Does Division of Labor Increase Productivity? Evidence from Primary Care*

Amanda Dahlstrand[1], Shan Huang[2], Guy Michaels[3], and Nestor Le Nestour[4]

[1] *University of Zurich,* [2] *Stockholm School of Economics; University of Copenhagen,* [3] *London School of Economics,* [4] *Stockholm University*

January 15, 2025

PRELIMINARY - PLEASE DO NOT POST OR CIRCULATE

Abstract

The idea that the division of labor increases productivity is central to economic analyses of countries, industries, social structures, and occupations within organizations. We study the division of labor within an online primary healthcare organization, where an algorithm assigns patient cases between two clinician occupations: nurses and doctors. We compare a knowledge hierarchy, in which less specialized nurses resolve some cases themselves and escalate others to doctors, with a direct-to-expert approach where cases are assigned directly to doctors. We use approximately 500,000 cases and an identification strategy that leverages temporary congestion which increases the odds of an assignment to the nurse-initiated knowledge hierarchy. We find that nurses resolve 70% of cases at the margin and send the remainder to doctors. Although the knowledge hierarchy slightly reduces the rates of meaningful diagnosis and prescription, it does not adversely impact patient satisfaction, acute care utilization, labor earnings, or mortality. The knowledge hierarchy lowers total costs to the healthcare system by up to 8% without compromising quality at the primary care organization or downstream. Finally, we explore how the knowledge hierarchy's comparative advantage varies across case types and assess the extent to which task allocation aligns with comparative advantage.

*We thank Ellen Donath for excellent research assistance. We also thank many colleagues and seminar and conference participants at the HEFUU workshop in Uppsala, the URPP Annual Retreat at the University of Zurich, the Joachim Herz Foundation Add-on Fellowship Meeting in Hamburg, the CEPR Paris Symposium, the Stockholm School of Economics, and the London School of Economics for their insightful comments and feedback. We are grateful to the healthcare company for granting access to their data. We thankfully acknowledge financial support from the ESRC's Centre for Economic Performance and the Swiss National Science Foundation, grant number 100018-228312, and support from the University of Zurich's Research Priority Program "Equality of Opportunity". The research project was approved by the Stockholm Ethics Review Board.

The greatest improvement in the productive powers of labour... seem to have been the effects of the division of labor. — (Smith 1776)

1 Introduction

The division of labor and ensuing occupational specialization have long been central to economic studies of productivity (Smith 1776). While occupations have been shaped by historical and institutional factors, technological advancements and reduced communication costs transform how work is organized. In modern knowledge-intensive industries, characterized by complex tasks and low communication costs, one way to divide labor is through *knowledge hierarchies* (KH), where less specialized workers handle routine tasks at the front end and escalate more complex cases to experts (Garicano 2000).[1] Although knowledge hierarchies are widespread, empirical evidence on their productivity impacts is limited. This paper asks: What are the productivity effects of a knowledge hierarchy compared to directly involving experts? And to the extent that the two routes differ, do organizations allocate tasks according to comparative advantage? We study these questions in the context of healthcare.

Understanding the productivity effects of the division of labor is important in healthcare, where costs are rising in aging societies (World Health Organization 2021), and there are acute shortages of specialized medical personnel (Maier and Aiken 2016).[2] Meanwhile, innovations in information technology reshape the costs of coordination and the distribution of expertise in healthcare (Dahlstrand 2024; Dahlstrand et al. 2024; Bronsoler et al. 2022; Rajpurkar et al. 2022). Our study focuses on one lever used to cope with the challenges of modern healthcare: reorganizing the allocation of tasks between healthcare occupations.

To compare the productivity of a knowledge hierarchy with that of sending tasks directly to experts, we need a common set of tasks to be sent via both routes; an empirical strategy to overcome sorting in the allocation of tasks across routes; and a way to measure health outcomes and costs. We

[1] Knowledge hierarchies arise where higher layers rely on lower ones to transfer tasks. Examples include kitchen workers and chefs, machine operators and mechanical supervisors, or frontline customer support and specialized technicians. Licensing restrictions often shape these knowledge hierarchies, as between drafters and architects, paralegals and attorneys, or technicians and professional engineers.

[2] The US, for example, faces a predicted shortage of 86,000 medical doctors by 2036, about one-tenth of the 2021 workforce. See https://www.aamc.org/advocacy-policy/addressing-physician-workforce-shortage (last accessed: 12 Nov 2024).

overcome these empirical challenges by studying a large organization in Sweden that provides primary healthcare online. Tasks are patient cases with various combinations of symptoms, demographics, and medical histories, and are handled by doctors and nurses working online. An algorithm allocates tasks either to doctors or to nurses, who may refer some cases to doctors. While some cases are always sent directly to doctors and some always to nurses, a set of marginal cases can be sent to either route. These marginal cases are normally assigned to doctors, but under temporary congestion to doctors, the algorithm assigns cases to nurses. Conditional on case characteristics, assignment for these tasks thus depends on congestion, which allows us to identify the relative effect of the knowledge hierarchy route. Such congestion fluctuates rapidly at the online organization, resulting in the allocation to nurses of tasks that would otherwise have been assigned to doctors.

We analyze approximately 500,000 patient cases at the margin of being assigned to a nurse. For each case, we link comprehensive data from the organization to administrative data on healthcare utilization outside the organization and patients' demographic and socioeconomic information, allowing us to trace a patients' full care journey. We use detailed data on patient health outcomes, labor market outcomes, and the costs of healthcare to measure the knowledge hierarchy's productivity effects and any potential trade-offs.

Our empirical strategy uses quasi-exogenous variation from congestion to doctors in two ways. In the first approach, we apply a simple linear regression (OLS) model with detailed case controls and narrow date-by-four-hour time windows. This approach assumes conditionally random assignment, as congestion determines the algorithm's assignment of otherwise similar cases arriving around the same time. However, as the algorithm's exact functional form is unknown, residual selective sorting of cases might remain. To address this, our second approach implements an instrumental variables (IV) strategy, using the nurse share among active clinicians in prior consultations as an instrument for the current case assignment. This approach assumes that patient demand relative to staffing is not fully predictable and that previous nurse activity captures an increase in doctor congestion which affects the current case. By leveraging variation in congestion, we causally identify the effects of the nurse-initiated knowledge hierarchy compared to a direct-to-doctor assignment, independent of patient health conditions or preferences that typically drive case assignments in healthcare settings.

Although the marginal cases we study are handled by both nurses and doctors, their medical specialization differs. Physicians undergo six years of medical school and up to ten years of residency and specialization, while nurses complete a three-year program, typically without further specialization. Nurses often handle triage and patient communication but cannot prescribe medications or refer to specialists. They may, however, take on tasks less common for doctors, such as offering socio-emotional support or self-care advice. In practice, cases are handled by nurses, doctors, or both, but such organizational decisions are a black box to the researcher, leaving unclear whether these task allocations represent the most efficient use of available expertise.

We provide novel evidence on a knowledge hierarchy in primary healthcare. Our findings show that the large majority of cases at the margin, about 70%, are resolved by nurses without involving doctors, while 30% are referred to a doctor. The nurse-initiated knowledge hierarchy moderately reduces access to specialized doctor expertise, resulting in lower diagnostic specificity and a decline in prescription rates of 10 percentage points (from 44%). However, about a third of patients still receives a new prescription, suggesting that referrals may largely reflect occupational licensing limits.

While the knowledge hierarchy limits access to doctor-specific services, this does not indicate lower quality unless patients are negatively affected. To evaluate quality, we assess patient satisfaction and the use of external care services. We find that patients are no less satisfied in the knowledge hierarchy – if anything, the rate of top scores at the organization increases. While the rate of external primary care visits in two large Swedish regions, Scania and Stockholm, rises by about 3 percentage points (from 12%), we find no consistent increase in high-stakes acute secondary care use, such as urgent, emergency or hospital care. Thus, the knowledge hierarchy does not appear to substantially increase patients' imminent healthcare needs resulting from worse initial care provision.

While a patient's care journey reveals utilization patterns, it may miss unreported health declines that the patient does not seek care for. Therefore, we also examine medium- and long-term impacts on patients' lives, including income loss and mortality. Our findings indicate that the knowledge hierarchy, if anything, slightly *reduces* major income losses after the initial consultation by 2.1 percentage points (from 22%), with no effect on three-year mortality. These results suggest that the knowledge hierarchy has no detrimental effects on long-term economic well-being or survival.

Our results show that the knowledge hierarchy in primary care has little negative impact on patients. In the final step, we examine the implications on costs, including the initial and any downstream healthcare services covered by the public payer. We estimate that the knowledge hierarchy reduces marginal costs by 7.4% (OLS) to 8% (IV), or up to 93 SEK relative to a baseline cost of 1,162 SEK (about 100 USD) per direct-to-doctor care episode. These savings are driven by nurses solving the majority of cases independently and the cost of nurse consultations being less than two-thirds of doctor consultations. Overall, the nurse-initiated knowledge hierarchy moderately reduces healthcare costs while maintaining care quality – thus suggesting productivity gains.

The concept of specialization and task alignment underpins much of the literature on skill allocation and productivity (Acemoglu and Autor 2011; Jones 2009; Becker and Murphy 1992; Katz and Murphy 1992). Garicano (2000) highlights how knowledge hierarchies within organizations can facilitate efficient task allocation by enabling firms to leverage individual, differentiated skills. We empirically test these principles, examining how a knowledge hierarchy formed by nurses and doctors within an organization impacts productivity in healthcare.

By providing empirical evidence within an organization, we contribute to the seminal literature on comparative advantage (Ricardo 1817; Heckscher 1919; Ohlin 1924; Dornbusch et al. 1977; Costinot 2009), a key concept in theories of economic transactions. While firms are generally assumed to follow comparative advantage principles, empirical evidence on their application within organizations is limited. Our setup offers a unique opportunity to reveal how organizations implement comparative advantage between occupations with varying specialization in practice.

Recent economic research highlights the productivity effects of improved allocative efficiency among managers (Weidmann et al. 2024), differentiated skills in education (Biasi et al. 2021), and healthcare (Dahlstrand 2024). However, evidence on task allocation between specialized occupations remains limited. Recent studies have begun addressing this gap by examining productivity differences between healthcare occupations when scopes of practice change. Comparing primary to specialty care providers, Kristiansen and Serena (2024) find that mental health treatment quality improves when the scope of less specialized primary care providers is restricted, and young patients are sent directly to specialty psychiatric care. In a comparison of nurse practitioners (a US occupational category between a nurse and a doctor) and doctors, Chan and Chen (2022) show that doctors have a comparative and absolute advantage in the emergency department setting, whereas Currie and Zhang (2023) find that in a primary care setting, nurse practitioners are more effective than doctors at reducing hospitalizations and emergency visits. Unlike US nurse practitioners, nurses in our setting differ in their scope of practice and do not compete with doctors, although tasks still overlap. For example, Almström et al. (2024) shows evidence of task-shifting between nurses and doctors in private healthcare organizations in Sweden in response to relative reimbursement changes. Instead of comparing occupations head-to-head, our study takes a distinct approach by exploring a knowledge hierarchy where nurses decide whether to solve or refer cases, examining how nurses can both complement and substitute doctors.

Within the broader health economics literature, we contribute to the literature on human capital as a key factor in healthcare productivity and inefficiency. Prior research emphasizes the importance of specialization in healthcare provision (Arrow 1963; Chandra and Staiger 2007; Doyle et al. 2010; Currie and MacLeod 2017), and identifies the misallocation of specialized resources as the main driver of inefficiencies and rising costs (Baicker and Chandra 2004; Chandra et al. 2023). How differentiated, specialized skills are coordinated has major impacts on efficiency in healthcare (Chan 2016; Silver 2021; Dahlstrand 2024; Kelly et al. 2023). We extend this discussion by studying how a knowledge hierarchy may reduce inefficiencies through the organization of clinical work between occupations. By focusing on the primary care context, we examine how task-skill alignment across occupations shapes the patient's clinical path already at the first contact with the healthcare system.

Finally, the literature on occupational licensing and scope-of-practice reforms provides context for changes in traditional task allocations (Shapiro 1986; Kleiner 2016; Dillender et al. 2024). Studies show that weakened licensing restrictions for nurse practitioners have improved access to primary care, mental health, substance abuse treatment, and reduced mortality (Traczynski and Udalova 2018; Alexander and Schnell 2019; Guo et al. 2024; McMichael 2023). Unlike these studies, we focus on tasks nurses have always been authorized to perform but have been historically entrenched as doctors' work.

The remainder of the paper is organized as follows. Section 2 introduces the institutional background and our data. Section 3 describes our empirical framework. Section 4 discusses the productivity effects of the nurse-initiated knowledge hierarchy. Section 5 presents additional evidence on the mechanisms of the knowledge hierarchy. Section 6 concludes.

2 Background and data

To study the effects of division of labor, we examine how patient cases are allocated at an organization providing online primary care in Sweden. In our setup, patients can either be assigned to doctors, or they can be managed in a knowledge hierarchy structure under a division of labor between nurses and doctors. Our empirical setting provides several advantages. First, the healthcare firm's granular data on case assignments can be linked to regional and national administrative databases that allow us to assess treatment quality and costs. Second, our setting accommodates a common support of tasks that can handled by both doctors and nurses. Finally, we make use of internally set organizational rules in order to causally identify the implications of task allocation to a knowledge hierarchy.

2.1 Institutional setting

In order to study task allocation within a firm, we focus on a large Swedish online healthcare organization that we refer to as *the firm* or *the provider*.[3] This healthcare provider employs two types of clinicians for primary care consultations: *nurses* and *doctors*. Through these consultations, both types of clinicians solve cases of patients seeking healthcare, which we also refer to as *tasks*.

Healthcare in Sweden is primarily publicly funded and has comprehensive and universal coverage. The healthcare system operates on a decentralized model, with regional and local authorities being responsible for organizing healthcare services and reimbursing providers for services delivered. The bulk of healthcare costs is carried by the regions and funded through taxes, whereas patient contributions are low.[4]

Primary care services are typically the initial point of contact between patients and the healthcare system, where clinicians diagnose and treat common ailments and conditions. When needed, patients are referred to medical specialists or hospital services. Patients select and register with a primary healthcare center as their primary care provider.

We study one of these large primary care centers. The provider offers in-person consultations for registered patients, but primarily focuses on delivering primary care online. Consultations at the provider are covered by Sweden's universal public health insurance. The firm provides us with

[3]Many primary care providers in Sweden are large organizations employing up to hundreds of individual clinicians, comparable to large-scale group medical practices in the US (Almström et al. 2024). Telehealth consultations as offered by the provider we study are an increasingly popular approach to healthcare delivery. In the US, half of Medicare and Medicaid patients accessed telehealth services in 2020 (Centers for Medicare & Medicaid Services 2024). In Sweden, there were 555 online primary care consultations per 1,000 inhabitants in 2021 (Sveriges Kommuner och Regioner 2024). The firm we study is one of the leading digital healthcare providers in Sweden, handling a substantial share of all online consultations (we withhold exact market shares to maintain the provider's anonymity).

[4]The national insurance scheme involves small copayments for outpatient visits, including consultations in primary, specialty, and emergency care. The total patient fees for healthcare visits are capped annually, during our sample period at around 1,150 SEK (approximately USD 125). Patient copayments for online primary care ranged between 100 and 200 SEK (about 11 to 22 USD in 2020) for adults, and consultations for children were free.

detailed records of its consultations and clinicians assigned to each patient case.

Patients book consultations for themselves or their children via a mobile application. To access the service, they log in using an electronic identification system linked to their unique personal identity number (personnummer), which is used across Sweden's healthcare and government systems to identify individuals. This system allows the provider to retrieve key patient information, such as age, gender, and region, from civil records. After completing the identification process, patients select their symptoms from a predefined list and can provide additional information through a questionnaire or free text input. Consultations may be booked immediately as a drop-in or be scheduled for a future time. Once the booking is confirmed, patients are prompted to wait while a clinician is assigned to their case.

Whether a patient is assigned to a nurse or a doctor is determined by an algorithm. Our analysis focuses on marginal cases where either a nurse or a doctor can be assigned for the initial consultation. Historically, all consultations at the provider were handled by doctors, while nurses were primarily responsible for triaging patients or managing patient tracking (specifically, for a chlamydia screening program). However, starting in early 2019, the firm began to expand the role of nurses. Although doctors continued to manage the majority of cases, some marginal patient cases were allocated to consultations with nurses during periods of temporary congestion to doctors. Both nurses and doctors follow the same treatment protocols during the consultations and, in principle, can resolve cases. Unlike most healthcare settings, the allocation of these marginal cases to either a nurse or doctor is not based on the patient's health status or preferences. Instead, we can exploit quasi-exogenous variation in clinician assignment that arises from the firm's congestion management rules.

The case allocation algorithm assigns patients to either a nurse or a doctor based on a score that primarily considers their age, gender, region, and reported symptoms. Additionally, the algorithm adjusts for current congestion levels among doctors, which are influenced by both staff capacity and dynamically predicted patient demand. The exact algorithm has evolved over time and is kept as a trade secret. Physicians are compensated by the hour, whereas most nurses work part-time, hourly, or are self-employed. Because staffing schedules are set in advance, excess demand is common, and during periods of congestion, the system directs patients to nurse consultations to minimize waiting times. When congestion peaks, the provider may also contact off-shift clinicians to handle cases. This process allows congestion to be managed within short time frames of at most a few hours. Our analysis accounts for the firm's dynamic assignment mechanism, patient demographics, and reported symptoms to ensure the comparisons of outcomes at the margin.

When assigned a case, clinicians receive an overview of the patient's demographics, symptoms, and medical history, such as active medications or allergies, reported through previous interactions with the provider. Moreover, when a patient is forwarded to a new clinician, a mandatory text field for notes has to be filled out by the previous clinician. The text field asks the clinician to report the patient's symptoms, the reason for forwarding the case, and other consultation notes. When clinicians refer a case, they typically book a consultation in an available time slot with a specified clinician type (such as a doctor), but do not choose the identity of the follow-up clinician. During and after a consultation, clinicians may spend additional time managing administrative tasks related to the case.

Importantly, the digital interface for clinicians centers on their personal schedules of booked consultations. Clinicians can see the number of patients currently waiting and the number of clinicians on staff. However, this information does not vary across the provider network and is not directly accessible during individual consultations. This setup contrasts with most in-person healthcare environments, where clinicians generally have greater visibility into patient queues. As a result, we expect clinicians' treatment behaviors in this digital setting to be less influenced by immediate congestion.

We study marginal cases that can be handled by both doctors and nurses. Yet, these occupations are trained very differently. In Sweden, physicians must complete an 11-semester medical degree program and then undergo a full-time supervised internship of at least 18 months (Allmäntjänstgöring, AT) to receive a medical license. To become a specialist doctor, including in general medicine, most physicians then undergo at least 5 more years of full-time residency training (Specialiseringstjänstgöring, ST). Doctors thus spend 7 or more years in education. In contrast, nurses complete a 3-year degree program in order to obtain a nursing license. While nurses with at least one year of professional experience can also pursue a 1.5-year specialty training to work as district nurse (Distriktssköterska), or other specializations such as midwife or anesthesia nurse, the majority of nurses do not seek further specialization.[5] Nurses frequently communicate with patients during

[5] In our analysis sample, around 10% of nurses are specialized, whereas more than 75% of the medical doctors are in or have completed specialty training. While the Swedish healthcare sector faces general labor shortages, the markets for specialized doctors and nurses are particularly tight (Socialstyrelsen 2022). The report also highlights that

medical check-ups and make triaging decisions, but their scope of practice is more limited. In particular, nurses in Swedish primary care are not authorized for most medical drug prescriptions or specialist referrals.

Despite the differences, primary care patients in Sweden may be consulted by nurses, doctors, or a combination of both clinician types. If a nurse cannot resolve a case, such as due to occupational restrictions, they can pass the case along to a doctor. At the organization we study, the nurse in such a case books a follow-up consultation for the patient with a doctor. We consider cases initially assigned to a nurse – who may either resolve the issue or refer it to a doctor – as being managed within a knowledge hierarchy. This contrasts with cases directly assigned to a doctor. While we will also examine how clinicians at the firm organize follow-up care internally, whenever we refer to a *task assignment*, we specifically mean the initial assignment of the case.

2.2 Sample construction

The definition of our analysis sample starts from the 1.8 million consultations with the primary healthcare provider in the full calendar years 2019 and 2020. Out of these, we only consider initial, unscheduled consultations online. Our time frame encompasses 1 April 2019, when the organization started to employ nurses for patient consultations, until 24 December 2020, such that we can observe subsequent visits to the provider within a week. We exclude patients who are registered with the provider as well as infants (patients strictly below the age of two). These basic restrictions ensure that the algorithmic decision rules described above may assign nurses to a given patient consultation, as different rules govern consultations with patients registered at the provider or infants. We exclude cases for which baseline characteristics, such as patient demographics or length of the initial consultation, are missing. Lastly, we restrict the analysis sample to marginal cases that can be allocated to both, a doctor or a nurse-initiated knowledge hierarchy. This requires us to remove a set of cases for which the assignment is a priori deterministic and, thus, common support does not hold. We do so by identifying conditions with sufficient consultations held by either clinician type. In particular, we only include cases with symptoms for which we observe at least 5% of initial consultations to be assigned to either nurse or doctor. These restrictions leave us with almost 500,000

national workforce sizes differ, with 41,000 doctors and 113,000 nurses, suggesting a larger absolute nurse workforce with potentially greater scope for task reallocation.

marginal patient cases.[6]

We link each case to comprehensive administrative data on patients' previous and subsequent healthcare utilization, demographic, and socio-economic information from the Swedish National Board of Health and Welfare (Socialstyrelsen), Statistics Sweden (SCB), Region Scania, and Region Stockholm. We obtain information on medical drug prescriptions, as well as utilization and diagnoses in hospital care, urgent care, or specialist care services from Socialstyrelsen for the years 2013–2023. We observe consultations in primary care outside of the provider within the region of Scania from 2013-2020 and in Stockholm for the period 2013-2023.[7] The data from SCB include demographic information, such as age, gender, education level, and immigration status and socioeconomic information, such as employment status and educational attainment, from 2013-2020. In addition, we observe labor market information for patients, including annual sickness benefits paid out by the national social insurance agency and monthly earnings from patients' primary employment. By using patients' national identifiers, we can link data from the primary healthcare organization to external events outside that provider's services.

2.3 Variable definitions

We examine the differences and trade-offs between two task allocation modes: the knowledge hierarchy, where cases are initially assigned to nurses, and a direct-to-doctor task assignment. Our data enables us to study patients' health outcomes and the subsequent costs incurred in the healthcare system for each case, while accounting for heterogeneity in case characteristics. We outline our outcomes of interest and key control variables below. Appendix E provides additional details on our data.

2.3.1 Outcomes

To assess the net efficiency gains from the division of labor in a knowledge hierarchy setup, we consider three distinct perspectives: the firm, the patient, and the public payer who bears the costs of healthcare provision.

Organization of tasks at the firm. First, we examine how cases are managed within the

[6] Appendix Table A1 shows the total number of observations after each sample restriction, as well as separately for cases initially assigned to a nurse or to a doctor.

[7] Stockholm county is the capital region of Sweden. Scania, the third-largest county, is part of the aggregated region of South Sweden, and represents about 10% of cases in our sample.

firm to understand its internal organization. The knowledge hierarchy structure enables a flexible coordination of tasks. However, nurses and doctors still operate within distinct scopes of practice, with nurses facing regulatory limitations. Our first set of outcomes focuses on subsequent consultations with the provider, which may occur when patients are dissatisfied with their initial consultation and seek a different clinician or when the original clinician refers the case to another colleague. To differentiate between these scenarios, we focus on the rate of internal referrals for a follow-up consultation with a doctor. For cases initially assigned to nurses, this reflects the knowledge hierarchy concept, where nurses escalate cases to doctors. Additionally, we assess the time clinicians spend on each case. These outcome measures seek to provide insight into how tasks are managed within the knowledge hierarchy.

