in health maintenance organizations. *Journal of Political Economy*, 112(4):915–931, 2004.

Michael Gerfin and Martin Schellhorn. Nonparametric bounds on the effect of deductibles in health care insurance on doctor visits–swiss evidence. *Health economics*, 15(9):1011–1020, 2006.

Libertad Gonzalez. Nonparametric bounds on the returns to language skills. *Journal of Applied Econometrics*, 20(6):771–795, 2005.

Kate Ho and Ariel Pakes. Physician payment reform and hospital referrals. *American Economic Review*, 104(5):200–205, 2014.

Kate Ho and Adam M. Rosen. Partial identification in applied research: Benefits and challenges. *NBER Working paper 21641*, 2015.

FD Richard Hobbs, Clare Bankhead, Toqir Mukhtar, Sarah Stevens, Rafael Perera-Salazar, Tim Holt, Chris Salisbury, et al. Clinical workload in uk primary care: a retrospective analysis of 100 million consultations in england, 2007–14. *The Lancet*, 387(10035):2323–2330, 2016.

Guido W Imbens and Charles F Manski. Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857, 2004.

Diwas S Kc and Christian Terwiesch. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science*, 55(9):1486–1498, 2009.

Song-Hee Kim, Carri W Chan, Marcelo Olivares, and Gabriel J Escobar. Association among icu congestion, icu admission decision, and patient outcomes. *Critical Care Medicine*, 44(10): 1814–1821, 2016.

Brent Kreider, John V Pepper, Craig Gundersen, and Dean Jolliffe. Identifying the effects of snap (food stamps) on child health outcomes when participation is endogenous and misreported. *Journal of the American Statistical Association*, 107(499):958–975, 2012.

Charles F Manski and John V Pepper. Monotone instrumental variables: with an application to the returns to schooling. *Econometrica*, 68(4):997–1010, 2000.

Justin McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, 142(2):698–714, 2008.

Magne Mogstad, Andres Santos, and Alexander Torgovitsky. Using instrumental variables for inference about policy relevant treatment effects. *Econometrica (forthcoming)*, 2018.

Hannah T Neprash. Better late than never? physician response to schedule disruptions. Technical report, Mimeo, 2016.

Olga Perdikaki, Saravanan Kesavan, and Jayashankar M Swaminathan. Effect of traffic on sales and conversion rates of retail stores. *Manufacturing & Service Operations Management*, 14 (1):145–162, 2012.

Adam Powell, Sergei Savin, and Nicos Savva. Physician workload and hospital reimbursement: Overworked physicians generate less revenue per patient. *Manufacturing & Service Operations Management*, 14(4):512–528, 2012.

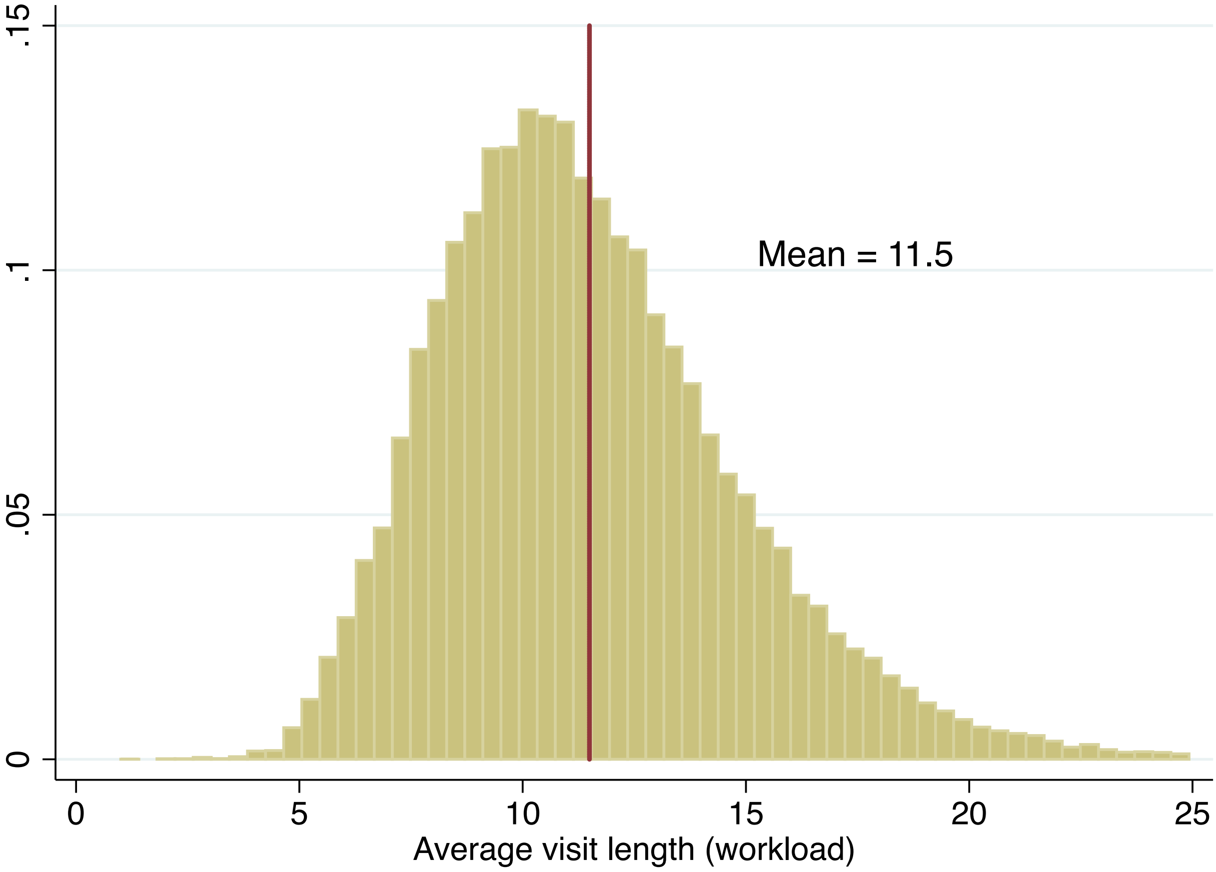
Anthony Scott. Chapter 22 economics of general practice. In *Handbook of Health Economics*, volume 1, Part B, pages 1175 – 1200. Elsevier, 2000.

Scott K Shriver, Harikesh S Nair, and Reto Hofstetter. Social ties and user-generated content: Evidence from an online social network. *Management Science*, 59(6):1425–1443, 2013.

Barbara Starfield, Leiyu Shi, and James Macinko. Contribution of primary care to health systems and health. *Milbank quarterly*, 83(3):457–502, 2005.

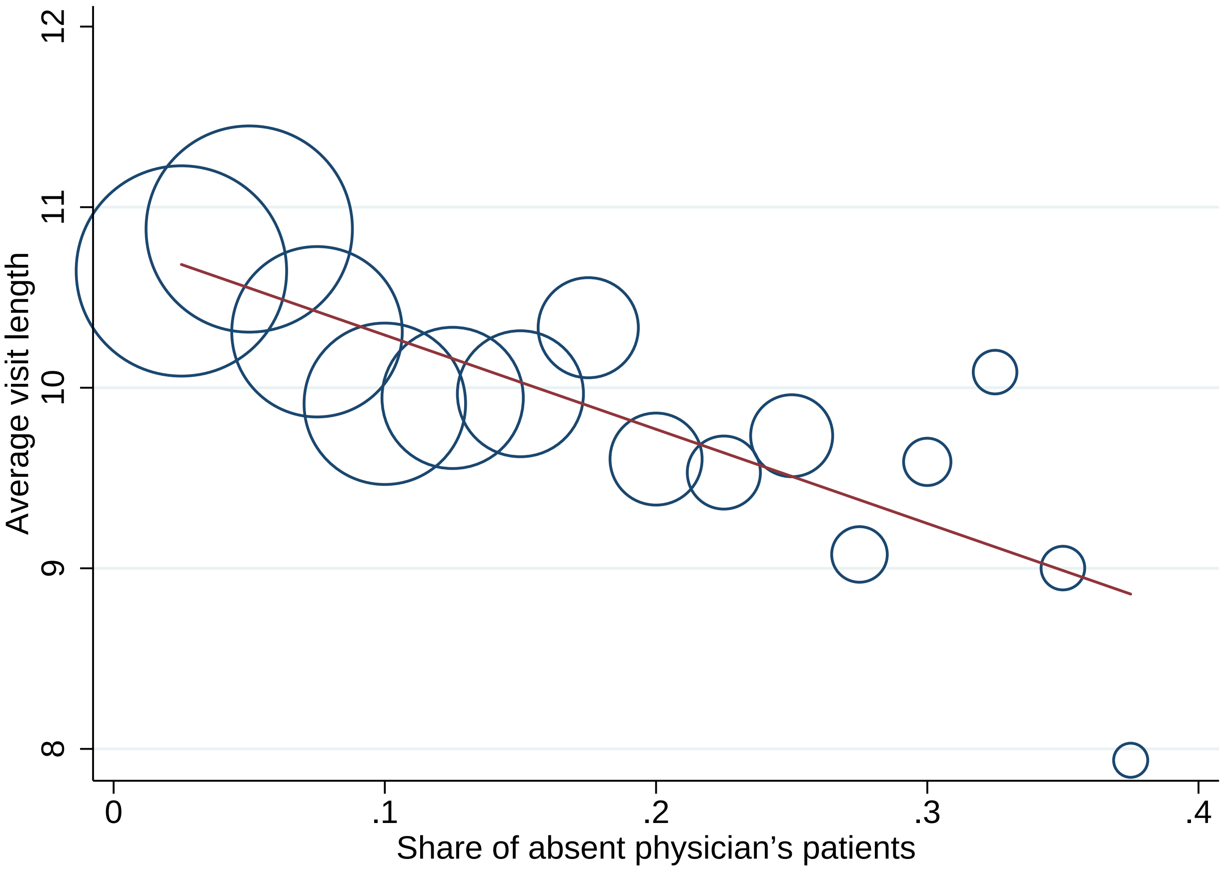
Tom Fangyun Tan and Serguei Netessine. When does the devil make work? an empirical study of the impact of workload on worker productivity. *Management Science*, 60(6):1574–1593, 2014.

