Tele-Triage, Care Substitution, and Health: Evidence from Quasi-Randomly Assigned Nurses

Ken Suzuki* Liam Rose† Linda Diem Tran‡ Anita Vashi§

October 10, 2025

Find the most current version here.

Abstract

Patients experiencing acute health symptoms often face uncertainty about how and where to receive care. We study patients who call a nurse advice line and receive one of four recommendations: emergency department (ED), urgent care (UC), primary care (PC), or self-care (Home). Leveraging an extension of examiner designs that recovers margin-specific effects for each pair of adjacent recommendations (ED-UC, UC-PC, PC-Home), we estimate the impact of nurse recommendations on both patient decisions and their subsequent health outcomes. We find that recommendations have large impacts on patient decisions at each margin. However, we find no evidence that recommending more acute care over less acute care results in survival gain. We further show that UC recommendations reduce costs relative to ED and PC recommendations by \$404 and \$247, respectively.

*Department of Veterans Affairs and Stanford Medicine ksuzuki4@stanford.edu

†Department of Veterans Affairs and Stanford Medicine liamrose@stanford.edu

‡Department of Veterans Affairs and Stanford Medicine diemtran@stanford.edu

§Department of Veterans Affairs, University of California, San Francisco, and Stanford Medicine anitavashi@gmail.com

We are grateful to Todd Wagner, Carlos Dobkin, Laura Giuliano, Jasmin Moshfegh, Hannah Bae, Nathaniel Breg, and seminar participants at UC Santa Cruz and Stanford University for helpful comments. This project was funded by the Department of Veteran Affairs (Office of Integrated Veteran Care and project EBP 22-108, 1I50HX003614-01). The content is solely the responsibility of the authors. All errors are our own. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

1 Introduction

Telephone triage services are increasingly prevalent across healthcare systems and insurance networks, including all major U.S. insurers, national health systems like the NHS, and state and federal health agencies. These lines provide individuals with access to medical personnel – most often a registered nurse – who can offer health-related information, advice, and guidance. For patients seeking unplanned care, these services are intended to serve as the first point of contact, where nurses assess symptoms and recommend an appropriate level of care. Beyond offering guidance, telephone triage has become a critical tool for managing rising healthcare demands and alleviating ED overcrowding by directing patients to the most suitable care setting.

Although telephone triage services can play a potentially important role in reducing information frictions and improving allocative efficiency, we know little about whether and how triage recommendations influence patient choices, healthcare costs, and health outcomes within healthcare delivery. Furthermore, there is little empirical evidence on the consequences of human variability in the upstream stage of access to care. Prior research highlights substantial variation in clinical decisions among healthcare professionals, suggesting that some patients may receive a higher or lower level of care than their symptoms warrant (Abualenain et al., 2013; Molitor, 2018; Chan and Gruber, 2020; Chan et al., 2022; Coussens and Ly, 2025). In the telephone triage context, some nurses may perform triage too defensively, funneling low-acuity patients into the ED unnecessarily. Other nurses may advise patients to seek less intensive care too optimistically, causing a delay in timely access to critical care.

We aim to address this gap using unique administrative data from a nurse advice line operated by the Veterans Affairs (VA) healthcare system. We apply an examiners design, leveraging the quasi-random assignment of calls to the next available nurse within call centers. In this context, triage nurses choose one of four follow-up recommendations: emergency department (ED), urgent care (UC), primary care (PC), or self-care (Home) according to symptom. Patients face low or no cost sharing that is equal in each possible care setting, and they may or may not follow the recommendation.

This institutional setting presents two challenges vis-a-vis the traditional examiner design. First, although the conventional examiners design literature often collapses multiple treatment alternatives into a binary treatment, recent research highlights that two-stage least squares (TSLS) may fail to identify interpretable causal effects when examiners affect potential treatments on sub-treatment margins (Mueller-Smith, 2015; Chyn et al., 2024). In our context, one might collapse triage recommendations into ED vs. non-ED (UC, PC, and Home) and use nurse ED recommendation propensity as a single instrument for the binarized ED recommendation indicator. However, if nurses with a high ED propensity also tend to recommend UC over PC, for example, this TSLS design does not recover interpretable effects of ED recommendation.[1]

To address this identification problem, we examine identification conditions under which TSLS recovers margin-specific causal effects on each of two adjacent recommendations (ED-UC, UC-PC, PC-Home), based on recent methodological development in the examiners design (Humphries et al., 2024). We use TSLS with a focal recommendation propensity as an instrument for the focal recommendation (e.g., ED), while controlling for non-focal recommendation propensities (e.g., PC and Home). Under the identification assumptions, this method allows us to recover the margin-specific local average treatment effect (LATE) on each margin (e.g., the effect of ED recommendation relative to UC recommendation).

Second, our setting differs from the conventional examiner setup in that patients

[1]In Appendix Section B2, we show that the TSLS identifies the sum of (i) a weighted average of sub-LATEs involving ED recommendation and (ii) bias terms, if we binarize recommendations into ED and non-ED.

need not follow the recommendation of the examiner. For this reason, our TSLS estimates recover the effects of receiving a particular recommendation (e.g., ED) – and not the effects of receiving a particular type of care. To see why, consider the patients whom nurses consider to be on the ED-UC margin. Some of these patients may choose self-care if they are recommended UC, while they may visit a PC when recommended ED.[2] For these patients, assignment to a nurse with a high propensity to recommend ED may affect downstream health outcomes through patient choices other than whether to visit the ED.[3]

We find substantial variation in triage recommendations among nurses on the high and middle acuity margins (ED-UC and UC-PC), after controlling for non-focal recommendation propensities and patient baseline characteristics (county of residence (FIPS), call center, call time, and algorithm disposition). Our first-stage coefficient indicates that reassigning a call from a nurse at the 5th percentile to one at the 95th percentile of the ED recommendation propensity distribution increases the likelihood of being recommended ED over UC by 12.6 percentage points. Similarly, call reassignment from the 5th to the 95th of the UC recommendation propensity distribution increases the likelihood of being recommended UC over PC by 19.3 percentage points. On the low acuity margin (PC-Home), calls assigned to the 5th and the 95th percentile nurses differ by 4.9 percentage points in their propensity to be recommended PC over Home.

Our results suggest that, on average, patients recommended ED care are 24.0 percentage points less likely to use UC and 25.3 percentage points more likely to use

[2]This shift is possible, for example, if a recommendation for more intensive care changes patient belief about the necessity of professional care (on extensive margin).

[3]Recent studies highlight similar identification problems due to subjects' non-compliance in the context of randomized field experiments with multiple treatment alternatives (Kirkeboen et al., 2016; Kline and Walters, 2016; Pinto, 2022). The literature seeks natural ways to impose identification restrictions on non-compliance. For instance, Kline and Walters (2016) and Pinto (2022) restrict certain response types, invoking revealed preferences. Exploring such restrictions in examiners design is beyond the scope of this paper, and we leave it for future research.

ED within three days of the call, compared to those recommended UC. In addition, the ED recommendation modestly shifts patient utilization on the extensive margin. Patients recommended ED are 6.9 percent more likely to shift their choice from self-care (Home) to any professional care (PC, UC, or ED), compared to patients recommended UC. However, we find no evidence that this overall shift in utilization toward more intensive care results in survival gain. Moreover, 28 days after the initial call, patients recommended ED care incur $404 greater cumulative health care spending than patients recommended UC.

For patients on the less acute UC-PC margin, we find that being recommended UC care leads to a 11.5 percentage point (32.7%) drop in the primary care visits and a 20.7 percentage point increase in urgent care use within three days of the call. Furthermore, patients recommended UC are 1.2 percentage points (13.8%) less likely to use ED within three days, compared to patients recommended PC. On the extensive margin, the UC recommendation modestly increases the probability of receiving any care by 4.8 percentage points relative to the PC recommendation. Overall, this mixture of shifts in utilization contributes to a reduction in health care spending. After 28 days, patients recommended UC have $247 less in costs, compared to patients recommended PC.

Lastly, on the least acute PC-Home margin, our results suggest that patient utilization shifts almost exclusively on the corresponding care margin. Patients recommended PC are 7.6 percentage points more likely to use primary care within three days, compared to patients recommended Home. By contrast, the effects of being recommended PC over Home on other utilization indicators (UC, ED, and hospital admission) are small and statistically insignificant.

This paper contributes to several strands of literature. First, our study is closely related to the literature on variation in clinical decisions among healthcare professionals (Abualenain et al., 2013; Molitor, 2018; Chan and Gruber, 2020; Coussens and

Ly, 2025) and research that exploits quasi-random assignment of healthcare providers and variations in their practice style (Chan et al., 2022; Silver and Zhang, 2022; Chan et al., 2023). Although healthcare providers often face multiple treatment alternatives, applications of the examiners design mostly focus on a binary (or binarized) treatment. We extend recent methodological developments in the judges design into a healthcare setting to address multiple treatments (Arteaga, 2023; Humphries et al., 2024; Kamat et al., 2024). Our application is most closely related to Humphries et al. (2024) that considers identification conditions to recover margin-specific causal effects (incarceration vs. conviction; conviction vs. dismissal) when judges make an ordered choice based on a single latent index (e.g., criminality).

Second, we contribute to a limited but growing literature on the impact of triage decisions on access to care within healthcare delivery (Chan and Gruber, 2020; Ferro and Serra, 2025; Islam et al., 2021; Sexton et al., 2022). The literature on triage in EDs documents substantial variation in triage decisions between and within triage nurses, with patients assigned a lower priority experiencing longer wait time or additional ED care afterward (Chan and Gruber, 2020; Ferro and Serra, 2025). We show that the telephone triage process and the variability among triage nurses affect patient choices and access to care further upstream, even before patients go to healthcare facilities.

Finally, a broad literature investigates the consequences of substituting care settings. While ED care accounts for over 5% of U.S. healthcare spending, studies suggest that more than 30% of ED visits could be managed in less acute settings (Weinick et al., 2010; Vashi et al., 2019).[4] Urgent care centers have been shown to substitute for EDs, particularly when examined through availability during extended hours (Allen et al., 2021) or when clinics become available in a local market (Alexander et al., 2019).

[4]Evaluations of efforts to move care out of EDs aside from urgent care have shown mixed results (Flores-Mateo et al., 2012; Raven et al., 2016). Increases in primary care access successfully reduced ED visits in uninsured and Medicaid populations (Sadowski et al., 2009; Retchin et al., 2009), but case management, individualized care plans, and information sharing were not consistently effective (Soril et al., 2015).

However, UCs may also contribute to higher overall healthcare spending (Wang et al., 2021; Currie et al., 2023). Currie et al. (2023) shows downstream hospital admissions as a potential cause of increased spending. Using a different source of variation, we show that for patients considered by a triage nurse to be on the margin of needing ED care, substitution toward UC can be cost-saving.

2 Background

The VA operates one of the nation's largest health care systems, providing care to approximately 10 million veterans at 171 medical centers and 1113 outpatient facilities distributed across the country. To receive VA health care, an individual must have served and been honorably discharged from the military and qualify under at least one of three broad categories: have a disability connected to their service, have income below a set threshold, or have been discharged within the last five years.[5] In a given year, VA provides care to about one-third of US veterans, providing extensive service in a vertically-integrated system that includes primary care, mental health care, specialty care, acute care, and long-term care.

VA medical centers and outpatient clinics generally operate on a "hub-and-spoke" model, where regional medical centers work together with a number of nearby outpatient clinics. Medical centers are then geographically divided into 18 regional care systems known as Veteran Integrated Service Networks (VISNs). Historically, medical centers and VISNs developed their own call centers to serve as entry points for veterans and their families. The call centers provide frequently used administrative and clinical services. While services have differed somewhat among call centers, they all provided some form of assistance with appointment scheduling, enrollment questions,

[5] VA uses the Department of Housing and Urban Development's annual geographic-based income limits, further allowing individuals to be 10% over the threshold if they agree to pay copays. Over 80% of enrolled veterans face no cost sharing.

pharmacy services, and nurse triage.

Nurse triage services allow patients to speak with a Registered Nurse (RN) for symptom evaluation and healthcare disposition. When a veteran calls for triage, the next available nurse is assigned to assess the patient's needs and recommend appropriate follow-up care. The triage process is standardized through a decision-support algorithm used across all nurses and call centers. First, the nurse gathers and inputs basic patient information—such as age, gender, chief complaint, and pain scale—into the algorithm. Second, the algorithm generates clinical questions based on these inputs. Third, the nurse communicates with the patient and enters the responses. Finally, the algorithm provides recommendations for disposition (e.g., ED, urgent care, primary care, dentist, or self-care) and follow-up timing (e.g., now-911, now, 2-8 hours, 12-24 hours, 2-3 days, 1-2 weeks). Appendix Figure B1 illustrates these steps.

While the decision-support algorithm standardizes the telephone triage process, nurses still exercise discretion by overriding the algorithm's triage recommendations. This study exploits cross-nurse variations in their discretion within algorithm recommendations. Appendix Figure B2 illustrates this by plotting each nurse's propensity to recommend ED care, conditional on whether the algorithm suggests ED care (y-axis) or non-ED care (x-axis), across call centers. Nurses positioned in the top-right corner of each panel (call center) consistently follow the algorithm, whereas those farther from this corner are more likely to override the algorithmic recommendations.

In addition to providing services directly, VA also purchases care from non-VA providers. Importantly, this includes emergency care, with more than one-third of ED visits involving VA occurring at non-VA facilities. VA encourages enrollees that consider their life or health to be in danger to seek immediate medical attention, and prior approval is not required. Further, VA maintains a network of non-VA urgent care centers that enrollees can utilize. Triage nurses are instructed to work with patients to direct them to the appropriate care location.

3 Data

3.1 Overview

We construct our analysis sample by linking multiple sources of administrative data from the VA, including records of nurse triage cases, healthcare utilization, and patient demographics. This section sketches the most relevant information about our analysis sample. Appendix Table B1 describes our data cleaning and sample construction in further detail.

3.2 Data Sources and Sample Construction

Our sample construction starts with the universe of telephone triage cases received in all call centers across the US from July 1, 2018, to December 31, 2022. The triage records have information at the call level, including triage date-time (year, month, day, hour, and minute), patient ID, triage nurse ID, station (call center) ID, triage disposition (recommended follow-up location and timing), chief complaint (symptom), pain scale (0-10), and duration of chief complaint. Each call is further linked to the patient's prior healthcare utilization events at VA facilities (within 365 days of the triage), prior diagnoses (31 Elixhauser comorbidity indices), VA benefits eligibility status (priority group indicators), and demographics (e.g., age, gender, marital status). We use these covariates for randomization and robustness checks as well as for profiling complier characteristics.

We define a patient is recommended "ED" by nurse if nurse follow-up location is emergency department. Likewise, we say that a patient is recommended "UC" if nurse follow-up location is urgent care. We define a patient to be recommended "PC" if nurse's follow-up location is clinic or other miscellaneous categories, such as virtual care and dentist. Lastly, we say that a patient is recommended "Home" if nurse's follow-up location is home. Appendix Table B2 describes triage recommendations in detail.

Our primary outcomes include healthcare utilization events after triage (primary care, urgent care, and ED), emergency hospital admission (those that came through the ED), mortality, and health care spending. We construct mortality indicators from the date of the triage call using VA vital status data, which are further supplemented with death records from both Medicare and the Social Security Administration. We also capture full costs of downstream care. Costs for VA-provided care come from Managerial Cost Accounting (MCA) data, which provides the official cost estimates for VA encounters. Costs for VA-paid care come from claims paid by VA to outside providers.

To construct our main sample, we impose the following key restrictions (See Appendix Table B1 for details). First, we drop the calls during non-business hours (before 8 am, after 4 pm, weekends, and holidays). Some call centers do not offer telephone triage during non-business hours and transfer calls to other call centers or non-VA contractors. Second, we remove calls from patients with the most recent prior triage call or ED visit within 30 days to define an index triage incident. Third, we only keep calls in nurse-by-call center-by-year-by-algorithm disposition cells with at least 10 observations.[6] From this we drop any cells that do not have at least two nurses remaining. With these restrictions, our baseline sample consists of 1,273,843 calls (from 836,420 patients) received by 1,725 nurses at 96 call centers.

Table 1 summarizes the characteristics of our sample of triage calls. The average caller is a near-elderly patient (average age = 63.1) with high rates of healthcare

[6]We impose this restriction since our identification strategy relies on cross-nurse variation within cells defined by algorithm disposition as described in Section 4.3. This restriction ensures that nurses' ED tendencies are estimated with a sufficiently large number of observations. Anecdotally, some nurse managers stated that some nurses would only work nurse triage for short periods, or that nurse managers themselves would occasionally step in to field calls when needed.

utilization in the previous year: 97.4% for primary care, 38.2% for ED visits, and 10.2% for hospital admissions. Nurses recommend primary care (PC) for 62.7% of calls, urgent care (UC) for 4.4%, and emergency department (ED) care for 27.7%. Utilization rates within three days of the call are low: 30.7% of patients visit PC, 6.3% visit UC, and 19.2% visit the ED. Only 49.9% receive any form of care (PC, UC, or ED) within three days.

4 Empirical Methods

4.1 Overview

Our primary goal is to analyze how nurse recommendations influence patient healthcare choices after the triage call. For simplicity, suppose that the patient has four alternative options: (i) self-care (or no care) at home ($Home$), (ii) primary care (PC), (iii) urgent care (UC), and (iv) emergency department (ED). Likewise, suppose that the nurse selects one of the four follow-up care recommendations: $Home$, PC, UC, or ED. We denote the joint distribution of the patient's healthcare utilization as follows:

$$F(Home_{id}, PC_{id}, UC_{id}, ED_{id} \mid R_i) \tag{1}$$

where ED_{id} indicates whether patient i visits an ED within d days of triage. UC_{id}, PC_{id}, and $Home_{id}$ are indicators of urgent care, primary care, and self-care at home, respectively. $R_i \in \{Home, PC, UC, ED\}$ indicates the health care option recommended by the nurse.

We would like to contrast this conditional distribution (1) evaluated at different recommendations $R_i \in \{Home, PC, UC, ED\}$. For example, consider a case where the nurse changes the triage recommendation from UC to ED. If recommendations were randomly determined, we would know the effect of this change on the patient choice probability by contrasting two distributions:

$$\begin{aligned} &F(Home_{id}, PC_{id}, UC_{id}, ED_{id} \mid R_i = ED) \\ &\quad \text{vs.} \\ &F(Home_{id}, PC_{id}, UC_{id}, ED_{id} \mid R_i = UC). \end{aligned} \tag{2}$$

In general, it is difficult to identify the effect of nurse recommendation with observational data, as the nurse does not select a follow-up care recommendation at random. The nurse is more likely to recommend an acute form of care when patients present more acute and severe health conditions. Patients with acute symptoms may use an acute form of care, regardless of the nurse's recommendations. Hence, the contrast (2) may only tell us differences in health conditions between patients when it is calculated with observational data.

We address this identification problem by exploiting the quasi-random assignment of triage nurses to calls and variations in triage recommendations across those nurses. We exploit that different nurses have different preferences on the appropriate level of follow-up care even for patients with the same health conditions, analogous to the findings in the literature on physicians' practice variations.[7]

The following subsections describe our research design. Although we analyze margin-specific causal effects for each pair of the two adjacent triage recommendations (ED-UC, UC-PC, and PC-Home), the following sections describe our empirical strategy for the ED-UC margin for compactness. Appendix Sections A1 and B1 detail our strategy. In what follows, we use "call" and "patient" interchangeably for notational simplicity unless otherwise noted, although the unit of observation i is call.