Quality of care for the patient. Second, we evaluate whether the nurse-initiated knowledge hierarchy impacts care quality, focusing on patient-centered outcomes. Our data allows us to track patients' healthcare journeys beyond the provider, enabling us to observe quality signals that may not be captured by the firm.

We begin by evaluating patients' access to services typically provided by doctors, such as diagnoses, drug prescriptions (from both the provider and external sources), and specialist referrals. Introducing an additional layer relying on nurses' judgments could reduce patient access to doctor-provided services. However, cases that require these services may still be referred up the knowledge hierarchy to a doctor, and a reduction in access may not be inefficient as long as it does not lead to adverse patient outcomes.

Next, we assess patient satisfaction through their ratings to capture whether patients' perceived care quality is affected by an assignment to the knowledge hierarchy. We also study the use of external primary care services outside the firm, using data from Scania and Stockholm and the subsample of patients registered in these regions. This analysis informs us about whether patients seek external care due to inconclusive treatment or dissatisfaction with the initial care provided.

In addition, to assess the impact of the knowledge hierarchy on health outcomes, we examine both immediate and medium- to long-term adverse events. We focus on high-cost imminent care events, such as urgent care visits, emergency department visits, and hospitalizations, which may indicate adverse health outcomes during the care journey. However, utilizing these services may also be beneficial if patients require a higher level of care, or patients may avoid seeking care even while health conditions deteriorate. To gain a more comprehensive understanding, we also study longer-term adverse events, including income reductions and patient mortality, that occur after the healthcare episode at the provider. These outcomes allow us to determine whether the mode of task organization affects patients' health outcomes beyond just healthcare utilization.

Costs incurred in healthcare. Lastly, we analyze the effects of the knowledge hierarchy on the cost of patients' care journeys, both at the primary care provider and downstream. While examining patient outcomes helps us assess the quality delivered under each work mode, evaluating costs allows us to determine the overall productivity effects of the nurse-initiated knowledge hierarchy compared to the direct-to-doctor route.

In Sweden's publicly funded healthcare system, most healthcare costs are borne by the public payer. We examine the expenses incurred by the public healthcare system, including a patient's consultations with our provider, prescriptions, specialist care, and any urgent or hospital care following the primary care visit. Table A16 presents the cost of each healthcare service and the sources of our data. Most cost estimates are derived from public sources and regional announcements on reimbursement rates, except for prescription costs and the cost of physical primary care consultations, which rely on data for our sample.

Thus, we evaluate multiple outcome sets: task organization within the firm, quality of care, and costs borne by the public system. For most outcomes, except for longer-term measures like income and mortality, we focus on the seven-day period following the initial consultation to clearly link outcomes to that event. Together, these outcomes allow us to draw conclusions on the broader efficiency trade-offs between the knowledge hierarchy and the direct-to-doctor route.

2.3.2 Control variables

While the assignment algorithm takes into account pre-determined case characteristics, such as patients' primary symptoms and demographic information, congestion changes over time. Our main analysis controls for patients' login time, which marks when they submit relevant case information and join the consultation queue, in addition to the pre-determined individual case characteristics. Below is an overview of the control variables included in the analysis.

Login-time. The time at which a patient logs in is crucial for case assignment, as both patient demand and clinician availability fluctuate throughout the day, affecting congestion. To account for this variation, we include fine-grained time controls by segmenting login times into 4-hour blocks starting at midnight of each calendar day.

Symptom categories. Before any clinician contact, patients select a primary symptom from a drop-down list. Appendix Figure A6 presents the symptom categories in our sample along with the number of initial nurse- and doctor consultations. The largest category, other health inquiries, covers about one-third of the sample and serves as a catch-all category for cases where patients did not choose a symptom from the pre-specified list. The remaining symptom categories in our data include common ailments, such as cold symptoms, infections, and Covid-19.[8]

Demographic information. We control for basic demographics by including, as control variables, patients' gender and indicators for various age categories. The age category variables reflect different life stages: toddlers (2-4 years), children (5-12 years), teens (13-19 years), adults (20-39 years), middle-aged (40-59 years), and seniors (60+ years). We also include indicators for the aggregated regional areas where patients are registered in the year prior to the initial consultation.

Health risk. Health risk is measured by patients' healthcare utilization over the three years prior to the consultation, excluding the 30 days immediately before the consultation to ensure risk is pre-determined. We use separate indicators for inpatient hospitalizations, emergency department (ED) visits, urgent care visits, and specialist visits. Additionally, we account for any comorbidities diagnosed in specialty or hospital care prior to 2019, the start of our analysis period.

Socio-economic background. We account for patients' socio-economic background using data from 2018, the year before the analysis period. Control variables include indicators for income above the sample median (calculated for patients aged 20 or older), employment status (categorized as employed, self-employed, or unemployed), and education level (including basic schooling, secondary education, and further post-secondary education of varying durations). Civil status is captured by indicators for marital status (married, unmarried, or previously married). As employment, education, income, and civil status data are unavailable for certain age groups (such as for underage patients), we include separate indicators for missing data on these dimensions. Finally, we consider migration status by including indicators for first-generation migrants, second-generation migrants, and patients with no foreign background.

[8] Appendix Table A2 lists the most common diagnosis codes in our analysis sample and in physical primary care consultations of patients in Scania. Our analysis sample excludes mental health conditions such as anxiety disorders or reactions to severe stress but otherwise includes conditions that are also common in physical primary care.

Table 1 presents summary statistics for the main analysis sample, along with characteristics of cases initially assigned to nurses versus those assigned directly to doctors. About 63% of the cases are female, and the average age is about 30 years. Cases assigned to nurses differ from those assigned to doctors most notably in terms of symptoms, reflecting the information used by the assignment algorithm. Nurse-assigned cases have a higher proportion of vaguely defined symptoms, such as "other health inquiries," abdominal pain, fever, or Covid-19. In contrast, cases directly assigned to doctors are more likely to involve cold-related symptoms like cold and flu, sore throat, or sinusitis, as well as specific infections, including eye infections and urinary tract infections. We also observe a higher proportion of women among cases directly assigned to doctors, along with minor imbalances across other characteristics.

Table 1. Characteristics of the analysis sample

	Full sample		Initially to nurse		Direct to doctor		Diff. p-value
Symptom categories							
Abdominal pain	0.030	[0.17]	0.045	[0.21]	0.025	[0.16]	0.00
Cold and flu	0.089	[0.28]	0.024	[0.15]	0.11	[0.31]	0.00
Cold sores	0.025	[0.16]	0.012	[0.11]	0.029	[0.17]	0.00
Constipation	0.0071	[0.084]	0.0062	[0.079]	0.0074	[0.086]	0.00
Covid-19	0.058	[0.23]	0.15	[0.36]	0.031	[0.17]	0.00
Diarrhea or vomiting	0.021	[0.14]	0.024	[0.15]	0.020	[0.14]	0.00
Eye infection	0.075	[0.26]	0.049	[0.22]	0.083	[0.28]	0.00
Fever	0.029	[0.17]	0.034	[0.18]	0.028	[0.16]	0.00
Headache	0.023	[0.15]	0.036	[0.19]	0.019	[0.14]	0.00
Nail problem	0.022	[0.15]	0.010	[0.10]	0.026	[0.16]	0.00
Other health inquiries	0.35	[0.48]	0.45	[0.50]	0.32	[0.47]	0.00
Bites and stings	0.054	[0.23]	0.029	[0.17]	0.062	[0.24]	0.00
Sinusitis	0.032	[0.17]	0.0093	[0.096]	0.038	[0.19]	0.00
Sore throat	0.076	[0.26]	0.057	[0.23]	0.081	[0.27]	0.00
Uncategorized	0.030	[0.17]	0.030	[0.17]	0.030	[0.17]	0.49
Urinary tract infection	0.079	[0.27]	0.033	[0.18]	0.093	[0.29]	0.00
Demographics							
Female	0.63	[0.48]	0.59	[0.49]	0.64	[0.48]	0.00
Patient age	29.2	[16.5]	29.3	[16.6]	29.2	[16.5]	0.51
West Sweden	0.20	[0.40]	0.20	[0.40]	0.20	[0.40]	0.07
Stockholm	0.44	[0.50]	0.45	[0.50]	0.44	[0.50]	0.00
Middle Sweden	0.19	[0.39]	0.18	[0.39]	0.19	[0.39]	0.00
South Sweden	0.095	[0.29]	0.093	[0.29]	0.095	[0.29]	0.02
Norrland	0.034	[0.18]	0.031	[0.17]	0.035	[0.18]	0.00
Småland + the islands	0.042	[0.20]	0.042	[0.20]	0.042	[0.20]	0.31
Health risk							
Any prior hospitalization	0.19	[0.39]	0.18	[0.39]	0.19	[0.39]	0.00
Any prior ED	0.33	[0.47]	0.34	[0.47]	0.33	[0.47]	0.00
Any prior urgent care	0.24	[0.42]	0.23	[0.42]	0.24	[0.43]	0.00
Any prior specialist	0.64	[0.48]	0.63	[0.48]	0.64	[0.48]	0.00
Any comorbidity	0.21	[0.41]	0.20	[0.40]	0.21	[0.41]	0.00
Socio-economic variables							
Income above median	0.32	[0.47]	0.31	[0.46]	0.33	[0.47]	0.00
Any benefits	0.11	[0.32]	0.11	[0.31]	0.11	[0.32]	0.00
Schooling < 9 years	0.049	[0.22]	0.052	[0.22]	0.048	[0.21]	0.00
Middle/High school	0.24	[0.42]	0.23	[0.42]	0.24	[0.42]	0.00
Further educ. < 3 years	0.11	[0.31]	0.10	[0.31]	0.11	[0.31]	0.00
Further educ. <= 3 years	0.19	[0.39]	0.18	[0.39]	0.20	[0.40]	0.00
Education n/a	0.42	[0.49]	0.43	[0.49]	0.41	[0.49]	0.00
Employed	0.57	[0.50]	0.55	[0.50]	0.57	[0.50]	0.00
Self-employed	0.042	[0.20]	0.042	[0.20]	0.042	[0.20]	0.51
Unemployed	0.050	[0.22]	0.052	[0.22]	0.049	[0.22]	0.00
Employment status n/a	0.34	[0.47]	0.36	[0.48]	0.34	[0.47]	0.00
Married	0.23	[0.42]	0.22	[0.41]	0.23	[0.42]	0.00
Divorced/Widowed	0.075	[0.26]	0.075	[0.26]	0.075	[0.26]	0.70
Unmarried	0.42	[0.49]	0.43	[0.49]	0.42	[0.49]	0.15
Civil status n/a	0.27	[0.45]	0.28	[0.45]	0.27	[0.45]	0.00
Not migrated	0.75	[0.43]	0.73	[0.45]	0.76	[0.43]	0.00
Immigrant 1st gen	0.16	[0.36]	0.17	[0.38]	0.15	[0.36]	0.00
Immigrant 2nd gen	0.096	[0.29]	0.100	[0.30]	0.095	[0.29]	0.00
Observations	490,505		111,707		378,798		490,505

Note: This table provides summary statistics for key case characteristics used as control variables in the main analysis sample. Symptom categories reflect the primary symptom reported by patients when requesting a consultation. Demographics include patients' gender and age at the time of consultation, and patients' registered aggregated region as of November 2018, before the start of the analysis period. Health risk variables capture healthcare utilization over the 3 years prior to the consultation, but excluding the 30 days immediately before, and include an indicator for any comorbidity diagnosed in specialist care. Socio-economic variables encompass indicators for above-median income (where the median is 294,700 SEK among patients over the age of 20), benefit receipt, employment status, education level, civil status, and migrant background, all recorded in November 2018. Standard deviations are shown in square brackets. The final column reports the p-value from a t-test comparing the means between patient cases initially assigned to a nurse and those directly assigned to a doctor.

3 Empirical framework

Our analysis examines the trade-offs between different work coordination structures. To causally identify the quality and cost effects of the knowledge hierarchy compared to a direct-to-doctor assignment, we must address a key identification challenge: Task allocation to occupations is not random, which likely leads to selective sorting of cases and, consequently, selection into different production processes. We tackle this challenge using two quasi-experimental approaches: (1) We exploit variation in congestion within a case's narrow time window along with detailed characteristics that influence the algorithm's case assignment; and (2) We employ the share of nurses among clinicians actively consulting previous marginal cases as a measure of congestion in a flexible instrumental variable (IV) setup, to account for unobserved residual selection in case assignment. Below, we outline our econometric approach and discuss the validity of our instrument.

3.1 Specification

To relate case outcomes to the impact of the knowledge hierarchy, we consider the following model structure:

$$Y_i = \delta_i Nurse_i + X_{i1}\beta_{1i} + X_{i2}\beta_{2i} + \epsilon_i, \tag{1}$$

where Y_i denotes the outcome of interest for case i and $Nurse_i$ denotes the initial assignment to a nurse instead of a doctor. X_{i1} is a vector of indicator variables capturing the baseline characteristics of a case, including login time (date-by-4 hours windows), the patient's main symptom, age and gender, and region. X_{i2} represents additional control variables, such as the patient's health risk and socioeconomic characteristics, as described in Section 2.3. ϵ_i denotes the error term.

δ_i captures the effect of the initial assignment to a nurse, and thus the knowledge hierarchy route, as compared to direct assignment to a doctor, and explicitly allows for heterogeneity in effects over cases. The unconditional assignment to a clinician type is not random: for instance, nurses may typically see patients with milder symptoms and less complex conditions. However, in our setting, case assignment is determined algorithmically by congestion at the time of case login and case characteristics. This allows us to exploit quasi-experimental variation in case assignment conditional on the case's login time and observable characteristics. We rely on an unconfoundedness assumption: $\mathbb{E}[\epsilon_i|X_{i1}, X_{i2}] = 0$, and estimate Equation 1 by Ordinary Least Squares (OLS) to obtain the average treatment effect on the treated (ATT), $\delta = \mathbb{E}[\delta_i|Nurse_i = 1]$. Since cases are not clearly assigned within clusters, we follow Abadie and Cattaneo (2018) and compute robust standard errors.[9]

However, the precise functional form of the assignment algorithm is proprietary to the firm and unknown to us. For example, the provider might incorporate complex interactions between patient age, symptoms, and congestion when assigning cases, and we cannot fully replicate the algorithm's rules. As a result, even with our rich set of control variables, we may not fully account for potential endogeneity in the initial assignment. To address this, we complement our OLS approach with an instrumental variable (IV) strategy that leverages exogenous variation in congestion.

Our IV strategy employs, as an instrument of the current case i's assignment, the *Share of nurses among active clinicians in the past 60 minutes.* The instrument is a leave-own-case-out measure of congestion among cases just prior to the assignment of a given case i. Our first stage thus takes the following form:

$$Nurse_i = \gamma_i Congestion_i + X_{i1}\theta_{1i} + X_{i2}\theta_{2i} + \nu_i, \tag{2}$$
$$\text{with } Congestion_i \equiv \frac{|\{Nurses\}_{t(i)}|}{|\{Nurses, Doctors\}_{t(i)}|}.$$

Here, $Congestion_i$ corresponds to the share of nurses as a fraction of clinicians (nurse or doctor) actively taking consultations in a time frame $t(i)$ of 60 minutes within our analysis sample prior to the login time of patient case i. $|\{Nurses\}_{t(i)}|$ ($|\{Nurses, Doctors\}_{t(i)}|$) define the set of active nurses (total active clinicians), and exclude case i's own clinician. We assume that no more nurses are staffed compared to doctors.[10] The vectors of time and case characteristics, X_{i1} and X_{i2}, are defined as above, and ν_i is an error term. In the IV specification, we estimate a two-stage least squares (2SLS) model represented by Equations 1 and 2 with robust standard errors.

[9] Section 4.4 shows that our conclusions are unchanged under alternative assumptions about the structure of the error terms.

[10] We impose this assumption based on discussions with the organization about their staffing, and given that we only observe active clinicians but not clinicians who do not end up taking consultations. The assumption implies that $\sup(Congestion_i) = 0.5$ and affects 0.78% of observations in the analysis sample.

3.2 Instrument validity

In the IV strategy, δ represents a local average treatment effect (LATE), that is, the average causal effect of being initially assigned to a nurse, as opposed to direct assignment to a doctor, for cases where the initial assignment is affected by congestion measured over prior cases. To interpret δ as the LATE, we require that our congestion instrument, the share of nurses among active clinicians in the past 60 minutes, satisfies the four standard IV assumptions: relevance, conditional independence, monotonicity, and exclusion. Below, we provide evidence supporting the validity of our congestion instrument.

Relevance. First, the instrumental variable should have a clear impact on the initial case assignment. We begin by presenting descriptive evidence in Figure 1 that our congestion instrument effectively captures variation in congestion that drives the initial assignment to a nurse. The left subfigure shows variation in the share of nurses among active clinicians in prior cases that we can leverage for our IV strategy. The right subfigure demonstrates an almost linear increase in the unconditional probability of an initial nurse assignment as the instrument increases. However, these figures do not account for temporal variation, as both congestion and staffing levels may fluctuate over time, both daily and within a given date. Appendix Figure A7 shows that congestion indeed varies, particularly by time of day and month. To address potential sorting over time correlated with congestion, we include granular fixed effects, using date-by-4-hour windows at the login time.[11]

Appendix Table A3 presents the first-stage regression estimates from Equation 2, when systematically expanding the set of control variables. Once we control for baseline case characteristics used by the algorithm (login date-by-4 hour windows, symptom, and patient demographics), the first-stage coefficient remains large and precisely estimated at 1.05. Moreover, the first-stage estimate remains stable when we include additional controls for patient health risk and socio-economic characteristics. The F-statistic for the first stage exceeds 11,000, confirming that our congestion instrument is both strong and predictive of initial assignment, even after accounting for detailed case characteristics.

Conditional independence. Second, we require that potential outcomes are independent of our congestion instrument, conditional on the control variables. If we had access to the assignment algorithm's exact score for each case based on its pre-determined characteristics, the realized

[11] In robustness checks, discussed in Section 4.4, we vary the size of login time windows and show that results remain largely unchanged.

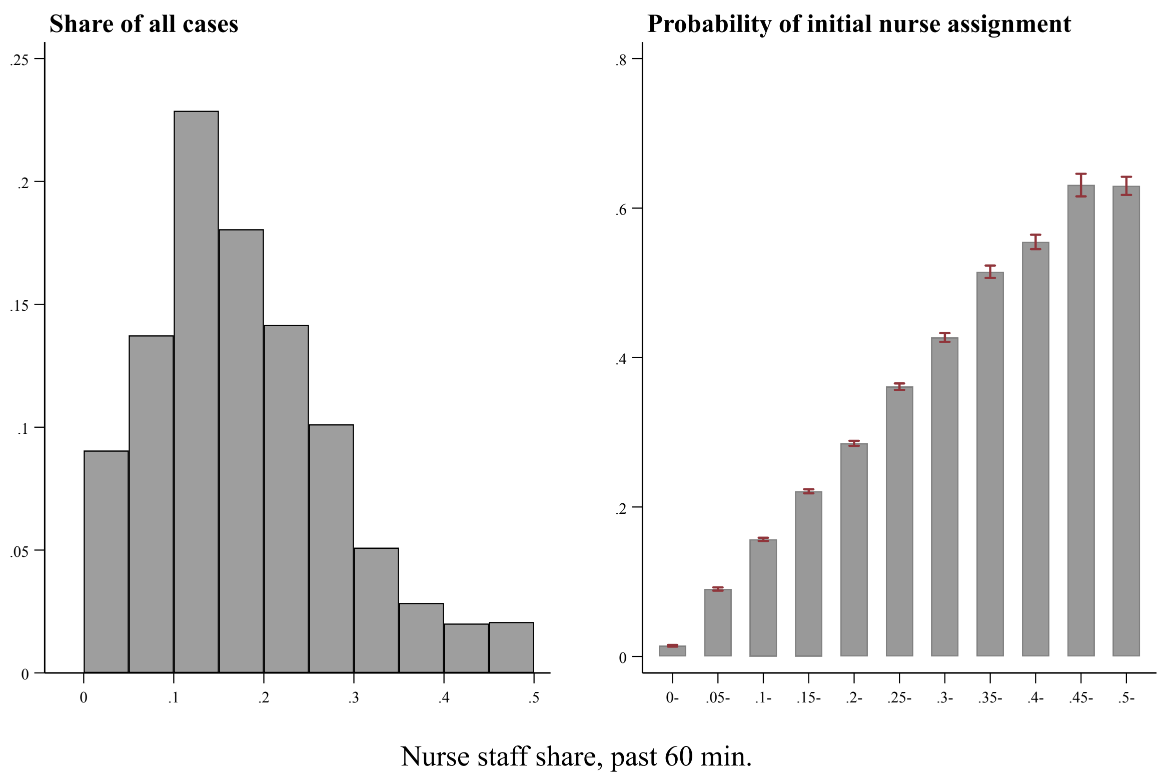


Figure 1. Variation in the congestion instrument

Notes: This figure shows descriptive figures for our congestion instrument, the share of nurses among active clinicians in the past 60 minutes of a given consultation. The left subfigure shows the distribution of the instrument. The right subfigure shows the average probability of an initial assignment to a nurse, our treatment, in categorized values of the instrument. Lines indicate 95% confidence intervals of the mean.

assignment would depend solely on congestion at the patient's login time. Our instrument would then be valid as long as congestion is uncorrelated with potential outcomes under nurse or doctor assignment. In the absence of the exact algorithm, we can condition on the baseline case characteristics that influence it: login time, symptom categories, and patient demographics. To ensure the validity of our empirical strategy, we assess whether additional case characteristics, such as patients' health risks or socioeconomic backgrounds, which could influence potential outcomes, are balanced relative to our congestion instrument after controlling for these baseline characteristics.

Figure 2 demonstrates that baseline case characteristics alone do not fully eliminate correlations between the initial case assignment and patients' health risks or socioeconomic status. However, once we condition on the baseline controls, the congestion instrument is largely balanced against these additional characteristics, both individually and jointly. Appendix Table A4 shows that congestion may be correlated with case characteristics and thus potential outcomes, but once we account for baseline observables used as algorithm inputs, the instrument is jointly balanced against additional case characteristics.

Monotonicity. Third, if treatment effects can be heterogeneous, we need to assume that the effects of congestion on treatment assignment are unidirectional across individual cases in order to interpret the IV estimate as the average causal effect for compliers. Specifically, our setup assumes that increased congestion does not reduce the likelihood of an initial nurse assignment, implying that cases sent to the nurse-initiated knowledge hierarchy under low congestion are also sent to a nurse under high congestion. This assumption aligns with our institutional context, where direct-to-doctor assignment was the default, and the nurse-initiated knowledge hierarchy route emerged later as a backup option during periods of high demand.

To verify testable implications of the monotonicity assumption, we first examine the correlation between our congestion instrument and treatment assignment across observable subgroups, which should be weakly positive in each subsamples (Bhuller et al. 2016; Dobbie et al. 2018). Appendix Table A5 presents the first-stage regressions results from Equation 2 for various splits along dimensions of health risk, demographics, and socioeconomic factors. Across these subgroups, we estimate positive and statistically significant first-stage coefficients that are quantitatively similar, indicating no evidence of a violation of monotonicity.