Figure 1: Distribution of average visit length (workload)



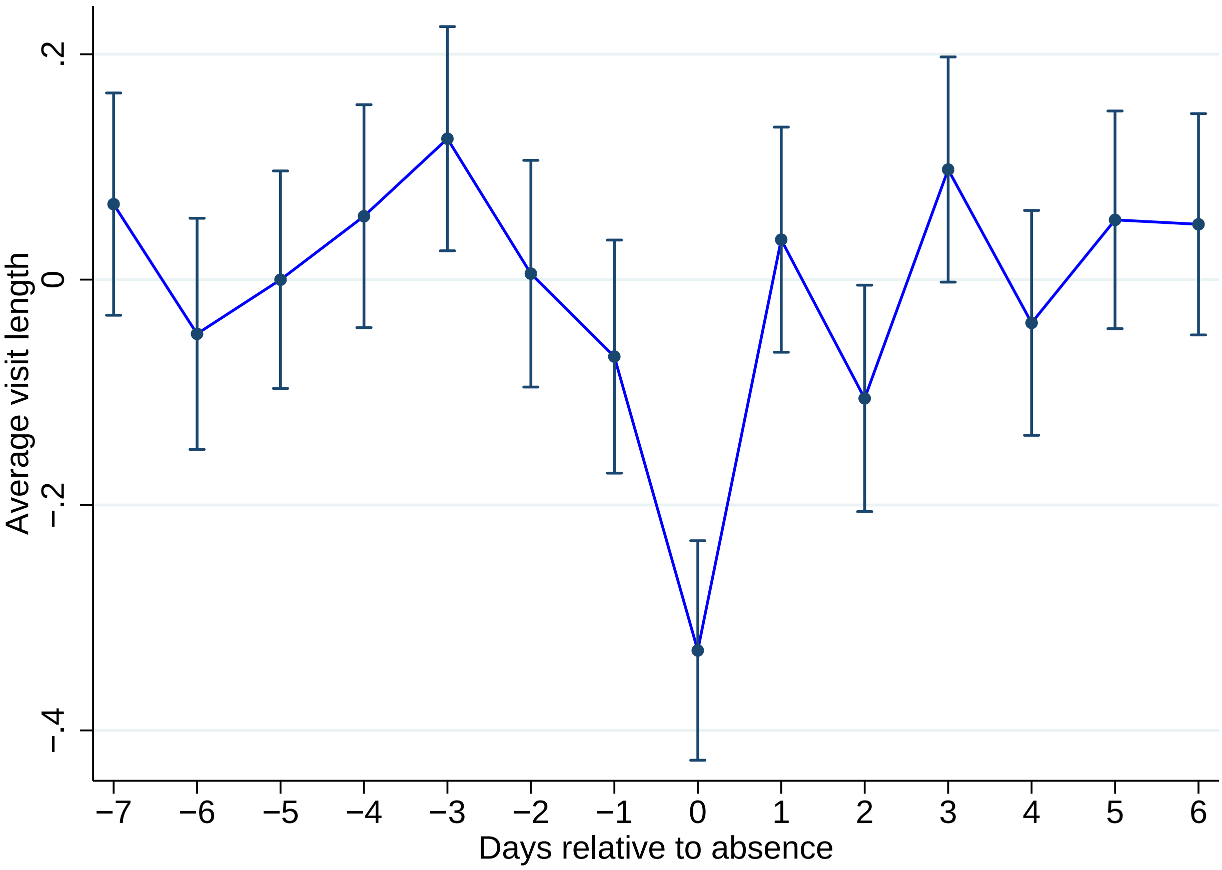

Note: The figure plots the average visit length–our workload measure. The (red) vertical line denotes the workload measure's mean. As the $99^{th}$ percentile of the distribution is 21.5, about 2000 observations pertaining to visits in excess of 25 minutes were excluded for illustrative purposes.

Figure 2: Share of missing physician's patients and workload



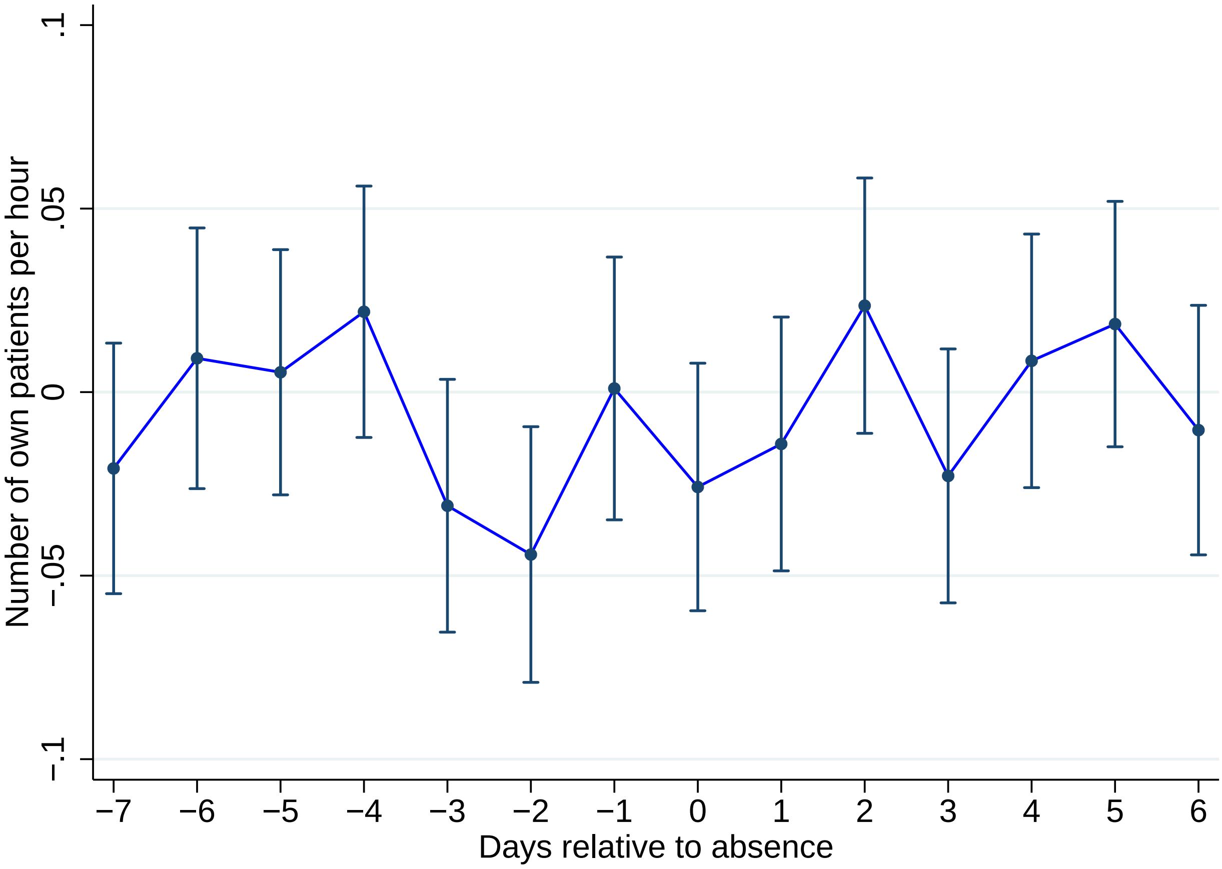

Note: The figure plots means of average visit length (workload) in 0.025 percentage point bins of the share of the missing physician's patients. The superimposed line is the predicted relation between absence and workload.