[7] Some examples include Coussens and Ly (2025), Abualenain et al. (2013), and Molitor (2018)

4.2 Empirical Model

Although empirical applications of examiners design often rely on examiners' propensities to apply a binary treatment as an instrument, recent research highlights that TSLS estimands may be biased and potentially wrong-signed when researchers collapse multiple treatment options into a binary treatment indicator (Mueller-Smith, 2015; Chyn et al., 2024). In our context, exclusion restriction may be violated when nurses with a high ED tendency also tend to recommend UC over PC or PC over Home.[8] To address this identification problem, we examine identification conditions that allow us to identify margin-specific causal effects, based on Humphries et al. (2024). Under the assumptions, we use TSLS with nurse ED tendency as an instrument for ED recommendation, while controlling for non-focal recommendation propensities (PC and Home). Intuitively, we recover the causal effects of recommending ED over UC by comparing nurses who differ in propensities to recommend ED over UC but have the same propensity to recommend PC and Home.

To describe our identification conditions, consider patient i triaged by nurse $J_i \in \mathcal{J}_{c(i),t(i)}$ who is among the set of available nurses $\mathcal{J}_{c,t}$ in call center c at time t. Our multiple treatments are the nurse's follow-up care recommendations, denoted by $R_i \in \{Home, PC, UC, ED\}$. Let $R_i^r = 1\{R_i = r\}$, $r \in \{Home, PC, UC, ED\}$ be an indicator of whether patient i is recommended health care opion r from the nurse. We define the nurse's propensity to recommend ED by $Z_i^{ED} = E\left[R_i^{ED} \mid J_i\right]$. We similarly define Z_i^{UC}, Z_i^{PC}, and Z_i^{Home}. We denote the potential recommendation by $R_i(z^{ED}, z^{PC}, z^{Home})$ and the potential recommendation indicator for option r by $R_i^r(z^{ED}, z^{PC}, z^{Home})$ to clarify that the patient would receive a different triage recom-

[8] In Appendix Section B2, we show that the examiners design estimand cannot be generally interpreted as causal effects of ED recommendation if we binarize recommendations into ED and non-ED (UC, PC, or Home) and use nurse propensity to recommend ED as a single instrument. Specifically, we demonstrate that it identifies the sum of (i) a weighted average of sub-LATEs involving ED recommendation and (ii) bias terms.

mendation if assigned to a nurse with a different ED tendency z^{ED}, conditional on PC and Home tendencies, z^{PC} and z^{Home}. The observed recommendation relates to the potential recommendation by $R_i = R_i(Z^{ED}, Z^{PC}, Z^{Home})$.

We examine the effect of being recommended ED over UC on several patient outcomes Y_i within 3 to 28 days of triage, including health care utilization (primary care, urgent care, ED, hospital admission), costs (outpatient, inpatient, and total costs), and mortality. The associated potential outcomes $Y_i(r)$, $r \in \{Home, PC, UC, ED\}$, are a function of triage recommendation that the patient receives from the nurse. We specify the TSLS model as follows:

$$Y_i = \beta_0 + \beta_1 R_i^{ED} + \beta_2 Z_i^{PC} + \beta_3 Z_i^{Home} + X_i'\pi + u_i \tag{3}$$

$$R_i^{ED} = \alpha_0 + \alpha_1 Z_i^{ED} + \alpha_2 Z_i^{PC} + \alpha_3 Z_i^{Home} + X_i'\delta + \nu_i \tag{4}$$

where β_1 is our parameter of interest that captures the effect of ED recommendation, relative to UC recommendation. X_i is a vector of patient characteristics, described in Section 4.3. This specification also controls for the nurse's propensities to recommend primary care Z_i^{PC} and self-care Z_i^{Home} to avoid violation of exclusion restriction.

Formally, our identification strategy uses the quasi-randomly assigned nurse's ED tendency Z_i^{ED} as an instrument for the patient's ED recommendation status R_i^{ED}, after controlling for X_i, Z_i^{PC}, and Z_i^{Home}. We consider the following conditions for IV validity (Imbens and Angrist, 1994; Humphries et al., 2024):

Condition 1 (IV Validity): For a random sample of triage calls i, the following conditions hold, conditional on X_i:

(i) Relevance: $\alpha_1 \neq 0$ in equation (4).

(ii) Independence and Exclusion: Potential outcomes and potential triage recommendations are independent of the nurse's overall tendencies in recommending

ED, PC, and Home. Specifically,

$\big(\{Y_i(r)\}_{r\in\{\text{ED, UC, PC, Home}\}}, R_i(z^{ED}, z^{PC}, z^{Home})\big) \perp \big(Z_i^{ED}, Z_i^{PC}, Z_i^{Home}\big)$,

$\forall z^{ED}, z^{PC}, z^{Home}$.

(iii) Unordered Partial Monotonicity: Consider a set of nurses with the same propensities to recommend PC (z^{PC}) and Home (z^{Home}) but differing tendencies to recommend UC versus ED care ($z^{ED} < z'^{ED}$). For all i,

(a) $R_i^{ED}(z'^{ED}, z^{PC}, z^{Home}) \geq R_i^{ED}(z^{ED}, z^{PC}, z^{Home})$,

(b) $R_i^{UC}(z'^{ED}, z^{PC}, z^{Home}) \leq R_i^{UC}(z^{ED}, z^{PC}, z^{Home})$,

(c) $R_i^{PC}(z'^{ED}, z^{PC}, z^{Home}) = R_i^{PC}(z^{ED}, z^{PC}, z^{Home})$, and

(d) $R_i^{Home}(z'^{ED}, z^{PC}, z^{Home}) = R_i^{Home}(z^{ED}, z^{PC}, z^{Home})$.

We examine each of those IV conditions in Section 4.4. Appendix Section A1 describes a behavioral model of nurse decision-making that aligns with our IV assumptions. Briefly, these conditions are satisfied, for example, if nurses (i) face the common acuity distribution, (ii) select a recommendation based on patient acuity (ED for the highest acuity, UC for upper middle, PC for lower middle, and Home for the lowest), and (iii) vary in their threshold levels of acuity that divide patients into two adjacent recommendations (ED-UC, UC-PC, PC-Home). Appendix Section B1 formally shows that our design identifies margin-specific LATE under Condition 1.

4.3 Constructing Leave-Out Nurse Recommendation Measures

To operationalize our IV strategy, we first calculate each nurse's ED recommendation propensity within cells defined by call center, call year, and algorithm disposition (follow-up location and interval), and we use the nurse's cell-specific ED propensity as an instrument for whether the patient is recommended UC or ED care. We focus on cross-nurse variation within these cells for two reasons. First, we restrict our comparison to nurses in the same environment, as local environments—such as call center norms and available healthcare resources—likely influence nurses' final recommendations (Islam et al., 2021). Second, nurses' treatment intensity relative to other nurses may vary across different algorithmic recommendations. A nurse could have a relatively high ED vs. UC tendency when the algorithm recommends ED, while having a relatively low ED vs. UC tendency when the algorithm recommends UC. This non-uniformity causes a violation of monotonicity if we use nurses' unconditional ED propensity as an instrument.[9]

We construct a leave-one-out instrument by averaging ED recommendation indicators of other patients triaged by the same nurse, following the examiner design literature. Specifically, for call i that is assigned to nurse j, we first obtain residual of ED recommendation status, denoted as R_i^{ED*}, before calculating the leave-one-out average, using the following linear regression:

$$R_i^{ED*} = R_i^{ED} - X_i^{0\prime}\Gamma - H_i'\Lambda = Z_{ij} + \epsilon_i \tag{5}$$

where X_i^0 is a vector of our baseline controls, including interactions of call center-by-call time (month-year, day-of-week, and AM/PM indicators), interactions of call center-by-algorithm disposition, and county of patient's residence (5-digit FIPS indicators). In addition to X_i^0, we also partial out a vector of additional controls H_i ("hold-out controls"), including age, sex, marital status, race and ethnicity, period of service, prior healthcare utilization, and prior diagnoses. Importantly, the hold-out

[9] While all nurses in our data use the same decision-support algorithm, Appendix Figure B2 shows that the extent to which nurses override the algorithm's ED (or non-ED) disposition varies markedly across call centers. For example, in call center 23, all nurses almost always recommend ED when the algorithm recommends ED, while these nurses substantially vary in their ED propensity when the algorithm recommends non-ED options. In fact, some VA call centers do not allow nurses to downgrade recommendations from ED to non-ED care when the algorithm recommends ED care. Hence, nurses' treatment intensity measured by unconditional ED propensity may incorrectly rank these nurses for some algorithm disposition cells. This mechanically introduces monotonicity violation (Mueller-Smith, 2015; Sigstad, 2023).

set H_i is not essential for balance, but is included for statistical precision.[10] The residuals R_i^{ED*} include nurse j's ED tendency Z_{ij} and idiosyncratic call-level error term ϵ_i.

Then we construct the leave-one-out measure for call i by averaging the residual ED recommendation of all other patients but patient $k(i)$ assigned to nurse j in call center c with algorithm disposition a:

$$Z_i^{ED} = \frac{1}{K_{jcya} - 1} \sum_{i'} \frac{1\left\{k(i') \neq k(i), j(i') = j, c(i') = c, y(i') = y, a(i') = a\right\} R_{i'}^{ED*}}{n_{k(i')jcya}} \tag{6}$$

where K_{jcya} is the number of patients assigned to nurse j in call center c in year y with algorithm disposition a and n_{kjcya} is the total number of calls from patient k received by nurse j in call center c in year y with algorithm disposition a. We call this Z_i^{ED} as the leave-one-patient-out measure and use it as an instrument for the patient's ED recommendation indicator R_i^{ED} in equation (4). We similarly construct leave-out measures of the nurse's propensity to recommend UC, PC, and Home (Z_i^{UC}, Z_i^{PC}, Z_i^{Home}).

As robustness checks, we implement our IV design by constructing the leave-one-out measures in several different ways and confirm that the results are unaffected. Appendix Section B3 details our alternative measures.

4.4 Instrument Validity

4.4.1 Relevance

We begin our empirical analysis by examining each of the IV assumptions in Condition 1 before presenting our TSLS estimation results. First, we test for instrument

[10] We check balance without controlling for the hold-out set H_i. We confirm that patients are balanced between nurses within the baseline controls X_i^0.

relevance (Condition 1(i)). This assumption requires that the leave-out measure of ED recommendations Z_i^{ED} meaningfully predicts the actual ED recommendation R_i^{ED} ($\alpha_1 > 0$), conditional on the leave-out PC and Home measures, Z_i^{PC} and Z_i^{Home}, and the baseline controls X_i^0.

Figure 1a presents a binned scatter plot of residualized ED recommendation R_i^{ED} on the y-axis against residualized leave-out measure Z_i^{ED} on the x-axis. The figure indicates that shifting the call assignment from the 5 percentile (leftmost) bin to the 95 percentile (rightmost) increases the probability of receiving an ED recommendation from 15.8% to 39.3%, holding the leave-out PC measure constant. The solid line represents the estimated first stage coefficient $\widehat{\alpha}_1 = 0.8595$ (SE $= 0.0067$), confirming the first-stage relationship between R_i^{ED} and Z_i^{ED} is statistically significant.

Similarly, Figure 1b presents the first-stage relationship between R_i^{UC} and Z_i^{UC}, conditional on the leave-out ED and Home propensities Z_i^{ED} and Z_i^{Home} ($\widehat{\alpha}_1 = 0.9196$, SE $= 0.0050$). Likewise, Figure 1c shows the first-stage relationship between R_i^{PC} and Z_i^{PC}, conditional on the leave-out ED and UC propensities Z_i^{ED} and Z_i^{UC} ($\widehat{\alpha}_1 = 0.9376$, SE $= 0.0057$).

4.4.2 Conditional Independence and Exclusion

We next consider the conditional independence assumption (Condition (ii)). We consider this assumption to be reasonable. The assignment of telephone triage nurses to patients is as good as random since incoming calls are allocated to the next available nurse within call centers, and neither patients nor nurses are able to select whom they speak with.

We empirically examine conditional independence by testing whether the leave-out ED propensity Z_i^{ED} is correlated with patient hold-out characteristics H_i. Following standard balance checks in the examiner design literature, we first estimate each call's predicted ED recommendation probability, $\widehat{R}_i^{ED}$, by regressing ED recommendation indicator, R_i^{ED}, on baseline controls, X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), and hold-out controls, H_i (age, sex, marital status, race and ethnicity, period of service, prior healthcare utilization, and prior diagnose). Next, we test whether $\widehat{R}_i^{ED}$ is correlated with Z_i^{ED}, conditional on X_i^0, Z_i^{PC}, and Z_i^{Home}.

Figure 1 shows that the leave-out measure, Z_i^{ED}, is not meaningfully related to patient characteristics, as indicated by the flat dashed line. Appendix Figure B3 presents a multivariate regression of Z_i^{ED} on hold-out controls, H_i, conditional on X_i^0. The F-test fails to reject the null at the 10% significance level, suggesting that hold-out variables H_i do not jointly predict nurses' ED recommendation tendency. Likewise, Appendix Figures B4-B5 show that nurses' UC and PC propensities are not meaningfully related to hold-out characteristics.

4.4.3 Monotonicity

We assess the monotonicity assumptions in Condition 1(iii) (a)-(b) to identify interpretable average treatment effects in the presence of heterogeneous treatment effects. In our context, monotonicity requires that, for any pair of nurses j, j' with the same PC and Home propensities (z^{PC}, z^{Home}), if nurse j' has a higher ED propensity than nurse j ($z'^{ED} > z^{ED}$), then all calls that would be recommended ED by nurse j must be also recommended ED by nurse j' ($R_i^{ED}(z'^{ED}, z^{PC}, z^{Home}) \geq R_i^{ED}(z^{ED}, z^{PC}, z^{Home})$). This assumption may fail for several potential reasons. For instance, nurse j' may have a lower ED propensity than nurse j for younger patients ($R_i^{ED}(z'^{ED}, z^{PC}, z^{Home}) = 0 < 1 = R_i^{ED}(z^{ED}, z^{PC}, z^{Home}) \mid Age < 65$), even when nurse j' has a higher ED propensity than nurse j overall.

To empirically assess Condition 1(iii) (a)-(b), we implement a conventional test for (average) monotonicity by examining whether the first-stage coefficients are positive across subsamples defined by observable characteristics (Frandsen et al., 2023).

Specifically, we split the sample based on several characteristics of the patients and estimate the coefficients of the first stage separately. Appendix Tables B5-B7 show that the leave-out ED (/UC/PC) propensity is positively associated with ED (/UC/PC) recommendation for all subsamples. In Section 5.5, we further show that the potential response estimates for binary outcomes are bounded between 0 and 1, suggesting stronger support for monotonicity.[11]

Condition (iii) (c)-(d) implies that potential PC and Home recommendations would not change when calls were reassigned between nurses who differ in their relative ED-UC propensities ($z'^{ED} > z^{ED}$) but have the same propensities to recommend PC and Home ($z^{PC} = z'^{PC}$ and $z^{Home} = z'^{Home}$). This assumption has a testable implication: on average, the observable characteristics of the patients recommended PC or Home should be balanced between the two nurses.

Following Humphries et al. (2024), we empirically test this implication using predicted outcomes (ED, UC, and PC visits) constructed in the same manner as the balance check in Section 4.4.2. Specifically, we regress the predicted outcomes $\widehat{Y}_i$ on Z_i^{ED}, controlling for Z_i^{PC}, Z_i^{Home}, and X_i^0, while restricting the sample to patients who are recommended PC or Home:

$$\widehat{Y}_i = \eta_0 + \eta_1 Z_i^{ED} + \eta_2 Z_i^{PC} + \eta_3 Z_i^{Home} + X_i^{0\prime}\Gamma + e_i. \tag{7}$$

where η_1 captures correlations between the nurse's propensity to recommend ED (Z_i^{ED}) and patient characteristics. In Appendix Table B8, we find that the nurse's propensity to recommend ED is not meaningfully associated with patient characteristics when holding the nurse's propensities to recommend PC and Home constant. The estimates $\widehat{\eta}_1$ are small in magnitude and not statistically significant. Similarly,

[11] Chan et al. (2022) shows that this potential outcome test is stronger than the conventional test for (average) monotonicity, following Kitagawa (2015).

Appendix Tables B9 and B10 repeat this exercise for UC-PC and PC-Home margins, again confirming that the nurse's focal recommendation tendency is not associated with patient characteristics.

5 Results

This section presents TSLS estimates of margin-specific causal effects of triage recommendations on patient utilization, health, and costs. We present results in separate tables and figures by recommendation margins (ED-UC, UC-PC, PC-Home). All outcomes are cumulative from day 1 (triage call day) through day x ($x = 3, ..., 28$).

5.1 Utilization

Table 2 and Appendix Figure B6 present the TSLS estimates from equations (3) and (4), capturing margin-specific causal effects of being recommended ED over UC among compliers—those for whom nurses vary in their recommendation between ED and UC. The TSLS estimates show that the ED recommendation significantly shifts patients' utilization from UC to ED. Patients recommended ED care are 24.0 percentage points less likely to use UC and 25.3 percentage points more likely to use ED within three days of the call, compared to patients recommended UC. This makes them more than three times more likely to go to the ED as the average caller recommended UC care, and more than twice more likely as the overall sample mean. With the large increase in ED utilization, we observe a rise in emergency admissions through the ED. Patients recommended ED care are 1.6 percentage points more likely to experience an emergency admission than those recommended UC, a 133 percent increase.

The ED recommendation modestly shifts patient utilization beyond the corresponding care margin. Patients recommended ED are 2.7 percentage points (13.7 percent) more likely to go to primary care in three days, compared to patients recommended UC. We also find that the ED recommendation increases both the probability of receiving any acute care (UC or ED) and any form of professional care (PC, UC, or ED) modestly, by 2.5 percentage points (5.2 percent) and 4.1 percentage points (6.9 percent), respectively.

In terms of the timing of the effect, overall, the TSLS estimates suggest that the ED recommendation influences patient utilization only in the very short run on day 1 through day 3. The TSLS estimates for UC and ED utilization on day 28 (-24.6 p.p. and +24.2 p.p., respectively) do not differ much from the estimates on day 3 (-24.0 p.p. and +25.3 p.p., respectively). This implies that the changes in patient UC and ED utilization mostly occur on day 1 through day 3, and then there is no additional recommendation effect thereafter.[12] By contrast, the effect on any care utilization shrinks and becomes insignificant on day 28.

Table 3 and Appendix Figure B7 show estimates for patients on the middle acute margin between urgent care and primary care recommendations. Unlike the ED-UC margin, the estimated effects on UC and PC utilization are not of a similar magnitude. Patients recommended UC care are 20.7 percentage points more likely to go to UC and 11.5 percentage points less likely to use primary care within three days of the triage call. These patients are also 1.2 percentage points less likely to use the ED in this time period, which translates 13.8 percent compared to the average ED utilization rate of patients recommended PC. The UC recommendation modestly increases the probability of receiving any professional care by 4.8 percentage points (10.6 percent). We find no difference in the probability of a hospital admission

Table 4 and Appendix Figure B8 show estimates for patients on the least acute margin between PC and Home recommendations. The TSLS estimates suggest that the PC recommendation changes patient utilization almost exclusively on the corre-

[12]We confirm this by estimating the average potential outcomes under each counterfactual recommendation in Section 5.5.

sponding care margin of PC and Home. Patients recommended PC are 7.6 percentage points (27.8 percent) more likely to use primary care within three days, compared to patients recommended Home. The effects on any care are 7.6 percentage points on day 3 and 6.3 percentage points on day 28. The effects on urgent care, ED utilization, and hospital admission are small and statistically insignificant.