In addition, we can visually inspect the conditional expectation function of the treatment,

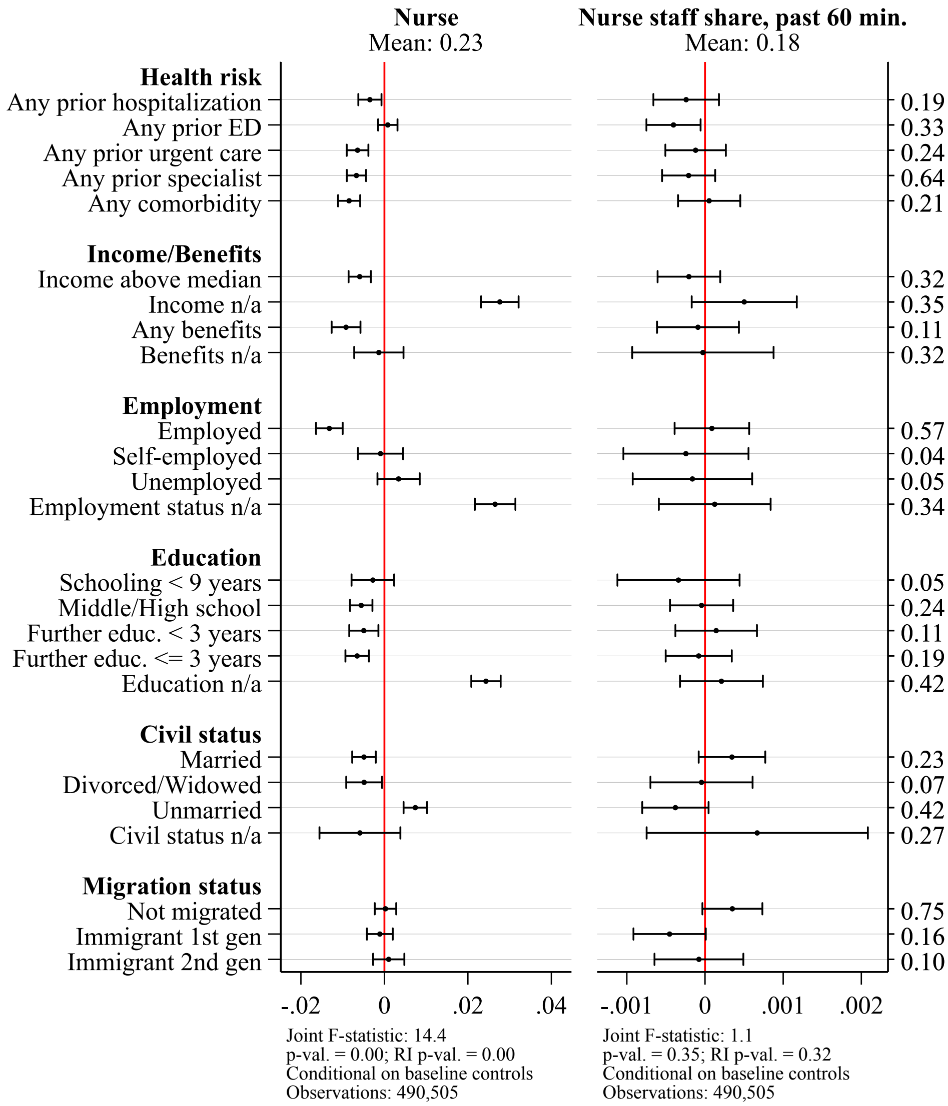


Figure 2. Balance of the treatment and congestion instrument in case characteristics

Notes: This figure presents balance tests for the treatment variable (*Nurse*) in the left and center columns, and for the congestion instrument (*Nurse staff share, past 60 min.*) in the right column. Each row shows the results of a bivariate regression where the treatment or instrument variable is regressed on a specific case characteristic. All regressions account for baseline case characteristics: login time, main symptom, and patient demographics. Login time indicates date-by-4 hours fixed effects. Symptom categories refer to indicators for the main symptom category reported by the patient when requesting a consultation. Demographics include indicators for patient gender, age categories, and aggregated regions. The horizontal lines represent 95% confidence intervals based on robust standard errors. The F-statistics reflect the joint F-tests for the treatment or instrument variable when regressed on all control variables, conditioned on login time. We report both, conventional as well as randomization inference (RI) p-values from 500 draws, based on Kerwin et al. (2024). The numbers on the far right display the sample means for each case characteristic.

initial nurse assignment, along the range of the congestion instrument in different subsamples using binscatter estimates (Cattaneo et al. 2024). Appendix Figure A8 shows the conditional expectation functions for subgroups defined by baseline characteristics, including symptom categories, age groups, and regions. We do not observe a downward slope or weak first stages in any case, further supporting the validity of the monotonicity assumption.

Exclusion restriction. Finally, in order to interpret the IV estimates as causal effects of the nurse-initiated knowledge hierarchy, we require that congestion affects case outcomes only through the initial case assignment. While the exclusion restriction assumption is untestable, we can assess potential violations.

The exclusion assumption is violated if clinicians adjust their work to high levels of congestion, which then affect patient outcomes. Although we cannot rule out clinician responses to congestion, we can examine the extent to which the congestion instrument correlates with consultation characteristics. Appendix Table A6 shows that patient waiting times increase (almost by definition) with congestion, but the differences are negligible compared to typical waiting times of patients seeking primary care and unlikely to affect outcomes: in the highest decile of congestion, waiting times are 29 minutes compared to 14 minutes in the lowest decile. Additionally, consultations lasting less than a minute, which may indicate clinicians checking in with patients due to long waiting times, are not more common during high congestion, and dropout rates are, in fact, lower. While clinicians spend slightly less time per case during high congestion, the difference (11.4 minutes vs. 11.6 minutes) is unlikely to affect patient care. Overall, we find little evidence that high congestion meaningfully impacts provider or patient outcomes.

As an additional falsification test, we construct a sample of cases where, by the logic of the case assignment algorithm, congestion does not affect assignment. These cases serve as a "zero first-stage" sample, where congestion should not affect any of the main outcomes (Angrist et al. 2010). We leverage that in a subset of symptom categories, the algorithm assigns cases to a doctor or nurse in a *strictly* deterministic fashion, regardless of congestion. Appendix Figure A9 shows these strictly deterministic cases, where fewer than 1% of cases are assigned to a nurse or a doctor. We create the congestion instrument using the share of nurses among active clinicians in marginal cases within the past 60 minutes of a strictly deterministic case.

Appendix Figure A10 provides descriptive evidence on the congestion instrument within the zero first-stage sample. The distribution is similar to that in the main analysis (Figure 1), suggesting that we do not capture time periods with markedly different congestion levels. Yet, we observe little systematic relation between the probability of a nurse assignment over the distribution of congestion. Appendix Table A7 presents first-stage F-statistics of at most 6.1, compared to over 11,000 in the main analysis. Additionally, the reduced-form regressions show no evidence of the congestion instrument affecting outcomes, and no significant effects from nurse assignments driven by congestion. This confirms that our congestion instrument operates as expected: when congestion does not influence the initial case assignment, it has no effects on outcomes.

4 Main results

We next present our results on the effects of the nurse-initiated knowledge hierarchy for the healthcare provider and patients. We study different outcome dimensions and compare cases initially assigned to a nurse with those assigned directly to doctors. We begin by examining how the healthcare provider organizes tasks at the margin and utilizes its primary resource: clinician time. Our findings show that in 70% of nurse-initiated cases at the margin, a case can be resolved by nurses themselves without referring it up the knowledge hierarchy. We then turn to patients and study their care journey outside the provider following an initial assignment to a nurse versus a doctor. We find that the nurse-initiated knowledge hierarchy impedes patients' access to high diagnostic quality and prescriptions. However, we find no consistent evidence that these effects translate into reduced care quality in terms of acute care events, income reductions, or mortality. Finally, we compare the marginal cost differences between the nurse-initiated knowledge hierarchy and the direct-to-doctor route. Considering the public payer's contribution to a case's care journey within and outside the organization, we estimate cost savings of up to 8% per care episode.

We provide both the OLS and IV estimates from the specifications outlined in Section 3.1. The OLS coefficients are informed by our understanding of the inputs into the assignment algorithm. However, since the exact functional form of the algorithm is unknown, we cannot fully rule out selective case sorting. The IV estimates rely on plausibly exogenous variation in congestion and impose weaker structural assumptions on the algorithm but may yield imprecise results.

4.1 Organization of tasks at the firm

We begin by focusing on the primary care provider and how patient consultations are organized within the knowledge hierarchy. Specifically, we examine the impact of an initial assignment to a nurse on internal follow-up consultations and clinician time spent on a case within up to seven days after the initial consultation date, compared to the baseline of a direct-to-doctor assignment. Table 2 presents OLS estimates from Equation 1, as well as IV and reduced-form estimates based on Equation 2, leveraging quasi-exogenous variation in initial case assignments due to congestion. All regressions include the full set of baseline and additional case controls described in Section 3.1. The reduced-form coefficients show that the congestion instrument significantly affects provider outcomes, validating its influence through the probability of initial nurse assignment.

We first analyze follow-up consultations with a doctor, which include referrals from nurses as well as other follow-up appointments that a clinician books, as long as they take place with a doctor. Table 2 indicates that an initial nurse assignment increases the probability of an internal referral to a doctor by about 30%, relative to a baseline follow-up rate of 1.4%. This implies that approximately 30% of marginal cases initiated by nurses are escalated to doctors, while follow-up consultations with doctors remain rare otherwise. However, these estimates also imply that nurses can solve most of their cases without referring to a doctor: About 70% of the cases on the margin are solved directly by the nurse.

Next, we consider all types of subsequent consultations, which may in addition capture repeat visits sought by patients. As shown in Table 2, the knowledge hierarchy increases the rate of any subsequent consultation by about 30%, relative to a baseline rate of about 12%. However, when we isolate repeat drop-ins, which could indicate that care was inconclusive, we find little evidence that subsequent consultations following an initial nurse assignment would be driven by unsatisfied patients seeking repeat care. The OLS estimate indicates a minor increase in repeat drop-in cases by 0.84 percentage points against a baseline of 5.5%, whereas the IV estimate is not statistically significant. Thus, while subsequent consultations increase in the knowledge hierarchy, this increase is driven by nurse referrals to doctors.

Finally, we examine how the knowledge hierarchy affects the total clinician time spent on a case during the care journey at the provider, accounting for any subsequent consultations. Table 2 reveals

Table 2. Effect of an initial nurse assignment on the organization of tasks at the provider

	Any referral to doctor (7d)			Any subseq. consultation (7d)			Any return drop-in (7d)		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	0.31***		0.30***	0.29***		0.26***	0.0084***		-0.0041
	(0.0015)		(0.0060)	(0.0017)		(0.0090)	(0.00091)		(0.0058)
Nurse staff share, past 60 min.		0.31***			0.27***			-0.0043	
		(0.0070)			(0.0098)			(0.0061)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505
First-stage K-P F-statistic			11,964			11,964			11,964
Baseline mean	0.014	0.014	0.014	0.12	0.12	0.12	0.055	0.055	0.055
	Clinician total time in min. (7d)			Doctor total time in min. (7d)			Nurse total time in min. (7d)		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	3.30***		3.19***	-7.98***		-8.41***	11.3***		11.6***
	(0.047)		(0.27)	(0.037)		(0.24)	(0.027)		(0.11)
Nurse staff share, past 60 min.		3.34***			-8.81***			12.2***	
		(0.28)			(0.26)			(0.16)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505
First-stage K-P F-statistic			11,964			11,964			11,964
Baseline mean	13.7	13.7	13.7	13.5	13.5	13.5	0.19	0.19	0.19

Note: This table presents regression results for our main specifications on the effects of initial case assignment to a nurse (the *knowledge hierarchy*) rather than directly to a doctor. Each subpanel corresponds to a different outcome as the dependent variable. Patient-provider interactions after the initial consultation are defined as follows: "Any referral to doctor" refers to any internal referrals or revisits with a doctor, "Any subsequent consultations" covers any interaction with the provider after an initial visit, and "Any return drop-in" indicates any unscheduled revisit for the same symptom. "Total time" defines the time that clinicians spend on a case, including consultation time spent with the patient and time on administrative work during the initial and all subsequent consultations. The left column in each subpanel reports the Ordinary Least Squares (OLS) estimate for the treatment variable (*Nurse*) based on Equation 1. The middle column presents the reduced-form regression estimates for the congestion instrument, measured as the share of nurses among active clinicians in the past 60 minutes (*Nurse staff share, past 60 min.*). The right column provides the Instrumental Variables specification, using *Nurse staff share, past 60 min.* as an instrument for *Nurse*, estimated via Two-Stage Least Squares (2SLS), following Equations 1 and 2. All regressions control for symptom categories, demographics, region, health risk factors, and socioeconomic variables, along with login date-by-4-hour fixed effects. The baseline mean represents the average of the dependent variable for cases directly assigned to a doctor. The first-stage K-P F-statistic refers to the Kleibergen-Paap F-statistic. Robust standard errors are shown in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

that total clinician time increases by about 20% in the knowledge hierarchy: Clinicians spend 3.2 (IV) to 3.3 (OLS) minutes longer on a case, from a baseline of 13.7 minutes that a directly assigned doctor spends on a case. When splitting by clinician type, we observe that the knowledge hierarchy reduces total doctor time by 7.89 (OLS) to 8.41 (IV) minutes, but nurses effectively overcompensate doctor time by spending 11.3 (OLS) to 11.6 (IV) minutes more.

Appendix Table A9 examines direct consultation time, showing that nurses in the knowledge hierarchy route also spend more time with patients rather than only increasing administrative time. Overall, the knowledge hierarchy shifts doctor workload to nurses.

We trace the patient's care journey over the seven calendar days following the consultation date. Appendix Table A8 provides additional results for provider outcomes defined within one day after the consultation date. The coefficient estimates are nearly identical, indicating that most cases are resolved by the next day.

4.2 Effects on the quality of care

We next turn to patient-centered outcomes to analyze how the knowledge hierarchy affects the quality of care provided. We consider a range of outcome measures: patients' access to diagnostic information, prescriptions, and specialists; their satisfaction with care at the provider; the use of external primary care; their utilization of acute, high-cost secondary care services; and the occurrence of adverse events. By evaluating a diverse set of patient outcomes across their care journey, both within and outside the primary care provider, we aim to provide a comprehensive assessment of any potential quality-of-care differences resulting from the nurse-initiated knowledge hierarchy at a patient's first point of contact with the healthcare system.

Table 3 presents the OLS estimates for our main patient outcomes, alongside the IV and reduced-form estimates based on the congestion instrument, with the full set of case characteristics controlled for as described in Section 3.1. Below, we detail the main results and their implications.

4.2.1 Diagnosis, prescriptions and specialist visits

First, we assess whether the nurse-initiated knowledge hierarchy impacts patients' access to services that doctors specialize in: diagnostic information, prescriptions, and referrals to specialists. Nurses are not trained in diagnosing and not authorized to provide prescriptions or specialist referrals.

Table 3. Effect of an initial nurse assignment on quality of care

	Informative diagnosis (7d)			Any new prescription (7d)			Any specialist (7d)		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	-0.064***		-0.068***	-0.10***		-0.083***	0.0030***		0.00029
	(0.0016)		(0.0095)	(0.0017)		(0.011)	(0.00073)		(0.0046)
Nurse staff share, past 60 min.		-0.071***			-0.087***			0.00031	
		(0.0100)			(0.012)			(0.0049)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505
First-stage K-P F-statistic			11,964			11,964			11,964
Baseline mean	0.81	0.81	0.81	0.44	0.44	0.44	0.036	0.036	0.036

	Rating: top score (7d)			Rating: physical replacement (7d)			Any external PCP consultation (7d)		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	0.045***		0.046***	0.015***		0.031***	0.029***		0.030***
	(0.0018)		(0.012)	(0.0018)		(0.012)	(0.0018)		(0.011)
Nurse staff share, past 60 min.		0.049***			0.033***			0.031***	
		(0.012)			(0.012)			(0.012)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	490,505	490,505	490,505	490,505	490,505	490,505	261,034	261,034	261,034
First-stage K-P F-statistic			11,964			11,964			6,028
Baseline mean	0.42	0.42	0.42	0.45	0.45	0.45	0.12	0.12	0.12

	Any urgent care (7d)			Any ED (7d)			Any hospitalization (7d)		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	0.0019***		0.0011	0.0061***		-0.0045	-0.00023		-0.0010
	(0.00051)		(0.0032)	(0.00074)		(0.0046)	(0.00039)		(0.0025)
Nurse staff share, past 60 min.		0.0011			-0.0047			-0.0011	
		(0.0034)			(0.0048)			(0.0026)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505
First-stage K-P F-statistic			11,964			11,964			11,964
Baseline mean	0.017	0.017	0.017	0.030	0.030	0.030	0.0087	0.0087	0.0087

	Income drop >20% (cal. mo. after)			Zero income (cal. month after)			Death excl. external causes (3y)		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	-0.021***		-0.0066	-0.0034***		-0.014*	-0.000096		-0.0010
	(0.0024)		(0.016)	(0.0013)		(0.0085)	(0.00016)		(0.00095)
Nurse staff share, past 60 min.		-0.0067			-0.014*			-0.0011	
		(0.016)			(0.0086)			(0.00100)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	213,425	213,425	213,425	213,425	213,425	213,425	490,505	490,505	490,505
First-stage K-P F-statistic			5,024			5,024			11,964
Baseline mean	0.22	0.22	0.22	0.050	0.050	0.050	0.0016	0.0016	0.0016

Note: This table presents regression results for our main specifications on the effects of initial case assignment to a nurse (the *knowledge hierarchy*) rather than directly to a doctor. Each subpanel corresponds to a different outcome as the dependent variable. Outcomes are defined as follows: "Informative diagnosis" excludes symptomatic/health status diagnoses (ICD R00-R99/Z00-Z99) and implies that any informative diagnosis was provided at the organization within seven days; "New prescription" refers to ATC codes unobserved in the previous 3 months; "Rating" implies that any top rating (top score or survey response that consultation replaced physical care) was given for any consultation at the provider within seven days; "External PCP consultation" refers to any visit to a primary care clinic other than at the provider (Region Stockholm) or an in-person primary care consultation with a doctor or nurse (Region Scania), and is defined for patients registered in Scania or Stockholm; "Specialist" refers to any specialist visit; "Urgent care center" refers to local care centers (Närakut); "ED" refers to emergency departments located at hospitals (Akutmottagning). "Income" refers to gross earnings from the main income source within a calendar year and are reported in each calendar month for employees. Income reductions capture any income drops in the calendar month following the consultation compared to the average in the three months prior, including sick leaves, as the employer replaces 80% of the regular salary in the first 14 days, and employer earnings become zero afterwards as the Swedish Social Insurance Agency then begins to pay out sickness benefits. To study income-related outcomes, we only consider a sample of cases for which patients are reported as employees with income in the three months prior to the consultation exceeding a small threshold of 3,533.33 SEK in every month (in 2010 values), based on the annual threshold of 42,400 SEK used by Saez et al. (2019). "Death excluding external causes" refers 3-year mortality excluding deaths due to accidents, self-inflicted harm, and other causes of death (ICD codes V01-Y89 but excluding X41-X42 and X44-X45, and including U12.9. The left column in each subpanel reports the Ordinary Least Squares (OLS) estimate for the treatment variable (*Nurse*) based on Equation 1. The middle column presents the reduced-form regression estimates for the congestion instrument, measured as the share of nurses among active clinicians in the past 60 minutes (*Nurse staff share, past 60 min.*). The right column provides the Instrumental Variables specification, using *Nurse staff share, past 60 min.* as an instrument for *Nurse*, estimated via Two-Stage Least Squares (2SLS), following Equations 1 and 2. All regressions control for symptom categories, demographics, region, health risk factors, and socioeconomic variables, along with login date-by-4-hour fixed effects. The baseline mean represents the average of the dependent variable for cases directly assigned to a doctor. The first-stage K-P F-statistic refers to the Kleibergen-Paap F-statistic. Robust standard errors are shown in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

To measure diagnostic quality, we consider whether any informative International Classification of Diseases (ICD) diagnosis code was provided in the patient records within seven days following the initial consultation.[12] Table 3 shows that initial assignment to a nurse decreases the likelihood of receiving an informative diagnosis by 6.4 (OLS) to 7 (IV) percentage points from a baseline rate of 81%, indicating worse diagnostic quality under the nurse-initiated knowledge hierarchy. Table A9 shows that the decrease in diagnostic specificity is driven by the initial consultation.

To examine access to prescription medications, we consider whether any new prescriptions, which patients do not regularly purchase for chronic conditions, are filled within seven days after the consultation.[13] Table 3 shows that the initial nurse assignment considerably decreases the odds of receiving a prescription by 8.3 (IV) to 10 (OLS) percentage points from a baseline of 44%. However, prescription rates in nurse-initiated cases still remain high, at least 34%, suggesting that referrals up in the knowledge hierarchy may be primarily due to occupational licensing restrictions, as nurses cannot prescribe.[14]

For access to specialists, we focus on overall specialist visits as we do not observe direct referrals. Table 3 presents mixed results on the effect of an initial nurse assignment on specialist service utilization within the next seven days: The OLS estimate indicates a small increase of 0.3 percentage points in specialist visits relative to a baseline of 3.6% for direct-to-doctor assignments, whereas the IV estimate is substantially smaller and insignificant.

These results indicate that the nurse-initiated knowledge hierarchy consistently reduces patients' access to high-quality diagnostic information and decreases the likelihood of receiving a prescription, with little effects on access to specialist care. However, these findings do not necessarily imply that patients receive less appropriate care.

[12] We consider ICD codes as *informative* if they are non-missing and exclude symptoms and health status updates (R00-R99 or Z00-Z99), which do not diagnose conditions. Diagnoses are rarely missing in our analysis sample (0.32% for initial consultations), as the provider is required to provide a diagnosis for reimbursement by the public payer. Uninformative ICD codes are also common in physical primary care: Only 52.59% of the diagnoses provided in nurse or doctor consultations in Scania are informative, or 66.88% conditional on non-missing diagnoses.

[13] We exclude any drugs from the Anatomical Therapeutic Chemical level 4 (chemical subgroups), which patients have already purchased at pharmacies within the 6 months prior to the initial consultation. External prescriptions are included in this analysis.

[14] Consistent with this interpretation, the new prescription rate for cases referred to doctors within the knowledge hierarchy is comparable to that in cases initially assigned to doctors, 41.87%. In contrast, only 17.91% of cases in the knowledge hierarchy that nurses resolve themselves result in a new prescription.

4.2.2 Patient satisfaction

Second, we evaluate patient-reported satisfaction with their care. After each consultation, patients are asked to rate how their case was handled in two ways: first, by providing an overall satisfaction score from one to five stars, and second, by indicating with "yes," "no," or "don't know" whether the consultation replaced the need for in-person care. These ratings allow us to assess whether patients in the nurse-initiated knowledge hierarchy are less satisfied with their care, potentially due to perceived quality differences or an expectation to speak with a doctor rather than a nurse.

We consider any top rating received over a case's full care journey within seven days of the initial consultation.[15] As shown in Table A9, an initial assignment to a nurse has, at most, a modest positive effect on patients' top score ratings. Patients are 4.5 (OLS) to 4.6 (IV) percentage points more likely to provide a top score during their initial or any follow-up consultation, relative to a baseline of 42% top score ratings. Additional analyses in Appendix Table A9 show that the knowledge hierarchy does not increase below top score ratings and that the rate of top ratings in the initial consultation is improved.

Our results regarding the question of whether online consultations replace physical care are less conclusive. The IV estimates suggest that the knowledge hierarchy increases positive responses by 3 percentage points from a baseline of 45%, while the OLS estimates indicate a smaller increase of 1.5 percentage points. However, Appendix Table A9 shows that the knowledge hierarchy increases negative responses to the physical care replacement question, and a decrease in the rate of positive responses when we only consider the initial consultation.

Overall, our results suggest an improvement in patient satisfaction ratings within the knowledge hierarchy, with no consistent impact on patients' perceptions of whether physical care can be replaced. Often, a direct-to-expert route is provided with the rationale that this increases satisfaction with a service – yet, we find little evidence that the knowledge hierarchy decreases patient satisfaction over the direct-to-doctor route.