Figure 3: Workload around days of a colleague's absence




Note: The figure plots the coefficients and standard errors from the event study model described in Equation (2). The dependent variable is average visit length (workload).

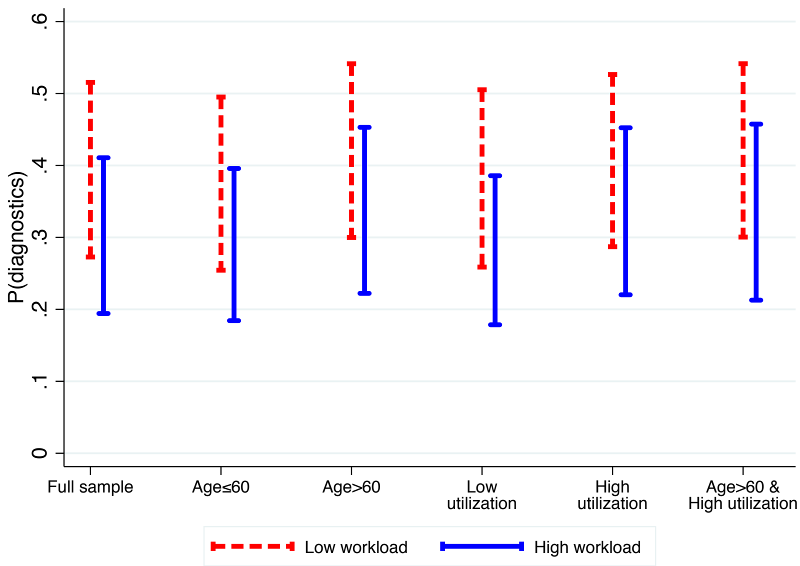
Figure 4: Number of own patients per hour around days of a colleague's absence



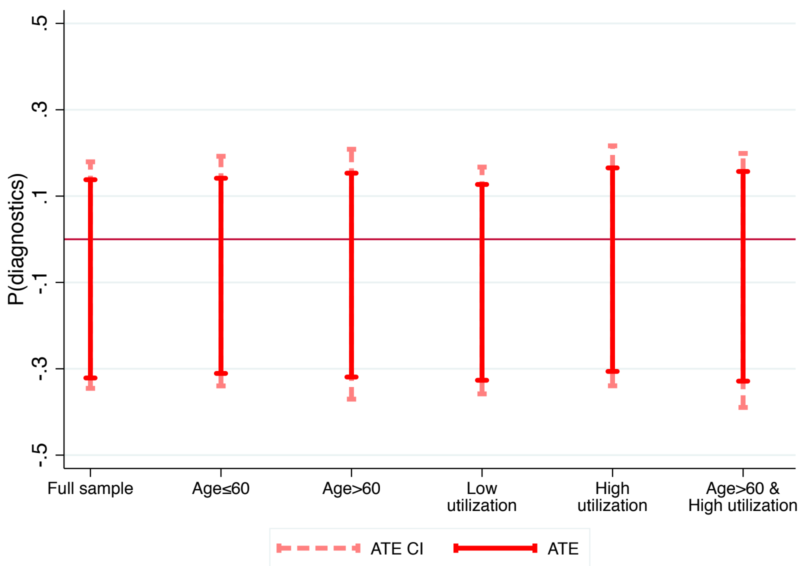

Note: The figure plots the coefficients and standard errors from the event study model akin to the model in Equation (2). The dependent variable is the number of a physician's own patients per hour.

Figure 5: Bounds on the effect of workload on the utilization of diagnostic inputs, IV

(a) Low and high workload



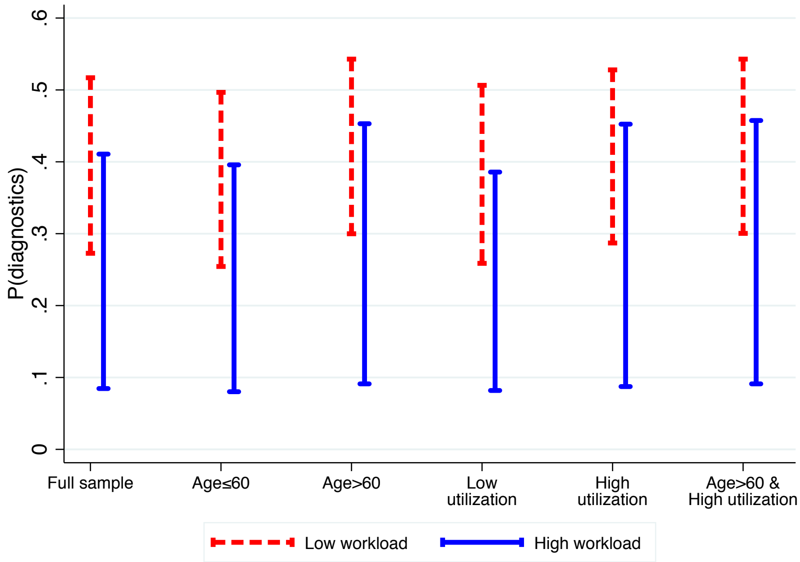

(b) ATE



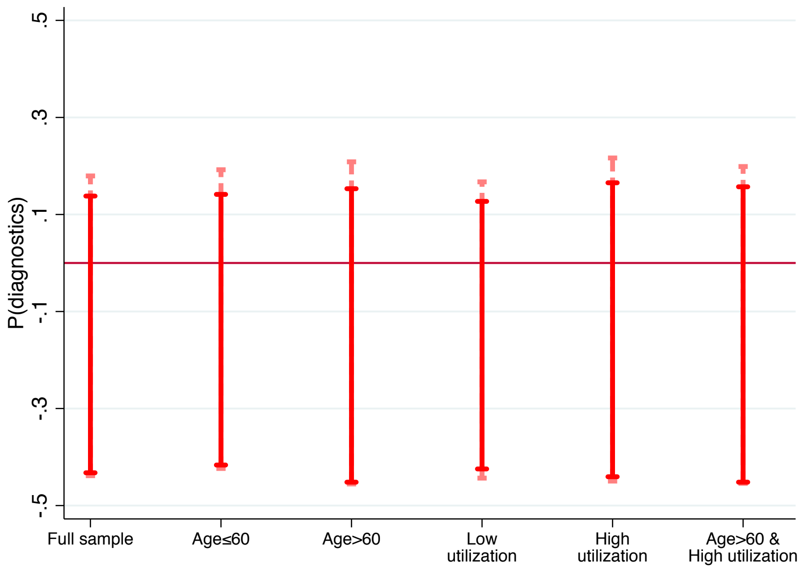

Note: Estimates from the bounds analysis under the IV assumption. Panel (a) presents estimated sets for the probability of using diagnostics under different workload levels. Panel (b) presents estimated sets and 95 percent Confidence Intervals for the ATE of a switch from low to high workload on this probability.

Figure 6: Bounds on the effect of workload on the utilization of diagnostic inputs, MIV