5.2 Mortality

For patients on the high acuity margin between ED and UC recommendations, although the ED recommendation substantially shifts patient utilization toward more intensive care, we do not find evidence that patients recommended ED experience any survival gain, compared to patients recommended UC. In Table 2 and Appendix Figure B6, we find that the effect of ED recommendation on 3-day mortality is +0.076 percentage points (SE = 0.041), a marginally significant positive estimate. The estimate for 28-day mortality is +0.164 percentage points (SE = 0.131) and is not statistically significant. In Appendix Table B12, we further investigated longer-term mortality up to one year. One year after the call, the mortality gap is +0.022 percentage points (SE = 0.407) between those recommended ED and UC.

Table 3 and Appendix Figure B7 show estimates for patients on the UC-PC recommendation margin. As discussed in Section 5.1, patients recommended UC are more likely to use UC and more likely to use professional care (extensive margin), compared to patients recommended PC. However, we do not find evidence that these intensive and extensive margin shifts in utilization significantly affect patient mortality. The estimate for 3-day mortality is -0.029 percentage points (SE = 0.021), and that for 28-day mortality is -0.097 percentage points (SE = 0.072). We do not find any significant mortality gap up to one year in Appendix Table B13.

Lastly, in Table 4, Appendix Table B14, and Appendix Figure B8, we do not find any significant difference in mortality between patients recommended PC and those recommended Home. The effects of PC recommendation on 3-day and 28-day mortality are +0.034 percentage points (SE = 0.028) and +0.122 (SE = 0.084), respectively.

5.3 Costs

In Table 2, we find that, on average, patients recommended ED care incur $110.7 higher outpatient costs than those recommended UC within three days. Although this net increase in outpatient costs can be driven by both the large shift in utilization from UC to ED and the modest shift from self-care to any professional outpatient care (PC, UC, or ED), Appendix Table B11a suggests that a large part of the increased costs is likely explained by the higher cost of an ED visit without admission compared to a UC visit. Additionally, Table 2 shows that patients recommended ED incurs $104.1 in additional inpatient costs within three days, compared to those recommended UC, consistent with a 1.6 percentage point higher hospital admission rate. Altogether, patients recommended ED care over UC care incur 22.9 percent ($404) higher cumulative costs after 28 days.

Interestingly, for patients on the UC-PC recommendation margin, Table 3 and Appendix Figure B7 show that patients recommended UC (more acute option) incur *lower* costs, compared to patients recommended PC (less acute option). Outpatient costs for patients recommended UC are $69.8 (30.1 percent) lower after three days. In Appendix Table B11b, we find that the UC recommendation decreases primary care costs (-$81.7) *more* than it increases urgent care costs (+$52.6), although the UC recommendation decreases primary care utilization (-11.5 p.p.) *less* than it increases urgent care utilization (+20.7 p.p.).[13]

[13]This finding can be explained by differences in the type and number of services provided. An urgent care visit tends to be more focused on addressing the chief complaint, while a primary care visit potentially addresses multiple concerns beyond the chief complaint. This difference can make an urgent care visit less costly than a primary care visit.

For patients on the PC-Home recommendation margin, we do not find significant differences in costs between patients recommended PC and those recommended Home in Table 4 and Appendix Figure B8. 28 days after the call, patients recommended PC incur only \$36 higher total costs than patients recommended Home, and the effect is statistically insignificant. The null effects on costs are consistent with our finding that the PC recommendation only has a small impact on the probability of having a primary care visit.

5.4 Robustness Checks

We examine the sensitivity of the TSLS estimates to the choice of control variables. Although our balance check shows that our leave-out measure is uncorrelated with patients' predetermined observable characteristics after controlling for baseline controls X_i^0, it may still be correlated with patients' potential outcomes through unobservable factors. To assess the severity of this threat to conditional independence, we test the sensitivity of our TSLS estimates to (i) the exclusion of hold-out controls from the conditioning set[14] and (ii) the inclusion of several symptomatic measures recorded at the time of the call. Appendix Table B15 reports estimates from the specification that controls only for the baseline control X_i^0. Appendix Table B16 presents TSLS estimates that additionally control for pain scale (0-10) and duration of symptoms (in 10 bins). Lastly, Appendix Table B17 shows TSLS results from the specification that further controls for chief complaint fixed effects. The TSLS estimates remain generally stable across different sets of controls.

Appendix Section B3 further examines the robustness of the TSLS estimates to alternative leave-out measures. Again, we confirm that our findings are not sensitive

[14] Recall that, in addition to baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), our preferred specification also controls for hold-out controls H_i (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for statistical precision as shown in equation (5).

to how we construct these measures.[15]

5.5 Evaluation of Mean Potential Outcomes

While the TSLS estimates recover margin-specific LATEs among compliers, these causal effect parameters do not provide insight into how the levels of potential outcomes evolve over time under each of the counterfactual recommendations. In Figures 2-4, we further illustrate evolutions of the average potential outcomes for compliers when recommended the more intensive option for care (black circles, ED when on ED-UC margin, UC when on UC-PC margin, and PC when on PC-Home margin) and when recommended the less intensive option (gray circles, UC when on ED-UC margin, PC when on UC-PC margin, and Home when on PC-Home margin).[16] Note that, in each panel, the gaps in estimated potential outcomes between the two counterfactual recommendations are numerically equivalent to our TSLS estimates.

As discussed in Section 5.1, for patients on ED-UC recommendation margin, our TSLS estimates suggest that the ED recommendation shifts patient utilization from UC to ED immediately after the call (day 1 through day 3) and then the recommendation has no additional effect thereafter. Figure 2 confirms this finding: the average potential ED (UC) utilization rates under the two counterfactual recommendations run in parallel from day 3 to day 28, with an almost constant difference (25 percentage points). By contrast, the differences in potential utilization of any care (PC, UC, or ED) between the two counterfactuals shrink over time, suggesting that the ED recommendation may modestly shift the timing of having any care. We find similar patterns in potential utilization for patients on UC-PC recommendation margin: po-

[15] In results not shown, we complete all robustness exercises on the UC-PC and PC-Home margins, again with little change in the size or precision of the estimates.

[16] We estimate $E\left[Y_i(r) \mid i \in \text{Complier}, X_i^0\right]$ for $r \in \{ED, UC\}$, following Abadie (2002). For $r = ED$, we regress $Y_i \cdot R_i^{ED}$ on R_i^{ED}, where the right-hand side R_i^{ED} is instrumented by the leave-out measure Z_i^{ED}. The TSLS estimate for the right-hand side R_i^{ED} identifies the average potential outcomes of compliers. For $r = UC$, we run the same regression after replacing R_i^{ED} with $1 - R_i^{ED}$. See Appendix Section B4 for details.

tential PC (UC/ED/UC or ED) utilization rates of the two counterfactuals run in parallel, and the gap in potential utilization of any care decreases over time.

Focusing on the levels of potential utilization rates, we find that patient healthcare take-up is generally low. On ED-UC recommendation margin, the potential ED utilization rate is only 43.6%,[17] and the potential utilization of any care (PC, UC, or ED) is only 59.8% within three days when the patient is recommended ED (Figure 2). 28 days after the call, the potential utilization of any care is around 80%, which means that one in five patients does not receive any professional care for four weeks after the call, even though they are recommended the most intensive form of care (ED). Similarly, on UC-PC recommendation margin, over one in five patients do not receive any professional care for four weeks after the call, even when recommended UC (Figure 3). With this high level of patient non-utilization as a benchmark, our TSLS estimates suggest that recommending more intensive care over less intensive care has a substantial impact on patient choices.

Methodologically, we note that, under the IV condition, the average potential responses for binary outcomes must be bounded between 0 and 1 (Chan et al., 2022; Kitagawa, 2015).[18] The estimated potential outcomes in Figures 2-4 are consistent with this implication, providing stronger support for the validity of our IV strategy.

5.6 Average Characteristics of Compliers

While the LATEs represent causal effects among compliers to whom different nurses would give a different recommendation, the underlying complier population differs between the three margins (ED-UC, UC-PC, PC-Home) and from the overall sample in observable and unobservable ways. To better interpret our TSLS estimates, we

[17]In contrast, Tran et al. (2017) reports that 68.6% of the patients recommended ED show up ED in an Australian telephone triage system.

[18]Chan et al. (2022) uses a bound between -1 and 0 for this test, as they run a TSLS regression of $outcome \cdot (1 - treatment)$ on $treatment$. In contrast, our potential outcomes are estimated by running a TSLS regression of $outcome \cdot treatment$ on $treatment$.

profile observable characteristics of our complier population on each margin.[19]

Table 5 contrasts the average characteristics of the compliers on the three margins and those of the overall sample. The estimated averages of prior diagnoses (Elixhauser comorbidity counts, max = 31) are 3.003 (ED-UC), 2.742 (UC-PC), 2.628 (PC-Home) (overall mean = 2.901). The estimates of complier age are 63.043 (ED-UC), 60.772 (UC-PC), and 60.477 (PC-Home) (overall mean = 63.054). Similarly, Appendix Tables B20-B23 further examine each of the 31 prior diagnoses. Again, we find that, overall, the prevalence of comorbidity decreases along the recommendation margins: highest on the ED–UC margin, lower on the UC–PC margin, and lowest on the PC–Home margin.

As described in Section 4.2, we derive the margin-specific causal interpretation of the TSLS estimates under the assumption that nurses select a recommendation based on patient acuity (ED for the highest acuity, UC for upper middle, PC for lower middle, and Home for the lowest), forming the high (ED-UC), middle (UC-PC), and low (PC-Home) acuity margins. The observed patterns of the average complier age and prior diagnoses – important correlates of patient acuity – are consistent with this framework.

6 Conclusion

This paper studies how triage recommendations affect patient utilization, healthcare costs, and health outcomes, exploiting unique data from the telephone triage system for US veterans. Extending the recent methodological developments in the examiners design literature into a healthcare setting, we estimate margin-specific causal effects of nurse recommendations on three acuity margins (ED-UC, UC-PC, PC-Home) using

[19] Following Abadie (2002), we estimate $E\left[h_{ki} \mid i \in \text{Complier}, X_i^0\right]$ for some hold-out characteristic h_{ki} by regressing $h_{ki} \cdot R_i^{ED}$ on R_i^{ED} with the right-hand side R_i^{ED} being instrumented by the leave-out measure Z_i^{ED}. See Appendix Section B5 for details.

a focal recommendation propensity as an instrument for the focal recommendation, while controlling for non-focal recommendation propensities.

Our results have three key takeaways. First, nurse recommendations are highly influential in determining and shifting patient choices upstream. Recommending more acute care over less acute care (e.g., ED over UC) substantially influences patient utilization on the corresponding margins, while also modestly influencing patient use of professional care on the extensive margin. Second, these induced shifts in care settings impact healthcare costs, while we do not find any evidence that these shifts result in survival gain. We find that recommending UC results in lower costs on both high (ED-UC) and middle (UC-PC) acuity margins, suggesting that for marginal patients, urgent care is cost-saving in addressing acute healthcare needs. Third, we find the high prevalence of non-utilization of professional care among the callers. With the high non-utilization as a benchmark, our results suggest that adjusting triage recommendations can play an important role in encouraging patients to see a healthcare provider.

As for real-world costs of human variability, the potential gain from eliminating nurse variation is likely modest. For concreteness, we perform a back-of-the-envelope calculation when re-assigning a patient on ED-UC margin to a nurse with a standard deviation lower ED tendency (SD = 7.3 p.p.). This patient is more likely to be recommended UC over ED by 6.3 percentage points (0.86 (FS) $\times$ 0.073). This change reduces total costs by \$25.5 (\$404 $\times$ 0.063) (1.4%). The implied cost reduction is likely a modest sum, especially compared to the potential costs of aligning nurse behavior.

There are some issues that remain for future research. First, our study does not compare patients who go through telephone triage and those who do not. To have a complete picture of telephone triage, future research will need to examine how telephone triage and the lack thereof influence patient outcomes. Second, our design does not allow us to study patients to whom the nurses do not vary in their recommendations. These subsets include, for example, patients with very high acuity to whom all nurses would unanimously recommend ED care. Triage and subsequent care likely play an important role in determining outcomes of these patients. Alternative research designs and identifying variations are required to study these patients. Lastly, since triage nurse assignment does not fully determine care in our setting, our design does not allow us to directly estimate the effect of receiving actual care (e.g., ED care). Future research may enrich the examiners design literature by exploring identification conditions for settings where the first decision-maker affects the second decision-maker's choice.

References

Abadie, Alberto, "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 2002, *97* (457), 284–292.

_ , "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, April 2003, *113* (2), 231–263.

Abualenain, Jameel, William J. Frohna, Robert Shesser, Ru Ding, Mark Smith, and Jesse M. Pines, "Emergency Department Physician-Level and Hospital-Level Variation in Admission Rates," *Annals of Emergency Medicine*, June 2013, *61* (6), 638–643.

Alexander, Diane, Janet Currie, and Molly Schnell, "Check up before you check out: Retail clinics and emergency room use," *Journal of Public Economics*, October 2019, *178*, 104050.

Allen, Lindsay, Janet R. Cummings, and Jason M. Hockenberry, "The impact of urgent care centers on nonemergent emergency department visits," *Health Services Research*, August 2021, *56* (4), 721–730.

Arteaga, Carolina, "Parental Incarceration and Children's Educational Attainment," *The Review of Economics and Statistics*, November 2023, *105* (6), 1394–1410.

Chan, David C. and Jonathan Gruber, "Provider Discretion and Variation in Resource Allocation: The Case of Triage Decisions," *AEA Papers and Proceedings*, May 2020, *110*, 279–283.

_ **, David Card, and Lowell Taylor**, "Is There a VA Advantage? Evidence from Dually Eligible Veterans," *American Economic Review*, 2023, *113* (11), 3003–43.

Chan, David C, Matthew Gentzkow, and Chuan Yu, "Selection with Variation in Diagnostic Skill: Evidence from Radiologists," *The Quarterly Journal of Economics*, May 2022, *137* (2), 729–783.

Chyn, Eric, Brigham Frandsen, and Emily C Leslie, "Examiner and Judge Designs in Economics: A Practitioner's Guide," *National Bureau of Economic Research Working Paper Series*, 2024, *No. 32348.*

Coussens, Stephen and Dan P. Ly, "Variation in Emergency Department Physician Admitting Practices and Subsequent Mortality," *JAMA Internal Medicine*, February 2025, *185* (2), 153–160.

Currie, Janet, Anastasia Karpova, and Dan Zeltzer, "Do urgent care centers reduce Medicare spending?," *Journal of Health Economics*, May 2023, *89*, 102753.

Ferro, Simone and Chiara Serra, "Triage at shift changes and distortions in the perception and treatment of emergency patients," *Journal of Health Economics*, January 2025, *99*, 102944.

Flores-Mateo, Gemma, Concepción Violan-Fors, Paloma Carrillo-Santisteve, Salvador Peiró, and Josep-Maria Argimon, "Effectiveness of Organizational Interventions to Reduce Emergency Department Utilization: A Systematic Review," *PLOS ONE*, May 2012, *7* (5), e35903. Publisher: Public Library of Science.

Frandsen, Brigham, Lars Lefgren, and Emily Leslie, "Judging Judge Fixed Effects," *American Economic Review*, January 2023, *113* (1), 253–77.

Heckman, James J. and Rodrigo Pinto, "Unordered Monotonicity," *Econometrica*, 2018, *86* (1), 1–35. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA13777.

Humphries, John Eric, Aurelie Ouss, Kamelia Stavreva, Megan T. Stevenson, and Winnie van Dijk, "Conviction, Incarceration, and Recidivism: Understanding the Revolving Door," August 2024.

Imbens, Guido W. and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica : journal of the Econometric Society*, 1994, *62* (2), 467–475.

Islam, Farah, Marc Sabbe, Pieter Heeren, and Koen Milisen, "Consistency of decision support software-integrated telephone triage and associated factors: a systematic review," *BMC Medical Informatics and Decision Making*, March 2021, *21* (1), 107.

Kamat, Vishal, Samuel Norris, and Matthew Pecenco, "Conviction, Incarceration, and Policy Effects in the Criminal Justice System," *SSRN Electronic Journal*, 2024.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad, "Field of Study, Earnings, and Self-Selection*," *The Quarterly Journal of Economics*, August 2016, *131* (3), 1057–1111.

Kitagawa, Toru, "A TEST FOR INSTRUMENT VALIDITY," *Econometrica*, 2015, *83* (5), 2043–2063. Publisher: [Wiley, The Econometric Society].

Kline, Patrick and Christopher R. Walters, "Evaluating Public Programs with Close Substitutes: The Case of Head Start*," *The Quarterly Journal of Economics*, November 2016, *131* (4), 1795–1848.

Molitor, David, "The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration," *American Economic Journal: Economic Policy*, February 2018, *10* (1), 326–356.

Mueller-Smith, Michael, "THE CRIMINAL AND LABOR MARKET IMPACTS OF INCARCERATION," 2015, p. 59.

Pinto, Rodrigo, "Beyond Intention-to-Treat: Using the Incentives of Moving to Opportunity to Identify Neighborhood Effects," 2022.

Raven, Maria C., Margot Kushel, Michelle J. Ko, Joanne Penko, and Andrew B. Bindman, "The Effectiveness of Emergency Department Visit Reduction Programs: A Systematic Review," *Annals of Emergency Medicine*, October 2016, *68* (4), 467–483.e15.

Retchin, Sheldon M., Sheryl L. Garland, and Emmanuel A. Anum, "The transfer of uninsured patients from academic to community primary care settings," *The American Journal of Managed Care*, April 2009, *15* (4), 245–252.

Sadowski, Laura S., Romina A. Kee, Tyler J. VanderWeele, and David Buchanan, "Effect of a Housing and Case Management Program on Emergency Department Visits and Hospitalizations Among Chronically Ill Homeless Adults: A Randomized Trial," *JAMA*, May 2009, *301* (17), 1771–1778.

Sexton, Vanashree, Jeremy Dale, Carol Bryce, James Barry, Elizabeth Sellers, and Helen Atherton, "Service use, clinical outcomes and user experience associated with urgent care services that use telephone-based digital triage: a systematic review," *BMJ Open*, January 2022, *12* (1), e051569.

Sigstad, Henrik, "Monotonicity among Judges: Evidence from Judicial Panels and Consequences for Judge IV Designs," *Working Paper*, 2023.

Silver, David and Jonathan Zhang, "Invisible Wounds: Health and Well-Being Impacts of Mental Disorder Disability Compensation on Veterans," *National Bureau of Economic Research Working Paper Series*, 2022, *No. 29877.*

Soril, Lesley J. J., Laura E. Leggett, Diane L. Lorenzetti, Tom W. Noseworthy, and Fiona M. Clement, "Reducing Frequent Visits to the Emergency Department: A Systematic Review of Interventions," *PLOS ONE*, April 2015, *10* (4), e0123660. Publisher: Public Library of Science.

Tran, Duong Thuy, Amy Gibson, Deborah Randall, Alys Havard, Mary Byrne, Maureen Robinson, Anthony Lawler, and Louisa R. Jorm, "Compliance with telephone triage advice among adults aged 45 years and older: an Australian data linkage study," *BMC Health Services Research*, August 2017, *17* (1), 512.

Vashi, Anita A., Tracy Urech, Brendan Carr, Liberty Greene, Theodore Warsavage Jr, Renee Hsia, and Steven M. Asch, "Identification of Emergency Care–Sensitive Conditions and Characteristics of Emergency Department Utilization," *JAMA Network Open*, August 2019, *2* (8), e198642.

Wang, Bill, Ateev Mehrotra, and Ari B. Friedman, "Urgent Care Centers Deter Some Emergency Department Visits But, On Net, Increase Spending," *Health Affairs*, April 2021, *40* (4), 587–595. Publisher: Health Affairs.