[15] Of initial consultations, 48.22% (45.64%) received any satisfaction score (physical replacement rating); when including subsequent consultations, 50.94% (48.27%) had any score (physical replacement rating).

4.2.3 External primary care services

Third, we investigate whether an initial assignment to a nurse affects patients' use of external primary care providers (PCPs). A substantial share of nurse-assigned cases at the margin are resolved without a referral to a doctor, yet patients may prefer consulting a doctor and thus seek a primary care physician outside the provider we study.

We observe external PCP contacts in Scania and Stockholm, but information outside these regions is missing. Thus, we restrict the sample to roughly 50% of cases from patients registered either in Scania or Stockholm in 2018.[16]

Our focus is on substitution to physical primary care, as it is associated with higher costs than online care. Table 3 shows that the assignment to the nurse-initiated knowledge hierarchy may increase patient demand for external PCP.[17] While 12% of baseline cases handled by doctors result in a consultation with other providers within 7 days, the rate of external PCP consultations increases by 2.9 (OLS) to 3 (IV) percentage points. These findings suggest that while the nurse-initiated knowledge hierarchy resolves many cases without requiring a doctor at the organization, it may modestly substitute patient to physical external primary care.

4.2.4 Acute care

Fourth, we examine the effect of the nurse-initiated knowledge hierarchy on patients' use of acute care services, including high-cost emergency and hospital services. Specifically, we consider visits to urgent care centers – out-of-hours primary care facilities implemented in some regions to alleviate pressure on emergency care – as well as visits to emergency departments (ED), and hospitalizations. These outcomes serve as direct measures of potential adverse health events that may arise during patients' care journeys.

Table 3 presents conflicting results regarding the effect of an initial nurse assignment on the

[16] Data from Regions Scania and Stockholm are administratively collected and include services beyond traditional primary care (e.g., maternity checkups, counseling). For Scania, we limit data to physical consultations with nurses or doctors at clinics which patients can register as their PCPs. For Stockholm, we include only visits to primary care branches. For both regions, we exclude all contacts with the provider. Since patients registered in one region may seek care in another, baseline odds of primary care visits may be underestimated, but this is unlikely to bias our treatment estimates.

[17] In Scania, we directly restrict to physical consultations. In Stockholm, we cannot distinguish whether external primary care visits are online, but Scania data shows that 1.2% are followed by any external online primary care consultation within 7 days. As online consultations are cheaper, we at most overestimate the follow-up care costs under the knowledge hierarchy.

utilization of higher-cost acute care services. For urgent care services, the OLS estimate indicates a slight increase of 0.19 percentage points from a baseline of 1.7%, whereas the IV estimate shows no effect. For ED visits, the OLS estimate again points to an increase of 0.61 percentage points from a baseline rate of 0.3%, while the IV estimate is insignificant and negative. We observe no effects of the initial nurse assignment in either specification on hospitalizations.[18]

Overall, we do not observe consistent and substantial increases in the uptake of downstream acute care services for cases that are handled in the knowledge hierarchy instead of being directly managed by doctors.

4.2.5 Adverse events

Lastly, we examine medium- and long-term patient outcomes, specifically income reductions and mortality, which may arise from the insufficient handling of a patient's case, and capture events that adversely affect patients' lives outside their direct healthcare utilization.

We begin by investigating the occurrence of income reductions in the calendar month following the initial consultation, relative to the average income in the three months prior, from the patient's main income source. Our data includes monthly gross earnings from the main employer, which is defined as the largest source of income at the end of each calendar year. To minimize measurement error, we restrict our analysis to cases where patients have reasonably stable employment over the three months prior to the initial consultation.[19] Notably, any observed income reductions also capture sick leave, as employers in Sweden replace only 80% of regular salary for the first 14 days of sick leave, and after 14 days, income from the employer falls to zero as the Swedish Social Insurance Agency begins to pay sickness benefits.

In Appendix Table A9, we consider an alternative definition for income reductions. We examine income drops corresponding to the pattern of a full month of sick pay along with sickness benefits received during the calendar year. These additional results reflect the same overall pattern as our

[18] Appendix Table A9 also considers avoidable hospitalizations, defined as conditions considered preventable in primary care, based on a list of ICD diagnostic codes provided in Page et al. (2007). The results are inconclusive, with a marginally significant (at the 10% level) positive OLS estimate and an insignificant and negative IV estimate.

[19] Specifically, we consider cases where patients are employed at the end of the year, and their income in the three months preceding the consultation exceeds 3,533.33 SEK each month (in 2010 values). This threshold is based on the annual threshold of 42,400 SEK used by Saez et al. (2019) to ensure we do not capture employees changing jobs or seasonal workers. We focus on employed patients because monthly income from self-employment is calculated as the annual income divided by twelve in our data.

main outcomes: If anything, the knowledge hierarchy slightly reduces the rate of income reductions.

Table 3 shows that, if anything, the nurse-initiated knowledge hierarchy *reduces* the rate of major income reductions. Our OLS estimates suggest a moderate decrease of 2.1 percentage points in the likelihood of any income reduction exceeding 20%, compared to a baseline rate of 22% under the direct-to-doctor assignment, whereas the IV estimate is insignificant. When considering a full reduction to zero income from the main employer, the OLS estimate indicates a small decrease of 0.34 percentage points in the rate of zero income compared to a baseline rate of 5%, and the IV estimate points to a marginally significant (10% level) larger 1.4 percentage points decrease.

Next, we examine the effects of the nurse-initiated knowledge hierarchy on patients' three-year mortality following the consultation date. We focus on mortality from health conditions, excluding external causes such as accidents or self-inflicted harm. Table 3 reveals no causal effect of initial nurse assignment on the three-year mortality rate.[20]

Taken together, our results suggest that the nurse-initiated knowledge hierarchy has no negative impact on medium- or long-term outcomes for cases at the margin, compared to a mode of work in which all cases are handled by doctors.

4.3 Cost analysis

Finally, we consider the costs associated with service provision throughout patients' journeys in the healthcare system. Given that healthcare in Sweden is publicly funded, the majority of costs are covered through taxpayers' contributions. We take on the perspective of the public payer and examine whether any cost differences, including those from the utilization of healthcare services downstream, arise between the nurse-initiated knowledge hierarchy and direct-to-doctor assignments.

To compute cost differences between the nurse-initiated knowledge hierarchy and direct-to-doctor assignments, we consider the cost categories listed in Table 4. These include the costs of consultations at the provider, along with additional costs from drug prescriptions, specialist visits, and acute care services. As we only observe external primary care for a subset of cases from patients in Scania or Stockholm, we account for these costs separately.

We first provide an estimate of the total costs for a baseline case assigned directly to a doctor at

[20]We exclude deaths due to accidents, self-inflicted harm, and other external causes of death (ICD codes V01-Y89, excluding X41-X42 and X44-X45, but including U12.9). Table A9 shows similar results in terms of all-cause mortality.

Table 4. Baseline costs of the direct-to-doctor assignment

Cost category	Corresponding outcome	Cost (SEK)	Factor	Estimate (SEK)
Online primary care consultation, doctor		500	1	500
Online primary care consultation, nurse[1]		275	0	0
External primary care consultation[2]	Any external PCP consultation	1452.35	0.12	174.28
Prescription	Any new prescription	260	0.44	114.4
Specialist visit	Any specialist	3594	0.04	143.76
Emergency department visit	Any ED	3911.5	0.03	119.745
Urgent care center visit	Any urgent care	2002	0.02	40.04
Hospitalization	Any hospitalization	7800	0.009	70.2
Total baseline cost in SEK, incl. physical primary care			1162.43	
Total baseline cost in SEK, excl. physical primary care			988.15	

Notes: Costs for each service are sourced from public reports, regional announcements, or from own calculations for the analysis sample. Additional details are provided in Appendix Table A16.
[1] We assume that a direct-to-doctor assignment involves no nurse consultations at the provider.
[2] External primary care is only observed within Scania (physical nurse or doctor primary care consultation) and Stockholm (primary care contact).

our healthcare provider. Table 4 shows that these costs amount to roughly 988 SEK (about 90 USD) when excluding external primary care, or 1,162 SEK (about 100 USD) when including them. These figures are derived from a back-of-the-envelope calculation, where each cost category is weighted by the baseline rate of service usage, based on sample averages under a direct-to-doctor assignment.[21]

Next, Table 5 presents estimates of the cost differential between the nurse-initiated knowledge hierarchy and the direct-to-doctor assignment. We compute total costs for each case based on the categories in Table 4, and then take the natural logarithm. Our results indicate moderate cost savings under the knowledge hierarchy. For patients registered in Scania or Stockholm, for whom we observe external primary care, we estimate savings ranging from 7.4% (OLS) to 8% (IV), or 86 SEK to 93 SEK, relative to a baseline cost of 1,162 SEK. In the full sample, we compute similar cost savings of 75 SEK (OLS) to 84 SEK (IV) when imputing the costs of external primary care use.[22]

As highlighted in previous results, we find little evidence of increased downstream healthcare utilization among patients in the knowledge hierarchy compared to the baseline. However, at the provider level, we observe that nurses are able to resolve up to 70% of marginal cases without referring them up in the knowledge hierarchy. In addition, nurse consultations are substantially

[21] As nurse consultations at the provider occur in only a small share of direct-to-doctor cases (with nurses spending 0.07 minutes on average), we simplify the cost calculations by excluding nurse consultations in these cases.

[22] Table A9 shows estimated savings of 12% (OLS) to 13% (IV), or 118.5 SEK to 128.4 SEK, relative to a baseline cost of 988 SEK when excluding external primary care costs. External primary care usage increases by 3 percentage points, with an associated cost of 1,452.36 SEK (see Tables 3 and 4), resulting in an additional expense of 43.57 SEK (0.03 * 1,452.36 SEK), which we subtract from the savings.

cheaper, costing only 55% of the cost of a doctor consultation, as shown in Table 4. Appendix Table A9 further supports this, showing that about half of the cost reductions we observe are driven by cost savings at the provider, where savings range from 11% (OLS) to 13% (IV), or at least 55 SEK at a baseline provider cost of 500 SEK.

Overall, our estimates suggest reductions in healthcare costs from the nurse-initiated knowledge hierarchy of up to 8%. We also find no evidence of decreases in patient satisfaction or an increased occurrence of adverse patient events. Thus, the cost savings from the nurse-initiated knowledge hierarchy indicate marginal efficiency gains from assigning cases initially to a nurse.

4.4 Sensitivity

We now present several sensitivity tests to assess the robustness of our main results. First, we confirm that there are no pre-existing level or trend differences between the initial case assignments prior to the consultation. Second, we demonstrate that our findings remain consistent when using two alternative measures of congestion as instrumental variables: one based on the number of active clinicians and another based on the number of nurse consultations within cases that are almost always assigned to doctors. Finally, we verify the robustness of our main results across alternative control variables, different time windows for login, and varying assumptions regarding the structure of standard errors.

4.4.1 Outcome dynamics

In Figure 3, we present OLS and IV estimates of the effects of the nurse-initiated knowledge hierarchy from separate regressions of patient outcomes in the weeks leading up to and following the initial consultation. Subfigure 3a shows that patients have consultations with the provider in a similar manner prior to the initial consultation. In the week of the consultation, the nurse-initiated knowledge hierarchy significantly increases the likelihood of a follow-up consultation. Similarly, subfigures 3b, 3c, and 3d show no differential pre-trends in the rates of prescribing, emergency department visits, or hospitalizations.[23] Furthermore, any effects of the knowledge hierarchy disappear after the week of the consultation.

[23] In the main analysis, we consider prescriptions that have not been filled in the 6 months prior to the initial consultation in order to capture the likelihood of an additional prescription. In order to test for general prescribing patterns, we include chronic medications and consider any prescription in this analysis.

Table 5. Effect of an initial nurse assignment on healthcare costs

	Log cost, incl. ext. PCP		
	OLS	Red.	IV
Initially to nurse	-0.074***		-0.080***
	(0.0043)		(0.025)
Nurse staff share, past 60 min.		-0.082***	
		(0.026)	
Baseline characteristics	✓	✓	✓
Additional controls	✓	✓	✓
Observations	261,034	261,034	261,034
First-stage K-P F-statistic			6,028
Baseline mean	6.80	6.80	6.80

Note: This table presents regression results for our main specifications on the effects of initial case assignment to a nurse (the *knowledge hierarchy*) rather than directly to a doctor. Each subpanel corresponds to a different outcome as the dependent variable. Outcomes are defined as follows: Outcomes are defined as follows: "Log costs, incl. ext. PCP" represents the natural logarithm of the estimated care costs, including costs at the provider and the downstream expenses of any prescriptions, specialist visits, external primary care consultations, emergency department visits, urgent care visits, and hospitalizations within seven days of the initial consultation. The sample includes patients in Scania or Stockholm, for whom external primary care is observed. Estimates for each cost category are provided in Table 4. The left column in each subpanel reports the Ordinary Least Squares (OLS) estimate for the treatment variable (*Nurse*) based on Equation 1. The middle column presents the reduced-form regression estimates for the congestion instrument, measured as the share of nurses among active clinicians in the past 60 minutes (*Nurse staff share, past 60 min.*). The right column provides the Instrumental Variables specification, using *Nurse staff share, past 60 min.* as an instrument for *Nurse*, estimated via Two-Stage Least Squares (2SLS), following Equations 1 and 2. All regressions control for symptom categories, demographics, region, health risk factors, and socioeconomic variables, along with login date-by-4-hour fixed effects. The baseline mean represents the average of the dependent variable for cases directly assigned to a doctor. The first-stage K-P F-statistic refers to the Kleibergen-Paap F-statistic. Robust standard errors are shown in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

These figures thus indicate that health outcomes in the weeks before the initial consultation do not significantly differ between patients initially assigned to a nurse and those directly assigned to a doctor. Moreover, any effects we observe occur within the week of the consultation, supporting our choice of defining a seven-day care episode.

4.4.2 Alternative instrumental variable

To further strengthen our main results regarding the effect of the knowledge hierarchy, we construct an additional congestion measures and estimate our IV specifications using this alternative measures as instrumental variable for the initial case assignment. The results are largely consistent with our main findings, supporting the robustness of our conclusions.

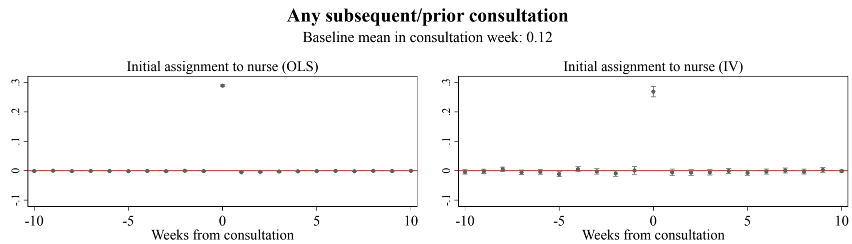
Our alternative instrument captures *doctor shortage*, measured by the count of nurse-initiated cases in the past 60 minutes that should almost deterministically be assigned to doctors. We define symptom categories with an initial consultation share above 95% as cases that are almost always assigned to doctors but still show some variation in their assignment. Given that this variable is right-skewed, we winsorize it at the 99th percentile to reduce noise. Appendix Figure A11 shows that, in over 60% of cases, no doctor-deterministic cases are assigned to a nurse. However, in about 10% of cases, this occurs 2 or more times, and there is a positive relationship between this instrument and nurse assignment.

Appendix Table A10 shows that, while the first-stage F-statistic is weaker and estimates are less precise with this alternative instrument, the IV results still largely support our main results. Specifically, we find similar effects of the knowledge hierarchy on the rate of referrals to a doctor, a larger reduction in the rate of new prescriptions, and marginally significant but larger reductions in costs. Additionally, there are no adverse effects on measures of patient health or adverse events.

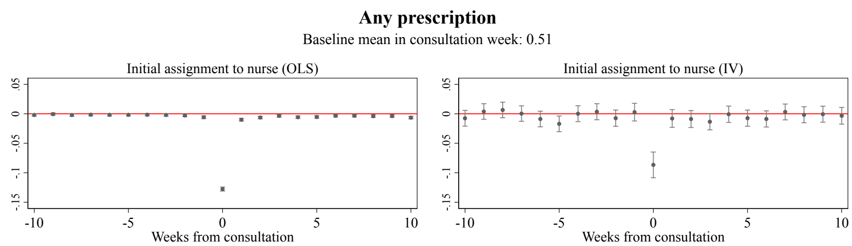
4.4.3 Robustness of the econometric specification

In Appendix Table A11, we assess the robustness of our results to systematic expansions of the set of control variables. The OLS and IV estimates for our main patient outcomes remain nearly unchanged across these specifications.

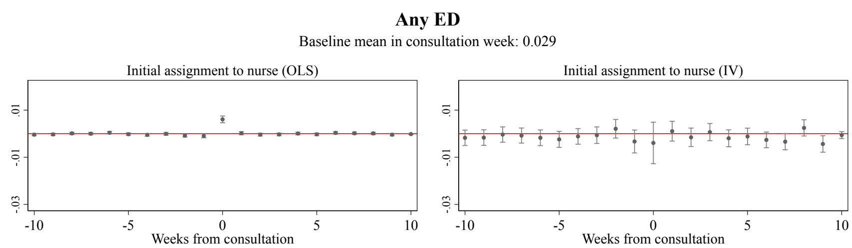
In Appendix Table A12, we consider alternative sets of fixed effects to control for the login-time of a case. Our conclusions remain unchanged when we use year, month, and weekday instead of



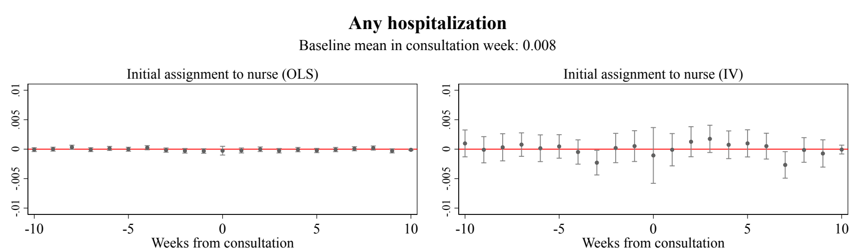
(a) Any consultation at the provider



(b) Any prescription



(c) Any emergency department visit



(d) Any hospitalization

Figure 3. Effect of an initial nurse assignment on lags and leads of patient outcomes

Notes: These figures show the estimated effect of an initial assignment to a nurse (the *knowledge hierarchy*) for outcomes in the 10 weeks prior to or after the initial consultation. "Week 0" marks the week starting with the initial consultation date. Each estimate is based on a separate regression of a lagged or lead outcome on the treatment variable (*Nurse*). Estimates are obtained by Ordinary Least Squares (OLS, left) based on Equation 1, or Two-Stage Least Squares (IV, right) using *Nurse staff share, past 60 min.* as an instrument for *Nurse* based on Equations 1 and 2. All regressions control for the full set of case characteristics, along with login date-by-4-hour fixed effects. The baseline mean represents the average of an outcome in week 0 for cases directly assigned to a doctor.

calendar date without 4-hour windows, or when we include even tighter date-by-hour fixed effects.

In Appendix Table A13, we present standard errors computed under alternative assumptions regarding their structure. These do not substantially differ when clustering by login date or by login date within 4-hour windows.

5 Mechanisms

The previous section discussed the causal effects of the nurse-initiated knowledge hierarchy on the provider, patients, and healthcare costs. In this section, we provide further insights into how different tasks are handled within the knowledge hierarchy. First, we descriptively characterize the types of cases that are more likely to be referred up the hierarchy. We find that more complex cases – such as those involving older patients or higher health risks – as well as clearly defined symptoms – which may require care outside of a nurse's scope of practice – are more often sent to doctors. Second, we show that more ambiguous cases are also where the knowledge hierarchy appears to be most effective. Finally, we highlight that more cost-efficient cases in the knowledge hierarchy are also more likely to be assigned to it on the margin.

5.1 Characterizing referrals in the knowledge hierarchy

In the previous section, we have shown that about 30% of marginal cases are referred up to doctors, which is the main driver of additional costs imposed on the healthcare system in the knowledge hierarchy. We now aim to better understand which cases are more likely to be referred, and which cases nurses are able to resolve on their own in the knowledge hierarchy.

Figure 4 examines the bivariate correlations between referrals up the knowledge hierarchy and various case characteristics. The figure highlights several key differences in referral rates based on symptom type, age, and health risk, allowing us to draw two main conclusions. First, cases involving older patients, those with comorbidities, or those requiring specialist visits or prior hospitalization are more likely to be referred to doctors. This is consistent with the theory behind the knowledge hierarchy, which suggests that more complex cases should be handled by more specialized experts. Appendix Table A14 supports this notion by showing that doctors spend more time on cases referred to them by nurses compared to cases directly allocated to them.

Second, Figure 4 reveals that certain more clearly defined symptoms – such as urinary tract infections, sore throats, or bites and stings – are more likely to be referred up the knowledge hierarchy. In contrast, less clearly defined symptoms, such as uncategorized symptoms, abdominal pain, or health inquiries, are less likely to be escalated. Additionally, cases involving children are less likely to be referred. These findings suggest a comparative advantage of the knowledge hierarchy in a subset of tasks defined by patient age and symptom clarity.

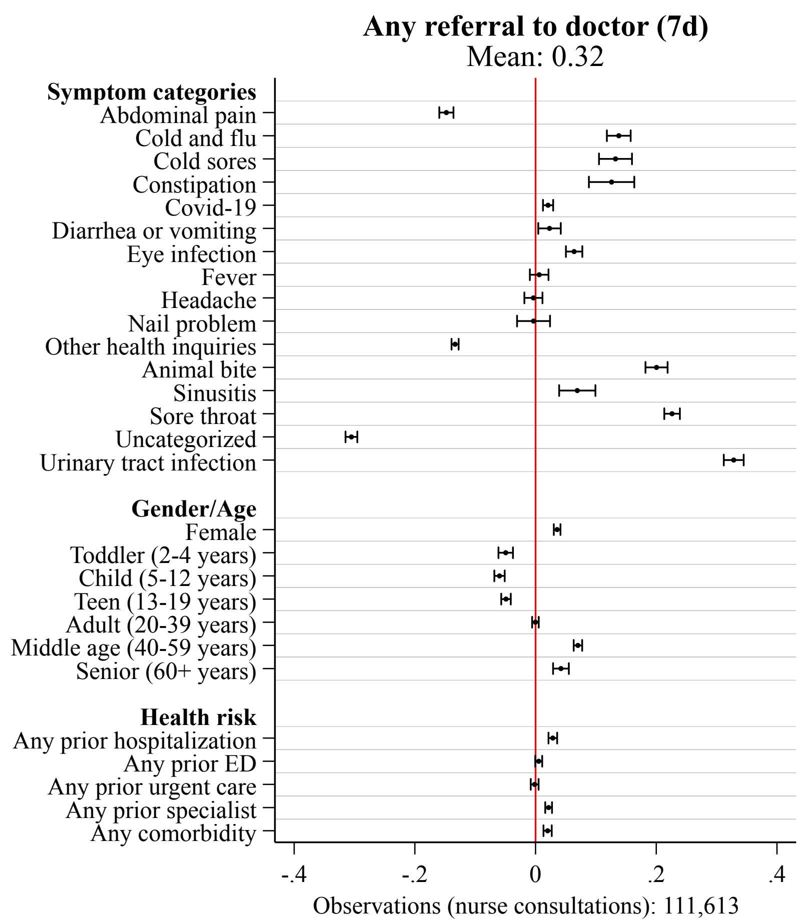


Figure 4. Correlates of referrals from nurses to doctors

Notes: This figure presents, for the subset of cases initially assigned to a nurse, correlates for referrals to doctors (*Referral to doctor*) and thus up in the knowledge hierarchy. Each row shows the results of a bivariate regression where *Referral to doctor* is regressed on a specific case characteristic. All regressions account for the login date-by-4 hours fixed effects. Symptom categories refer to indicators for the main symptom category reported by the patient when requesting a consultation. The horizontal lines represent 95% confidence intervals based on robust standard errors. Mean refers to the average rate at which cases are forwarded from a nurse to a doctor in the sample of nurse-initiated cases.