(a) Low and high workload



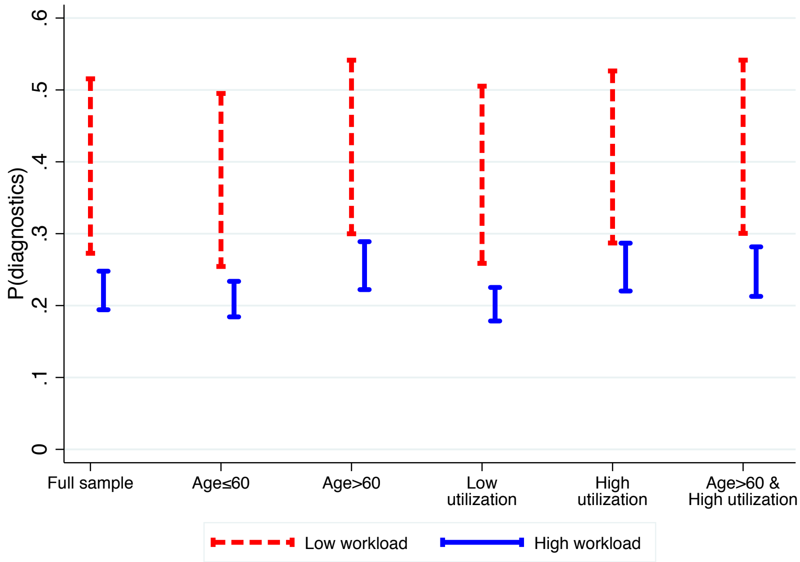

(b) ATE



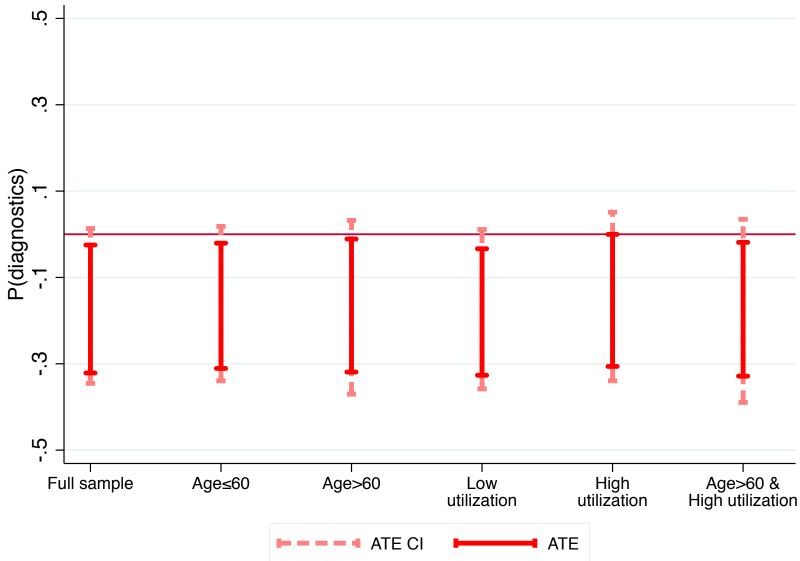

Note: Estimates from the bounds analysis under the MIV assumption. Panel (a) presents estimated sets for the probability of using diagnostics under different workload levels. Panel (b) presents estimated sets and 95 percent Confidence Intervals for the ATE of a switch from low to high workload on this probability.

Figure 7: Bounds on the effect of workload on the utilization of diagnostic inputs, IV-MTS

(a) Low and high workload



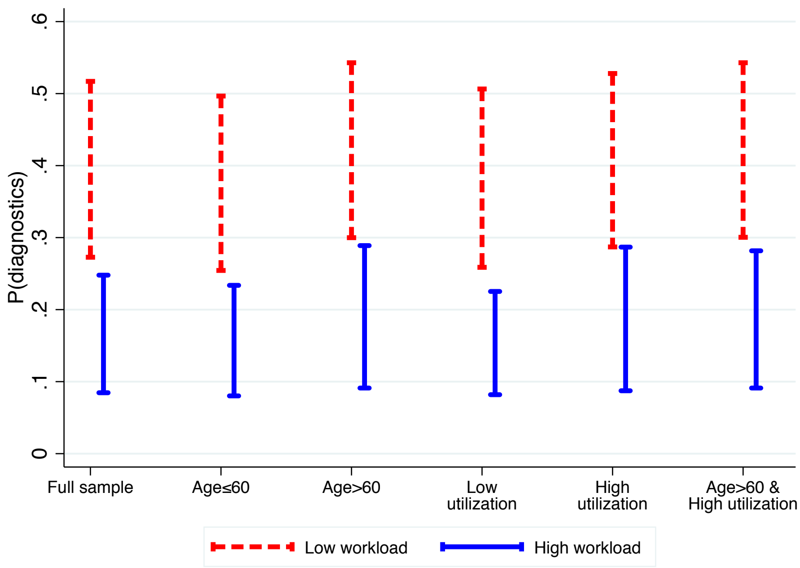

(b) ATE



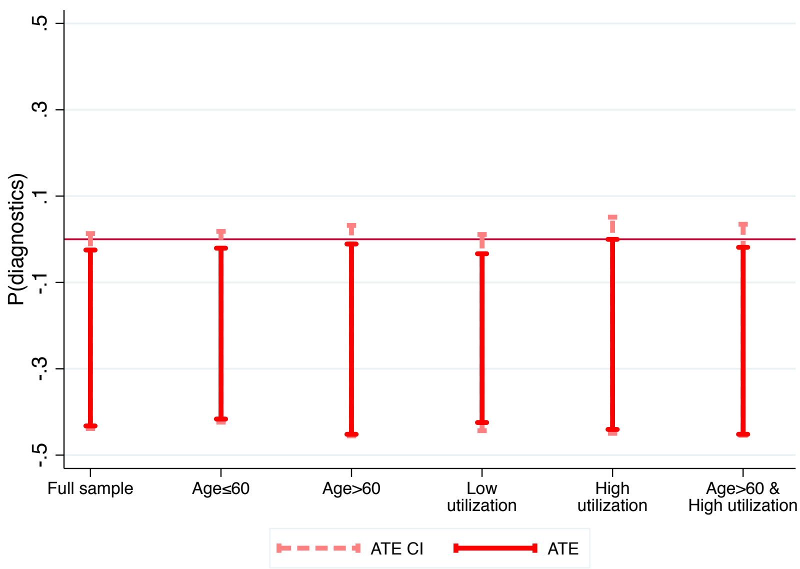

Note: Estimates from the bounds analysis under the IV-MTS assumption. Panel (a) presents estimated sets for the probability of using diagnostics under different workload levels. Panel (b) presents estimated sets and 95 percent Confidence Intervals for the ATE of a switch from low to high workload on this probability.

Figure 8: Bounds on the effect of workload on the utilization of diagnostic inputs, MIV-MTS

(a) Low and high workload




(b) ATE




Note: Estimates from the bounds analysis under the MIV-MTS assumption. Panel (a) presents estimated sets for the probability of using diagnostics under different workload levels. Panel (b) presents estimated sets and 95 percent Confidence Intervals for the ATE of a switch from low to high workload on this probability.

Table 1: Summary statistics of visits data

| | Mean<br>(1) | SD<br>(2) |
|---|---|---|
| **Patient characteristics** | | |
| Mean age | 47.60 | 26.68 |
| Share women | 0.58 | 0.49 |
| Share born in Israel | 0.61 | 0.49 |
| Share smokers | 0.30 | 0.46 |
| Share obese | 0.26 | 0.44 |
| Share hypertension | 0.34 | 0.47 |
| Share hyperlipidemia | 0.45 | 0.50 |
| Share ischemic heart disease | 0.15 | 0.36 |
| **Office visits characteristics** | | |
| Visit length | 11.56 | 9.31 |
| Share referral to specialist | 0.14 | 0.35 |
| Share referral to imaging | 0.08 | 0.27 |
| Share referral to lab tests | 0.20 | 0.40 |
| Share referrals to ER | 0.01 | 0.11 |
| Share Painkiller | 0.05 | 0.21 |
| Share antibiotics | 0.10 | 0.30 |
| Number of patients | 78,959 | |
| Number of physicians | 93 | |
| Observations | 823,349 | |

Notes: The table includes Sunday-Thursday face-to-face visits in the clinics used in this study in the period 2011-2014 (see text).

Table 2: The effect of absences on workload

| | (1) | (2) |
|---|---|---|
| A. IV 1 | | |
| Share of absent physician's patients of all patients | -4.82** | -4.82** |
| | (0.26) | (0.26) |
| F-statistic | 356 | 357 |
| p-value | 0.000 | 0.000 |
| B. IV 2 | | |
| Seeing absent physician's patients | -0.63** | -0.62** |
| | (0.05) | (0.05) |
| F-statistic | 188 | 188 |
| p-value | 0.000 | 0.000 |
| Year-month, day & physician FE | Yes | Yes |
| Patient age, gender & condition controls | No | Yes |
| Observations | 823,349 | 823,349 |

Notes: All columns report estimates of effect of absence on workload, as per Equation (3). The Year-month fixed effects consist of a dummy variable for each of the calendar months in our data. Patient age, gender & condition controls include: age, age squared, a gender dummy, and 113 indicators for chronic conditions. Additionally, we include dummy variables for visits for which the main reason is: issue a medical certificate, prescription renewal, filling out forms, and an administrative visit. Standard errors clustered by physician-day are reported in parentheses. One or two asterisks indicate significance at 5% or 1%, respectively.