Weinick, Robin M., Rachel M. Burns, and Ateev Mehrotra, "Many Emergency Department Visits Could Be Managed At Urgent Care Centers And Retail Clinics," *Health Affairs*, September 2010, *29* (9), 1630–1636. Publisher: Health Affairs.

7 Tables and Figures

Table 1. Characteristics of Baseline Sample

Variable	Mean
Age	63.054
Male	0.863
Married	0.513
White	0.691
Black	0.182
Hispanic	0.061
Rural County	0.206
Comorbidity Count	2.901
ED Visit in Prior Year	0.382
Admission in Prior Year	0.102
Inpatient in Prior Year	0.101
Primary Care in Prior Year	0.974
Primary Care Visit 3d	0.307
Urgent Care Visit 3d	0.063
ED Visit 3d	0.192
Admission 3d	0.027
Any Care 3d	0.499
Inpatient Cost 3d	153.232
Outpatient Cost 3d	286.668
Total Cost 3d	439.900
Nurse Recommended PC	0.627
Nurse Recommended UC	0.044
Nurse Recommended ED	0.277
Algorithm Recommended PC	0.667
Algorithm Recommended UC	0.004
Algorithm Recommended ED	0.263
Calls	1,273,843
Patients	836,420
Nurses	1,725
Call Centers	96

Notes: This table presents characteristics of calls in the baseline sample.

Table 2. IV Results: Effect of Nurse ED vs. UC Recommendation

	Primary Care	Urgent Care	ED	UC or ED	Any Care	Hospital Admission	Mortality	Outpatient Cost	Inpatient Cost	Total Cost
Panel A: 3-Day Outcomes										
Nurse Recommends ED	2.662	-24.025	25.342	2.506	4.137	1.603	0.076	110.745	104.120	214.865
	(0.859)	(0.637)	(0.816)	(0.965)	(0.984)	(0.355)	(0.041)	(35.548)	(38.501)	(62.305)
Outcome Mean (UC)	19.588	38.646	11.392	48.278	58.877	1.214	0.028	227.906	93.061	320.967
Outcome Mean (All)	30.672	6.311	19.187	25.159	49.857	2.671	0.041	290.885	157.584	448.469
Panel B: 28-Day Outcomes										
Nurse Recommends ED	1.267	-24.587	24.203	2.196	0.930	2.198	0.164	160.101	243.998	404.099
	(0.966)	(0.669)	(0.863)	(0.979)	(0.839)	(0.450)	(0.131)	(82.975)	(132.806)	(176.258)
Outcome Mean (UC)	52.305	42.171	18.192	55.688	77.640	2.788	0.276	1217.306	546.884	1764.190
Outcome Mean (All)	63.081	7.977	25.912	32.782	74.960	4.668	0.416	1406.952	855.097	2262.050

Notes: This table reports TSLS estimates of the effect of ED recommendation. All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. In the third row of each panel, the sample average for each outcome conditional on UC recommendation is reported as a benchmark ($E[Y_i \mid R_i = UC]$). In the fourth row, the overall (unconditional) sample average is also reported ($E[Y_i]$). In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. Standard errors are clustered at the call center-by-call time level.

Table 3. IV Results: Effect of Nurse UC vs. PC Recommendation

	Primary Care	Urgent Care	ED	UC or ED	Any Care	Hospital Admission	Mortality	Outpatient Cost	Inpatient Cost	Total Cost
Panel A: 3-Day Outcomes										
Nurse Recommends UC	-11.484	20.693	-1.225	18.634	4.812	-0.091	-0.029	-69.782	-23.403	-93.185
	(0.544)	(0.464)	(0.452)	(0.589)	(0.622)	(0.178)	(0.021)	(19.522)	(22.351)	(35.774)
Outcome Mean (PC)	35.228	5.306	8.677	13.746	45.423	0.973	0.022	225.842	52.582	278.425
Outcome Mean (All)	30.672	6.311	19.187	25.159	49.857	2.671	0.041	290.885	157.584	448.469
Panel B: 28-Day Outcomes										
Nurse Recommends UC	-7.884	21.009	-1.348	17.447	2.140	-0.382	-0.097	-137.188	-109.740	-246.929
	(0.604)	(0.482)	(0.516)	(0.608)	(0.538)	(0.243)	(0.072)	(48.074)	(72.106)	(99.177)
Outcome Mean (PC)	65.879	6.956	15.602	21.731	73.368	2.656	0.267	1261.896	530.963	1792.860
Outcome Mean (All)	63.081	7.977	25.912	32.782	74.960	4.668	0.416	1406.952	855.097	2262.050

Notes: This table reports TSLS estimates of the effect of UC recommendation. All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. In the third row of each panel, the sample average for each outcome conditional on PC recommendation is reported as a benchmark ($E[Y_i \mid R_i = PC]$). In the fourth row, the overall (unconditional) sample average is also reported ($E[Y_i]$). In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. Standard errors are clustered at the call center-by-call time level.

Table 4. IV Results: Effect of Nurse PC vs. Home Recommendation

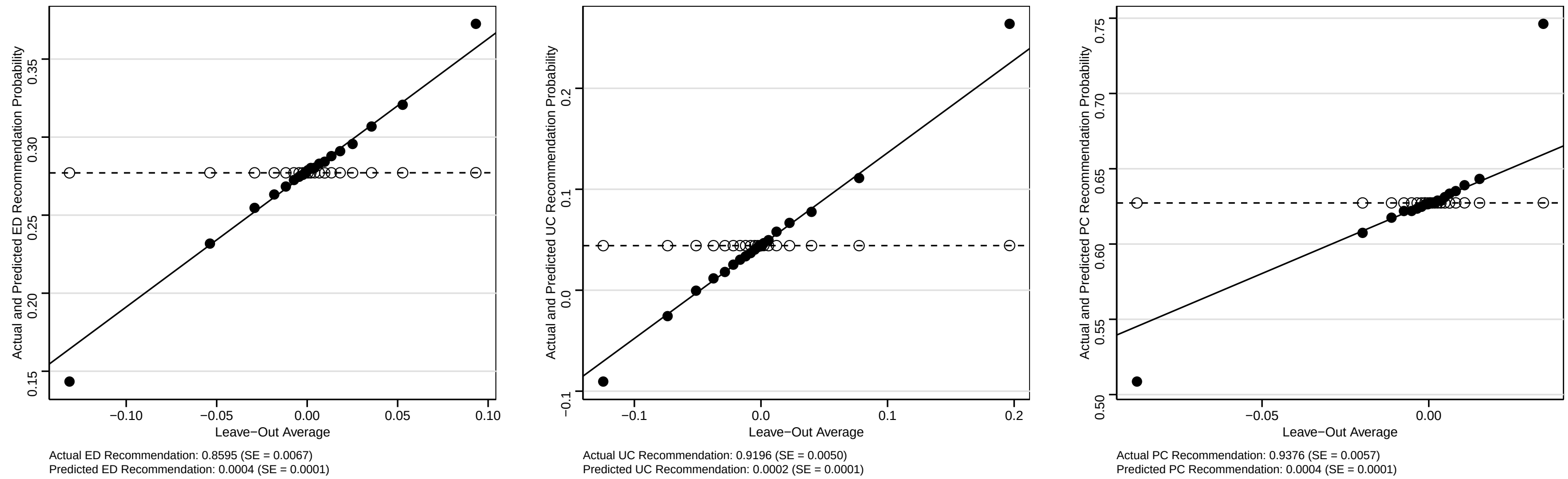
	Primary Care	Urgent Care	ED	UC or ED	Any Care	Hospital Admission	Mortality	Outpatient Cost	Inpatient Cost	Total Cost
Panel A: 3-Day Outcomes										
Nurse Recommends PC	7.598	-0.249	0.738	0.414	7.627	0.072	0.034	16.921	-5.174	11.747
	(0.718)	(0.348)	(0.398)	(0.508)	(0.769)	(0.131)	(0.028)	(11.940)	(10.491)	(19.226)
Outcome Mean (Home)	27.277	3.447	4.937	8.276	33.401	0.536	0.017	164.209	26.862	191.072
Outcome Mean (All)	30.672	6.311	19.187	25.159	49.857	2.671	0.041	290.885	157.584	448.469
Panel B: 28-Day Outcomes										
Nurse Recommends PC	6.443	-0.398	0.265	-0.007	6.259	0.144	0.122	-9.933	46.234	36.301
	(0.778)	(0.398)	(0.531)	(0.621)	(0.744)	(0.232)	(0.084)	(46.162)	(67.379)	(93.704)
Outcome Mean (Home)	56.330	5.163	11.749	16.352	62.204	2.130	0.279	1100.355	463.917	1564.272
Outcome Mean (All)	63.081	7.977	25.912	32.782	74.960	4.668	0.416	1406.952	855.097	2262.050

Notes: This table reports TSLS estimates of the effect of PC recommendation. All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. In the third row of each panel, the sample average for each outcome conditional on Home recommendation is reported as a benchmark ($E\left[Y_i \mid R_i = Home\right]$). In the fourth row, the overall (unconditional) sample average is also reported ($E\left[Y_i\right]$). In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. Standard errors are clustered at the call center-by-call time level.

Table 5. Complier Characteristics

		Compliers (ED-UC)		Compliers (UC-PC)		Compliers (PC-Home)	
Variable	Overall Mean	Mean	Ratio	Mean	Ratio	Mean	Ratio
Comorbidity Count	2.901	3.003	1.035	2.742	0.945	2.628	0.906
		(0.025)	[1.018 - 1.052]	(0.013)	[0.936 - 0.954]	(0.018)	[0.894 - 0.918]
Age	63.054	63.646	1.009	60.772	0.964	60.477	0.959
		(0.183)	[1.004 - 1.015]	(0.118)	[0.960 - 0.967]	(0.162)	[0.954 - 0.964]
Male	0.863	0.868	1.005	0.828	0.959	0.850	0.985
		(0.004)	[0.996 - 1.014]	(0.003)	[0.953 - 0.966]	(0.004)	[0.977 - 0.993]
Married	0.513	0.514	1.003	0.505	0.986	0.501	0.977
		(0.006)	[0.979 - 1.026]	(0.004)	[0.971 - 1.000]	(0.005)	[0.958 - 0.996]
White	0.691	0.679	0.982	0.607	0.878	0.655	0.948
		(0.006)	[0.966 - 0.999]	(0.004)	[0.866 - 0.890]	(0.005)	[0.933 - 0.962]
Black	0.182	0.181	0.993	0.239	1.312	0.190	1.043
		(0.005)	[0.941 - 1.044]	(0.004)	[1.272 - 1.353]	(0.004)	[0.999 - 1.087]
Hispanic	0.061	0.073	1.196	0.086	1.424	0.087	1.441
		(0.003)	[1.092 - 1.301]	(0.003)	[1.336 - 1.512]	(0.003)	[1.335 - 1.547]

Notes: This table presents the average characteristics among compliers, along with the overall sample averages. Following Abadie (2003), we estimate $E[H_i \mid \text{Compliers}]$ by regressing an interaction between each patient characteristic H and ED recommendation indicator $(H \cdot R^{ED})$ on ED recommendation indicator (R^{ED}) with the right-hand-side R^{ED} instrumented by the leave-out ED recommendation propensity on ED-UC margin. The average characteristics among compliers on UC-PC and PC-Home margins are estimated similarly. All regressions control for the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county) described in Appendix Table B3. Standard errors are clustered at the call center-by-call time level.



(a) ED - UC   (b) UC - PC   (c) PC - Home

Figure 1. First Stage and Balance

Notes: The binned scatter plot in Panel (a) represents the first-stage regression in equation (4) and the corresponding balance regression that replaces actual ED recommendation with predicted ED recommendation. The bin averages of the actual ED recommendation R_i^{ED} are shown in solid circles, whereas the bin averages of the predicted ED recommendation $\widehat{R}_i^{ED}$ are shown in hollow circles. Both R_i^{ED} and $\widehat{R}_i^{ED}$ are residualized for the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), and are centered at the average ED recommendation rate (27.7%) on the y-axis. Those baseline controls are described in Appendix Table B3. Panels (b) and (c) represent the corresponding analysis on UC-PC and PC-Home margins.

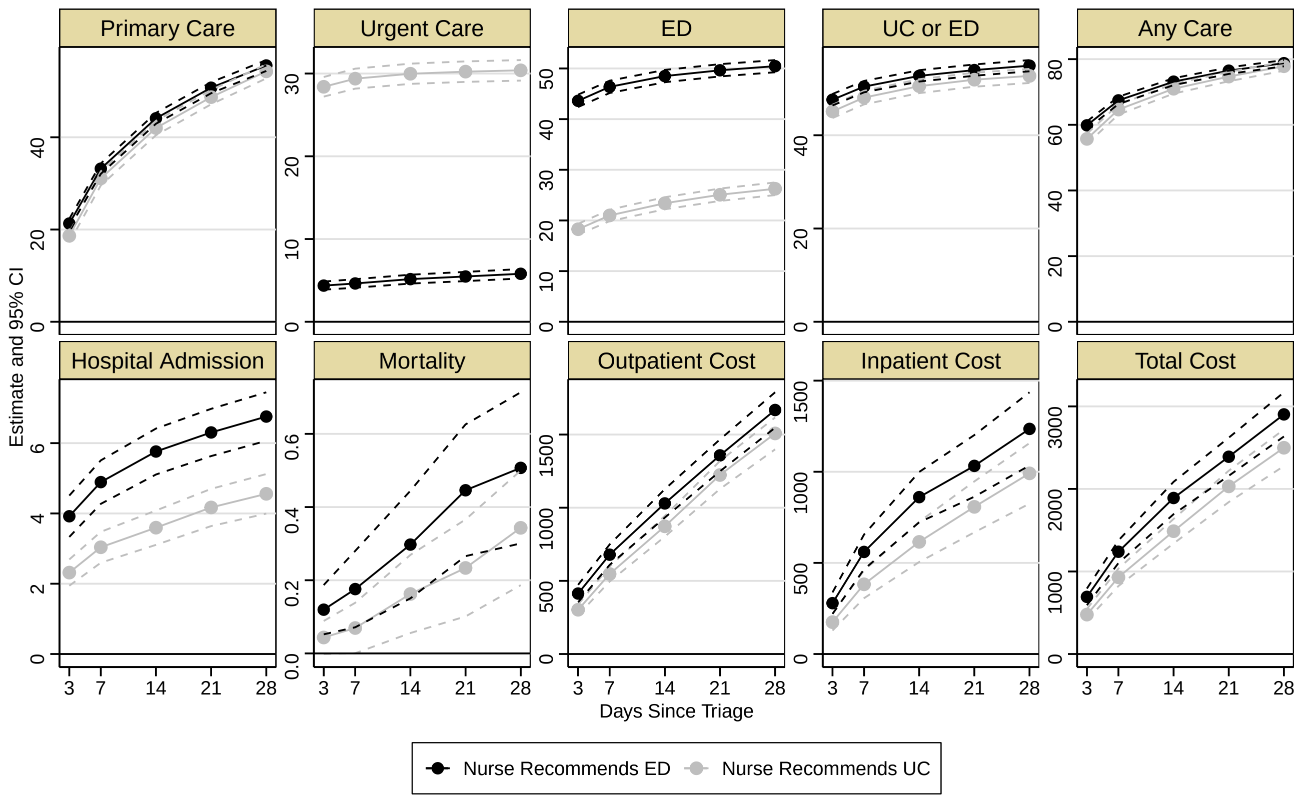


Figure 2. Average Potential Outcomes among Compliers (ED - UC)

Notes: This figure presents the averages of potential outcomes among compliers. The black circles are the counterfactuals under ED recommendation ($E[Y(ED) \mid \text{Compliers}]$), while the gray circles are under UC recommendation ($E[Y(UC) \mid \text{Compliers}]$). All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. Following Abadie (2003), we estimate $E[Y(ED) \mid \text{Compliers}]$ by regressing an interaction between each outcome and ED recommendation indicator ($Y \cdot R^{ED}$) on ED recommendation indicator (R^{ED}) with the right-hand-side R^{ED} instrumented by the leave-out measure. $E[Y(UC) \mid \text{Compliers}]$ is similarly estimated by replacing R^{ED} with $1 - R^{ED}$. In addition to the baseline controls X_i^0, all regressions control for the hold-out variables. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. The dashed lines show the 95% confidence intervals based on standard errors clustered at the call center-by-call time level.

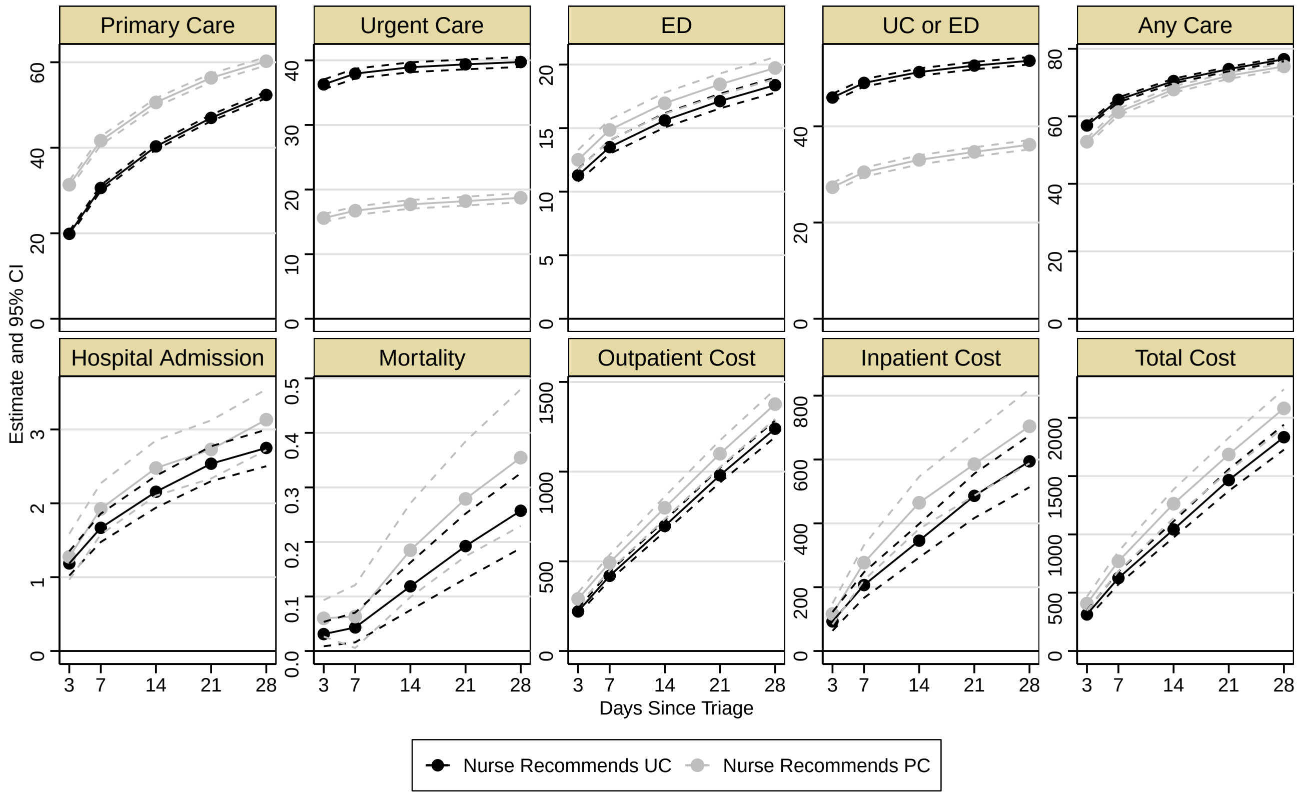


Figure 3. Average Potential Outcomes among Compliers (UC - PC)

Notes: This figure presents the averages of potential outcomes among compliers. The black circles are the counterfactuals under UC recommendation ($E[Y(UC) \mid \text{Compliers}]$), while the gray circles are under PC recommendation ($E[Y(PC) \mid \text{Compliers}]$). All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. Following Abadie (2003), we estimate $E[Y(UC) \mid \text{Compliers}]$ by regressing an interaction between each outcome and UC recommendation indicator ($Y \cdot R^{UC}$) on UC recommendation indicator (R^{UC}) with the right-hand-side R^{UC} instrumented by the leave-out measure. $E[Y(PC) \mid \text{Compliers}]$ is similarly estimated by replacing R^{UC} with $1 - R^{UC}$. In addition to the baseline controls X_i^0, all regressions control for the hold-out variables. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. The dashed lines show the 95% confidence intervals based on standard errors clustered at the call center-by-call time level.