5.2 Task heterogeneity

Building on our finding that referral rates up the knowledge hierarchy are lower for cases with ambiguous symptoms or involving children, we now explore whether this reflects a comparative advantage of the knowledge hierarchy in these case types. Specifically, we ask: Are these the cases where the knowledge hierarchy is most effective in reducing healthcare costs?

To this end, we estimate the following specification:

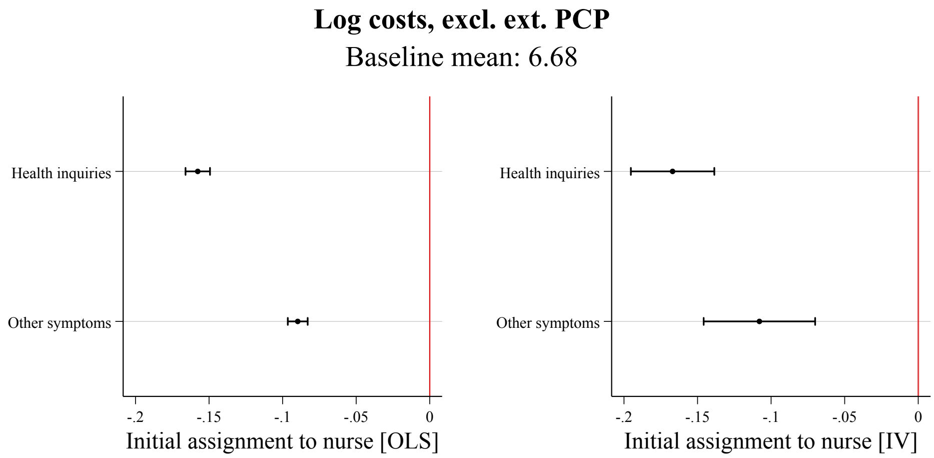
$$Y_i = \sum_g \delta_g (I_{g(i)=g} \times Nurse_i) + \kappa_g + X_{1i}\tilde{\beta_1} + X_{2i}\tilde{\beta_2} + \tilde{\epsilon}_i, \tag{3}$$

where g represents a subgroup, $I_{g(i)=g}$ indicates that case i belongs to subgroup g, and the other variables are defined as in 1. In the IV approach, we instrument $(I_{g(i)=g} \times Nurse_i)$ by $(I_{g(i)=g} \times Congestion_i)$, as described in 2. Estimates of δ_g provide insight into the subgroup-specific effects of the knowledge hierarchy on costs.[24]

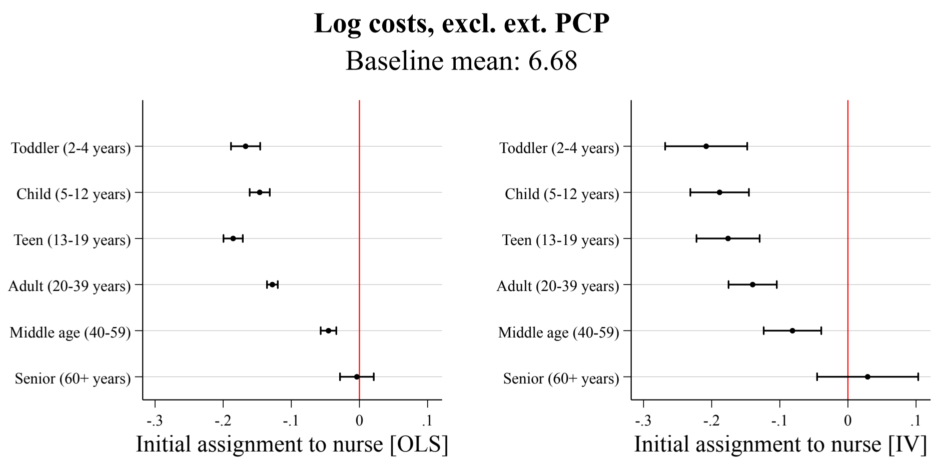
Figure 5 presents the estimates of δ_g, comparing subgroups g based on symptom categories and across age groups. The figure supports the idea that cost savings from the knowledge hierarchy are most pronounced for cases with ambiguous symptoms or for younger patients. In subfigure 5a, our OLS estimates show that the knowledge hierarchy reduces costs for "other health inquiries" by nearly twice as much as for other symptoms (16% vs. 9%). Subfigure 5b reveals an age gradient in cost reductions: Among older patients, the knowledge hierarchy shows no effect on costs, while the cost reductions are strongest for toddlers, children, and teens (up to 18% in the OLS and 26% in the IV approach). Appendix Figure A12 further shows that these cost savings are driven by lower rates of referrals up the knowledge hierarchy.

One potential explanation for the observed heterogeneity across symptoms and age is that some cases may involve greater uncertainty about the patient's condition. For example, children brought in through their parents may not be able to clearly articulate their symptoms, limiting the information available. In these cases, where the severity of the condition is uncertain, nurses may be capable of providing appropriate care and prevent unnecessary doctor visits, which results in cost savings.

[24] In order to make use of our full sample, this part of the analysis considers healthcare costs excluding external primary care.



(a) Heterogeneity by main symptom category



(b) Heterogeneity by age group

Figure 5. Heterogeneity across cases in total costs, excluding external primary care

Note: These figures show the heterogeneous effects of an initial assignment to a nurse (the *knowledge hierarchy*) on costs, broken down by subgroups defined either by whether the main reported symptom is "Other health inquiries" (subfigure 5a) or by age category (subfigure 5b). The outcome "Log costs, excl. ext. PCP" represents the natural logarithm of the estimated care costs. These costs include any downstream expenses from online doctor or nurse primary care consultations at the provider, such as prescriptions, specialist visits, emergency department visits, urgent care visits, and hospitalizations that occur within seven days of the initial consultation, but exclude the costs of external primary care consultations. Each subfigure presents estimates from a regression of the outcome on all interactions of the initial nurse assignment (*Nurse*) with each subgroup, while controlling for subgroup fixed effects, the full set of case characteristics, and login date-by-4-hour fixed effects, following Equation 3. The estimates are obtained by Ordinary Least Squares (OLS, left) or Two-Stage Least Squares (IV, right), where interactions of *Nurse staff share, past 60 min.* with each subgroup serving as instruments for the interactions between *Nurse* and each subgroup. Each row represents the estimated subgroup-specific effect of an initial nurse assignment, with horizontal lines representing 95% confidence intervals based on robust standard errors.

5.3 Complier characteristics

Finally, we seek to better understand which cases are moved to the nurse-initiated knowledge hierarchy under increasing congestion. Our IV estimates, as presented in Section 4, represent a local average treatment effect (LATE), that is, the average causal effect of the nurse-initiated knowledge hierarchy on the subgroup of compliers whose probability of initial assignment to a nurse is influenced by congestion. We now characterize this group of complier cases to investigate whether the allocation of tasks corresponds to the heterogeneous effects of the knowledge hierarchy on costs.

Appendix Table A15 compares the mean characteristics of the complier group with those of the overall analysis sample, following (Frandsen et al. 2023) and (Abadie 2003). Complier cases are more likely to seek a consultation for "Other health inquiries" (a catch-all category) or Covid-19. In contrast, the complier group has a lower share of urinary tract infections, sore throats, or eye infections, which may require a prescription and tend to be more specific. Complier cases are also slightly less likely to be female, but are barely different in other risk or socio-economic characteristics from the full sample. Overall, these results suggest that cases assigned to the knowledge hierarchy are less precisely defined and correspond to the subset of cases with larger cost savings.

6 Conclusion

In knowledge-intensive sectors where expertise is scarce and costs are escalating, optimizing task allocation across differentiated skills is becoming increasingly critical. We study the effects of the division of labor when tasks are organized in a knowledge hierarchy. We focus on an industry particularly affected by high costs and scarce resources: healthcare.

Our study provides evidence that a nurse-initiated knowledge hierarchy in primary care reduces costs imposed on the public healthcare system without compromising care quality. We find that this approach achieves cost savings of up to 8%, driven by nurses' ability to resolve 70% of cases on their own, with little evidence of lower quality. The knowledge hierarchy maintains high levels of patient satisfaction and does not adversely affect health outcomes, as measured by acute care needs, labor market outcomes, or patient mortality. We find that the knowledge hierarchy particularly reduces healthcare costs for cases with ambiguous symptoms or involving younger patients.

Our findings contribute to the understanding of how the division of labor can enhance productivity by aligning tasks to occupational competencies in healthcare, within the organization, and downstream. Beyond healthcare, our results have implications for similar knowledge-intensive sectors with highly specialized occupations, suggesting that the organization of tasks in a knowledge hierarchy rather than under rigid occupational norms may improve efficiency when differentiated tasks require varying levels of expertise.

References

Abadie, Alberto, "Semiparametric instrumental variable estimation of treatment response models," *Journal of Econometrics*, April 2003, *113* (2), 231–263.

_ **and Matias D. Cattaneo**, "Econometric Methods for Program Evaluation," *Annual Review of Economics*, August 2018, *10* (1), 465–503.

Acemoglu, Daron and David Autor, "Skills, Tasks and Technologies: Implications for Employment and Earnings," in "Handbook of Labor Economics," Vol. 4, Elsevier, 2011, pp. 1043–1171.

Alexander, Diane and Molly Schnell, "Just what the nurse practitioner ordered: Independent prescriptive authority and population mental health," *Journal of Health Economics*, July 2019, *66*, 145–162.

Almström, Axel Hellbom, Lina Maria Ellegård, Klara Stromberg, and Andreea Enache, "Mixed Payment and Mixed Objectives: Insights from the Ownership Structure in Swedish Primary Care," 2024.

Angrist, Joshua, Victor Lavy, and Analia Schlosser, "Multiple Experiments for the Causal Link between the Quantity and Quality of Children," *Journal of Labor Economics*, October 2010, *28* (4), 773–824. Publisher: The University of Chicago Press.

Arrow, Kenneth J., "Uncertainty and the Welfare Economics of Medical Care," *The American Economic Review*, 1963, *53* (5), 941–973.

Baicker, Katherine and Amitabh Chandra, "The Productivity of Physician Specialization: Evidence from the Medicare Program," *American Economic Review*, April 2004, *94* (2), 357–361.

Becker, Gary S. and Kevin M. Murphy, "The Division of Labor, Coordination Costs, and Knowledge," *The Quarterly Journal of Economics*, 1992, *107* (4), 1137–1160. Publisher: Oxford University Press.

Bhuller, Manudeep, Gordon B Dahl, Katrine V Løken, and Magne Mogstad, "Incarceration, Recidivism, and Employment," *Journal of Political Economy*, 2016.

Biasi, Barbara, Chao Fu, and John Stromme, "Equilibrium in the Market for Public School Teachers: District Wage Strategies and Teacher Comparative Advantage," *NBER Working Paper*, 2021.

Bronsoler, Ari, Joseph Doyle, and John Van Reenen, "The Impact of Health Information and Communication Technology on Clinical Quality, Productivity, and Workers," *Annual Review of Economics*, August 2022, *14* (Volume 14, 2022), 23–46. Publisher: Annual Reviews.

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng, "On Binscatter," *American Economic Review*, May 2024, *114* (5), 1488–1514.

Centers for Medicare & Medicaid Services, "Medicare Telehealth Trends," 2024. Accessed: 2024-10-28.

Chan, David and Yiqun Chen, "The Productivity of Professions: Evidence from the Emergency Department," October 2022.

Chan, David C., "Teamwork and Moral Hazard: Evidence from the Emergency Department," *Journal of Political Economy*, June 2016, *124* (3), 734–770.

Chandra, Amitabh and Douglas O. Staiger, "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks," *Journal of Political Economy*, February 2007, *115* (1), 103–140. Publisher: The University of Chicago Press.

_ , Carrie Colla, and Jonathan Skinner, "Productivity Variation and Input Misallocation: Evidence from Hospitals," Technical Report w31569, National Bureau of Economic Research, Cambridge, MA August 2023.

Costinot, Arnaud, "An Elementary Theory of Comparative Advantage," *Econometrica*, 2009, *77* (4), 1165–1192.

Currie, Janet and Jonathan Zhang, "Doing More with Less: Predicting Primary Care Provider Effectiveness," *The Review of Economics and Statistics*, February 2023, pp. 1–45.

_ and W Bentley MacLeod, "Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians," *Journal of labor economics*, 2017, *35* (1), 1–43.

Dahlstrand, Amanda, "Defying Distance? The provision of services in the digital age," 2024.

_ , Nestor Le Nestour, and Guy Michaels, "Online versus In-Person Services: Effects on Patients and Providers," 2024.

Dillender, Marcus, Anthony T. Lo Sasso, Brian J. Phelan, and Michael Richards, "Occupational Licensing and the Healthcare Labor Market," *Journal of Human Resources*, February 2024, pp. 0722–12450R2.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang, "The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review*, February 2018, *108* (2), 201–240.

Dornbusch, R., S. Fischer, and P. A. Samuelson, "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods," *The American Economic Review*, 1977, *67* (5), 823–839.

Doyle, Joseph J., Steven M. Ewer, and Todd H. Wagner, "Returns to physician human capital: Evidence from patients randomized to physician teams," *Journal of Health Economics*, December 2010, *29* (6), 866–882.

Frandsen, Brigham, Lars Lefgren, and Emily Leslie, "Judging Judge Fixed Effects," *American Economic Review*, January 2023, *113* (1), 253–277.

Garicano, Luis, "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 2000, *108* (5), 874–904. Publisher: The University of Chicago Press.

Guo, Jiapei, Angela E. Kilby, and Mindy S. Marks, "The impact of scope-of-practice restrictions on access to medical care," *Journal of Health Economics*, March 2024, *94*, 102844.

Heckscher, Eli, "The Effect of Foreign Trade on the Distribution of Income," *Ekonomisk Tidskrift*, 1919, *21*, 497–512.

Jones, Benjamin F., "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?," *The Review of Economic Studies*, January 2009, *76* (1), 283–317.

Katz, Lawrence F. and Kevin M. Murphy, "Changes in Relative Wages, 1963–1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, February 1992, *107* (1), 35–78.

Kelly, Elaine, Carol Propper, and Ben Zaranko, "Team Composition and the Returns to Human Capital: Evidence from Nursing Teams," *Centre for Economic Policy Research*, 2023.

Kerwin, Jason, Nada Rostom, and Olivier Sterck, "Striking the Right Balance: Why Standard Balance Tests Over-Reject the Null, and How to Fix it," *SSRN Electronic Journal*, 2024.

Kleiner, Morris M., "Battling over Jobs: Occupational Licensing in Health Care," *American Economic Review*, May 2016, *106* (5), 165–170.

Kristiansen, Ida Lykke and Benjamin Ly Serena, "De-specialization in health care: Evidence from mental health treatments," 2024.

Maier, Claudia B. and Linda H. Aiken, "Task shifting from physicians to nurses in primary care in 39 countries: a cross-country comparative study," *European Journal of Public Health*, December 2016, *26* (6), 927–934.

McMichael, Benjamin J., "Supply-side health policy: The impact of scope-of-practice laws on mortality," *Journal of Public Economics*, June 2023, *222*, 104901.

Ohlin, Bertil, "Handelns teori (The Theory of Trade)." PhD dissertation, Stockholm University, Stockholm, Sweden 1924.

Page, Anthea, Sarah Ambrose, John Glover, and Diana Hetzel, *Atlas of Avoidable Hospitalisations in Australia: Ambulatory Care-Sensitive Conditions*, Adelaide: PHIDU, University of Adelaide, 2007.

Rajpurkar, Pranav, Emma Chen, Oishi Banerjee, and Eric J. Topol, "AI in health and medicine," *Nature Medicine*, January 2022, *28* (1), 31–38. Publisher: Nature Publishing Group.

Ricardo, David, "The Theory of Comparative Advantage," in Pierro with Collaboration of M.H. Dobb Sraffa, ed., *Principles of Political Economy and Taxation*, Vol. 1, Cambridge, London: Cambridge University Press, 1817.

Saez, Emmanuel, Benjamin Schoefer, and David Seim, "Payroll Taxes, Firm Behavior, and Rent Sharing: Evidence from a Young Workers' Tax Cut in Sweden," *American Economic Review*, May 2019, *109* (5), 1717–1763.

Shapiro, Carl, "Investment, Moral Hazard, and Occupational Licensing," *The Review of Economic Studies*, 1986, *53* (5), 843–862. Publisher: [Oxford University Press, Review of Economic Studies, Ltd.].

Silver, David, "Haste or Waste? Peer Pressure and Productivity in the Emergency Department," *The Review of Economic Studies*, May 2021, *88* (3), 1385–1417.

Smith, Adam, *An Inquiry Into the Nature and Causes of the Wealth of Nations* 1776.

Socialstyrelsen, "Bedömning av tillgång och efterfrågan på legitimerad personal i hälso- och sjukvård samt tandvård," February 2022.

Sveriges Kommuner och Regioner, "Ekonomi- och verksamhetsstatistik inom hälso- och sjukvården," 2024. Accessed on October 28, 2024.

Traczynski, Jeffrey and Victoria Udalova, "Nurse practitioner independence, health care utilization, and health outcomes," *Journal of Health Economics*, March 2018, *58*, 90–109.

Weidmann, Ben, Joseph Vecci, Farah Said, David J. Deming, and Sonia R. Bhalotra, "How Do You Find a Good Manager?," July 2024.

World Health Organization, "Global expenditure on health: Public spending on the rise?," December 2021.

Appendix

Does Division of Labor Increase Productivity? Evidence from Primary Care

A Additional sample descriptives

Table A1. Restrictions on the main analysis sample

	Doctor	**Nurse**	**Total**
All observed consultations			1,814,706
+ Restrict to unscheduled online consultations	80.5%	19.5%	1,229,383
+ Restrict to analysis time window	78.8%	21.2%	1,118,738
+ Exclude patients registered with the provider	79.4%	20.6%	1,047,474
+ Exclude infants	78.7%	21.3%	988,042
+ Exclude follow-ups within two hours	78.8%	21.2%	963,734
+ Exclude consultations with missing characteristics	78.8%	21.2%	959,244
+ Exclude rare symptom categories	79.0%	21.0%	953,359
+ Restrict to marginal symptom categories	77.2%	22.8%	490,507
+ Exclude singleton observations in date-shift cells	77.2%	22.8%	490,505

Note: This table presents the number of observations after applying our sample restrictions. The columns show the share of consultations with a doctor or a nurse, as well as the total sample size. Each row introduces an additional sample restriction: We restrict the sample to unscheduled online consultations with doctors or nurses from 1 April 2019 to 24 December 2020, during which nurse consultations were active and follow-up consultations within 7 days can be observed. We exclude patients registered with the provider as well as infant patients (age below one year), as different internal protocols apply to them. Consultations are excluded if they involve the same patient within two hours, or if data is missing for baseline characteristics (login time, patient age, gender, region), or migration status. Symptom categories with fewer than 1,000 observations are excluded, and we focus on marginal categories handled by both doctors and nurses, where nurses manage between 5% and 95% of cases. From the remaining sample, we drop login-time cells with singleton observations in our baseline specification. After imposing all restrictions, we are left with 490,505 observations of initial consultations, which we refer to as the main analysis sample.

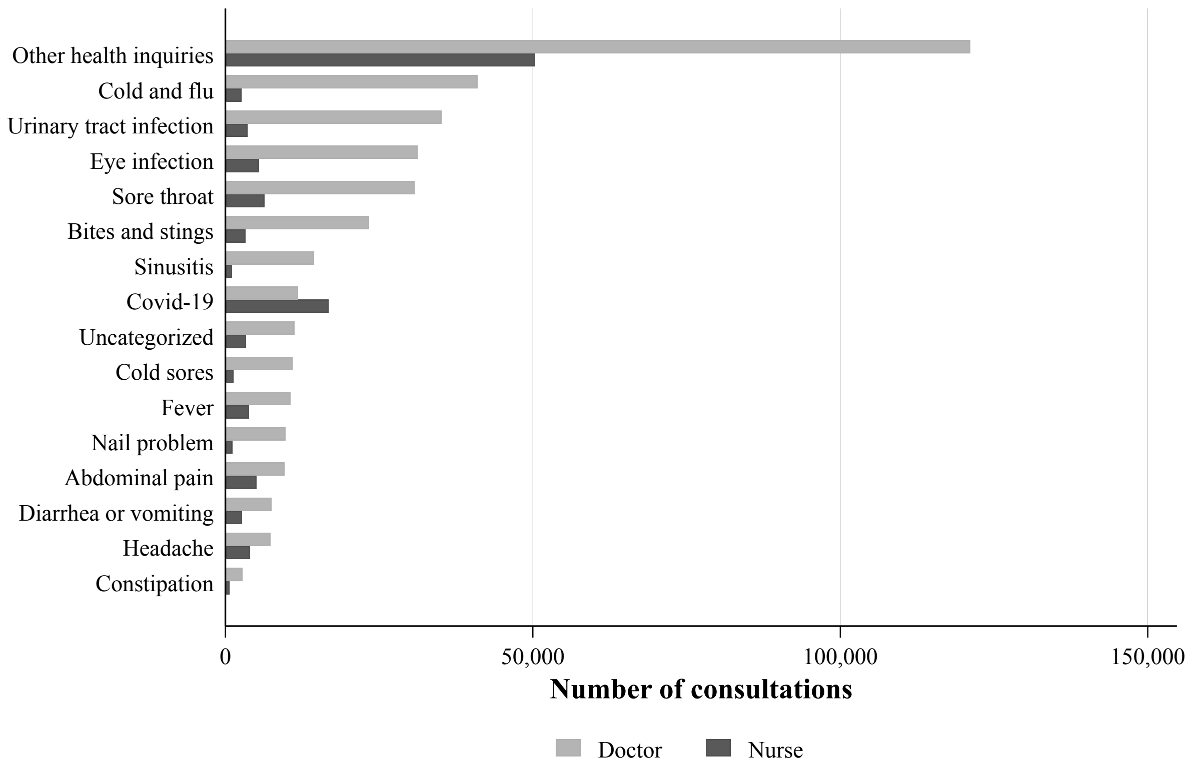


Figure A6. Symptom categories in the analysis sample

Notes: This figure shows the number of initial consultations across symptom categories as reported by patients in our analysis sample, separately for consultations assigned to a doctor or a nurse.

Table A2. Ten most common ICD-10 codes in the analysis sample and in Scania

Analysis sample			
ICD-10 code	Description	Count	Percentage
J06	Acute upper respiratory infections of multiple and unspecified sites	51,358	10.5
N30	Cystitis (bladder inflammation)	25,898	5.3
J03	Acute tonsillitis (upper throat inflammation)	22,187	4.5
T14	Injury of unspecified body region	21,326	4.4
Z71	Other counseling/medical advice, not elsewhere classified	20,815	4.3
H10	Conjunctivitis (eye membrane inflammation)	15,949	3.3
R52	Pain, unspecified	13,227	2.7
N39	Other disorders of urinary system	12,805	2.6
J01	Acute sinusitis	11,610	2.4
Z53	Other specific treatment, not carried out	11,269	2.3
Consultations in Scania			
ICD-10 code	Description	Count	Percentage
F41	Other anxiety disorders	28,728	4.3
J06	Acute upper respiratory infections of multiple and unspecified sites	27,246	4.0
Z00	General medical examination	26,313	3.9
R52	Pain, unspecified	20,703	3.1
R10	Abdominal and pelvic pain	17,824	2.6
F43	Reaction to severe stress, and adjustment disorders	15,516	2.3
R05	Cough	14,971	2.2
H66	Suppurative and unspecified otitis media	13,406	2.0
Z71	Other counseling/medical advice, not elsewhere classified	13,165	2.0
N30	Cystitis (bladder inflammation)	13,006	1.9

Note: This table reports the ten most frequent ICD-10 codes in our main sample (upper panel) and among in-person primary care consultations with nurses or doctors in Scania from 2019-2020 (lower panel). All ICD-10 codes were truncated to be on the three-character level.