Table 3: The effect of workload on utilization of diagnostic inputs

| | OLS | | IV 1 | | IV 2 | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **A. Dependent Variable: all diagnostic inputs (mean =0.36 )** | | | | | | |
| Average visit length | 0.48** | 0.45** | 1.64** | 1.62** | 1.86** | 1.84** |
| | (0.02) | (0.02) | (0.36) | (0.36) | (0.44) | (0.44) |
| **B. Dependent Variable: referral to a specialist (mean =0.14 )** | | | | | | |
| Mean visit length | 0.33** | 0.30** | 1.13** | 1.11** | 1.34** | 1.32** |
| | (0.02) | (0.02) | (0.25) | (0.25) | (0.32) | (0.32) |
| **C. Dependent Variable: referral to a lab test (mean =0.20 )** | | | | | | |
| Average visit length | 0.16** | 0.16** | 0.76* | 0.77* | 0.89* | 0.93** |
| | (0.02) | (0.02) | (0.31) | (0.31) | (0.36) | (0.36) |
| **D. Dependent Variable: referral to imaging (mean =0.08 )** | | | | | | |
| Mean visit length | 0.22** | 0.20** | 0.11 | 0.09 | 0.15 | 0.12 |
| | (0.01) | (0.01) | (0.18) | (0.18) | (0.22) | (0.22) |
| Year-month, day & physician FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Patient age, gender & condition controls | No | Yes | No | Yes | No | Yes |
| Observations | 823,349 | 823,349 | 823,349 | 823,349 | 823,349 | 823,349 |

Notes: Panels (A), (B), (C) and (D) of this table report estimates of effect of workload on the probability of use of any of the diagnostic inputs, referral to a specialist, referral to a lab test and referral to imaging, respectively, as per Equation (4). The Year-month fixed effects consist of a dummy variable for each of the calendar months in our data. Patient age, gender & condition controls include: age, age squared, a gender dummy, and 113 indicators for chronic conditions. Additionally, we include dummy variables for visits for which the main reason is: issue a medical certificate, prescription renewal, filling out forms, and an administrative visit. Standard errors clustered by physician-day are reported in parentheses. One or two asterisks indicate significance at 5% or 1%, respectively.

Table 4: The effect of workload on utilization of diagnostic inputs, by patient group

| | Age>60 | | | Age≤60 | | | High utilization | | | Low utilization | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | OLS | IV1 | IV2 | OLS | IV1 | IV2 | OLS | IV1 | IV2 | OLS | IV1 | IV2 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| **A. Dependent Variable: all diagnostic inputs** | | | | | | | | | | | | |
| | (mean=0.39 ) | | | (mean =0.33 ) | | | (mean=0.34 ) | | | (mean=0.34 ) | | |
| Average visit length | 0.44** | 2.04** | 1.78* | 0.45** | 1.39** | 1.81** | 0.44** | 1.86** | 2.13** | 0.44** | 1.45** | 1.57** |
| | (0.04) | (0.63) | (0.70) | (0.03) | (0.41) | (0.52) | (0.03) | (0.52) | (0.62) | (0.03) | (0.47) | (0.57) |
| **B. Dependent Variable: referral to a specialist** | | | | | | | | | | | | |
| | (mean=0.17 ) | | | (mean=0.13 ) | | | (mean=0.17 ) | | | (mean=0.12 ) | | |
| Average visit length | 0.35** | 1.82** | 1.94** | 0.27** | 0.81** | 0.95** | 0.34** | 1.47** | 1.72** | 0.27** | 0.88** | 1.01** |
| | (0.03) | (0.49) | (0.53) | (0.02) | (0.26) | (0.35) | (0.03) | (0.40) | (0.48) | (0.02) | (0.29) | (0.37) |
| **C. Dependent Variable: referral to a lab test** | | | | | | | | | | | | |
| | (mean=0.21 ) | | | (mean=0.19 ) | | | (mean=0.19 ) | | | (mean=0.20 ) | | |
| Average visit length | 0.12** | 1.28* | 0.76 | 0.18** | 0.46 | 0.92* | 0.12** | 1.06* | 1.08* | 0.18** | 0.52 | 0.73 |
| | (0.03) | (0.54) | (0.57) | (0.03) | (0.35) | (0.44) | (0.03) | (0.43) | (0.50) | (0.03) | (0.40) | (0.48) |
| **D. Dependent Variable: referral to imaging** | | | | | | | | | | | | |
| | (mean=0.09 ) | | | (mean=0.07 ) | | | (mean=0.09 ) | | | (mean=0.07 ) | | |
| Average visit length | 0.22** | 0.24 | 0.36 | 0.19** | 0.05 | -0.01 | 0.22** | 0.34 | 0.55 | 0.18** | -0.08 | -0.23 |
| | (0.02) | (0.38) | (0.39) | (0.02) | (0.19) | (0.26) | (0.02) | (0.29) | (0.35) | (0.02) | (0.22) | (0.29) |
| Year-month, day & physician FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Patient age, gender & condition controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 332,911 | 332,911 | 332,911 | 490,438 | 490,438 | 490,438 | 407,465 | 407,465 | 407,465 | 415,884 | 415,884 | 415,884 |

Notes: Panels (A), (B), (C) and (D) of this table report estimates of effect of workload on the probability of use of any of the diagnostic inputs, referral to a specialist, referral to a lab test and referral to imaging, respectively, as per Equation (4), by subsamples. Columns (1)-(6) report the results by age and Columns (7)-(12) report the results by patient condition. The Year-month fixed effects consist of a dummy variable for each of the calendar months in our data. Patient age, gender & condition controls include: age, age squared, a gender dummy, and 113 indicators for chronic conditions. Additionally, we include dummy variables for visits for which the main reason is: issue a medical certificate, prescription renewal, filling out forms, and an administrative visit. Standard errors clustered by physician-day are reported in parentheses. One or two asterisks indicate significance at 5% or 1%, respectively.

Table 5: The effect of workload on treatment decision

| | OLS | | IV 1 | | IV 2 | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **A. Dependent Variable: all treatments (mean =0.15 )** | | | | | | |
| Average visit length | 0.04* | 0.01 | -0.49 | -0.52 | -0.81* | -0.88** |
| | (0.02) | (0.02) | (0.27) | (0.27) | (0.32) | (0.32) |
| **B. Dependent Variable: referral to the emergency room (mean =0.01 )** | | | | | | |
| Mean visit length | 0.05** | 0.04** | 0.00 | -0.00 | -0.07 | -0.07 |
| | (0.01) | (0.01) | (0.07) | (0.08) | (0.09) | (0.09) |
| **C. Dependent Variable: prescription of pain killers (mean =0.05 )** | | | | | | |
| Mean visit length | 0.03** | 0.02* | -0.17 | -0.17 | -0.29 | -0.29 |
| | (0.01) | (0.01) | (0.15) | (0.14) | (0.18) | (0.18) |
| **D. Dependent Variable: prescription of antibiotics (mean =0.10 )** | | | | | | |
| Mean visit length | -0.03 | -0.04** | -0.32 | -0.32 | -0.51 | -0.55* |
| | (0.02) | (0.01) | (0.23) | (0.23) | (0.26) | (0.26) |
| Year-month, day & physician FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Patient age, gender & condition controls | No | Yes | No | Yes | No | Yes |
| Observations | 823,349 | 823,349 | 823,349 | 823,349 | 823,349 | 823,349 |

Notes: Panels (A), (B), (C) and (D) of this table report estimates of effect of workload on the probability of any of the treatments, referral to the emergency room, prescription of pain killers and prescription of antibiotics, respectively, as per Equation (4). The Year-month fixed effects consist of a dummy variable for each of the calendar months in our data. Patient age, gender & condition controls include: age, age squared, a gender dummy, and 113 indicators for chronic conditions. Additionally, we include dummy variables for visits for which the main reason is: issue a medical certificate, prescription renewal, filling out forms, and an administrative visit. Standard errors clustered by physician-day are reported in parentheses. One or two asterisks indicate significance at 5% or 1%, respectively.