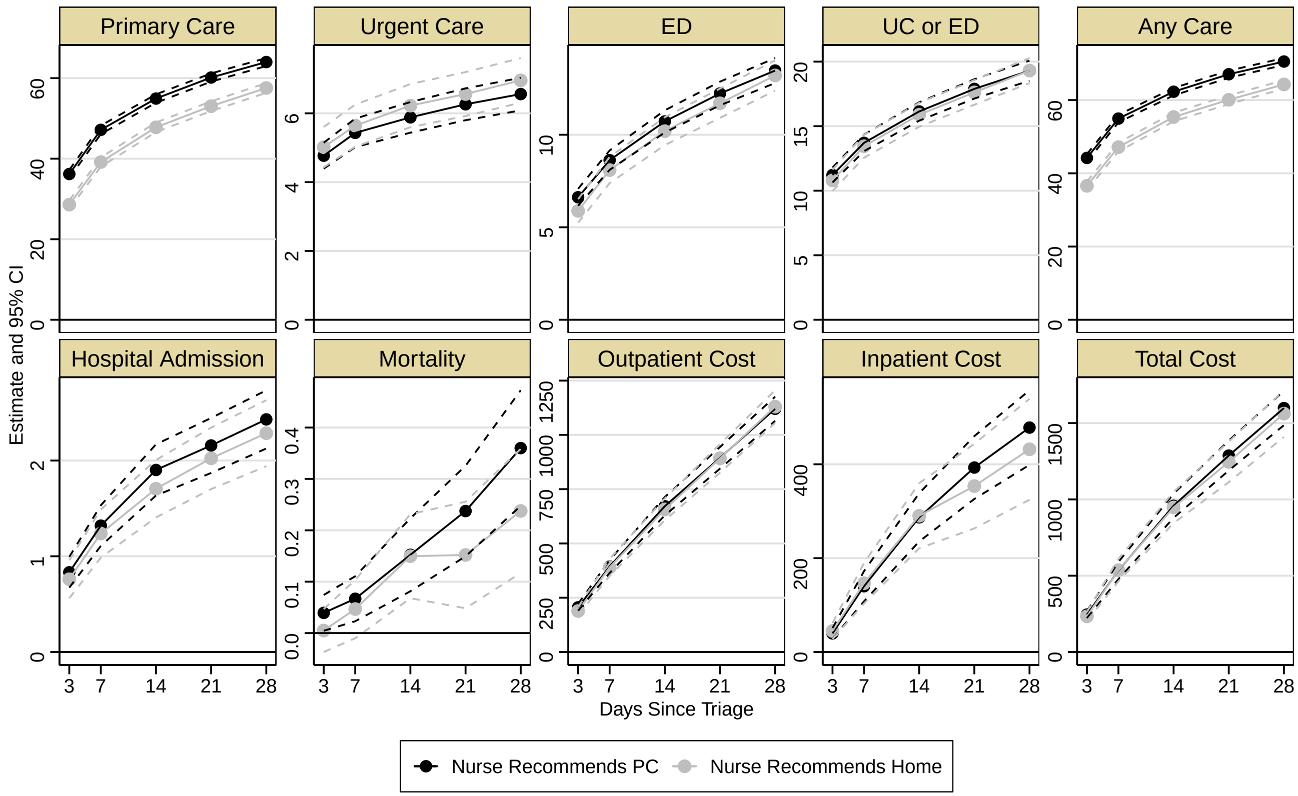


Figure 4. Average Potential Outcomes among Compliers (PC - Home)

Notes: This figure presents the averages of potential outcomes among compliers. The black circles are the counterfactuals under UC recommendation ($E[Y(PC) \mid \text{Compliers}]$), while the gray circles are under PC recommendation ($E[Y(Home) \mid \text{Compliers}]$). All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. Following Abadie (2003), we estimate $E[Y(PC) \mid \text{Compliers}]$ by regressing an interaction between each outcome and PC recommendation indicator ($Y \cdot R^{PC}$) on PC recommendation indicator (R^{PC}) with the right-hand-side R^{PC} instrumented by the leave-out measure. $E[Y(Home) \mid \text{Compliers}]$ is similarly estimated by replacing R^{PC} with $1 - R^{PC}$. In addition to the baseline controls X_i^0, all regressions control for the hold-out variables. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. The dashed lines show the 95% confidence intervals based on standard errors clustered at the call center-by-call time level.

Supplemental Appendix A: Behavioral Model of Nurse Triage

A1 Ordered Choice Model

This section describes behavioral models that satisfy the unordered pairwise monotonicity (UPM) assumption (Condition 1(iii)). One such model is an ordered choice framework in which all examiners make decisions based on a one-dimensional unobserved heterogeneity (Humphries et al., 2024). In our telephone triage context, the ordered choice model assumes that (i) all nurses select a triage recommendation (Home, PC, UC, or ED) based on a single latent index (i.e., the acuity of patient health conditions), (ii) rank patients in the same manner from the lowest to the highest level of acuity, and (iii) recommend ED care for those with the highest acuity, UC for those with upper middle acuity, UC for those with lower middle acuity, and Home for those with the lowest acuity. The only reason nurses may select a different recommendation for the same patient is variations in their preferences for the most appropriate level of care for a given level of acuity.

Appendix Figure A0a illustrates an example in which two arbitrary nurses $j \in \{1, 2\}$ select one of four follow-up care recommendations $R(j) \in \{Home, PC, UC, ED\}$. The two nurses face the same acuity distribution, normalized to $W_i \sim Unif(0, 1) \mid X_i^0$. Their preferences are represented by thresholds π_j^r, $r \in \{PC, UC, ED\}$, $j \in \{1, 2\}$. Nurse j's propensity to recommend care r is denoted by $Z_j^r = Pr\left[R(j) = r\right]$. They select a recommendation as follows:

$$R_i(j) = \begin{cases} ED, & \text{if } W_i \geq \pi_j^{ED} \\ UC, & \text{if } \pi_j^{UC} \leq W_i < \pi_j^{ED} \\ PC, & \text{if } \pi_j^{PC} \leq W_i < \pi_j^{UC} \\ Home, & \text{if } W_i < \pi_j^{PC}. \end{cases} \tag{A1}$$

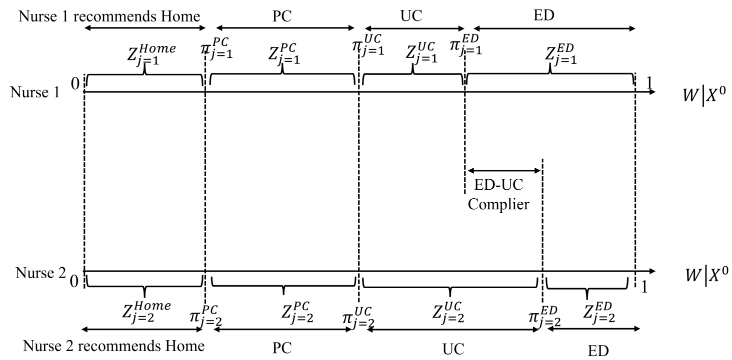
This ordered choice model allows us to identify margin-specific treatment effects on ED-UC margin. In Appendix Figure A0a, the two nurses $j = 1, 2$ have the same threshold for UC-PC and PC-Home margins ($\pi^{UC}_{j=1} = \pi^{UC}_{j=2}$ and $\pi^{PC}_{j=1} = \pi^{PC}_{j=2}$) but differ in their thresholds for ED-UC margin ($\pi^{ED}_{j=1} < \pi^{ED}_{j=2}$). This implies that nurse 1 has a higher propensity to recommend ED (and a lower propensity to recommend UC) than nurse 2, while both nurses exhibit the same propensity to recommend PC and Home. A hypothetical reassignment from nurse 1 to nurse 2 affects only patients with acuity $W_i \in \left[\pi^{ED}_{j=1}, \pi^{ED}_{j=2}\right]$, shifting their recommendation from ED to UC (ED-UC compliers). This motivates our identification strategy that compares nurses with the same propensities to recommend PC ($z^{PC} = z'^{PC}$) and Home ($z^{Home} = z'^{Home}$) but different propensities to recommend ED vs. UC ($z^{ED} < z'^{ED}$):

$$\begin{aligned}
&\frac{E\left[Y \mid z'^{ED}, z^{PC}, z^{Home}\right] - E\left[Y \mid z^{ED}, z^{PC}, z^{Home}\right]}{E\left[R^{ED} \mid z'^{ED}, z^{PC}, z^{Home}\right] - E\left[R^{ED} \mid z^{ED}, z^{PC}, z^{Home}\right]} \\
=&E\left[Y(ED) - Y(UC) \mid R(z^{ED}, z^{PC}, z^{Home}) = UC, R(z'^{ED}, z^{PC}, z^{Home}) = ED\right]
\end{aligned} \tag{A2}$$

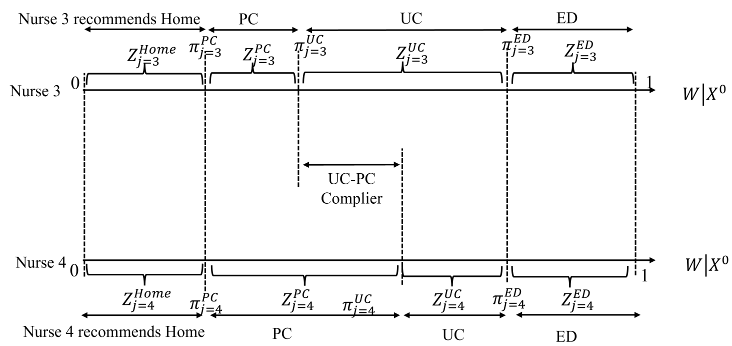
where the Wald estimand in the left-hand side identifies the margin-specific local average treatment effect among ED-UC compliers in the right-hand side. Similarly, the margin-specific treatment effects on UC-PC (PC-Home) margin can be identified by comparing nurses with the same propensities to recommend ED and Home (ED and UC) and differing propensities to recommend UC and PC (PC and Home) as illustrated in Appendix Figures A0b and A0c. In Appendix Section B1, we show that under Condition 1 (IV Validity), the Wald ratio identifies the same margin-specific LATE as the one we graphically derive in equation (A2) under the ordered choice framework.

We consider the ordered choice model based on a one-dimensional latent index (acuity) to be a good approximation of telephone triage nurses' decision-making.

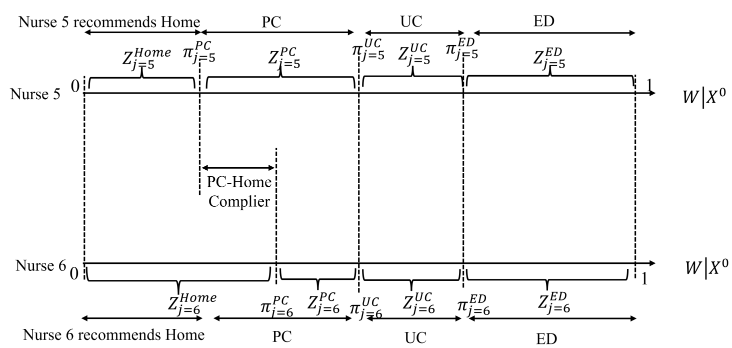
Unlike triage in emergency departments, disaster zones, or battlefields, VA telephone triage focuses on advising patients on the appropriate level of care for their symptoms, rather than prioritizing them based on real-time resource availability. While VA nurses may consider average local resource availability (e.g., patients' proximity to an ED), this factor is absorbed by baseline controls X_i^0, which include place-by-time fixed effects (call center, call time, and patient county). Within these fixed effects, we assume nurses primarily focus on patient acuity for their decisions. As discussed in Section 4.4.3, we also provide evidence that nurses rank patient acuity homogeneously and that differences in nurses' ED vs. UC tendencies do not affect potential recommendations on non-focal margins (UC-PC and PC-Home), conditional on their PC and Home tendencies (Z_i^{PC} and Z_i^{Home}), following Humphries et al. (2024).



(a) Nurse ED vs. UC Recommendation



(b) Nurse UC vs. PC Recommendation



(c) Nurse PC vs. Home Recommendation

Figure A0. Ordered Choice Model

Notes: Panel (a) illustrates how two arbitrary nurses $j \in \{1, 2\}$ decide a follow-up care recommendation $R(j) \in \{Home, PC, UC, ED\}$ under the ordered choice model. We assume that (i) all nurses rank patients' acuity, denoted by a single latent index W, in the same manner, but (ii) the nurses differ in their preferences for the level of care, denoted by thresholds π_j^r, $r \in \{PC, UC, ED\}$, $j \in \{1, 2\}$. Nurse j's propensity to recommend care r is denoted by $Z_j^r = Pr\left[R(j) = r \mid X^0\right]$. This figure visualizes an example where two nurses $j = 1, 2$ have the same thresholds for UC-PC and PC-Home margins but differ in their threshold for ED-UC margin. Similarly, Panels (b) and (c) illustrate two-nurse examples on UC-PC and PC-Home margins.

Supplemental Appendix B: Methods

B1 Derivation of Margin-Specific LATE under Condition 1

This section shows that under Condition 1 (IV Validity), our instrumental variable design identifies the margin-specific LATE.

Suppose that patients are quasi-randomly assigned to one of two nurses $J_i \in \{j, j'\}$. Nurse j' has a higher ED propensity than nurse j ($z'^{ED} > z^{ED}$), while the two nurses have the same propensity to recommend PC and Home ($z'^{PC} = z^{PC}$, $z'^{Home} = z^{Home}$). We denote a vector of nurse recommendation propensities by $Z_i = (Z_i^{ED}, Z_i^{PC}, Z_i^{Home})$. For compactness, we write the propensity vector for nurse j' by $z' = (z'^{ED}, z^{PC}, z^{Home})$ and for nurse j by $z = (z^{ED}, z^{PC}, z^{Home})$. Abstracting covariates away, we consider the following Wald ratio:

$$\frac{RF}{FS} = \frac{E\left[Y_i \mid Z_i = z'\right] - E\left[Y_i \mid Z_i = z\right]}{E\left[R_i^{ED} \mid Z_i = z'\right] - E\left[R_i^{ED} \mid Z_i = z\right]} \tag{B1}$$

In what follows, we show that, under Condition 1 (IV Validity), this Wald ratio in equation (B1) identifies the LATE of patients who receive an ED recommendation if assigned to nurse j' and receive a UC recommendation if assigned to nurse j.

Under Condition 1 (IV Validity), we can rewrite the reduced form (RF) as follows:

$$
\begin{aligned}
RF = & E\left[Y_i \mid Z_i = z'\right] - E\left[Y_i \mid Z_i = z\right] \\
= & E\left[\sum_{r\in\{ED,UC,PC,Home\}} Y_i(r)\cdot 1\{R_i(z')=r\} \mid Z_i = z'\right] \\
& - E\left[\sum_{r\in\{ED,UC,PC,Home\}} Y_i(r)\cdot 1\{R_i(z)=r\} \mid Z_i = z\right] \\
= & E\left[\sum_{r\in\{ED,UC,PC,Home\}} Y_i(r)\cdot 1\{R_i(z')=r\}\right] - E\left[\sum_{r\in\{ED,UC,PC,Home\}} Y_i(r)\cdot 1\{R_i(z)=r\}\right] \\
= & E\left[E\left[\sum_{r\in\{ED,UC,PC,Home\}} Y_i(r)\cdot 1\{R_i(z')=r\} \mid R_i(z'), R_i(z)\right]\right] \\
& - E\left[E\left[\sum_{r\in\{ED,UC,PC,Home\}} Y_i(r)\cdot 1\{R_i(z)=r\} \mid R_i(z'), R_i(z)\right]\right] \\
= & E\left[Y_i(ED) - Y_i(UC) \mid R_i(z') = ED, R_i(z) = UC\right]\cdot Pr\left[R_i(z') = ED, R_i(z) = UC\right]
\end{aligned}
\tag{B2}
$$

where the second equality follows from the relationship between observed and potential outcomes, the third from independence, the fourth from the law of iterated expectations, and the last from unordered partial monotonicity.

Similarly, we can rewrite the first stage (FS) as follows

$$
\begin{aligned}
FS = & E\left[R_i^{ED} \mid Z_i = z'\right] - E\left[R_i^{ED} \mid Z_i = z\right] \\
= & E\left[R_i^{ED}(z') \mid Z_i = z'\right] - E\left[R_i^{ED}(z) \mid Z_i = z\right] \\
= & E\left[R_i^{ED}(z')\right] - E\left[R_i^{ED}(z)\right] \\
= & E\left[E\left[R_i^{ED}(z') \mid R_i(z'), R_i(z)\right]\right] - E\left[E\left[R_i^{ED}(z) \mid R_i(z'), R_i(z)\right]\right] \\
= & Pr\left[R_i(z') = ED, R_i(z) = UC\right]
\end{aligned}
\tag{B3}
$$

where the second equality follows from the relationship between observed and potential recommendations, the third from independence, the fourth from the law of iterated expectations, and the last from unordered partial monotonicity.

Combining RF and FS, the Wald estimand can be rewritten as follows:

$$\frac{RF}{FS} = E\left[Y_i(ED) - Y_i(UC) \mid R_i(z') = ED, R_i(z) = UC\right] \tag{B4}$$

where the right-hand side of equation (B4) is the LATE of being recommended ED over UC among compliers whose recommendation switches from UC to ED when the assignment changes from nurse j to nurse j'.

B2 Bias in Binarized ED Recommendation Approach

Although the traditional examiners design applications mostly focus on a binary (binarized) treatment, our main paper examines identifying conditions to recover margin-specific causal effects without binarizing four alternative recommendations, builiding on Humphries et al. (2024). This section derives a bias expression if we instead binarize nurse recommendations into ED vs. non-ED (UC, PC, or Home) and use nurse propensity to recommend ED as a single instrument.

Suppose that patients are quasi-randomly assigned to one of two nurses $J_i \in \{j, j'\}$ with nurse j' having a higher ED propensity than nurse j. Abstracting covariates away, the examiners design estimand for the binary ED recommendation reduces to the following Wald estimand:

$$\frac{RF}{FS} = \frac{E\left[Y_i \mid J_i = j'\right] - E\left[Y_i \mid J_i = j\right]}{E\left[1\{R_i = ED\} \mid J_i = j'\right] - E\left[1\{R_i = ED\} \mid J_i = j\right]} \tag{B5}$$

In what follows, we show that this Wald ratio in equation (B5) does not generally recover interpretable causal effects of ED recommendation. Specifically, we demonstrate that it identifies a sum of (i) a weighted average of sub-LATEs involving ED recommendation and (ii) bias terms.