B Tests on the instrument validity

Table A3. First stage

	Initial assignment to nurse					
	(1)	(2)	(3)	(4)	(5)	(6)
Nurse staff share, past 60 min.	1.35	1.08	1.05	1.05	1.05	1.05
	(0.0055)	(0.0099)	(0.0096)	(0.0096)	(0.0096)	(0.0096)
Login time		✓	✓	✓	✓	✓
Symptom categories			✓	✓	✓	✓
Demographics				✓	✓	✓
Health risk					✓	✓
Socio-economic variables						✓
Observations	490,505	490,505	490,505	490,505	490,505	490,505
R-squared	0.12	0.13	0.20	0.21	0.21	0.21
K-P F-statistic	61,527	11,841	11,967	11,961	11,963	11,964

Notes: This table reports the first stage from regressing the treatment (*Initial assignment to nurse*) on the congestion instrument, the share of nurses among active clinicians in the past 60 minutes (*Nurse staff share, past 60 min.*). Login time indicates login date-by-4-hours fixed effects. Symptom categories refer to indicators for the main symptom category reported by the patient when requesting a consultation. Demographics include indicators for patient gender, age categories, and aggregated regions. Health risk includes indicators for any prior hospitalization, ED visit, urgent care center visit, and specialist visit in the 3 years prior to the consultation, but excluding the 30 days immediately before, as well as an indicator for any comorbidity. Socio-economic variables include indicators for above-median income, benefit receipt, employment type, education level, civil status, and migrant background. Robust standard errors are in parentheses.

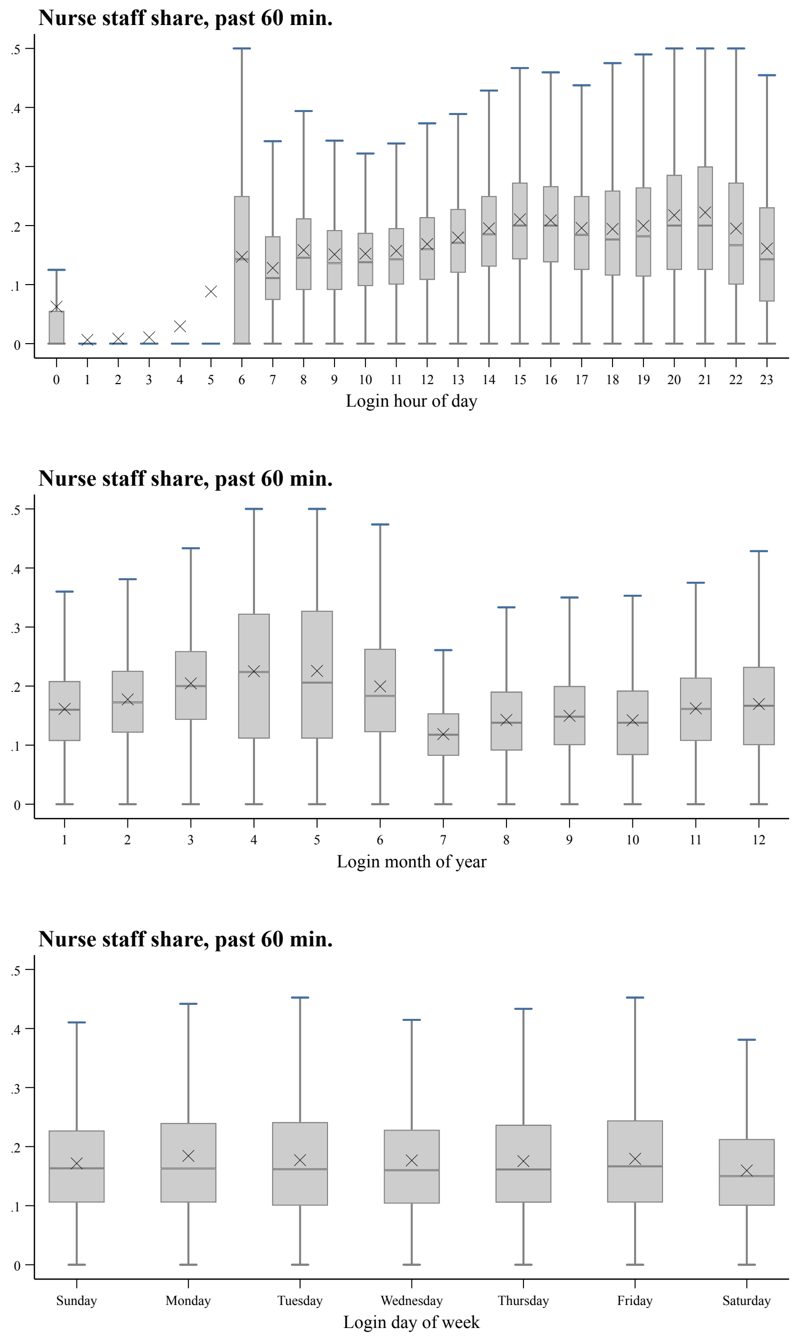


Figure A7. Distribution of the congestion instrument across different time intervals

Notes: This figure presents the distribution of the congestion instrument, the share of nurses among active clinicians in the past 60 minutes, across three dimensions: hour of the day (top panel), month of the year (middle panel), and day of the week (bottom panel). The box plots represent the interquartile range for each time period. The median is shown as a horizontal line inside the box, and the mean is indicated by a cross. The whiskers extend to the upper and lower adjacent values.

Table A4. Conditional balance of the congestion instrument

	Nurse staff share, past 60 min.			
	(1)	(2)	(3)	(4)
Login time	✓	✓	✓	✓
Symptom categories		✓	✓	✓
Demographics			✓	✓
Health risk				✓
Socio-economic variables				
Observations	490,505	490,505	490,505	490,505
R-squared	0.7	0.7	0.7	0.7
Joint F-statistic	4.13	1.77	1.09	1.01
Joint F-test p-val.	0.00	0.00	0.35	0.44
Joint F-test RI p-val.	0.00	0.00	0.32	0.45

Notes: This table reports balance tests of the congestion instrument, the share of nurses among active clinicians in the past 60 minutes (*Nurse staff share, past 60 min.*), conditional on varying sets of case characteristics. Control characteristics that the congestion instrument is conditioned on are marked by checkmarks. Missing checkmarks indicate case characteristics that the congestion instrument is balanced against. Login time indicates date-by-4 hours fixed effects. Symptom categories refer to indicators for the main symptom category reported by the patient when requesting a consultation. Demographics include indicators for patient gender, age categories, and aggregated regions. Additional case characteristics include patient health risk and socio-economic characteristics. Health risk includes indicators for any prior hospitalization, ED visit, urgent care center visit, and specialist visit in the 3 years prior to the consultation, but excluding the 30 days immediately before, as well as an indicator for any comorbidity. Socio-economic variables include indicators for above-median income, benefit receipt, employment type, education level, civil status, and migrant background. The test statistics are based on robust standard errors. Based on Kerwin et al. (2024), randomization inference p-values from 500 repetitions are provided alongside conventional ones.

Table A5. Monotonicity subgroup test

	Prior hospitalization		Prior ED visit		Prior specialist visit	
	Any	None	Any	None	Any	None
Nurse staff share, past 60 min.	1.04***	1.05***	1.07***	1.04***	1.03***	1.08***
	(0.022)	(0.011)	(0.016)	(0.012)	(0.012)	(0.016)
Baseline characteristics	✓	✓	✓	✓	✓	✓
Observations	92,562	397,778	164,242	326,205	312,315	178,108
K-P F-statistic	2,258	9,682	4,211	7,723	7,482	4,452
	Prior urgent care		**Comorbidity**		**Gender**	
	Any	None	Any	None	Female	Male
Nurse staff share, past 60 min.	1.04***	1.06***	1.03***	1.05***	0.99***	1.15***
	(0.020)	(0.011)	(0.021)	(0.011)	(0.012)	(0.016)
Baseline characteristics	✓	✓	✓	✓	✓	✓
Observations	115,792	374,587	102,329	388,024	307,543	182,881
K-P F-statistic	2,729	9,193	2,389	9,481	6,902	5,078
	Income		**Migration background**		**Further education**	
	$>$ median	$\leq$ median	Any	None	Any	None
Nurse staff share, past 60 min.	1.03***	1.06***	1.09***	1.03***	1.05***	1.04***
	(0.017)	(0.012)	(0.019)	(0.011)	(0.011)	(0.018)
Baseline characteristics	✓	✓	✓	✓	✓	✓
Observations	159,220	331,156	123,212	367,218	351,374	139,023
K-P F-statistic	3,745	8,207	3,345	8,539	8,485	3,388

Note: This table reports the results of Equation 2, which represents the first stage of the IV regression, across various subsamples, including the baseline case characteristics as control variables. The instrument is the share of nurses among active clinicians in the past 60 minutes (*Nurse staff share, past 60 min.*). The baseline case characteristics include the symptom categories, demographics, region, along with login date-by-4 hour fixed effects. *Further educ.* refers to any post-secondary education reported for patients aged 25 or older, while income is reported for patients over the age of 20. The subsamples split by further education and median income exclude patients with missing or undefined information. Descriptions of all variables are provided in Appendix Table A17. The baseline mean refers to the mean of the variable *Initial assignment to nurse*. The First-stage K-P F-statistic refers to the Kleibergen-Paap F-statistic. Robust standard errors are shown in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

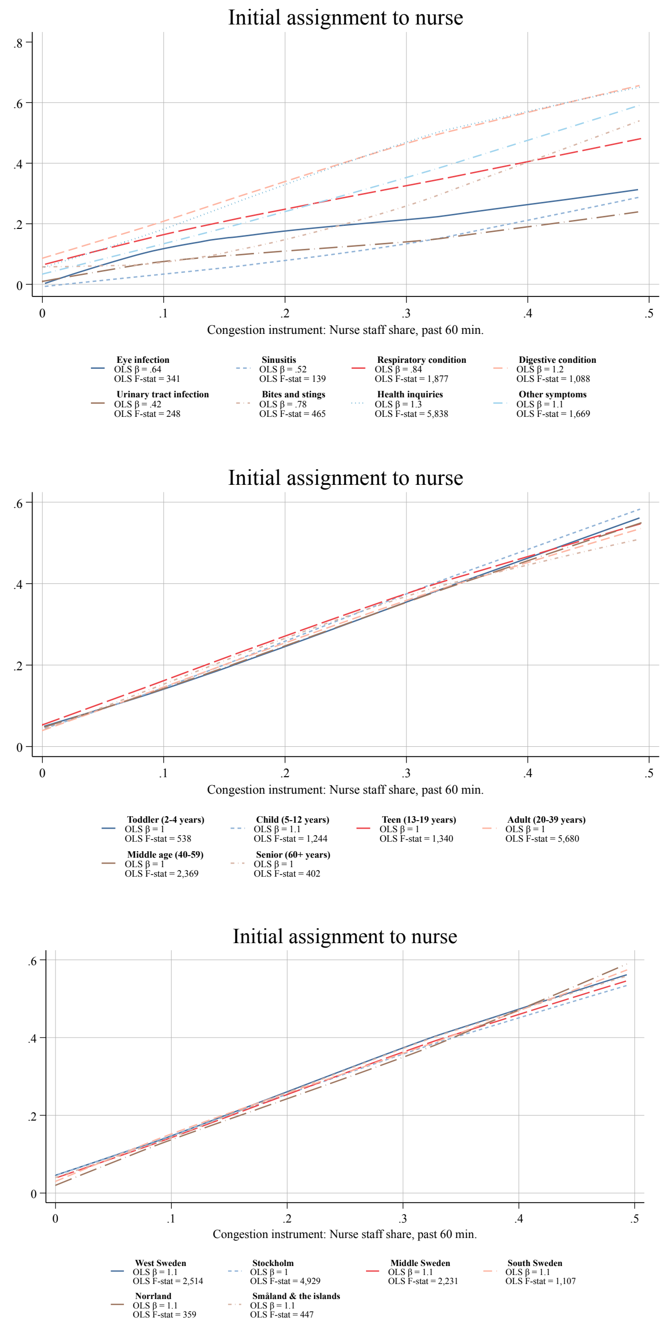


Figure A8. Instrument monotonicity test

Notes: We show non-parametric estimates within subsamples of the conditional expectation function (CEF) for the treatment, initial nurse assignment, given our congestion instrument, *Nurse staff share, past 60 min.*. The CEF is estimated within each subsample as a smoothed line over a binscatter. The estimates are conditional on our baseline set of controls (login time, symptom, and demographic controls), and bins are selected to minimize the integrated mean square error (Cattaneo et al. 2024). The OLS β and F-statistics are estimated with linear regressions in each subsample conditional on our baseline set of controls. The following symptom categories are aggregated: Respiratory Condition includes Cold and Flu, Sore Throat, and Covid-19; Digestive Condition includes Abdominal Pain, Constipation, and Diarrhea or Vomiting; Other Symptoms include Nail Problem, Headache, Fever, Cold Sores, and Uncategorized Symptoms.

Table A6. Differences in consultation characteristics under low and high congestion

	High congestion		Low congestion		T-test	
	Mean	SD	Mean	SD	Diff.	p-val.
Patient waiting time in min.	29.4	44.3	14.4	22.7	14.9	0.00
Consultation time below one minute	0.045	0.21	0.043	0.20	0.0030	0.05
Drop out	0.017	0.13	0.038	0.19	-0.021	0.00
Clinician total time in min.	11.4	6.84	11.6	6.71	-0.20	0.00
Observations	49346		49326			

Notes: This table presents summary statistics for various consultation characteristics during low and high congestion. For low (high) congestion, we consider the lower (upper) decile of the congestion instrument, *Nurse share, past 60 min.*. The last two columns perform t-tests on the difference in means between consultation characteristics at low and high congestion.

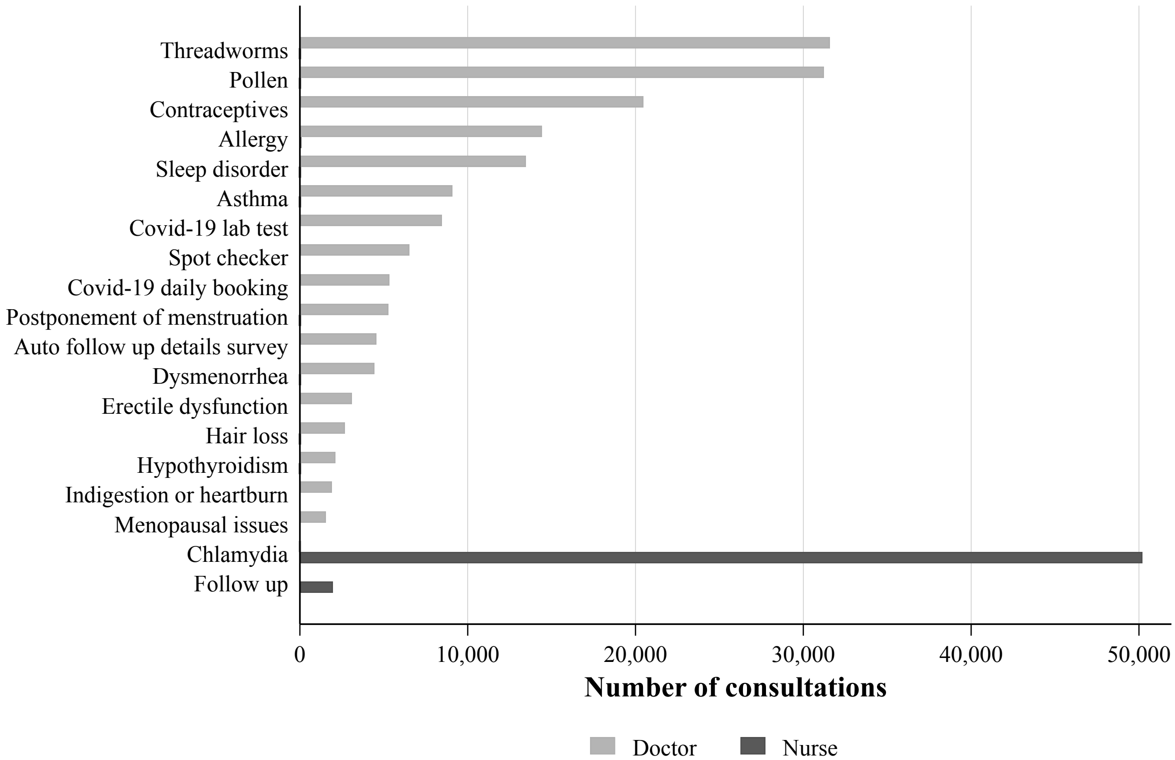


Figure A9. Strictly deterministic symptom categories in the sample with no compliers

Notes: This figure shows the number of initial consultations across strictly deterministically assigned symptom categories as reported by patients, separately for consultations exclusively assigned to a doctor or a nurse. In these symptom categories, fewer than 1% of cases are assigned to either clinician type and congestion should not influence the case assignment.

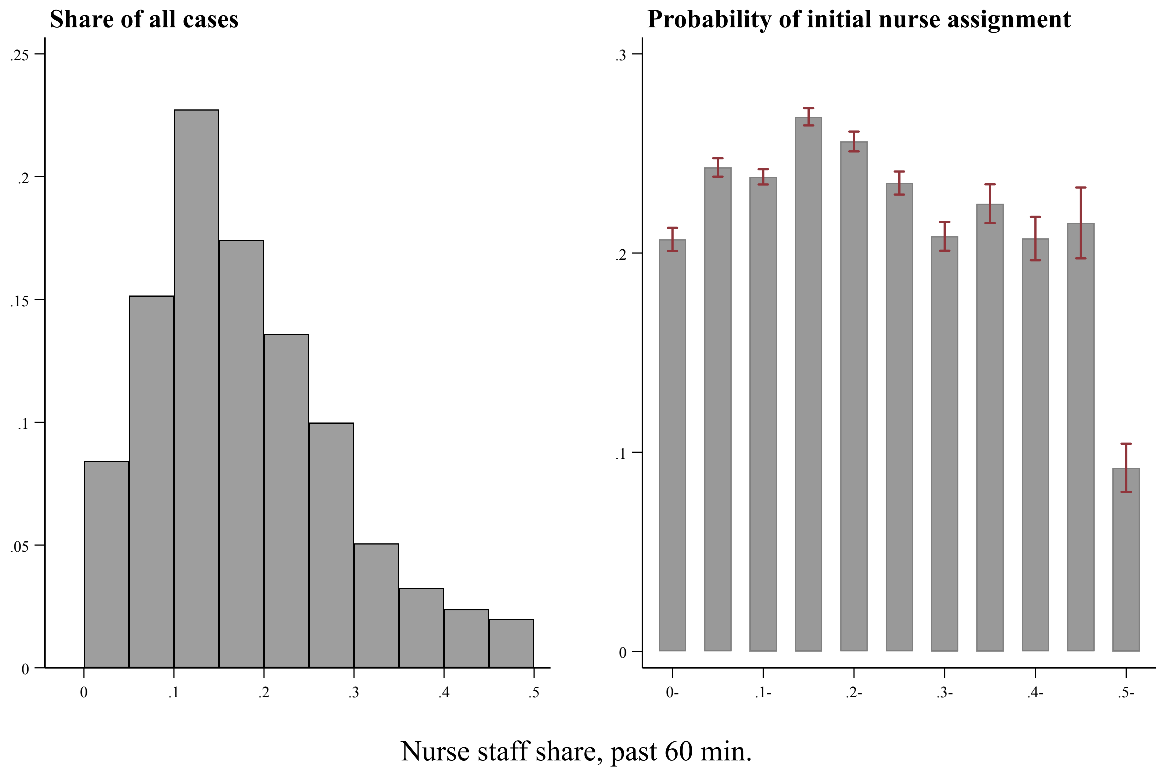


Figure A10. Variation in the congestion instrument in a sample with strictly deterministic clinician assignment

Notes: This figure shows descriptive figures for the congestion instrument, *Nurse staff share, past 60 min.*, as constructed in our analysis sample based on Equation 2. However, the descriptive figures are shown for a sample of cases where symptoms are strictly deterministically assigned to either a nurse or a doctor, with fewer than 1% of cases assigned to either clinician type. In these cases, congestion should not influence case assignment. The left subfigure shows the distribution of the instrument. The right subfigure shows the average probability of an initial assignment to a nurse, our treatment, plotted against categorized values of the instrument. Lines indicate 95% confidence intervals of the mean.