Table 6: The effect of workload on treatment decision, by patient group

| | Age>60 | | | Age≤60 | | | High utilization | | | Low utilization | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | OLS | IV1 | IV2 | OLS | IV1 | IV2 | OLS | IV1 | IV2 | OLS | IV1 | IV2 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| **A. Dependent Variable: all treatments** | | | | | | | | | | | | |
| | (mean=0.16 ) | | | (mean=0.15 ) | | | (mean=0.16 ) | | | (mean=0.15 ) | | |
| Average visit length | -0.02 | -0.89 | -1.25* | 0.03 | -0.30 | -0.57 | -0.04 | -0.41 | -0.76 | 0.05* | -0.58 | -0.98* |
| | (0.03) | (0.48) | (0.53) | (0.02) | (0.31) | (0.39) | (0.03) | (0.40) | (0.47) | (0.02) | (0.34) | (0.42) |
| **B. Dependent Variable: referral to the emergency room** | | | | | | | | | | | | |
| | (mean=0.01 ) | | | (mean=0.01 ) | | | (mean=0.02 ) | | | (mean=0.01 ) | | |
| Average visit length | 0.05** | -0.15 | -0.15 | 0.03** | 0.07 | -0.03 | 0.06** | -0.06 | -0.13 | 0.02** | 0.04 | -0.02 |
| | (0.01) | (0.16) | (0.17) | (0.01) | (0.08) | (0.11) | (0.01) | (0.13) | (0.16) | (0.01) | (0.08) | (0.11) |
| **C. Dependent Variable: prescription of pain killers** | | | | | | | | | | | | |
| | (mean=0.07 ) | | | (mean=0.03 ) | | | (mean=0.07 ) | | | (mean=0.03 ) | | |
| Average visit length | 0.02 | -0.65 | -0.72* | 0.03* | 0.06 | 0.00 | 0.00 | -0.17 | -0.29 | 0.04** | -0.20 | -0.33 |
| | (0.02) | (0.33) | (0.36) | (0.01) | (0.14) | (0.19) | (0.02) | (0.26) | (0.31) | (0.01) | (0.15) | (0.20) |
| **D. Dependent Variable: prescription of antibiotics** | | | | | | | | | | | | |
| | (mean=0.08 ) | | | (mean=0.11 ) | | | (mean=0.08 ) | | | (mean=0.11 ) | | |
| Average visit length | -0.08** | -0.13 | -0.43 | -0.02 | -0.42 | -0.61 | -0.10** | -0.22 | -0.41 | -0.00 | -0.40 | -0.68 |
| | (0.02) | (0.35) | (0.38) | (0.02) | (0.28) | (0.34) | (0.02) | (0.31) | (0.35) | (0.02) | (0.31) | (0.37) |
| Year-month, day & physician FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Patient age, gender & condition controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 332,911 | 332,911 | 332,911 | 490,438 | 490,438 | 490,438 | 407,465 | 407,465 | 407,465 | 415,884 | 415,884 | 415,884 |

Notes: Panels (A), (B), (C) and (D) of this table report estimates of effect of workload on the probability of any of the treatments, referral to the emergency room, prescription of pain killers and prescription of antibiotics, respectively, as per Equation (4), by subsamples. Columns (1)-(6) report the results by age and Columns (7)-(12) report the results by patient condition. The Year-month fixed effects consist of a dummy variable for each of the calendar months in our data. Patient age, gender & condition controls include: age, age squared, a gender dummy, and 113 indicators for chronic conditions. Additionally, we include dummy variables for visits for which the main reason is: issue a medical certificate, prescription renewal, filling out forms, and an administrative visit. Standard errors clustered by physician-day are reported in parentheses. One or two asterisks indicate significance at 5% or 1%, respectively.

Table 7: The effect of workload on subsequent encounters

| | OLS | | IV 1 | | IV 2 | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **A. Dependent Variable: subsequent visit within 15 days (mean =0.26 )** | | | | | | |
| Average visit length | 0.03 | 0.01 | -0.45 | -0.49 | -0.59 | -0.59 |
| | (0.02) | (0.02) | (0.35) | (0.35) | (0.43) | (0.43) |
| **B. Dependent Variable: subsequent visit within 30 days (mean =0.40 )** | | | | | | |
| Average visit length | 0.05* | 0.03 | -0.13 | -0.19 | -0.16 | -0.16 |
| | (0.03) | (0.03) | (0.39) | (0.38) | (0.48) | (0.47) |
| **C. Dependent Variable: subsequent visit within 60 days (mean =0.57 )** | | | | | | |
| Average visit length | 0.04 | 0.01 | -0.44 | -0.51 | -0.26 | -0.26 |
| | (0.03) | (0.03) | (0.40) | (0.39) | (0.45) | (0.44) |
| **D. Dependent Variable: subsequent visit within 90 days (mean =0.66 )** | | | | | | |
| Average visit length | 0.03 | 0.00 | -0.22 | -0.30 | -0.31 | -0.32 |
| | (0.03) | (0.02) | (0.39) | (0.38) | (0.43) | (0.42) |
| Year-month, day & physician FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Patient age, gender & condition controls | No | Yes | No | Yes | No | Yes |
| Observations | 823,349 | 823,349 | 823,349 | 823,349 | 823,349 | 823,349 |

Notes: Panels (A), (B), (C) and (D) of this table report estimates of effect of workload on the likelihood of a subsequent visit. The Year-month fixed effects consists of a dummy variable for each of the calendar months in our data. Patient age, gender & condition controls include: age, age squared, a gender dummy, and 113 indicators for chronic conditions. Additionally, we include dummy variables for visits for which the main reason is: issue a medical certificate, prescription renewal, filling out forms, and an administrative visit. Standard errors clustered by physician-day are reported in parentheses. One or two asterisks indicate significance at 5% or 1%, respectively.

Table 8: The effect of workload on subsequent encounters, by patient group

| | Age>60 | | | Age≤60 | | | High utilization | | | Low utilization | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | OLS | IV1 | IV2 | OLS | IV1 | IV2 | OLS | IV1 | IV2 | OLS | IV1 | IV2 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| **A. Dependent Variable: subsequent visit within 15 days** | | | | | | | | | | | | |
| | (mean=0.30 ) | | | (mean =0.22 ) | | | (mean=0.31 ) | | | (mean=0.20 ) | | |
| Average visit length | 0.07 | -1.14 | -1.36 | -0.01 | -0.15 | -0.08 | 0.02 | -1.14* | -1.21 | 0.02 | -0.01 | -0.06 |
| | (0.04) | (0.65) | (0.71) | (0.03) | (0.37) | (0.48) | (0.03) | (0.53) | (0.64) | (0.03) | (0.41) | (0.49) |
| **B. Dependent Variable: subsequent visit within 30 days** | | | | | | | | | | | | |
| | (mean=0.49 ) | | | (mean =0.34 ) | | | (mean=0.50 ) | | | (mean=0.31 ) | | |
| Average visit length | 0.11** | -0.92 | -0.42 | -0.00 | 0.20 | 0.04 | 0.06 | -0.78 | -0.70 | 0.01 | 0.25 | 0.30 |
| | (0.04) | (0.68) | (0.76) | (0.03) | (0.41) | (0.54) | (0.04) | (0.58) | (0.69) | (0.03) | (0.47) | (0.57) |
| **C. Dependent Variable: subsequent visit within 60 days** | | | | | | | | | | | | |
| | (mean=0.68 ) | | | (mean =0.49 ) | | | (mean=0.69 ) | | | (mean=0.45 ) | | |
| Average visit length | 0.07 | -0.75 | -0.72 | -0.00 | -0.40 | -0.00 | 0.02 | -0.34 | -0.23 | 0.00 | -0.69 | -0.32 |
| | (0.04) | (0.65) | (0.67) | (0.03) | (0.45) | (0.54) | (0.03) | (0.53) | (0.59) | (0.04) | (0.50) | (0.59) |
| **D. Dependent Variable: subsequent visit within 90 days** | | | | | | | | | | | | |
| | (mean=0.77 ) | | | (mean =0.59 ) | | | (mean=0.78 ) | | | (mean=0.55 ) | | |
| Average visit length | 0.05 | -0.29 | -0.19 | 0.00 | -0.31 | -0.42 | -0.01 | -0.15 | -0.23 | 0.03 | -0.45 | -0.42 |
| | (0.03) | (0.56) | (0.60) | (0.03) | (0.47) | (0.54) | (0.03) | (0.47) | (0.52) | (0.04) | (0.52) | (0.58) |
| Year-month, day & physician FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Patient age, gender & condition controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 332,911 | 332,911 | 332,911 | 490,438 | 490,438 | 490,438 | 407,465 | 407,465 | 407,465 | 415,884 | 415,884 | 415,884 |

Notes: Panels (A), (B), (C) and (D) of this table report estimates of effect of workload on the likelihood of a subsequent visit, by subsamples. Columns (1)-(6) report the results by age and Columns (7)-(12) report the results by patient condition. The Year-month fixed effects consist of a dummy variable for each of the calendar months in our data. Standard errors clustered by physician-day are reported in parentheses. Patient age, gender & condition controls include: age, age squared, a gender dummy, and 113 indicators for chronic conditions. Additionally, we include dummy variables for visits for which the main reason is: issue a medical certificate, prescription renewal, filling out forms, and an administrative visit. One or two asterisks indicate significance at 5% or 1%, respectively.