We start our derivation by noting that binarizing recommendations does not change the underlying data-generating process: the nurses still select one of four follow-up care recommendations $R_i \in \{Home, PC, UC, ED\}$. With the two-nurse setting, the pair of potential recommendations $(R_i(j'), R_i(j)) \in \{Home, PC, UC, ED\} \times \{Home, PC, UC, ED\}$ can take up to 16 different values.

Appendix Table B0 specifies and labels all possible values for $(R_i(j'), R_i(j))$. We call these pairs as *response types*, following Heckman and Pinto (2018). What the Wald ratio in equation (B5) identifies depends on restrictions over these response types. For instance, we assume that nurse j' (weakly) prefers more acute recommendations than nurse j for any patient i. This assumption eliminates response types (5), (9), (10), (13), (14), and (15). Not essentially, we further eliminate response types (4), (6), (8), and (12) for expositional simplicity.[20] After imposing these restrictions, we are left with the six response types: (1) always ED taker, (2) ED-UC complier, (3) ED-PC complier, (7) UC-PC complier, (11) always PC taker, and (16) always Home taker. Appendix Figure B0 shows a behavioral model of the two nurses j' and j consistent with the six response types, using the ordered choice framework discussed in Appendix Section A1.

Under these restrictions on response types and IV assumptions, we rewrite the

[20]Retaining these response types adds one more sub-LATE term and two more bias terms in equation (B8).

reduced-form in equation (B5) as follows:

$$
\begin{aligned}
RF = & E\left[Y_i \mid J_i = j'\right] - E\left[Y_i \mid J_i = j\right] \\
=& E\left[\sum_{r \in \{ED, UC, PC, Home\}} Y_i(r) \cdot 1\{R_i(j') = r\} \mid J_i = j'\right] \\
& - E\left[\sum_{r \in \{ED, UC, PC, Home\}} Y_i(r) \cdot 1\{R_i(j) = r\} \mid J_i = j\right] \\
=& E\left[\sum_{r \in \{ED, UC, PC, Home\}} Y_i(r) \cdot 1\{R_i(j') = r\}\right] - E\left[\sum_{r \in \{ED, UC, PC, Home\}} Y_i(r) \cdot 1\{R_i(j) = r\}\right] \\
=& E\left[E\left[\sum_{r \in \{ED, UC, PC, Home\}} Y_i(r) \cdot 1\{R_i(j') = r\} \mid R_i(j'), R_i(j)\right]\right] \\
& - E\left[E\left[\sum_{r \in \{ED, UC, PC, Home\}} Y_i(r) \cdot 1\{R_i(j) = r\} \mid R_i(j'), R_i(j)\right]\right] \\
=& E\left[Y_i(ED) - Y_i(UC) \mid R_i(j') = ED, R_i(j) = UC\right] \cdot Pr\left[R_i(j') = ED, R_i(j) = UC\right] \\
& + E\left[Y_i(ED) - Y_i(PC) \mid R_i(j') = ED, R_i(j) = PC\right] \cdot Pr\left[R_i(j') = ED, R_i(j) = PC\right] \\
& + E\left[Y_i(UC) - Y_i(PC) \mid R_i(j') = UC, R_i(j) = PC\right] \cdot Pr\left[R_i(j') = UC, R_i(j) = PC\right]
\end{aligned}
\tag{B6}
$$

where the second equality follows from the relationship between observed and potential outcomes and exclusion, the third from independence, the fourth from the law of iterated expectations, and the last from our restrictions on the response types.

Similarly, we rewrite the first stage in equation (B5) as follows:

$$\begin{aligned}
FS = & E\left[1\{R_i = ED\} \mid J_i = j'\right] - E\left[1\{R_i = ED\} \mid J_i = j\right] \\
= & E\left[1\{R_i(j') = ED\} \mid J_i = j'\right] - E\left[1\{R_i(j) = ED\} \mid J_i = j\right] \\
= & E\left[1\{R_i(j') = ED\}\right] - E\left[1\{R_i(j) = ED\}\right] \\
= & E\left[E\left[1\{R_i(j') = ED\} \mid R_i(j'), R_i(j)\right]\right] - E\left[E\left[1\{R_i(j) = ED\} \mid R_i(j'), R_i(j)\right]\right] \\
= & Pr\left[R_i(j') = ED, R_i(j) \neq ED\right]
\end{aligned} \tag{B7}$$

where the second equality follows from the relationship between observed and potential recommendations, the third from independence, the fourth from the law of iterated expectations, and the last from our restrictions on the response types.

Combining RF and FS, the Wald estimand can be rewritten as follows:

$$\begin{aligned}
\frac{RF}{FS} = & E\left[Y_i(ED) - Y_i(UC) \mid R_i(j') = ED, R_i(j) = UC\right] \frac{Pr\left[R_i(j') = ED, R_i(j) = UC\right]}{Pr\left[R_i(j') = ED, R_i(j) \neq ED\right]} \\
& + E\left[Y_i(ED) - Y_i(PC) \mid R_i(j') = ED, R_i(j) = PC\right] \frac{Pr\left[R_i(j') = ED, R_i(j) = PC\right]}{Pr\left[R_i(j') = ED, R_i(j) \neq ED\right]} \\
& + \underbrace{E\left[Y_i(UC) - Y_i(PC) \mid R_i(j') = UC, R_i(j) = PC\right] \frac{Pr\left[R_i(j') = UC, R_i(j) = PC\right]}{Pr\left[R_i(j') = ED, R_i(j) \neq ED\right]}}_{\text{Bias}}
\end{aligned} \tag{B8}$$

In the right-hand side of equation (B8), the sum of the first two terms can be interpreted as a weighted average of two sub-LATEs involving ED recommendation: (i) the causal effect of being recommended ED over UC and (ii) the causal effect of being recommended ED over PC. Under our restriction on response types, the weight sums to one ($Pr\left[R_i(j') = ED, R_i(j) \neq ED\right] = Pr\left[R_i(j') = ED, R_i(j) = UC\right]$ $+Pr\left[R_i(j') = ED, R_i(j) = PC\right]$). The last term is proportional to the LATE of being recommended UC over PC, which biases our estimand, as our interest is to recover an interpretable causal effect of ED recommendation.

The bias term becomes zero if (i) patient potential outcomes (e.g., patient choices) stay the same regardless of whether recommended UC or PC ($Y_i(UC) = Y_i(PC)$), or (ii) the hypothetical re-assignment from nurse j to nurse j' never changes recommendation from PC to UC ($Pr\left[R_i(j') = UC, R_i(j) = PC\right] = 0$). Either condition is unlikely to hold in our institutional context, and hence the Wald ratio in equation (B5) is generally biased.

Table B0. Response Type

	$R_i(j')$	$R_i(j)$	Label
(1)	ED	ED	Always ED taker
(2)	ED	UC	ED-UC complier
(3)	ED	PC	ED-PC complier
(4)	ED	Home	ED-Home complier
(5)	UC	ED	UC-ED defier
(6)	UC	UC	Always UC taker
(7)	UC	PC	UC-PC complier
(8)	UC	Home	UC-Home complier
(9)	PC	ED	PC-ED defier
(10)	PC	UC	PC-UC defier
(11)	PC	PC	Always PC taker
(12)	PC	Home	PC-Home complier
(13)	Home	ED	Home-ED defier
(14)	Home	UC	Home-UC defier
(15)	Home	PC	Home-PC defier
(16)	Home	Home	Always Home taker

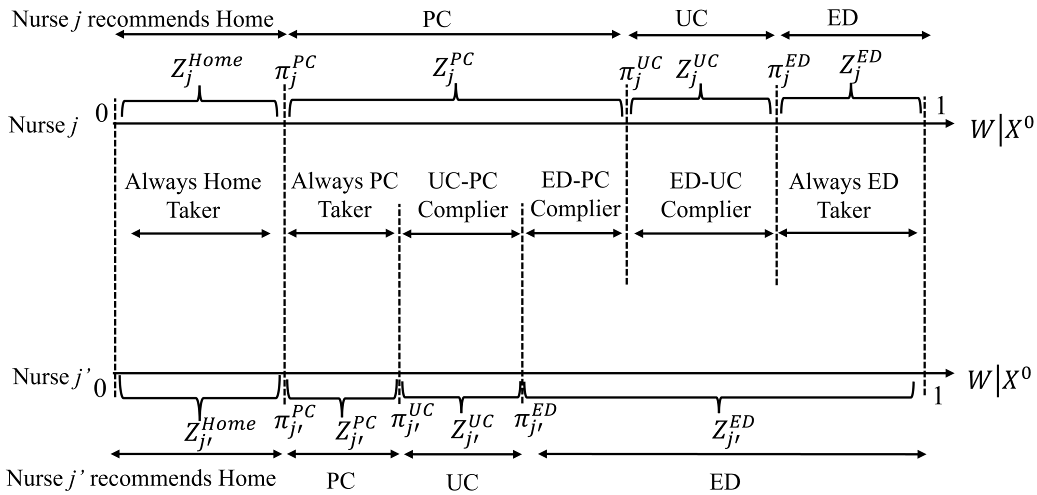


Figure B0. Ordered Choice and Response Types

B3 Robustness Check: Alternative Leave-Out Measures

As described in Section 4.3, our main analysis uses the leave-one-out average of residual ED recommendation of all other patients but the patient of call i. As robustness checks, we estimate the same linear TSLS model using alternative leave-out measures that are constructed differently.

First, we examine the sensitivity of the TSLS estimates to the choice of control variables. Recall that our main leave-out measure is constructed after we residualize nurse ED recommendation R_i^{ED} with respect to both baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county) and hold-out controls H_i (age, sex, marital status, race/ethnicity, period of service, prior health-care utilization, and prior diagnoses) using the linear regression in equation (5). As robustness checks, we examine the sensitivity of our estimates to (i) the exclusion of hold-out controls and (ii) the inclusion of several symptomatic measures at the time of the call in H_i. Appendix Table B15 reports the estimates from the specification that only partials out the baseline control X_i. Appendix Table B16 presents the

TSLS estimates that control for pain scale (0-10) and duration of symptoms (in 10 bins), whereas Appendix Table B17 shows the results that further control for chief complaint fixed effects.

Second, while our main analysis uses the leave-one-*patient*-out measure to account for rare events that the same patient is triaged by the same nurse multiple times during the analysis period, in Appendix Table B18 we present the TSLS estimates based on a leave-one-*call*-out average of all other calls but call i:

$$Z_i^{ED} = \frac{1}{n_{jcya} - 1} \sum_{i'} 1\left\{i' \neq i, j(i') = j, c(i') = c, y(i') = y, a(i') = a\right\} R_{i'}^{ED*} \tag{B9}$$

Lastly, while our preferred leave-out measure uses residuals of ED recommendation to better isolate nurse-specific effect from idiosyncratic variation in ED recommendation, in Appendix Table B19 we report the TSLS estimates calculated using a leave-one-patient-out average of raw ED recommendation:

$$Z_i^{ED} = \frac{1}{K_{jcya} - 1} \sum_{i'} \frac{1\left\{k(i') \neq k(i), j(i') = j, c(i') = c, y(i') = y, a(i') = a\right\} R_{i'}^{ED}}{n_{k(i')jcya}} \tag{B10}$$

We find that the estimates in Appendix Tables B16-B19 have the same sign as the main results reported in Table 2, and are also similar in magnitude.

B4 Local Average Potential Responses of Compliers

This section sketches our average potential outcome estimation on ED-UC margin.[21] Formally, abstracting the baseline covariates X_i^0 away, let us first consider two nurses who differ in their propensity to recommend ED relative to UC (${z'}^{ED} > z^{ED}$) while

[21] The average potential outcomes among the compliers on UC-PC (PC-Home) margin can be similarly identified using Z_i^{UC} (Z_i^{PC}) as an instrument and Z_i^{ED} and Z_i^{Home} (Z_i^{ED} and Z_i^{UC}) as controls.

having the same propensity to recommend PC. We denote the patient's potential outcomes by $Y_i(r)$, $r \in \{ED, UC, PC, Home\}$, to clarify that those potential outcomes depend on triage recommendation. Under IV condition, we can identify the average potential outcomes under ED recommendation as follows (Abadie, 2002):

$$\begin{aligned} &E\left[Y_i(ED) \mid R_i(z^{ED}, z^{PC}, z^{Home}) = UC, R_i(z'^{ED}, z^{PC}, z^{Home}) = ED\right] \\ =&\frac{E\left[Y_i \cdot R_i^{ED} \mid z'^{ED}, z^{PC}, z^{Home}\right] - E\left[Y_i \cdot R_i^{ED} \mid z^{ED}, z^{PC}, z^{Home}\right]}{E\left[R_i^{ED} \mid z'^{ED}, z^{PC}, z^{Home}\right] - E\left[R_i^{ED} \mid z^{ED}, z^{PC}, z^{Home}\right]}. \end{aligned} \tag{B11}$$

Likewise, we can identify the average potential outcomes among the compliers when they are recommended UC care as follows:

$$\begin{aligned} &E\left[Y_i(UC) \mid R_i(z^{ED}, z^{PC}, z^{Home}) = UC, R_i(z'^{ED}, z^{PC}, z^{Home}) = ED\right] \\ =&\frac{E\left[Y_i \cdot (1 - R_i^{ED}) \mid z'^{ED}, z^{PC}, z^{Home}\right] - E\left[Y_i \cdot (1 - R_i^{ED}) \mid z^{ED}, z^{PC}, z^{Home}\right]}{E\left[(1 - R_i^{ED}) \mid z'^{ED}, z^{PC}, z^{Home}\right] - E\left[(1 - R_i^{ED}) \mid z^{ED}, z^{PC}, z^{Home}\right]}. \end{aligned} \tag{B12}$$

We estimate the average outcomes of compliers who are recommended ED care in equation (B11) using the following TSLS regression:

$$Y_i \cdot R_i^{ED} = \beta_0 + \beta_1 R_i^{ED} + \beta_2 Z_i^{PC} + \beta_3 Z_i^{Home} + X_i^{0\prime} \pi + u_i \tag{B13}$$

$$R_i^{ED} = \alpha_0 + \alpha_1 Z_i^{ED} + \alpha_2 Z_i^{PC} + \alpha_3 Z_i^{Home} + X_i^{0\prime} \delta + \nu_i \tag{B14}$$

The average potential outcomes of compliers who are recommended UC care in equation (B12) can be estimated similarly:

$$Y_i \cdot (1 - R_i^{ED}) = \beta_0 + \beta_1 (1 - R_i^{ED}) + \beta_2 Z_i^{PC} + \beta_3 Z_i^{Home} + X_i^{0\prime} \pi + u_i \tag{B15}$$

$$(1 - R_i^{ED}) = \alpha_0 + \alpha_1 Z_i^{ED} + \alpha_2 Z_i^{PC} + \alpha_3 Z_i^{Home} + X_i^{0\prime} \delta + \nu_i. \tag{B16}$$

B5 Complier Characteristics

This section describes our approach to profiling the average characteristics of the compliers on ED-UC margin. Let H_i denote a vector of patient characteristics with h_{ki} being the k-th variable (e.g., patient age). Our goal is to estimate the mean h_{ki} among the complier population and contrast it to the mean h_{ki} among the overall patients in our sample. Formally, abstracting the baseline covariates X_i^0 away, let us first consider two nurses j' and j with j' having a higher ED tendency than j. Abadie (2002) shows that under independence and (strict) monotonicity assumptions, the mean h_{ki} among the complier can be identified by instrumental variable regression:

$$\begin{aligned} & E\left[h_{ki} \mid R_i(z^{ED}, z^{PC}, z^{Home}) = UC, R_i(z'^{ED}, z^{PC}, z^{Home}) = ED\right] \\ =& \frac{E\left[h_{ki} \cdot R_i^{ED} \mid z'^{ED}, z^{PC}, z^{Home}\right] - E\left[h_{ki} \cdot R_i^{ED} \mid z^{ED}, z^{PC}, z^{Home}\right]}{E\left[R_i^{ED} \mid z'^{ED}, z^{PC}, z^{Home}\right] - E\left[R_i^{ED} \mid z^{ED}, z^{PC}, z^{Home}\right]} \end{aligned} \tag{B17}$$

We estimate the Wald estimand in equation (B17) with the following TSLS regression:

$$h_{ki}R_i^{ED} = \beta_{0k} + \beta_{1k}R_i^{ED} + \beta_{2k}Z_i^{PC} + \beta_{3k}Z_i^{Home} + X_i^{0\prime}\Gamma_k + u_{ki} \tag{B18}$$

$$R_i^{ED} = \alpha_{0k} + \alpha_{1k}Z_i^{ED} + \alpha_{2k}Z_i^{PC} + \alpha_{3k}Z_i^{Home} + X_i^{0\prime}\gamma_k + e_{ki} \tag{B19}$$

where Z_i^{ED} and Z_i^{PC} are the leave-out measure of the assigned nurse's ED and PC tendencies, respectively. With more than two nurses, $\widehat{\beta}_{1k,TSLS}$ converges to a weighted average of the pairwise Wald estimands in equation (B17) of all pairs of the nurses under IV condition (Imbens and Angrist, 1994). Similarly, the average characteristics of the compliers on UC-PC (PC-Home) margin can be identified using Z_i^{UC} (Z_i^{PC}) as an instrument and Z_i^{ED} and Z_i^{Home} (Z_i^{ED} and Z_i^{UC}) as controls.

B6 Appendix Figures and Tables

Table B1. Selection of Analysis Sample

Step	Description	Calls	Patients	Nurses	Call Centers
0	All symptom calls with complete IDs and dates.	4,930,371	2,044,446	6,866	101
1	Drop calls with missing values in triage disposition, demographics, combat history, benefit eligibility, and comorbidity indices.	2,630,268	1,267,400	3,658	101
2	Drop calls from patients younger than 20 or older than 99; Drop calls from patients in inpatient facilities.	2,628,755	1,266,783	3,655	101
3	Drop calls received during non-business hours (before 8 am; after 4 pm; weekends; holidays).	1,781,248	1,003,221	3,320	100
4	Drop calls from patients with the most recent prior call within 30 days.	1,522,441	958,400	2,939	100
5	Drop calls from patients with the most recent prior ED visit within 30 days.	1,411,731	913,514	2,916	100
6	Drop calls received by nurse with less than 10 calls per algorithm-location-by-algorithm-interval-by-call-center-by-year cell. Drop calls in algorithm-location-by-algorithm-interval-by-call-center-by-year cell with only one remaining nurse.	1,273,843	836,420	1,725	96

Notes: This table describes sample restriction steps to construct the analysis sample. The table lists the number of calls, patients, nurses, and call centers at each step. Table 1 presents average call characteristics.

Table B2. Algorithm Location-by-Interval Disposition

Location	Interval	Calls	Nurse Recommended Home	Nurse Recommended PC	Nurse Recommended UC	Nurse Recommended ED
ED	Now-911	58,736	0.000	0.019	0.003	0.977
ED	Now	274,527	0.001	0.044	0.017	0.938
ED	2-8 hours	2,075	0.000	0.116	0.018	0.866
Urgent Care	Now	97	0.010	0.010	0.794	0.186
Urgent Care	2-8 hours	5,451	0.003	0.127	0.730	0.141
Primary Care	2-8 hours	181,343	0.001	0.824	0.089	0.085
Primary Care	12-24 hours	344,415	0.002	0.899	0.059	0.041
Primary Care	2-3 days	219,883	0.004	0.947	0.032	0.018
Primary Care	1-2 weeks	103,512	0.003	0.974	0.013	0.010
Home	Self-Care	82,804	0.766	0.192	0.031	0.011
Dentist	12-24 hours	778	0.000	0.916	0.062	0.022
Dentist	2-3 days	222	0.000	0.986	0.005	0.009

Notes: This table describes the list of algorithm dispositions (recommendations). As described in equation (6), our leave-out measure Z_i^{ED} captures cross-nurse variation in the propensity to recommend ED (UC, or PC) within cells defined by the locations and intervals recommended by the algorithm.