Table A7. Instrumental variable regressions in a sample with strictly deterministic clinician assignment

	Any referral to doctor (7d)			Informative diagnosis (7d)			Any new prescription (7d)		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initial assignment to nurse	0.47***		-0.52	-0.055		-6.65	-0.15***		1.55
	(0.048)		(1.98)	(0.037)		(6.95)	(0.046)		(8.04)
Nurse staff share, past 60 min.		-0.0012			-0.015			0.0035	
		(0.0044)			(0.015)			(0.018)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	218,071	218,071	218,071	218,071	218,071	218,071	218,071	218,071	218,071
First-stage K-P F-statistic			6.1			6.1			6.1
Baseline mean	0.012		0.012	0.81		0.81	0.54		0.54
	Rating: top score (7d)			**Any external PCP consultation (7d)**			**Any ED (7d)**		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initial assignment to nurse	-0.024		0.24	-0.038*		-8.34	-0.0074		-2.10
	(0.047)		(8.70)	(0.022)		(10.2)	(0.0097)		(2.12)
Nurse staff share, past 60 min.		0.00055			-0.013			-0.0047	
		(0.020)			(0.011)			(0.0044)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	218,071	218,071	218,071	113,930	113,930	113,930	218,071	218,071	218,071
First-stage K-P F-statistic			6.1			1.3			6.1
Baseline mean	0.48		0.48	0.042		0.042	0.011		0.011
	Any hospitalization (7d)			**Income drop >20% (cal. mo. after)**			**Log cost, incl. ext. PCP**		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initial assignment to nurse	-0.0033***		0.51	0.068		-9.40	-0.14		-3.10
	(0.00093)		(1.05)	(0.18)		(16.9)	(0.087)		(21.0)
Nurse staff share, past 60 min.		0.0011			-0.015			-0.0047	
		(0.0023)			(0.026)			(0.031)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	218,071	218,071	218,071	92,190	92,190	92,190	113,930	113,930	113,930
First-stage K-P F-statistic			6.1			2.3			1.3
Baseline mean	0.0031		0.0031	0.19		0.19	6.67		6.67

Note: This table presents regression results for a sample of cases where symptoms are strictly deterministically assigned to either a nurse or a doctor, with fewer than 1% of cases assigned to either clinician type. In these cases, congestion should not influence case assignment. Each subpanel corresponds to a different outcome as the dependent variable. The left column in each subpanel reports the Ordinary Least Squares (OLS) estimate for the treatment variable (*Nurse*) based on Equation 1. The middle column presents the reduced-form regression estimates for the congestion instrument, measured as the share of nurses among active clinicians in the past 60 minutes (*Nurse staff share, past 60 min.*). The right column provides the Instrumental Variables (IV) specification, using *Nurse staff share, past 60 min.* as an instrument for *Nurse*, estimated via Two-Stage Least Squares (2SLS), following Equations 1 and 2, but applied the sample of strictly deterministically assigned cases. All regressions control for symptom categories, demographics, health risk factors, and socioeconomic variables, along with along with login date-by-4-hour shifts fixed effects. The baseline mean represents the average of the dependent variable for cases directly assigned to a doctor. The first-stage K-P F-statistic refers to the Kleibergen-Paap F-statistic. Robust standard errors are shown in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

C Robustness of the main results

Table A8. Effect of an initial nurse assignment on the organization of tasks at the provider (within +1 day)

	Any referral to doctor (1d)			Any subseq. consultation (1d)			Any return drop-in (1d)		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	0.32***		0.31***	0.32***		0.30***	0.010***		0.0031
	(0.0015)		(0.0057)	(0.0016)		(0.0075)	(0.00069)		(0.0044)
Nurse staff share, past 60 min.		0.32***			0.31***			0.0032	
		(0.0067)			(0.0084)			(0.0046)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505
First-stage K-P F-statistic			11,964			11,964			11,964
Baseline mean	0.0059	0.0059	0.0059	0.053	0.053	0.053	0.028	0.028	0.028
	Clinician total time in min. (1d)			Doctor total time in min. (1d)			Nurse total time in min. (1d)		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	3.28***		3.25***	-7.83***		-8.24***	11.1***		11.5***
	(0.039)		(0.22)	(0.031)		(0.19)	(0.024)		(0.10)
Nurse staff share, past 60 min.		3.40***			-8.63***			12.0***	
		(0.23)			(0.21)			(0.15)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505
First-stage K-P F-statistic			11,964			11,964			11,964
Baseline mean	12.5	12.5	12.5	12.4	12.4	12.4	0.065	0.065	0.065

Note: This table presents regression results for our main specifications on the effects of initial case assignment to a nurse (the *knowledge hierarchy*) rather than directly to a doctor. Each subpanel corresponds to a different outcome as the dependent variable. Patient-provider interactions after the initial consultation are defined as follows: "Any referral to doctor" refers to any internal referrals or revisits with a doctor, "Any subsequent consultations" covers any interaction with the provider after an initial visit, and "Any return drop-in" indicates any unscheduled revisit for the same symptom. "Total time" defines the time that clinicians spend on a case, including consultation time spent with the patient and time on administrative work during the initial and all subsequent consultations. The left column in each subpanel reports the Ordinary Least Squares (OLS) estimate for the treatment variable (*Nurse*) based on Equation 1. The middle column presents the reduced-form regression estimates for the congestion instrument, measured as the share of nurses among active clinicians in the past 60 minutes (*Nurse staff share, past 60 min.*). The right column provides the Instrumental Variables specification, using *Nurse staff share, past 60 min.* as an instrument for *Nurse*, estimated via Two-Stage Least Squares (2SLS), following Equations 1 and 2. All regressions control for symptom categories, demographics, region, health risk factors, and socioeconomic variables, along with login date-by-4-hour fixed effects. The baseline mean represents the average of the dependent variable for cases directly assigned to a doctor. The first-stage K-P F-statistic refers to the Kleibergen-Paap F-statistic. Robust standard errors are shown in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A9. Effect of an initial nurse assignment on additional outcomes

	Doctor consultation time in min. (7d)			Nurse consultation time in min. (7d)			Informative diagnosis (initial)		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	-3.13***		-3.23***	4.51***		4.56***	-0.14***		-0.13***
	(0.018)		(0.11)	(0.011)		(0.048)	(0.0017)		(0.0098)
Nurse staff share, past 60 min.		-3.38***			4.77***			-0.14***	
		(0.12)			(0.066)			(0.010)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505
First-stage K-P F-statistic			11,964			11,964			11,964
Baseline mean	5.12	5.12	5.12	0.078	0.078	0.078	0.79	0.79	0.79

	Any prescription (7d)			Rating: below top score (7d)			Rating: no physical replacement (7d)		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	-0.13***		-0.091***	0.00098		0.0040	0.0080***		-0.00037
	(0.0017)		(0.011)	(0.0011)		(0.0070)	(0.00081)		(0.0049)
Nurse staff share, past 60 min.		-0.095***			0.0042			-0.00039	
		(0.012)			(0.0073)			(0.0051)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505
First-stage K-P F-statistic			11,964			11,964			11,964
Baseline mean	0.52	0.52	0.52	0.096	0.096	0.096	0.039	0.039	0.039

	Rating: top score (initial)			Rating: physical replacement (initial)			Any avoid. hospitalization (7d)		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	0.010***		0.017	-0.025***		-0.0040	0.00026*		-0.00053
	(0.0018)		(0.012)	(0.0018)		(0.012)	(0.00015)		(0.00088)
Nurse staff share, past 60 min.		0.018			-0.0041			-0.00056	
		(0.012)			(0.012)			(0.00093)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505
First-stage K-P F-statistic			11,964			11,964			11,964
Baseline mean	0.40	0.40	0.40	0.43	0.43	0.43	0.0013	0.0013	0.0013

	Sick pay 1m (cal. month after)			Log cost, excl. ext. PCP			Log cost of provider visits		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	-0.0043***		0.0013	-0.12***		-0.13***	-0.11***		-0.11***
	(0.00066)		(0.0042)	(0.0028)		(0.016)	(0.0016)		(0.0073)
Nurse staff share, past 60 min.		0.0013			-0.14***			-0.11***	
		(0.0042)			(0.017)			(0.0077)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	213,425	213,425	213,425	490,505	490,505	490,505	490,505	490,505	490,505
First-stage K-P F-statistic			5,024			11,964			11,964
Baseline mean	0.011	0.011	0.011	6.68	6.68	6.68	6.37	6.37	6.37

Note: This table presents regression results for our main specifications on the effects of initial case assignment to a nurse (the *knowledge hierarchy*) rather than directly to a doctor. Each subpanel corresponds to a different outcome as the dependent variable. Outcomes are defined as follows: "Consultation time" for the time clinicians spend directly with the patient during the initial and all subsequent consultations; "Admin time" for time spend on administrative work without patient contact; "Rating: below top score" and "Rating: no physical replacement" imply that any negative survey response was given to any consultation at the provider; "Rating: top score" and "Rating: physical replacement" imply that any top rating was given for the initial consultation at the provider; "Informative diagnosis" excludes symptomatic/health status diagnoses (ICD R00-R99/Z00-Z99) and implies that any informative diagnosis was provided at the initial consultation; "Sick pay 1m (post-1m)" refers to the receipt of sickness benefits in a given calendar year, as well as a drop in income in the calendar month following the consultation such that a full month of sick pay is captured: Income drops to below 37% of the average income in the three months prior, which corresponds to an employer-paid replacement rate of 80% for 14/30 days and no income from employment for 16/30 days, followed by a rebound to more than 37% of the pre-consultation income. "Log costs, incl. ext. PCP" represents the natural logarithm of the estimated care costs, including costs at the provider and the downstream expenses of any prescriptions, specialist visits, emergency department visits, urgent care visits, and hospitalizations within seven days of the initial consultation, but exclude the costs of external primary care consultations. "Log costs of provider visits" includes only the costs associated with online consultations at the provider. The left column in each subpanel reports the Ordinary Least Squares (OLS) estimate for the treatment variable (*Nurse*) based on Equation 1. The middle column presents the reduced-form regression estimates for the congestion instrument, measured as the share of nurses among active clinicians in the past 60 minutes (*Nurse staff share, past 60 min.*). The right column provides the Instrumental Variables specification, using *Nurse staff share, past 60 min.* as an instrument for *Nurse*, estimated via Two-Stage Least Squares (2SLS), following Equations 1 and 2. All regressions control for symptom categories, demographics, region, health risk factors, and socioeconomic variables, along with login date-by-4-hour fixed effects. The baseline mean represents the average of the dependent variable for cases directly assigned to a doctor. The first-stage K-P F-statistic refers to the Kleibergen-Paap F-statistic. Robust standard errors are shown in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

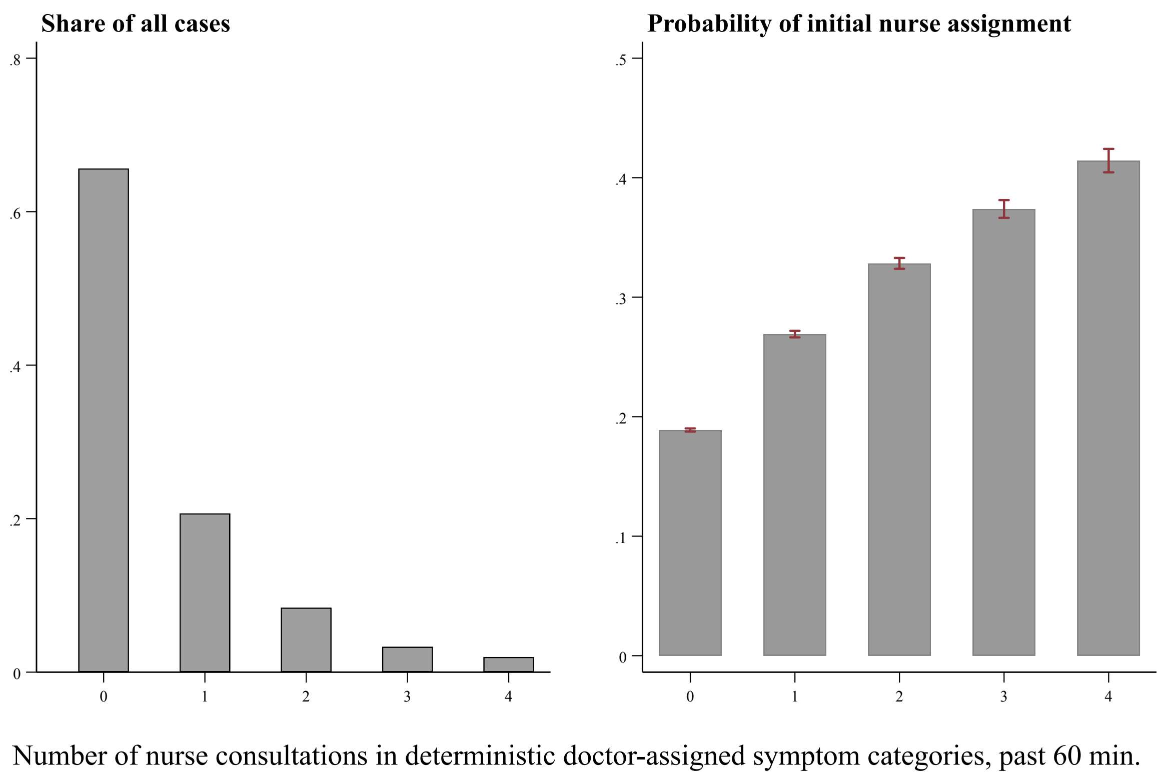


Figure A11. Variation in the doctor shortage instrument

Notes: This figure shows descriptive figures for an alternative instrument based on doctor shortages relative to patient demand. This doctor shortage instrument is constructed in two steps. First, it considers the number of cases in the past 60 minutes of a given consultation assigned to nurses, even though their symptoms are otherwise deterministically assigned to doctors (share of direct-to-doctor assignment above 95%). Then, this number is winsorized at the 99th percentile. The left subfigure shows the distribution of the instrument. The right subfigure shows the average probability of an initial assignment to a nurse, our treatment, in categorized values of the instrument. Lines indicate 95% confidence intervals of the mean.

Table A10. Main outcomes: Alternative doctor shortage instrument

	Any referral to doctor (7d)			Informative diagnosis (7d)			Any new prescription (7d)		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	0.31***		0.26***	-0.064***		-0.053	-0.10***		-0.18***
	(0.0015)		(0.041)	(0.0016)		(0.058)	(0.0017)		(0.066)
Doctor shortage, past 60 min.		0.0039***			-0.00080			-0.0027***	
		(0.00069)			(0.00088)			(0.0010)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505
First-stage K-P F-statistic			280			280			280
Baseline mean	0.014	0.014	0.014	0.81	0.81	0.81	0.44	0.44	0.44
	Any specialist (7d)			**Rating: top score (7d)**			**Rating: physical replacement (7d)**		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	0.0030***		-0.016	0.045***		0.083	0.015***		0.053
	(0.00073)		(0.027)	(0.0018)		(0.071)	(0.0018)		(0.071)
Doctor shortage, past 60 min.		-0.00024			0.0013			0.00081	
		(0.00041)			(0.0011)			(0.0011)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505
First-stage K-P F-statistic			280			280			280
Baseline mean	0.036	0.036	0.036	0.42	0.42	0.42	0.45	0.45	0.45
	Any external PCP consultation (7d)			**Any ED (7d)**			**Any hospitalization (7d)**		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	0.029***		-0.068	0.0061***		-0.018	-0.00023		0.0068
	(0.0018)		(0.064)	(0.00074)		(0.026)	(0.00039)		(0.015)
Doctor shortage, past 60 min.		-0.0010			-0.00027			0.00010	
		(0.00095)			(0.00040)			(0.00022)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	261,034	261,034	261,034	490,505	490,505	490,505	490,505	490,505	490,505
First-stage K-P F-statistic			149			280			280
Baseline mean	0.12	0.12	0.12	0.030	0.030	0.030	0.0087	0.0087	0.0087
	Any urgent care (7d)			**Income drop >20% (cal. mo. after)**			**Zero income (cal. month after)**		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	0.0019***		-0.013	-0.021***		-0.12	-0.0034***		-0.0018
	(0.00051)		(0.018)	(0.0024)		(0.100)	(0.0013)		(0.049)
Doctor shortage, past 60 min.		-0.00020			-0.0016			-0.000026	
		(0.00027)			(0.0014)			(0.00068)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	490,505	490,505	490,505	213,425	213,425	213,425	213,425	213,425	213,425
First-stage K-P F-statistic			280			107			107
Baseline mean	0.017	0.017	0.017	0.22	0.22	0.22	0.050	0.050	0.050
	Death excl. external causes (3y)			**Log cost, excl. ext. PCP**			**Log cost, incl. ext. PCP**		
	OLS	Red.	IV	OLS	Red.	IV	OLS	Red.	IV
Initially to nurse	-0.000096		-0.0020	-0.12***		-0.15	-0.074***		-0.26*
	(0.00016)		(0.0064)	(0.0028)		(0.095)	(0.0043)		(0.15)
Doctor shortage, past 60 min.		-0.000031			-0.0023			-0.0039*	
		(0.000097)			(0.0015)			(0.0022)	
Baseline characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	490,505	490,505	490,505	490,505	490,505	490,505	261,034	261,034	261,034
First-stage K-P F-statistic			280			280			149
Baseline mean	0.0016	0.0016	0.0016	6.68	6.68	6.68	6.80	6.80	6.80

Note: This table presents results for the effects of initial case assignment to a nurse (the *knowledge hierarchy*) rather than directly to a doctor with an alternative instrumental variable: The number of nurse consultations in almost deterministically doctor-assigned symptom categories in the past 60 minutes, winsorized at the 99 percentile (*Doctor shortage, past 60 min.*). Each subpanel corresponds to a different outcome as the dependent variable. Outcomes are defined as in Tables 2, 3, and 5. The left column in each subpanel reports the Ordinary Least Squares (OLS) estimate for the treatment variable (*Nurse*) based on Equation 1. The middle column presents the reduced-form regression estimates for the alternative instrument. The right column provides the Instrumental Variables specification, using *Doctor shortage, past 60 min.* as an instrument for *Nurse*, estimated via Two-Stage Least Squares (2SLS), following Equations 1 and 2, replaced by the alternative instrument. All regressions control for symptom categories, demographics, region, health risk factors, and socioeconomic variables, along with login date-by-4-hour fixed effects. The baseline mean represents the average of the dependent variable for cases directly assigned to a doctor. The first-stage K-P F-statistic refers to the Kleibergen-Paap F-statistic. Robust standard errors are shown in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A11. Alternative set of controls

	Any referral to doctor (7d)		Informative diagnosis (7d)		Any new prescription (7d)		Rating: top score (7d)		Any ext. PCP cons. (7d)		Any ED (7d)		Any hospitalization (7d)		Income drop >20% (cal. mo. after)		Log cost, incl. ext. PCP	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Panel A: Login time																		
Initially to nurse	0.30***	0.30***	-0.14***	-0.088***	-0.18***	-0.11***	0.028***	0.041***	0.032***	0.032***	0.016***	-0.0028	0.0025***	-0.00044	0.0067***	0.0020	-0.079***	-0.087***
	(0.0014)	(0.0059)	(0.0016)	(0.0099)	(0.0016)	(0.011)	(0.0018)	(0.011)	(0.0017)	(0.011)	(0.00071)	(0.0045)	(0.00037)	(0.0024)	(0.0023)	(0.015)	(0.0041)	(0.025)
Panel B: Login time + Symptoms																		
Initially to nurse	0.31***	0.30***	-0.065***	-0.067***	-0.11***	-0.086***	0.044***	0.047***	0.028***	0.030***	0.0060***	-0.0062	-0.00032	-0.0013	-0.020***	-0.0058	-0.080***	-0.085***
	(0.0015)	(0.0060)	(0.0016)	(0.0095)	(0.0017)	(0.011)	(0.0019)	(0.012)	(0.0018)	(0.011)	(0.00075)	(0.0046)	(0.00039)	(0.0025)	(0.0024)	(0.016)	(0.0043)	(0.026)
Panel C (Baseline controls): Login time + Symptoms + Demographics																		
Initially to nurse	0.31***	0.30***	-0.065***	-0.067***	-0.10***	-0.083***	0.045***	0.048***	0.028***	0.030***	0.0060***	-0.0052	-0.00029	-0.0012	-0.020***	-0.0057	-0.077***	-0.083***
	(0.0015)	(0.0060)	(0.0016)	(0.0095)	(0.0017)	(0.011)	(0.0019)	(0.012)	(0.0018)	(0.011)	(0.00074)	(0.0046)	(0.00039)	(0.0025)	(0.0024)	(0.016)	(0.0043)	(0.025)
Panel D: Login time + Symptoms + Demographics + Health risk																		
Initially to nurse	0.31***	0.30***	-0.065***	-0.067***	-0.10***	-0.083***	0.045***	0.047***	0.029***	0.030***	0.0061***	-0.0046	-0.00023	-0.0011	-0.020***	-0.0058	-0.075***	-0.080***
	(0.0015)	(0.0060)	(0.0016)	(0.0095)	(0.0017)	(0.011)	(0.0019)	(0.012)	(0.0018)	(0.011)	(0.00074)	(0.0046)	(0.00039)	(0.0025)	(0.0024)	(0.016)	(0.0043)	(0.025)
Panel E (Full controls): Login time + Symptoms + Demographics + Health risk + Socio-economic variables																		
Initially to nurse	0.31***	0.30***	-0.064***	-0.068***	-0.10***	-0.083***	0.045***	0.046***	0.029***	0.030***	0.0061***	-0.0045	-0.00023	-0.0010	-0.021***	-0.0066	-0.074***	-0.080***
	(0.0015)	(0.0060)	(0.0016)	(0.0095)	(0.0017)	(0.011)	(0.0018)	(0.012)	(0.0018)	(0.011)	(0.00074)	(0.0046)	(0.00039)	(0.0025)	(0.0024)	(0.016)	(0.0043)	(0.025)
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	261,034	261,034	490,505	490,505	490,505	490,505	213,425	213,425	261,034	261,034
Baseline mean	0.014	0.014	0.81	0.81	0.44	0.44	0.42	0.42	0.12	0.12	0.030	0.030	0.0087	0.0087	0.22	0.22	6.80	6.80

Notes: This table presents regression results for different sets of control variables. We show the effects of initial case assignment to a nurse rather than directly to a doctor, with each column corresponding to a different outcome as the dependent variable. Each table panel then increases the number of control variables included in the specification, successively including symptom categories, demographics, health risk factors, and socioeconomic variables. All regressions control for login time based on date-by-4 hour fixed effects. Each cell shows estimates for *Nurse* from the Ordinary Least Squares (OLS) or the Instrumental Variables (IV) specification using *Nurse staff share, past 60 min.* as an instrument and estimated via Two-Stage Least Squares (2SLS), following Equations 1 and 2. The baseline mean represents the average of the dependent variable for cases directly assigned to a doctor. The first-stage K-P F-statistic refers to the Kleibergen-Paap F-statistic. Robust standard errors are shown in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A12. Alternative time fixed effects

	Any referral to doctor (7d)		Informative diagnosis (7d)		Any new prescription (7d)		Rating: top score (7d)		Any ext. PCP cons. (7d)		Any ED (7d)		Any hospitalization (7d)		Income drop >20% (cal. mo. after)		Log cost, incl. ext. PCP	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Panel A: Date fixed effects																		
Initially to nurse	0.31***	0.28***	-0.064***	-0.080***	-0.10***	-0.11***	0.045***	0.038***	0.026***	-0.012	0.0057***	0.0037	-0.00026	0.000047	-0.021***	-0.0034	-0.078***	-0.12***
	(0.0015)	(0.0038)	(0.0016)	(0.0061)	(0.0016)	(0.0070)	(0.0018)	(0.0075)	(0.0018)	(0.0072)	(0.00073)	(0.0030)	(0.00038)	(0.0016)	(0.0023)	(0.0100)	(0.0042)	(0.016)
First-stage K-P F-statistic		31,052		31,052		31,052		31,052		15,351		31,052		31,052		12,530		15,351
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	261,034	261,034	490,505	490,505	490,505	490,505	213,496	213,496	261,034	261,034
Panel B: Date and 4h-window fixed effects																		
Initially to nurse	0.31***	0.29***	-0.063***	-0.050***	-0.10***	-0.070***	0.046***	0.057***	0.028***	0.017**	0.0057***	-0.0050	-0.00026	-0.00085	-0.021***	-0.0054	-0.075***	-0.093***
	(0.0015)	(0.0042)	(0.0016)	(0.0066)	(0.0016)	(0.0076)	(0.0018)	(0.0082)	(0.0018)	(0.0078)	(0.00073)	(0.0032)	(0.00038)	(0.0017)	(0.0024)	(0.011)	(0.0042)	(0.018)
First-stage K-P F-statistic		24,730		24,730		24,730		24,730		12,440		24,730		24,730		10,117		12,440
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	261,034	261,034	490,505	490,505	490,505	490,505	213,496	213,496	261,034	261,034
Panel C: Date and hour fixed effects																		
Initially to nurse	0.31***	0.29***	-0.063***	-0.048***	-0.10***	-0.070***	0.046***	0.062***	0.028***	0.018**	0.0060***	-0.0028	-0.00018	0.000085	-0.021***	-0.0030	-0.074***	-0.085***
	(0.0015)	(0.0043)	(0.0016)	(0.0067)	(0.0016)	(0.0078)	(0.0018)	(0.0084)	(0.0018)	(0.0080)	(0.00073)	(0.0032)	(0.00038)	(0.0017)	(0.0024)	(0.011)	(0.0042)	(0.018)
First-stage K-P F-statistic		23,624		23,624		23,624		23,624		11,940		23,624		23,624		9,722		11,940
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	261,034	261,034	490,505	490,505	490,505	490,505	213,496	213,496	261,034	261,034
Panel D (Baseline): Date-by-4 hour fixed effects																		
Initially to nurse	0.31***	0.30***	-0.064***	-0.068***	-0.10***	-0.083***	0.045***	0.046***	0.029***	0.030***	0.0061***	-0.0045	-0.00023	-0.0010	-0.021***	-0.0066	-0.074***	-0.080***
	(0.0015)	(0.0060)	(0.0016)	(0.0095)	(0.0017)	(0.011)	(0.0018)	(0.012)	(0.0018)	(0.011)	(0.00074)	(0.0046)	(0.00039)	(0.0025)	(0.0024)	(0.016)	(0.0043)	(0.025)
First-stage K-P F-statistic		11,964		11,964		11,964		11,964		6,028		11,964		11,964		5,024		6,028
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	261,034	261,034	490,505	490,505	490,505	490,505	213,425	213,425	261,034	261,034
Panel E: Date-by-hour fixed effects																		
Initially to nurse	0.31***	0.33***	-0.065***	-0.074***	-0.11***	-0.14***	0.046***	0.091***	0.029***	0.042**	0.0066***	-0.00081	-0.00017	-0.0066	-0.022***	-0.0086	-0.074***	-0.12**
	(0.0015)	(0.011)	(0.0016)	(0.017)	(0.0017)	(0.020)	(0.0019)	(0.022)	(0.0018)	(0.021)	(0.00075)	(0.0084)	(0.00039)	(0.0045)	(0.0025)	(0.028)	(0.0044)	(0.047)
First-stage K-P F-statistic		3,441		3,441		3,441		3,441		1,697		3,441		3,441		1,502		1,697
Observations	489,742	489,742	489,742	489,742	489,742	489,742	489,742	489,742	259,792	259,792	489,742	489,742	489,742	489,742	212,104	212,104	259,792	259,792
Baseline mean	0.014	0.014	0.81	0.81	0.44	0.44	0.42	0.42	0.12	0.12	0.030	0.030	0.0087	0.0087	0.22	0.22	6.80	6.80