Table 9: The effect of workload on non face-to-face encounters with patients

| | OLS | IV 1 | IV 2 |
|---|---|---|---|
| | (1) | (2) | (3) |
| **A. Dependent Variable:** | | | |
| **Response to online patient queries per hour (mean = 1.76 )** | | | |
| Average visit length | -0.06*** | 0.07*** | 0.08*** |
| | (0.00) | (0.03) | (0.03) |
| **B. Dependent Variable:** | | | |
| **Phone calls with patients per hour (mean = 0.31 )** | | | |
| Average visit length | -0.01** | 0.03** | 0.03** |
| | (0.00) | (0.01) | (0.01) |
| Observations | 43,487 | 43,487 | 43,487 |

Notes: Panels (A) and (B) of this table report estimates of effect of workload on the number of online patients queries and phone calls with patients per hour, respectively, as per Equation (4). All specifications include Year-month fixed effects that consist of a dummy variable for each of the calendar months in our data. One or two asterisks indicate significance at 5% or 1%, respectively.

Table 10: The effect of workload on the use of diagnostic inputs: IV and MIV bounds

| | Low workload | | High workload | | ATE (Low to High) | | ATE CI | | Obs. |
|---|---|---|---|---|---|---|---|---|---|
| | (Lower) | (Upper) | (Lower) | (Upper) | (Lower) | (Upper) | (Lower) | (Upper) | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| **A. IV** | | | | | | | | | |
| Full sample | 0.273 | 0.516 | 0.194 | 0.411 | -0.321 | 0.138 | -0.345 | 0.179 | 822,416 |
| Age $\leq$ 60 | 0.254 | 0.495 | 0.184 | 0.396 | -0.311 | 0.141 | -0.340 | 0.192 | 489,669 |
| Age>60 | 0.300 | 0.541 | 0.222 | 0.453 | -0.319 | 0.153 | -0.370 | 0.208 | 332,747 |
| Low utilization | 0.259 | 0.505 | 0.179 | 0.386 | -0.327 | 0.127 | -0.358 | 0.167 | 415,265 |
| High utilization | 0.287 | 0.526 | 0.220 | 0.452 | -0.306 | 0.165 | -0.339 | 0.216 | 407,151 |
| Age>60 & high U | 0.300 | 0.541 | 0.213 | 0.457 | -0.329 | 0.157 | -0.390 | 0.199 | 286,836 |
| **B. MIV** | | | | | | | | | |
| Full sample | 0.273 | 0.517 | 0.085 | 0.411 | -0.432 | 0.138 | -0.438 | 0.179 | 822,416 |
| Age $\leq$ 60 | 0.254 | 0.497 | 0.080 | 0.396 | -0.416 | 0.141 | -0.424 | 0.192 | 489,669 |
| Age>60 | 0.300 | 0.543 | 0.091 | 0.453 | -0.452 | 0.153 | -0.455 | 0.208 | 332,747 |
| Low utilization | 0.259 | 0.506 | 0.082 | 0.386 | -0.425 | 0.127 | -0.443 | 0.167 | 415,265 |
| High utilization | 0.287 | 0.528 | 0.087 | 0.452 | -0.441 | 0.165 | -0.450 | 0.216 | 407,151 |
| Age>60 & high U | 0.300 | 0.543 | 0.091 | 0.457 | -0.452 | 0.157 | -0.454 | 0.199 | 286,836 |

Notes: Panels (A) and (B) report the estimates from the bounds analysis under the IV and the MIV assumptions, respectively. The ATE is defined in (5) and pertains to the average change in the probability with which diagnostic inputs are used following a switch from a low to a high workload state. Columns (7) and (8) report the lower and upper endpoints of 95 percent confidence intervals computed following Imbens and Manski (2004). See text.

Table 11: The effect of workload on the use of diagnostic inputs: MTS-IV and MTS-MIV bounds

| | Low workload | | High workload | | ATE (Low to High) | | ATE CI | | Obs. |
|---|---|---|---|---|---|---|---|---|---|
| | (Lower) | (Upper) | (Lower) | (Upper) | (Lower) | (Upper) | (Lower) | (Upper) | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| **A. IV-MTS** | | | | | | | | | |
| Full sample | 0.273 | 0.516 | 0.194 | 0.248 | -0.321 | -0.025 | -0.345 | 0.013 | 822,416 |
| Age $\leq$ 60 | 0.254 | 0.495 | 0.184 | 0.234 | -0.311 | -0.021 | -0.340 | 0.018 | 489,669 |
| Age>60 | 0.300 | 0.541 | 0.222 | 0.289 | -0.319 | -0.011 | -0.370 | 0.032 | 332,747 |
| Low utilization | 0.259 | 0.505 | 0.179 | 0.225 | -0.327 | -0.034 | -0.358 | 0.011 | 415,265 |
| High utilization | 0.287 | 0.526 | 0.220 | 0.287 | -0.306 | -0.000 | -0.339 | 0.051 | 407,151 |
| Age>60 & high U | 0.300 | 0.541 | 0.213 | 0.282 | -0.329 | -0.019 | -0.390 | 0.035 | 286,836 |
| **B. MIV-MTS** | | | | | | | | | |
| Full sample | 0.273 | 0.517 | 0.085 | 0.248 | -0.432 | -0.025 | -0.438 | 0.013 | 822,416 |
| Age $\leq$ 60 | 0.254 | 0.497 | 0.080 | 0.234 | -0.416 | -0.021 | -0.424 | 0.018 | 489,669 |
| Age>60 | 0.300 | 0.543 | 0.091 | 0.289 | -0.452 | -0.011 | -0.455 | 0.032 | 332,747 |
| Low utilization | 0.259 | 0.506 | 0.082 | 0.225 | -0.425 | -0.034 | -0.443 | 0.011 | 415,265 |
| High utilization | 0.287 | 0.528 | 0.087 | 0.287 | -0.441 | -0.000 | -0.450 | 0.051 | 407,151 |
| Age>60 & high U | 0.300 | 0.543 | 0.091 | 0.282 | -0.452 | -0.019 | -0.454 | 0.035 | 286,836 |

Notes: Panels (A) and (B) report the estimates from the bounds analysis under the IV-MTS and the MIV-MTS assumptions, respectively. The ATE is defined in (5) and pertains to the average change in the probability with which diagnostic inputs are used following a switch from a low to a high workload state. Columns (7) and (8) report the lower and upper endpoints of 95 percent confidence intervals computed following Imbens and Manski (2004). See text.

# A Appendix A

## A.1 The reduced form regressions

Here we report the reduced form regression. Namely, we estimate the following equation:

$$y_{jsti} = \alpha + \nu_j + \nu_t + \beta_1 \cdot sa_{jst} + x_{jsti} \cdot \beta_2 + \epsilon_{jsti} \tag{A1}$$

Table A.1 reports the results for the diagnostic outcomes. As the table shows, consistent with the analysis in Section 4.3.1, the estimates in Panels (A) - (C) are negative and significant and the results in Panel (D) - referral to imaging are negative yet statistically insignificant.

Table A.2 reports the results for treatment outcomes. All estimates are statistically insignificant, in line with the estimates in Section 4.3.2, that do not show that a strong relationship between workload and treatment choice exists.

Table A.3 reports the results for the likelihood of subsequent visits. All estimates are statistically insignificant. This is quite consistent with the finding in Section 4.4 which indicate a small and mostly insignificant effect on subsequent visits.

Table A.4 reports the non face-to-face encounter results. The estimates are all negative and significant, in line with the results in section 4.5.

## A.2 Clinic level instrumental variable

Here, we report the results of the clinic level instrument analysis. Column (1) of Table A.5 shows the first stage results. Absence at the clinic decreases the daily average visit length by 0.37 minutes. These results are statistically significant and they are not sensitive to the inclusion of the patient level controls, as Column (2) shows.

Panel (A) of Table A.6, provides the corresponding IV estimates for the effect of workload on overall utilization of diagnostic inputs. The estimate in Column (1) indicate that a 1 minute decrease in average visit length decreases the utilization of diagnostic inputs by 1.12 percentage points. With patient level controls the effect remains roughly similar at 1.19 percentage points and it becomes significant at the 5% level. These results are smaller, yet comparable to those obtained using IV1 and IV2. Panel (B) of Table A.6, provides the estimates for the effect of workload on treatment choices. The estimates in Columns (1) and (2) are both negative and statistically insignificant with point estimate of -0.56 percentage points. The results are similar to those obtained using IV1 and IV2.