Table B3. Baseline Control Variables

Variable	Number of Categories
Call Center	96
FIPS	3,150
Day of Week	5
Month-by-Year	50
AM/PM	2
Algorithm Disposition (Location-by-Interval)	12

Notes: This table describes the baseline control variables, denoted as X_i^0 in Condition 1 and throughout the paper. We assume that the assigned nurse's ED tendency Z_i^{ED} is independent of the patient's potential outcomes and potential recommendation (treatment) status, conditional on the baseline controls.

Table B4. Hold-Out Control Variables

Variable	Number of Categories
Age Bin (5-year bins from 20-64, 2-year bins from 65-100)	27
Male	2
Married	2
White	2
Black	2
Hispanic	2
Period of Service	7
ED Visit in Prior Year (0 vs. 1+)	2
Admission in Prior Year (0 vs. 1+)	2
Inpatient in Prior Year (0 vs. 1+)	2
Primary Care Visit in Prior Year (0 vs. 1+)	2
Elixhauser Comorbidity Indicators (31 dummies)	2 (in each)

Notes: This table describes hold-out control variables, denoted by H_i throughout the paper. These variables are used to test whether patients are balanced between nurses within the baseline conditioning cell X_i^0, as shown in Figure B3. After confirming conditional independence, we include those hold-out variables H_i, in addition to the baseline controls X_i^0, in all regressions for statistical precision. We also use H_i to define subsamples for our heterogeneity analysis.

Table B5. Test for Monotonicity (ED - UC)

Variable	Subsample	Coef (SE)	Nurse Recommends ED	Observations
Age	65 and Over	0.8533 (0.0089)	0.2907	671,824
Age	Below 65	0.8645 (0.0091)	0.2620	602,019
Sex	Female	0.8509 (0.0162)	0.2608	173,971
Sex	Male	0.8605 (0.0071)	0.2797	1,099,872
Race	Asian/Other	0.8857 (0.0370)	0.2765	42,633
Race	Black	0.8723 (0.0137)	0.3017	231,989
Race	Hispanic	0.8403 (0.0235)	0.2923	77,337
Race	White	0.8569 (0.0078)	0.2691	880,795
Residence County	Rural	0.8450 (0.0143)	0.2733	262,214
Residence County	Urban	0.8619 (0.0075)	0.2781	1,011,629
Priority Group	Group 1,4 Highly disabled	0.8716 (0.0089)	0.2869	630,887
Priority Group	Group 2,3,6 Low/moderate disability	0.8782 (0.0135)	0.2495	261,235
Priority Group	Group 5 Low-Income	0.8258 (0.0146)	0.3012	225,860
Priority Group	Group 7-8 Non-Disabled, copayment required	0.8332 (0.0177)	0.2491	155,861
Comorbidity Count	1	0.8721 (0.0120)	0.2351	332,629
Comorbidity Count	2	0.8710 (0.0126)	0.2579	321,921
Comorbidity Count	3	0.8717 (0.0138)	0.2823	246,589
Comorbidity Count	4	0.8251 (0.0170)	0.3005	157,488
Comorbidity Count	5+	0.8334 (0.0152)	0.3477	215,212

Notes: This table reports first-stage coefficients on different subsamples of patients. Using all observations in our sample, we construct the leave-one-patient-out measure Z_i^{ED} in equation (6) after partialling out the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county) from ED recommendation indicator R_i^{ED}. Then we split the sample on various covariates (in the hold-out set H_i) and estimate first-stage coefficients separately. All first-stage coefficients are expected to be positive since monotonicity implies positive correlations between nurses' overall ED tendency and subgroup-specific ED tendencies (Frandsen et al., 2023). All regressions control for the baseline controls X_i^0 described in Appendix Table B3. Standard errors are clustered at the call center-by-call time level.

Table B6. Test for Monotonicity (UC - PC)

Variable	Subsample	Coef (SE)	Nurse Recommends UC	Observations
Age	65 and Over	0.8687 (0.0066)	0.0387	671,824
Age	Below 65	0.9674 (0.0067)	0.0502	602,019
Sex	Female	0.9961 (0.0116)	0.0541	173,971
Sex	Male	0.9054 (0.0053)	0.0426	1,099,872
Race	Asian/Other	0.8994 (0.0273)	0.0406	42,633
Race	Black	0.9256 (0.0095)	0.0473	231,989
Race	Hispanic	0.9643 (0.0165)	0.0513	77,337
Race	White	0.9126 (0.0061)	0.0427	880,795
Residence County	Rural	0.9411 (0.0108)	0.0463	262,214
Residence County	Urban	0.9148 (0.0055)	0.0436	1,011,629
Priority Group	Group 1,4 Highly disabled	0.9317 (0.0065)	0.0468	630,887
Priority Group	Group 2,3,6 Low/moderate disability	0.9308 (0.0104)	0.0428	261,235
Priority Group	Group 5 Low-Income	0.8853 (0.0113)	0.0402	225,860
Priority Group	Group 7-8 Non-Disabled, copayment required	0.8958 (0.0130)	0.0415	155,861
Comorbidity Count	1	0.9631 (0.0089)	0.0484	332,629
Comorbidity Count	2	0.9421 (0.0091)	0.0456	321,921
Comorbidity Count	3	0.9134 (0.0101)	0.0439	246,589
Comorbidity Count	4	0.8873 (0.0125)	0.0426	157,488
Comorbidity Count	5+	0.8418 (0.0113)	0.0369	215,212

Notes: This table reports first-stage coefficients on different subsamples of patients. Using all observations in our sample, we construct the leave-one-patient-out measure Z_i^{UC} in equation (6) after partialling out the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county) from UC recommendation indicator R_i^{UC}. Then we split the sample on various covariates (in the hold-out set H_i) and estimate first-stage coefficients separately. All first-stage coefficients are expected to be positive since monotonicity implies positive correlations between nurses' overall UC tendency and subgroup-specific UC tendencies (Frandsen et al., 2023). All regressions control for the baseline controls X_i^0 described in Appendix Table B3. Standard errors are clustered at the call center-by-call time level.

Table B7. Test for Monotonicity (PC - Home)

Variable	Subsample	Coef (SE)	Nurse Recommends PC	Observations
Age	65 and Over	0.9087 (0.0084)	0.6249	671,824
Age	Below 65	0.9626 (0.0076)	0.6300	602,019
Sex	Female	0.9677 (0.0146)	0.6305	173,971
Sex	Male	0.9326 (0.0062)	0.6268	1,099,872
Race	Asian/Other	0.8800 (0.0336)	0.6260	42,633
Race	Black	0.9668 (0.0124)	0.5996	231,989
Race	Hispanic	0.9671 (0.0197)	0.5954	77,337
Race	White	0.9265 (0.0070)	0.6379	880,795
Residence County	Rural	0.9430 (0.0143)	0.6373	262,214
Residence County	Urban	0.9362 (0.0062)	0.6247	1,011,629
Priority Group	Group 1,4 Highly disabled	0.9435 (0.0081)	0.6164	630,887
Priority Group	Group 2,3,6 Low/moderate disability	0.9391 (0.0119)	0.6545	261,235
Priority Group	Group 5 Low-Income	0.9211 (0.0141)	0.6095	225,860
Priority Group	Group 7-8 Non-Disabled, copayment required	0.9317 (0.0156)	0.6516	155,861
Comorbidity Count	1	0.9553 (0.0099)	0.6573	332,629
Comorbidity Count	2	0.9319 (0.0108)	0.6423	321,921
Comorbidity Count	3	0.9372 (0.0132)	0.6247	246,589
Comorbidity Count	4	0.9218 (0.0173)	0.6093	157,488
Comorbidity Count	5+	0.9250 (0.0162)	0.5746	215,212

Notes: This table reports first-stage coefficients on different subsamples of patients. Using all observations in our sample, we construct the leave-one-patient-out measure Z_i^{PC} in equation (6) after partialling out the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county) from PC recommendation indicator R_i^{PC}. Then we split the sample on various covariates (in the hold-out set H_i) and estimate first-stage coefficients separately. All first-stage coefficients are expected to be positive since monotonicity implies positive correlations between nurses' overall PC tendency and subgroup-specific PC tendencies (Frandsen et al., 2023). All regressions control for the baseline controls X_i^0 described in Appendix Table B3. Standard errors are clustered at the call center-by-call time level.

Table B8. Test for Unordered Pairwise Monotonicity (ED - UC)

	Predicted Visit ED 3d	Predicted Visit UC 3d	Predicted Visit PCP 3d
Nurse Propensity to Recommend ED	-0.0006	0.0008	-0.0005
	(0.0006)	(0.0003)	(0.0008)
Outcome mean	0.084	0.052	0.346
Observations	864,587	864,587	864,587

Notes: This table reports estimated coefficients $\widehat{\eta}_1$ in equation (7). We construct predicted utilization measures $\widehat{Y}_i$ using a linear regression of ED (UC, or PC) visit indicator on the baseline controls X_i^0 and hold-out controls H_i. Then we restrict the sample to calls for which nurses recommended a primary care visit or home care, and estimate the model in equation (7) to examine whether the nurse ED propensity is correlated with observable characteristics of the patient, following Humphries et al. (2024). The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. Standard errors are clustered at the call center-by-call time level.

Table B9. Test for Unordered Pairwise Monotonicity (UC - PC)

	Predicted Visit ED 3d	Predicted Visit UC 3d	Predicted Visit PCP 3d
Nurse Propensity to Recommend UC	-0.0003	0.0002	0.0008
	(0.002)	(0.0003)	(0.0008)
Outcome mean	0.403	0.039	0.235
Observations	418,509	418,509	418,509

Notes: This table reports estimated coefficients $\widehat{\eta}_1$ in equation (7). We construct predicted utilization measures $\widehat{Y}_i$ using a linear regression of ED (UC, or PC) visit indicator on the baseline controls X_i^0 and hold-out controls H_i. Then we restrict the sample to calls for which nurses recommended an ED visit or home care, and estimate the model in equation (7) to examine whether the nurse UC propensity is correlated with observable characteristics of the patient, following Humphries et al. (2024). The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. Standard errors are clustered at the call center-by-call time level.

Table B10. Test for Unordered Pairwise Monotonicity (PC - Home)

	Predicted Visit ED 3d	Predicted Visit UC 3d	Predicted Visit PCP 3d
Nurse Propensity to Recommend PC	0.004	-6.74e-5	0.002
	(0.004)	(0.001)	(0.002)
Outcome mean	0.420	0.087	0.223
Observations	409,256	409,256	409,256

Notes: This table reports estimated coefficients $\widehat{\eta}_1$ in equation (7). We construct predicted utilization measures $\widehat{Y}_i$ using a linear regression of ED (UC, or PC) visit indicator on the baseline controls X_i^0 and hold-out controls H_i. Then we restrict the sample to calls for which nurses recommended an ED or UC visit, and estimate the model in equation (7) to examine whether the nurse PC propensity is correlated with observable characteristics of the patient, following Humphries et al. (2024). The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. Standard errors are clustered at the call center-by-call time level.

Table B11. IV Results: Effect of Nurse Recommendations on Outpatient Costs Breakdown

(a) ED - UC

	PC Cost	UC Cost	ED Cost
Panel A: 3-Day Outcomes			
Nurse Recommends ED	13.630	-94.435	195.603
	(4.871)	(4.471)	(8.629)
Outcome Mean (UC)	83.454	155.700	84.018
Outcome Mean (All)	155.866	30.349	157.502

(b) UC - PC

	PC Cost	UC Cost	ED Cost
Panel A: 3-Day Outcomes			
Nurse Recommends UC	-81.693	52.572	2.035
	(3.065)	(2.666)	(4.517)
Outcome Mean (PC)	188.629	24.329	75.663
Outcome Mean (All)	155.866	30.349	157.502

(c) PC - Home

	PC Cost	UC Cost	ED Cost
Panel A: 3-Day Outcomes			
Nurse Recommends PC	38.879	3.823	11.868
	(4.010)	(1.784)	(4.050)
Outcome Mean (Home)	139.349	14.515	44.831
Outcome Mean (All)	155.866	30.349	157.502

Notes: Panel (a) reports TSLS estimates of the effect of ED recommendation on ED-UC margin. Costs are measured in US dollars. In the third row of each panel, the sample average for each outcome conditional on UC recommendation is reported as a benchmark ($E[Y_i \mid R_i = UC]$). In the fourth row, the overall (unconditional) sample average is also reported ($E[Y_i]$). In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. We break down costs based on stop code for VA care and place of service ID for care purchased from non-VA providers. Standard errors are clustered at the call center-by-call time level. Panels (b) and (c) represent the corresponding analysis on UC-PC and PC-Home margins.

Table B12. IV Results: Effect of Nurse ED vs. UC Recommendations on Long-Term Mortality

	Mortality
Panel A: 60-Day Outcomes	
Nurse Recommends ED	0.388
	(0.184)
Outcome Mean (UC)	0.588
Outcome Mean (All)	0.832
Panel B: 90-Day Outcomes	
Nurse Recommends ED	0.336
	(0.220)
Outcome Mean (UC)	0.882
Outcome Mean (All)	1.203
Panel C: 180-Day Outcomes	
Nurse Recommends ED	0.207
	(0.300)
Outcome Mean (UC)	1.692
Outcome Mean (All)	2.251
Panel D: 360-Day Outcomes	
Nurse Recommends ED	0.022
	(0.407)
Outcome Mean (UC)	3.303
Outcome Mean (All)	4.204

Notes: This table reports TSLS estimates of the effect of ED recommendation. All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. In the third row of each panel, the sample average for each outcome conditional on UC recommendation is reported as a benchmark ($E[Y_i \mid R_i = UC]$). In the fourth row, the overall (unconditional) sample average is also reported ($E[Y_i]$). In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. Standard errors are clustered at the call center-by-call time level.

Table B13. IV Results: Effect of Nurse UC vs. PC Recommendation on Long-Term Mortality

	Mortality
Panel A: 60-Day Outcomes	
Nurse Recommends UC	-0.154
	(0.108)
Outcome Mean (PC)	0.579
Outcome Mean (All)	0.832
Panel B: 90-Day Outcomes	
Nurse Recommends UC	-0.149
	(0.129)
Outcome Mean (PC)	0.873
Outcome Mean (All)	1.203
Panel C: 180-Day Outcomes	
Nurse Recommends UC	-0.178
	(0.177)
Outcome Mean (PC)	1.759
Outcome Mean (All)	2.251
Panel D: 360-Day Outcomes	
Nurse Recommends UC	-0.124
	(0.240)
Outcome Mean (PC)	3.485
Outcome Mean (All)	4.204

Notes: This table reports TSLS estimates of the effect of ED recommendation. All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. In the third row of each panel, the sample average for each outcome conditional on UC recommendation is reported as a benchmark ($E[Y_i \mid R_i = UC]$). In the fourth row, the overall (unconditional) sample average is also reported ($E[Y_i]$). In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. Standard errors are clustered at the call center-by-call time level.

Table B14. IV Results: Effect of Nurse PC vs. Home Recommendation on Long-Term Mortality

	Mortality
Panel A: 60-Day Outcomes	
Nurse Recommends PC	0.176
	(0.114)
Outcome Mean (Home)	0.508
Outcome Mean (All)	0.832
Panel B: 90-Day Outcomes	
Nurse Recommends PC	0.144
	(0.130)
Outcome Mean (Home)	0.725
Outcome Mean (All)	1.203
Panel C: 180-Day Outcomes	
Nurse Recommends PC	-0.067
	(0.179)
Outcome Mean (Home)	1.444
Outcome Mean (All)	2.251
Panel D: 360-Day Outcomes	
Nurse Recommends PC	0.110
	(0.249)
Outcome Mean (Home)	2.834
Outcome Mean (All)	4.204

Notes: This table reports TSLS estimates of the effect of ED recommendation. All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. In the third row of each panel, the sample average for each outcome conditional on UC recommendation is reported as a benchmark ($E[Y_i \mid R_i = UC]$). In the fourth row, the overall (unconditional) sample average is also reported ($E[Y_i]$). In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. Standard errors are clustered at the call center-by-call time level.

Table B15. IV Results: Effect of Nurse ED vs. UC Recommendation (Without Hold-Out Controls)

	Primary Care	Urgent Care	ED	UC or ED	Any Care	Hospital Admission	Mortality	Outpatient Cost	Inpatient Cost	Total Cost
Panel A: 3-Day Outcomes										
Nurse Recommends ED	2.721	-24.096	25.554	2.646	4.269	1.752	0.080	116.337	111.697	228.035
	(0.861)	(0.637)	(0.818)	(0.968)	(0.985)	(0.358)	(0.041)	(35.522)	(38.568)	(62.382)
Outcome Mean (UC)	19.588	38.646	11.392	48.278	58.877	1.214	0.028	227.906	93.061	320.967
Outcome Mean (All)	30.672	6.311	19.187	25.159	49.857	2.671	0.041	290.885	157.584	448.469
Panel B: 28-Day Outcomes										
Nurse Recommends ED	1.509	-24.659	24.609	2.518	1.155	2.466	0.199	188.464	291.924	480.387
	(0.972)	(0.670)	(0.870)	(0.985)	(0.842)	(0.458)	(0.131)	(83.615)	(133.834)	(178.304)
Outcome Mean (UC)	52.305	42.171	18.192	55.688	77.640	2.788	0.276	1217.306	546.884	1764.190
Outcome Mean (All)	63.081	7.977	25.912	32.782	74.960	4.668	0.416	1406.952	855.097	2262.050

Notes: This table reports TSLS estimates of the effect of ED recommendation. All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. In the third row of each panel, the sample average for each outcome conditional on UC recommendation is reported as a benchmark ($E[Y_i \mid R_i = UC]$). In the fourth row, the overall (unconditional) sample average is also reported ($E[Y_i]$). In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. Standard errors are clustered at the call center-by-call time level.

Table B16. IV Results: Effect of Nurse ED vs. UC Recommendation (With Pain Scale and Symptom Duration Controls)

	Primary Care	Urgent Care	ED	UC or ED	Any Care	Hospital Admission	Mortality	Outpatient Cost	Inpatient Cost	Total Cost
Panel A: 3-Day Outcomes										
Nurse Recommends ED	2.653	-23.961	24.991	2.243	3.930	1.484	0.081	117.839	100.040	217.879
	(0.862)	(0.635)	(0.813)	(0.959)	(0.982)	(0.357)	(0.042)	(35.678)	(38.670)	(62.709)
Outcome Mean (UC)	19.588	38.646	11.392	48.278	58.877	1.214	0.028	227.906	93.061	320.967
Outcome Mean (All)	30.672	6.311	19.187	25.159	49.857	2.671	0.041	290.885	157.584	448.469
Panel B: 28-Day Outcomes										
Nurse Recommends ED	1.582	-24.562	23.887	1.938	1.214	2.077	0.164	159.162	213.474	372.636
	(0.970)	(0.669)	(0.863)	(0.977)	(0.843)	(0.453)	(0.131)	(83.983)	(133.868)	(177.916)
Outcome Mean (UC)	52.305	42.171	18.192	55.688	77.640	2.788	0.276	1217.306	546.884	1764.190
Outcome Mean (All)	63.081	7.977	25.912	32.782	74.960	4.668	0.416	1406.952	855.097	2262.050

Notes: This table reports TSLS estimates of the effect of ED recommendation. All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. In the third row of each panel, the sample average for each outcome conditional on UC recommendation is reported as a benchmark ($E\left[Y_i \mid R_i = UC\right]$). In the fourth row, the overall (unconditional) sample average is also reported ($E\left[Y_i\right]$). In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. Standard errors are clustered at the call center-by-call time level.