Notes: This table presents regression results for different fixed effects to control for login time. We show the effects of initial case assignment to a nurse rather than directly to a doctor, with each column corresponding to a different outcome as the dependent variable. Each table panel then progressively narrows the time cells, with our baseline specification presented by Panel D which includes date-by-4 hour fixed effects. All regressions control for the full set of case characteristics beside the time fixed effects. Each cell shows estimates for *Nurse* from the Ordinary Least Squares (OLS) or the Instrumental Variables (IV) specification using *Nurse staff share, past 60 min.* as an instrument and estimated via Two-Stage Least Squares (2SLS), following Equations 1 and 2. The baseline mean represents the average of the dependent variable for cases directly assigned to a doctor. The first-stage K-P F-statistic refers to the Kleibergen-Paap F-statistic. Robust standard errors are shown in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A13. Alternative standard error specifications

	Any referral to doctor (7d)		Informative diagnosis (7d)		Any new prescription (7d)		Rating: top score (7d)		Any ext. PCP cons. (7d)		Any ED (7d)		Any hospitalization (7d)		Income drop >20% (cal. mo. after)		Log cost, incl. ext. PCP	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Panel A: No adjustments																		
Initially to nurse	0.31***	0.30***	-0.064***	-0.068***	-0.10***	-0.083***	0.045***	0.046***	0.029***	0.030***	0.0061***	-0.0045	-0.00023	-0.0010	-0.021***	-0.0066	-0.074***	-0.080***
	(0.00092)	(0.0059)	(0.0015)	(0.0094)	(0.0017)	(0.011)	(0.0018)	(0.012)	(0.0017)	(0.011)	(0.00067)	(0.0043)	(0.00036)	(0.0023)	(0.0024)	(0.015)	(0.0037)	(0.025)
(Baseline) Panel B: Robust standard errors																		
Initially to nurse	0.31***	0.30***	-0.064***	-0.068***	-0.10***	-0.083***	0.045***	0.046***	0.029***	0.030***	0.0061***	-0.0045	-0.00023	-0.0010	-0.021***	-0.0066	-0.074***	-0.080***
	(0.0015)	(0.0060)	(0.0016)	(0.0095)	(0.0017)	(0.011)	(0.0018)	(0.012)	(0.0018)	(0.011)	(0.00074)	(0.0046)	(0.00039)	(0.0025)	(0.0024)	(0.016)	(0.0043)	(0.025)
Panel C: Clustering by login date																		
Initially to nurse	0.31***	0.30***	-0.064***	-0.068***	-0.10***	-0.083***	0.045***	0.046***	0.029***	0.030***	0.0061***	-0.0045	-0.00023	-0.0010	-0.021***	-0.0066	-0.074***	-0.080***
	(0.0032)	(0.0069)	(0.0028)	(0.010)	(0.0023)	(0.011)	(0.0021)	(0.012)	(0.0020)	(0.011)	(0.00082)	(0.0045)	(0.00039)	(0.0025)	(0.0026)	(0.016)	(0.0056)	(0.025)
Panel D: Clustering by login date-by-4 hour windows																		
Initially to nurse	0.31***	0.30***	-0.064***	-0.068***	-0.10***	-0.083***	0.045***	0.046***	0.029***	0.030***	0.0061***	-0.0045	-0.00023	-0.0010	-0.021***	-0.0066	-0.074***	-0.080***
	(0.0023)	(0.0068)	(0.0023)	(0.0100)	(0.0019)	(0.011)	(0.0020)	(0.012)	(0.0018)	(0.011)	(0.00076)	(0.0045)	(0.00039)	(0.0025)	(0.0025)	(0.016)	(0.0047)	(0.025)
Full controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	490,505	490,505	490,505	490,505	490,505	490,505	490,505	490,505	261,034	261,034	490,505	490,505	490,505	490,505	213,425	213,425	261,034	261,034
Baseline mean	0.014	0.014	0.81	0.81	0.44	0.44	0.42	0.42	0.12	0.12	0.030	0.030	0.0087	0.0087	0.22	0.22	6.80	6.80

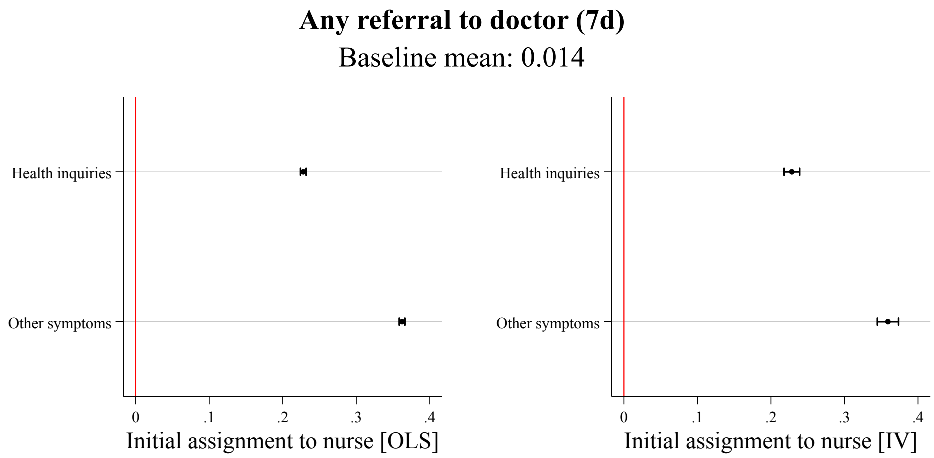
Notes: This table presents regression results for alternative standard errors estimators. We show the effects of initial case assignment to a nurse rather than directly to a doctor, with each column corresponding to a different outcome as the dependent variable. Each table panel uses a different specification to estimate standard errors. Each cell shows estimates for *Nurse* from the Ordinary Least Squares (OLS) or the Instrumental Variables (IV) specification using *Nurse staff share, past 60 min.* as an instrument and estimated via Two-Stage Least Squares (2SLS), following Equations 1 and 2. All regressions control for symptom categories, demographics, health risk factors, and socioeconomic variables, along with login time fixed effects based on date and 4-hour windows. The baseline mean represents the average of the dependent variable for cases directly assigned to a doctor. The first-stage K-P F-statistic refers to the Kleibergen-Paap F-statistic. Robust standard errors are shown in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

D Additional results

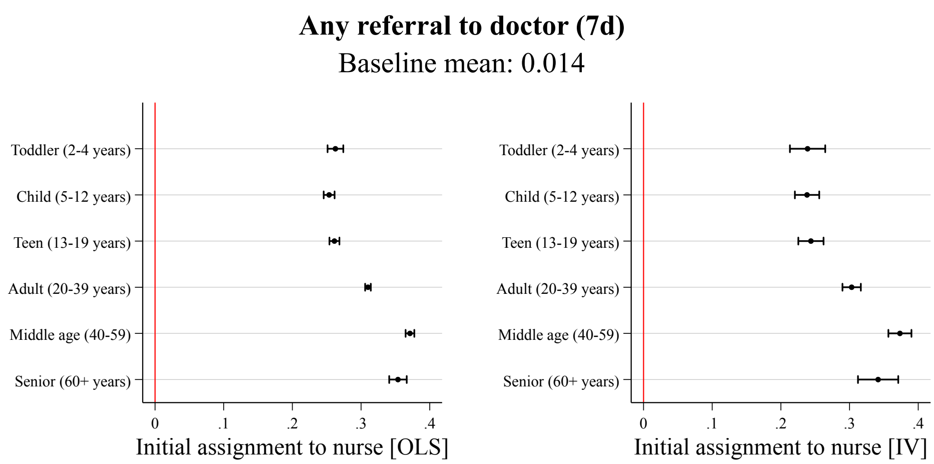
Table A14. Time spent on patients seen by a doctor

	Doctor time spent in min. (7d)		
	Total time	Consultation time	Admin time
Case referred by nurse	1.34***	0.53***	0.81***
	(0.044)	(0.026)	(0.036)
Baseline controls	✓	✓	✓
Additional controls	✓	✓	✓
Observations	414352	414352	414352
Baseline mean	11.7	4.44	7.30

Note: This table presents additional results on cases that are seen by a doctor, which include either cases directly assigned to a doctor or cases referred to a doctor by a nurse. We consider the closest consultation with a doctor within +7 days for cases that are referred from a nurse. The dependent variables measure the amount of time a doctor spends on a case, defined as: "Total time" for overall case duration, "consultation time" for time spent with the patient, and "admin time" for administrative work without patient contact. Doctor time is regressed on an indicator for whether a case was referred by a nurse, and estimated by Ordinary Least Squares (OLS). Robust standard errors are shown in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.



(a) Heterogeneity by main symptom category



(b) Heterogeneity by age group

Figure A12. Heterogeneity across cases in the propensity to be referred to a doctor

Note: These figures show the heterogeneous effects of an initial assignment to a nurse (the *knowledge hierarchy*) on the propensity to refer to a doctor, broken down by subgroups defined either by whether the main reported symptom is "Other health inquiries" (subfigure A12a) or by age category (subfigure A12b). The outcome "Any referral to doctor" refers to any internal referrals or revisits with a doctor. Each subfigure presents estimates from a regression of the outcome on all interactions of the initial nurse assignment (*Nurse*) with each subgroup, while controlling for subgroup fixed effects, the full set of case characteristics, and login date-by-4-hour fixed effects, following Equation 3. The estimates are obtained by Ordinary Least Squares (OLS, left) or Two-Stage Least Squares (IV, right), where interactions of *Nurse staff share, past 60 min.* with each subgroup serving as instruments for the interactions between *Nurse* and each subgroup. Each row represents the estimated subgroup-specific effect of an initial nurse assignment, with horizontal lines representing 95% confidence intervals based on robust standard errors.

Table A15. Mean characteristics of compliers

	Sample mean	**Complier mean**
Symptom categories		
Abdominal pain	0.030	0.051
Cold and flu	0.089	0.025
Cold sores	0.025	0.013
Constipation	0.0071	0.0089
Covid-19	0.058	0.11
Diarrhea or vomiting	0.021	0.029
Eye infection	0.075	0.046
Fever	0.029	0.042
Headache	0.023	0.036
Nail problem	0.022	0.017
Other health inquiries	0.35	0.46
Bites and stings	0.054	0.036
Sinusitis	0.032	0.012
Sore throat	0.076	0.065
Uncategorized	0.030	0.022
Urinary tract infection	0.079	0.032
Demographics		
Female	0.63	0.59
Patient age	29.2	28.5
West Sweden	0.20	0.20
Stockholm	0.44	0.42
Middle Sweden	0.19	0.20
South Sweden	0.095	0.10
Norrland	0.034	0.033
Småland + the islands	0.042	0.044
Health risk		
Any prior hospitalization	0.19	0.19
Any prior ED	0.33	0.34
Any prior urgent care	0.24	0.23
Any prior specialist	0.64	0.63
Any comorbidity	0.21	0.21
Socio-economic variables		
Income above median	0.32	0.30
Any benefits	0.11	0.11
Schooling < 9 years	0.049	0.051
Middle/High school	0.24	0.23
Further educ. < 3 years	0.11	0.10
Further educ. <= 3 years	0.19	0.18
Education n/a	0.42	0.44
Employed	0.57	0.54
Self-employed	0.042	0.042
Unemployed	0.050	0.054
Employment status n/a	0.34	0.36
Married	0.23	0.21
Unmarried	0.075	0.079
Divorced/Widowed	0.42	0.42
Civil status n/a	0.27	0.29
Not migrated	0.75	0.72
Immigrant 1st gen	0.16	0.17
Immigrant 2nd gen	0.096	0.11

Note: This table reports average characteristics for the overall analysis sample and the population of compliers. The first column reprints the overall sample mean. The second column shows the estimated average of a case characteristic x_i in the complier population. We follow the procedure described in Frandsen et al. (2023) based on Abadie (2003) and report the mean characteristic among compliers as the estimate of $\frac{E[\omega_i X_i]}{E[\omega_i]}$, where ω_i is the weight given to case i by the IV. We estimate complier characteristics by Two-Stage Least Squares (2SLS) regressions of the interaction between a predetermined case characteristic and treatment *Initial assignment to nurse* ($x_i \times Nurse_i$) on to the treatment ($Nurse_i$), instrumented by our congestion instrument *Nurse staff share, past 60 min.* and the baseline time controls (login date-by-4-hours fixed effects).

E Data Appendix

The primary data is supplied by a large healthcare firm in Sweden. Their data contains comprehensive records of all online primary care consultations within the firm for the years 2019 and 2020. For each consultation, we observe the date, duration, and purpose of the consultation, the clinician type, the contact method, and primary diagnoses recorded in the form of the ICD-10 codes, the standardized system of disease classification. In addition, we observe patients' demographic characteristics, such as age and gender. For clinicians, demographic information is limited. Each consultation record includes patients' and clinicians' unique personal identifiers, which have been pseudonymized for us.

The personal identifiers enable us to link patients' consultations to additional individual-level information from Statistics Sweden (SCB). SCB provides us with basic information about patients' immigration background. In addition, we obtain information about patients from three key datasets from SCB: the Population Statistics Register (RTB), the Longitudinal Integrated Database for Health Insurance and Labour Market Studies (LISA), and the Education Register (UREG). RTB covers patients' civil status and municipality of residence from 2013 to 2018. LISA provides socio-economic and demographic data for patients, including income, rehabilitation benefits, employment status, paid sickness, and income-related benefits, for the years 2013 to 2020, with additional income and employment data available for 2019 to 2022. Yearly data at LISA is typically updated in November of each year. Finally, UREG offers data on patients' educational backgrounds for the years 2013 to 2020.

We also observe patients' healthcare utilization by linking their unique identifiers from the consultation records to information from the National Board of Health and Welfare (Socialstyrelsen, NBHW). Data from NBHW contains prescriptions, mortality, outpatient, and inpatient healthcare data. The prescription data, covering 2013 to 2023, includes medications collected by patients at pharmacies but excludes medical drugs administered within healthcare facilities. Mortality data, available for 2019 to 2023, includes the date and cause of death. Outpatient and inpatient data, spanning 2013 to 2023, include visit and admission dates, ICD-10 diagnostic codes, and, for out-patient data, additional information on contact types.

In addition, we observe primary care visits, but those data are limited to primary care provided in two regions: Scania and Stockholm. The data from Region Scania and Region Stockholm contain all primary care visits (not only within the firm we study but also for other providers) in Scania, a region in southern Sweden with about 13% of the country's population, and Stockholm, the densely populated capital region with about 20% of Sweden's population. Data from Region Scania cover the period from 2013 to 2020 and data from Region Stockholm contain the years 2013 to 2023. These primary care data are linked to our other data sources through the unique patient identifiers. For primary care visits in Scania, we observe the date of the contact, clinician type, contact type, healthcare clinic, and main diagnoses in the form of the ICD-10 code. For primary care visits in Stockholm, we observe the date of the contact, the care branch, the pseudonymized identifier of the healthcare clinic, and main ICD-10 diagnoses.

Our cost estimates are obtained from various resources, primarily based on public reports. We present an overview of the source material in Appendix Table A16.

Finally, Appendix Table A17 presents an overview of the key case characteristics we use in our main analysis.

Table A16. Costs of healthcare services covered by Sweden's public healthcare system

Service	Cost in SEK	Year	Source	Link
Online primary care consultation				
Doctor	500	2019	Reimbursement rate for digital care services in primary care.	vardanalys.se[1] Last access: 5 Nov 2024
Nurse	275	2020	Recommendation on remuneration for digital healthcare services to healthcare providers (SKR).	skr.se[2] Last access: 5 Nov 2024
External primary care consultation				
Physical consultation	1452.35	2019/2020	Weighted average cost: 0.396×614 SEK+0.604×2002 SEK = 1452.35 SEK. Cost is calculated based on: A doctor visit cost of 2002 SEK (national average for 2019–2020: 0.5×1838 SEK + 0.5×2166 SEK = 2002 SEK), a nurse visit cost of 614 SEK (Region Östergötland estimate), and weights of 39.6% nurse and 60.4% doctor consultations (own calculations from Region Scania data).	vardanalys.se[1] Last access: 5 Nov 2024 Own calculation. Data from Region Scania.
Other healthcare				
Prescription	260	2019/2020	Average cost to the region over all prescriptions in 2019 or 2020 among patients in our analysis sample.	Own calculation Data from Socialstyrelsen.
Specialist visit	3594	2019	Unweighted average cost of a doctor's visit across various specialties in Northern Sweden (Norra Sjukvårdsregionen): medicine, pulmonology, infectious diseases, dermatology, urology, orthopedics, ophthalmology, otolaryngology, pediatrics, and gynecology. We consider specialties for acute and routine specialty services for an average one-time cost and exclude, for example, oncology and psychiatry which likely involve ongoing therapies.	norrasjukvardsregionforbundet.se[3] Last access: 5 Nov 2024
Emergency department visit	3991.5	2019/2020	Average costs in Southern Sweden, $(3,963 + 4,020) * 0.5 = 3,991.5$.	sodrasjukvardsregionen.se[4] Last access: 5 Nov 2024
Urgent care center visit	2002	2019/2020 information from 2024	Using the information provided—"Without an EU card, you must pay the full cost of the care yourself. [...] An appointment with a doctor at a vårdcentralen (healthcare centre) or a visit to the närakuten (urgent care centre) costs SEK 2,093."—we base our analysis on 2019/2020 costs, assuming that urgent care visits have the same cost as regular primary care visits.	1177.se[5] Last access: 5 Nov 2024
Hospitalization	7800	2020	Stated average cost of a hospitalization.	kristianstadsbladet.se[6] Last access: 5 Nov 2024

[1] Full source: https://www.vardanalys.se/rapporter/besok-via-natet/
[2] Full source: https://skr.se/download/18.32563d7d1784aa279ece294c/1618741364556/11-2020-WEBB-Rek-om-ersattning-for-digitala-vardtjanster.pdf
[3] Full source: https://www.norrasjukvardsregionforbundet.se/halso-och-sjukvard/avtal-och-priser/arkiv/
[4] Full source: https://sodrasjukvardsregionen.se/verksamhet/avtal-priser/regionala-priser-och-ersattningar-foregaende-ar/
[5] Full source: https://www.1177.se/en/Stockholm/other-languages/other-languages/soka-vard/hitta-ratt-vard-nara-dig—andra-sprak—stockholms-lan/
[6] Full source: https://mosaik.kristianstadsbladet.se/nyheter/varden-kostar-mycket-mer-an-du-betalar-for/

Table A17. Description of key case characteristics

Variable	Description	Data source
Treatment and instrument		
Initial assignment to a nurse; Nurse	Treatment: Indicator variable that is one if the initial consultation of a case is assigned to a nurse, and zero if the case is directly assigned to a doctor.	Provider
Nurse share, past 60 min	Congestion instrument: The share of initial nurse consultations among cases in the past 60 minutes of a given case	Provider
Login time		
Login date-by-4 hours	Fixed effects for the login time of a case when a consultation is requested. Login time is constructed as calendar date-by-4 hour window: 0 am - 4 am; 4 am - 8 am; 8 am - 12 pm; 12 pm - 4 pm; 4 pm - 8 pm; 8 pm - 12 am of a given date, such as April 1, 2020, 12 am - 4 am.	Provider
Symptoms		
Symptom category	A set of indicator variables for the main symptom category of the case, as one of 16 symptoms included in our analysis sample. The full list of symptoms is provided in A6.	Provider
Demographics		
Gender	Indicator variable that is one if the patient is female.	Provider
Age group	A set of indicator variables for the patient's age group, as one of the following categories: Infant (0-1 year); Toddler (2-4 years); Child (5-12 years); Teen (13-19 years); Adult (20-39 years); Middle age (40-59 years); Senior (60+ years).	Provider
Region	A set of indicator variables for the patient's registered region as reported in November of the year prior to the initial consultation. We consider six regional areas based on Sweden's national areas (riksområden): West Sweden, Stockholm, Middle Sweden (combining East Middle Sweden and West Middle Sweden), South Sweden, Norrland (combining Middle Norrland and Upper Norrland), and Småland and the islands.	SCB
Health risk		
Any prior hospitalization	Indicator for any inpatient hospitalization within the 3 years prior to the initial consultation, excluding visits in the 30 days immediately preceding it.	NBHW
Any prior ED	Indicator for any emergency department (akutmottagning) visit within the 3 years prior to the initial consultation, excluding visits in the 30 days immediately preceding it.	NBHW
Any prior urgent care	Indicator for any urgent care center (närakuter) visit within the 3 years prior to the initial consultation, excluding visits in the 30 days immediately preceding it.	NBHW
Any prior specialist	Indicator for any specialist visit within the 3 years prior to the initial consultation, excluding visits in the 30 days immediately preceding it.	NBHW
Any comorbidity	Indicator variable for an Elixhauser comorbidity index of 1 or higher, as diagnosed within specialty or hospital care in 2018 or earlier.	NBHW
Socio-economic variables		
Income above median	Indicator variable for the patient's annual income reported in Nov 2018 being above the median (non-missing) income in-sample: Income above 294,700 SEK (approximately 28,000 USD). Income is reported for patients $\geq$ age 20 in Nov 2018.	SCB
Income missing	Indicator variable for information on income being unavailable (either missing or unreported).	SCB
Any benefits	Indicator variable for the receipt of any social security benefits in 2018. Benefits are reported for patients $\geq$ age 20 in Nov 2018.	SCB
Benefits missing	Indicator variable for information on benefits being unavailable (either missing or unreported).	SCB
Employment status	Indicator variable for the main employment type reported in Nov 2018, as one of the following categories: employed; self-employed; unemployed. Employment status is reported for patients aged 20 to 67.	SCB
Employment status missing	Indicator variable for information on employment status being unavailable (either missing or unreported).	SCB
Education level	Indicator variable for the highest educational degree reported in Nov 2018, as one of the following categories: schooling below 9 years; middle/high school; secondary education $<$ 3 years (further educ. $<$ 3 years); secondary education $\geq$ 3 years (further educ. $\geq$ 3 years). Educational degrees are reported for patients $\geq$ age 25.	SCB
Education level missing	Indicator variable for information on education being unavailable (either missing or unreported).	SCB
Civil status	Indicator variable for the civil status reported in Nov 2018, as one of the following categories: married; divorced or widowed; unmarried. Civil status is reported for patients $>$ age 18.	SCB
Civil status missing	Indicator variable for information on civil status being unavailable (either missing or unreported).	SCB
Migration status	A set of indicator variables for the migration status of the patient, as one of the following categories: Not migrated; 1st generation immigrant (born outside of Sweden); 2nd generation immigrant (born in Sweden to two foreign-born parents). Migration information is available for the entire sample.	SCB

Notes: This table presents the variable definition and main data source for key case characteristics.