Table A.1: The effect of workload on utilization of diagnostic inputs, reduced form

| | IV 1 | | IV 2 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **A. Dependent Variable: all diagnostic inputs** | | | | |
| | -7.90** | -7.93** | -1.16** | -1.16** |
| | (1.71) | (1.71) | (0.27) | (0.27) |
| **B. Dependent Variable: referral to a specialist** | | | | |
| | -5.49** | -5.48** | -0.85** | -0.84** |
| | (1.20) | (1.19) | (0.20) | (0.20) |
| **C. Dependent Variable: referral to a lab test** | | | | |
| | -3.65* | -3.65* | -0.56* | -0.56* |
| | (1.46) | (1.46) | (0.22) | (0.22) |
| **D. Dependent Variable: referral to imaging** | | | | |
| | -0.51 | -0.56 | -0.10 | -0.10 |
| | (0.87) | (0.87) | (0.14) | (0.14) |
| Year-month, day & physician FE | Yes | Yes | Yes | Yes |
| Patient age, gender & condition controls | No | Yes | No | Yes |
| Observations | 823,349 | 823,349 | 823,349 | 823,349 |

Notes: Panels (A), (B), (C) and (D) of this table report estimates of effect of absence on the probability of referral to any of the diagnostic inputs, referral to a specialist, referral to a lab test and referral to imaging, respectively , as per Equation (A1). The Year-month fixed effects consists of a dummy variable for each of the calendar months in our data. Patient age, gender & condition controls include: age, age squared, a gender dummy, and 113 indicators for chronic conditions. Additionally, we include dummy variables for visits for which the main reason is: issue a medical certificate, prescription renewal, filling out forms, and an administrative visit. Standard errors clustered by physician-day are reported in parentheses. One or two asterisks indicate significance at 5% or 1%, respectively.

Table A.2: The effect of workload on treatment choice, reduced form

| | IV 1 | | IV 2 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **A. Dependent Variable: all treatments** | | | | |
| | 2.34 | 2.35 | 0.51** | 0.52** |
| | (1.29) | (1.29) | (0.20) | (0.20) |
| **B. Dependent Variable: referral to the emergency room** | | | | |
| | 0.01 | -0.03 | 0.04 | 0.04 |
| | (0.36) | (0.36) | (0.06) | (0.06) |
| **C. Dependent Variable: prescription of pain killers** | | | | |
| | 0.78 | 0.93 | 0.18 | 0.19 |
| | (0.70) | (0.70) | (0.11) | (0.11) |
| **D. prescription of antibiotics** | | | | |
| | 1.58 | 1.47 | 0.32* | 0.33* |
| | (1.10) | (1.10) | (0.16) | (0.16) |
| Year-month, day & physician FE | Yes | Yes | Yes | Yes |
| Patient age, gender & condition controls | No | Yes | No | Yes |
| Observations | 823,349 | 823,349 | 823,349 | 823,349 |

Notes: Panels (A), (B), (C) and (D) of this table report estimates of effect of absence on the probability of referral to any of the treatments, referral to the emergency room, prescription of pain killers and prescription of antibiotics, respectively, as per Equation (A1). The Year-month fixed effects consists of a dummy variable for each of the calendar months in our data. Patient age, gender & condition controls include: age, age squared, a gender dummy, and 113 indicators for chronic conditions. Additionally, we include dummy variables for visits for which the main reason is: issue a medical certificate, prescription renewal, filling out forms, and an administrative visit. Standard errors clustered by physician-day are reported in parentheses. One or two asterisks indicate significance at 5% or 1%, respectively.

Table A.3: The effect of workload on the likelihood of subsequent encounters, reduced form

| | IV 1 | | IV 2 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **A. Dependent Variable: subsequent visit within 15 days** | | | | |
| | 2.18 | 2.19 | 0.37 | 0.34 |
| | (1.69) | (1.68) | (0.27) | (0.27) |
| **B. Dependent Variable: subsequent visit within 30 days** | | | | |
| | 0.61 | 0.74 | 0.10 | 0.08 |
| | (1.86) | (1.85) | (0.30) | (0.29) |
| **C. Dependent Variable: subsequent visit within 60 days** | | | | |
| | 2.11 | 2.38 | 0.16 | 0.15 |
| | (1.92) | (1.89) | (0.28) | (0.27) |
| **D. Dependent Variable: subsequent visit within 90 days** | | | | |
| | 1.08 | 1.40 | 0.19 | 0.19 |
| | (1.89) | (1.85) | (0.27) | (0.26) |
| Year-month, day & physician FE | Yes | Yes | Yes | Yes |
| Patient age, gender & condition controls | No | Yes | No | Yes |
| Observations | 823,349 | 823,349 | 823,349 | 823,349 |

Notes: Panels (A), (B), (C) and (D) of this table report estimates of effect of absence on the likelihood of a subsequent visit. The Year-month fixed effects consists of a dummy variable for each of the calendar months in our data. Patient age, gender & condition controls include: age, age squared, a gender dummy, and 113 indicators for chronic conditions. Additionally, we include dummy variables for visits for which the main reason is: issue a medical certificate, prescription renewal, filling out forms, and an administrative visit. Standard errors clustered by physician-day are reported in parentheses. One or two asterisks indicate significance at 5% or 1%, respectively.

Table A.4: The effect of workload on non face-to-face encounters, reduced form

| | IV 1 | IV 2 |
|---|---|---|
| | (1) | (2) |
| **A. Dependent Variable:** | | |
| **Response to online patients queries per hour (mean =1.76 )** | | |
| | -0.340** | -0.061** |
| | (0.119) | (0.021) |
| **B. Dependent Variable:** | | |
| **Phone calls to patients per hour (mean =0.31 )** | | |
| | -0.149** | -0.024** |
| | (0.039) | (0.007) |
| Year-month, day & physician FE | Yes | Yes |
| Observations | 43,487 | 43,487 |

Panels (A) and (B) of this table report estimates of effect of absence on the number of online patients queries and phone calls with patients per hour, respectively, as per Equation (A1). All specifications include Year-month fixed effects that consists of a dummy variable for each of the calendar months in our data. One or two asterisks indicate significance at 5% or 1%, respectively.

Table A.5: The effect of absences on workload, clinic level

| | (1) | (2) |
|---|---|---|
| A. IV 3 | | |
| Absence at the clinic | -0.37** | -0.37** |
| | (0.04) | (0.04) |
| F-statistic | 84 | 84 |
| p-value | 0.000 | 0.000 |
| Year-month, day & physician FE | Yes | Yes |
| Patient age, gender & condition controls | No | Yes |
| Observations | 823,349 | 823,349 |

Notes: All columns report estimates of effect of absence at the clinic level on workload, as per Equation (3). The Year-month fixed effects consist of a dummy variable for each of the calendar months in our data. Patient age, gender & condition controls include: age, age squared, a gender dummy, and 113 indicators for chronic conditions. Additionally, we include dummy variables for visits for which the main reason is: issue a medical certificate, prescription renewal, filling out forms, and an administrative visit. Standard errors clustered by clinic-day are reported in parentheses. One or two asterisks indicate significance at 5% or 1%, respectively.

Table A.6: The effects of workload, clinic level instrument

| | IV 3 | |
|---|---|---|
| | (1) | (2) |
| **A. Dependent Variable: all diagnostic inputs (mean =0.36 )** | | |
| Average visit length | 1.12 | 1.19* |
| | (0.57) | (0.57) |
| **B. Dependent Variable: all treatments (mean =0.15 )** | | |
| Average visit length | -0.56 | -0.56 |
| | (0.43) | (0.42) |
| | | |
| Year-month, day & physician FE | Yes | Yes |
| Patient age, gender | | |
| & condition controls | No | Yes |
| Observations | 823,349 | 823,349 |

Notes: Panels (A) and (B) of this table report, using the clinic level instrumental variable, estimates of the effects of workload on utilization of diagnostic inputs and treatment decisions, respectively, as per Equation (4). The Year-month fixed effects consist of a dummy variable for each of the calendar months in our data. Patient age, gender & condition controls include: age, age squared, a gender dummy, and 113 indicators for chronic conditions. Additionally, we include dummy variables for visits for which the main reason is: issue a medical certificate, prescription renewal, filling out forms, and an administrative visit. Standard errors clustered by clinic-day are reported in parentheses. One or two asterisks indicate significance at 5% or 1%, respectively.

# B Appendix B

To create the "utilization score", we analyze the data at the patient year level. for each patient and each year we count the number of visits that patient made to a physician. This is the $\# - of - Visits_{it}$, the left hand side variable. We regress this measure of utilization against the set of patient personal characteristics using a model of the form:

$$\text{(A2)} \qquad \# - of - Visits_{it} = \alpha + X_{it} \cdot \beta + \epsilon_{it}$$

Based on the results of the regression we create a predicted $\# - of - Visits_{it}$ for each patient. We use this variable as the "utilization score" for each patient in the data. As we noted above, this score reflects the number of visits per time period that a patient with these characteristics would have, on average, in a given time period.