Table B17. IV Results: Effect of Nurse ED vs. UC Recommendation (With Chief Complaint Controls)

	Primary Care	Urgent Care	ED	UC or ED	Any Care	Hospital Admission	Mortality	Outpatient Cost	Inpatient Cost	Total Cost
Panel A: 3-Day Outcomes										
Nurse Recommends ED	3.039	-24.111	24.272	1.440	3.967	0.912	0.053	133.188	91.965	225.153
	(0.928)	(0.669)	(0.862)	(1.018)	(1.046)	(0.379)	(0.043)	(37.591)	(41.553)	(66.568)
Outcome Mean (UC)	19.588	38.646	11.392	48.278	58.877	1.214	0.028	227.906	93.061	320.967
Outcome Mean (All)	30.672	6.311	19.187	25.159	49.857	2.671	0.041	290.885	157.584	448.469
Panel B: 28-Day Outcomes										
Nurse Recommends ED	1.805	-24.680	23.304	1.320	1.532	1.421	0.129	135.735	98.697	234.435
	(1.040)	(0.704)	(0.914)	(1.039)	(0.900)	(0.481)	(0.138)	(90.626)	(144.857)	(191.357)
Outcome Mean (UC)	52.305	42.171	18.192	55.688	77.640	2.788	0.276	1217.306	546.884	1764.190
Outcome Mean (All)	63.081	7.977	25.912	32.782	74.960	4.668	0.416	1406.952	855.097	2262.050

Notes: This table reports TSLS estimates of the effect of ED recommendation. All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. In the third row of each panel, the sample average for each outcome conditional on UC recommendation is reported as a benchmark ($E[Y_i \mid R_i = UC]$). In the fourth row, the overall (unconditional) sample average is also reported ($E[Y_i]$). In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. Standard errors are clustered at the call center-by-call time level.

Table B18. IV Results: Effect of Nurse ED vs. UC Recommendation (Leave-One-Call-Out Measure)

	Primary Care	Urgent Care	ED	UC or ED	Any Care	Hospital Admission	Mortality	Outpatient Cost	Inpatient Cost	Total Cost
Panel A: 3-Day Outcomes										
Nurse Recommends ED	2.648	-24.016	25.350	2.524	4.133	1.602	0.076	111.041	103.962	215.003
	(0.859)	(0.636)	(0.815)	(0.965)	(0.983)	(0.355)	(0.041)	(35.528)	(38.473)	(62.271)
Outcome Mean (UC)	19.588	38.646	11.392	48.278	58.877	1.214	0.028	227.906	93.061	320.967
Outcome Mean (All)	30.672	6.311	19.187	25.159	49.857	2.671	0.041	290.885	157.584	448.469
Panel B: 28-Day Outcomes										
Nurse Recommends ED	1.248	-24.580	24.212	2.208	0.916	2.196	0.164	160.529	242.771	403.299
	(0.966)	(0.669)	(0.863)	(0.979)	(0.839)	(0.450)	(0.131)	(82.893)	(132.761)	(176.175)
Outcome Mean (UC)	52.305	42.171	18.192	55.688	77.640	2.788	0.276	1217.306	546.884	1764.190
Outcome Mean (All)	63.081	7.977	25.912	32.782	74.960	4.668	0.416	1406.952	855.097	2262.050

Notes: This table reports TSLS estimates of the effect of ED recommendation. All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. In the third row of each panel, the sample average for each outcome conditional on UC recommendation is reported as a benchmark ($E[Y_i \mid R_i = UC]$). In the fourth row, the overall (unconditional) sample average is also reported ($E[Y_i]$). In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. Standard errors are clustered at the call center-by-call time level.

Table B19. IV Results: Effect of Nurse ED vs. UC Recommendation (Leave-Out Average of Raw ED Recommendation)

	Primary Care	Urgent Care	ED	UC or ED	Any Care	Hospital Admission	Mortality	Outpatient Cost	Inpatient Cost	Total Cost
Panel A: 3-Day Outcomes										
Nurse Recommends ED	2.699	-23.870	25.535	2.822	4.467	1.676	0.076	111.471	104.502	215.972
	(0.855)	(0.631)	(0.808)	(0.956)	(0.977)	(0.352)	(0.041)	(35.187)	(38.125)	(61.697)
Outcome Mean (UC)	19.588	38.646	11.392	48.278	58.877	1.214	0.028	227.906	93.061	320.967
Outcome Mean (All)	30.672	6.311	19.187	25.159	49.857	2.671	0.041	290.885	157.584	448.469
Panel B: 28-Day Outcomes										
Nurse Recommends ED	1.142	-24.483	24.413	2.472	1.056	2.258	0.145	159.452	261.258	420.710
	(0.960)	(0.663)	(0.856)	(0.970)	(0.834)	(0.446)	(0.130)	(81.671)	(131.510)	(174.263)
Outcome Mean (UC)	52.305	42.171	18.192	55.688	77.640	2.788	0.276	1217.306	546.884	1764.190
Outcome Mean (All)	63.081	7.977	25.912	32.782	74.960	4.668	0.416	1406.952	855.097	2262.050

Notes: This table reports TSLS estimates of the effect of ED recommendation. All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. In the third row of each panel, the sample average for each outcome conditional on UC recommendation is reported as a benchmark ($E[Y_i \mid R_i = UC]$). In the fourth row, the overall (unconditional) sample average is also reported ($E[Y_i]$). In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. Standard errors are clustered at the call center-by-call time level.

Table B20. Complier Characteristics

		Compliers (ED-UC)		Compliers (UC-PC)		Compliers (PC-Home)	
Variable	Overall Mean	Mean	Ratio	Mean	Ratio	Mean	Ratio
Congestive heart failure	0.074	0.083	1.118	0.065	0.873	0.055	0.732
		(0.004)	[1.024 - 1.213]	(0.002)	[0.825 - 0.922]	(0.002)	[0.673 - 0.791]
Cardiac arrhythmias	0.156	0.168	1.079	0.135	0.869	0.134	0.863
		(0.005)	[1.021 - 1.137]	(0.003)	[0.838 - 0.901]	(0.003)	[0.820 - 0.907]
Valvular disease	0.037	0.040	1.093	0.031	0.849	0.029	0.794
		(0.002)	[0.965 - 1.221]	(0.001)	[0.781 - 0.918]	(0.002)	[0.703 - 0.886]
Pulmonary circulation disorders	0.018	0.023	1.327	0.014	0.791	0.012	0.709
		(0.002)	[1.124 - 1.531]	(0.001)	[0.690 - 0.892]	(0.001)	[0.585 - 0.834]
Peripheral vascular disorders	0.084	0.096	1.138	0.071	0.845	0.063	0.746
		(0.004)	[1.055 - 1.221]	(0.002)	[0.799 - 0.891]	(0.003)	[0.687 - 0.804]
Hypertension, uncomplicated	0.541	0.545	1.007	0.519	0.960	0.515	0.952
		(0.006)	[0.986 - 1.029]	(0.004)	[0.946 - 0.973]	(0.005)	[0.934 - 0.971]
Hypertension, complicated	0.077	0.083	1.077	0.069	0.902	0.063	0.814
		(0.003)	[0.988 - 1.166]	(0.002)	[0.853 - 0.951]	(0.002)	[0.751 - 0.877]
Paralysis	0.007	0.003	0.520	0.005	0.729	0.007	1.038
		(0.001)	[0.224 - 0.817]	(0.001)	[0.573 - 0.885]	(0.001)	[0.775 - 1.302]

Notes: This table presents the average characteristics among compliers, along with the overall sample averages. Following Abadie (2003), we estimate $E\left[H_i \mid \text{Compliers}\right]$ by regressing an interaction between each patient characteristic H and ED recommendation indicator $(H \cdot R^{ED})$ on ED recommendation indicator (R^{ED}) with the right-hand-side R^{ED} instrumented by the leave-out ED recommendation propensity on ED-UC margin. The average characteristics among compliers on UC-PC and PC-Home margins are estimated similarly. All regressions control for the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county) described in Appendix Table B3. Standard errors are clustered at the call center-by-call time level.

Table B21. Complier Characteristics (Continued)

Variable	Overall Mean	Compliers (ED-UC) Mean	Compliers (ED-UC) Ratio	Compliers (UC-PC) Mean	Compliers (UC-PC) Ratio	Compliers (PC-Home) Mean	Compliers (PC-Home) Ratio
Other neurological disorders	0.066	0.067	1.016	0.056	0.839	0.059	0.884
		(0.003)	[0.920 - 1.111]	(0.002)	[0.789 - 0.890]	(0.002)	[0.815 - 0.953]
Chronic pulmonary disease	0.204	0.218	1.067	0.175	0.855	0.162	0.792
		(0.005)	[1.019 - 1.116]	(0.003)	[0.828 - 0.883]	(0.004)	[0.755 - 0.828]
Diabetes, uncomplicated	0.124	0.130	1.047	0.126	1.020	0.117	0.944
		(0.004)	[0.984 - 1.110]	(0.003)	[0.980 - 1.060]	(0.003)	[0.891 - 0.998]
Diabetes, complicated	0.196	0.200	1.024	0.191	0.977	0.164	0.836
		(0.005)	[0.973 - 1.074]	(0.003)	[0.947 - 1.007]	(0.004)	[0.798 - 0.873]
Hypothyroidism	0.093	0.094	1.010	0.090	0.971	0.083	0.894
		(0.004)	[0.934 - 1.086]	(0.002)	[0.925 - 1.017]	(0.003)	[0.833 - 0.955]
Renal failure	0.086	0.091	1.059	0.075	0.877	0.069	0.805
		(0.004)	[0.976 - 1.142]	(0.002)	[0.832 - 0.922]	(0.003)	[0.747 - 0.863]
Liver disease	0.057	0.058	1.015	0.051	0.900	0.052	0.906
		(0.003)	[0.915 - 1.116]	(0.002)	[0.843 - 0.958]	(0.002)	[0.827 - 0.984]
Peptic ulcer disease, excluding bleeding	0.005	0.003	0.609	0.004	0.868	0.005	0.993
		(0.001)	[0.237 - 0.982]	(0.001)	[0.670 - 1.066]	(0.001)	[0.738 - 1.247]

Notes: This table presents the average characteristics among compliers, along with the overall sample averages. Following Abadie (2003), we estimate $E[H_i \mid \text{Compliers}]$ by regressing an interaction between each patient characteristic H and ED recommendation indicator $(H \cdot R^{ED})$ on ED recommendation indicator (R^{ED}) with the right-hand-side R^{ED} instrumented by the leave-out ED recommendation propensity on ED-UC margin. The average characteristics among compliers on UC-PC and PC-Home margins are estimated similarly. All regressions control for the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county) described in Appendix Table B3. Standard errors are clustered at the call center-by-call time level.

Table B22. Complier Characteristics (Continued)

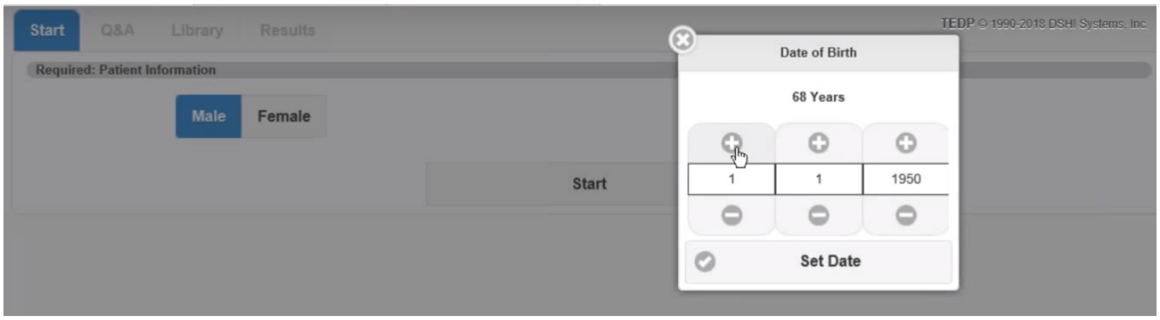
		Compliers (ED-UC)		Compliers (UC-PC)		Compliers (PC-Home)	
Variable	Overall Mean	Mean	Ratio	Mean	Ratio	Mean	Ratio
AIDS/HIV	0.005	0.005	1.141	0.006	1.344	0.004	0.855
		(0.001)	[0.770 - 1.512]	(0.001)	[1.098 - 1.589]	(0.001)	[0.590 - 1.121]
Lymphoma	0.008	0.009	1.105	0.007	0.850	0.006	0.744
		(0.001)	[0.845 - 1.364]	(0.001)	[0.707 - 0.994]	(0.001)	[0.541 - 0.947]
Metastatic cancer	0.008	0.008	1.020	0.006	0.740	0.007	0.855
		(0.001)	[0.710 - 1.330]	(0.001)	[0.596 - 0.883]	(0.001)	[0.645 - 1.066]
Solid tumor without metastasis	0.073	0.080	1.099	0.066	0.903	0.063	0.866
		(0.003)	[1.011 - 1.186]	(0.002)	[0.853 - 0.952]	(0.002)	[0.799 - 0.933]
Rheumatoid arthritis/collagen vascular diseases	0.033	0.036	1.103	0.033	1.020	0.033	1.018
		(0.002)	[0.973 - 1.234]	(0.001)	[0.940 - 1.099]	(0.002)	[0.915 - 1.122]
Coagulopathy	0.020	0.022	1.111	0.019	0.927	0.019	0.952
		(0.002)	[0.929 - 1.292]	(0.001)	[0.827 - 1.026]	(0.001)	[0.815 - 1.090]
Obesity	0.235	0.229	0.972	0.239	1.016	0.248	1.054
		(0.005)	[0.929 - 1.016]	(0.003)	[0.990 - 1.042]	(0.004)	[1.017 - 1.091]
Weight loss	0.026	0.027	1.058	0.020	0.793	0.021	0.834
		(0.002)	[0.896 - 1.221]	(0.001)	[0.710 - 0.876]	(0.001)	[0.722 - 0.946]

Notes: This table presents the average characteristics among compliers, along with the overall sample averages. Following Abadie (2003), we estimate $E[H_i \mid \text{Compliers}]$ by regressing an interaction between each patient characteristic H and ED recommendation indicator ($H \cdot R^{ED}$) on ED recommendation indicator (R^{ED}) with the right-hand-side R^{ED} instrumented by the leave-out ED recommendation propensity on ED-UC margin. The average characteristics among compliers on UC-PC and PC-Home margins are estimated similarly. All regressions control for the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county) described in Appendix Table B3. Standard errors are clustered at the call center-by-call time level.

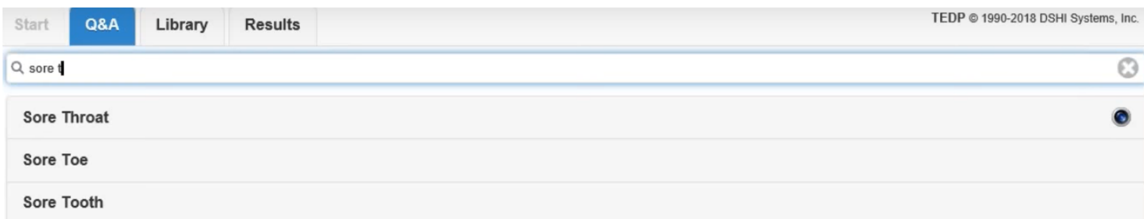
Table B23. Complier Characteristics (Continued)

Variable	Overall Mean	Compliers (ED-UC) Mean	Compliers (ED-UC) Ratio	Compliers (UC-PC) Mean	Compliers (UC-PC) Ratio	Compliers (PC-Home) Mean	Compliers (PC-Home) Ratio
Fluid and electrolyte disorders	0.060	0.065	1.089	0.047	0.779	0.047	0.791
		(0.003)	[0.985 - 1.193]	(0.002)	[0.727 - 0.832]	(0.002)	[0.722 - 0.860]
Blood loss anemia	0.007	0.007	1.064	0.007	1.003	0.005	0.705
		(0.001)	[0.759 - 1.370]	(0.001)	[0.831 - 1.175]	(0.001)	[0.508 - 0.901]
Deficiency anemia	0.050	0.049	0.972	0.043	0.855	0.042	0.837
		(0.003)	[0.866 - 1.079]	(0.002)	[0.795 - 0.915]	(0.002)	[0.755 - 0.918]
Alcohol abuse	0.109	0.107	0.985	0.106	0.976	0.108	0.991
		(0.004)	[0.916 - 1.053]	(0.002)	[0.936 - 1.016]	(0.003)	[0.934 - 1.049]
Drug abuse	0.068	0.067	0.989	0.067	0.981	0.063	0.931
		(0.003)	[0.897 - 1.081]	(0.002)	[0.928 - 1.034]	(0.002)	[0.860 - 1.002]
Psychoses	0.026	0.024	0.930	0.023	0.879	0.026	0.988
		(0.002)	[0.780 - 1.080]	(0.001)	[0.793 - 0.964]	(0.002)	[0.867 - 1.109]
Depression	0.351	0.364	1.037	0.375	1.069	0.348	0.992
		(0.006)	[1.005 - 1.070]	(0.004)	[1.048 - 1.089]	(0.005)	[0.965 - 1.019]

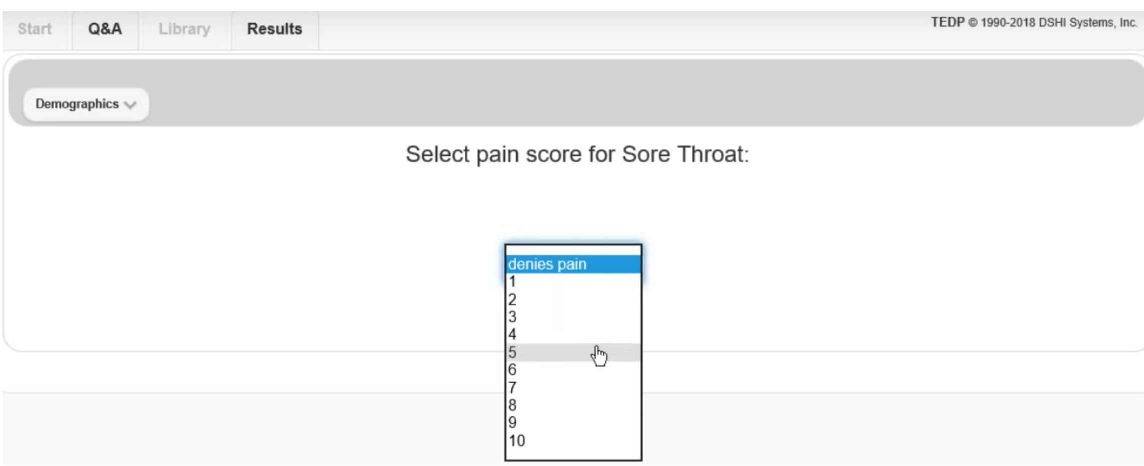
Notes: This table presents the average characteristics among compliers, along with the overall sample averages. Following Abadie (2003), we estimate $E[H_i \mid \text{Compliers}]$ by regressing an interaction between each patient characteristic H and ED recommendation indicator ($H \cdot R^{ED}$) on ED recommendation indicator (R^{ED}) with the right-hand-side R^{ED} instrumented by the leave-out ED recommendation propensity on ED-UC margin. The average characteristics among compliers on UC-PC and PC-Home margins are estimated similarly. All regressions control for the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county) described in Appendix Table B3. Standard errors are clustered at the call center-by-call time level.



(a) Initial Inputs (Demographics)

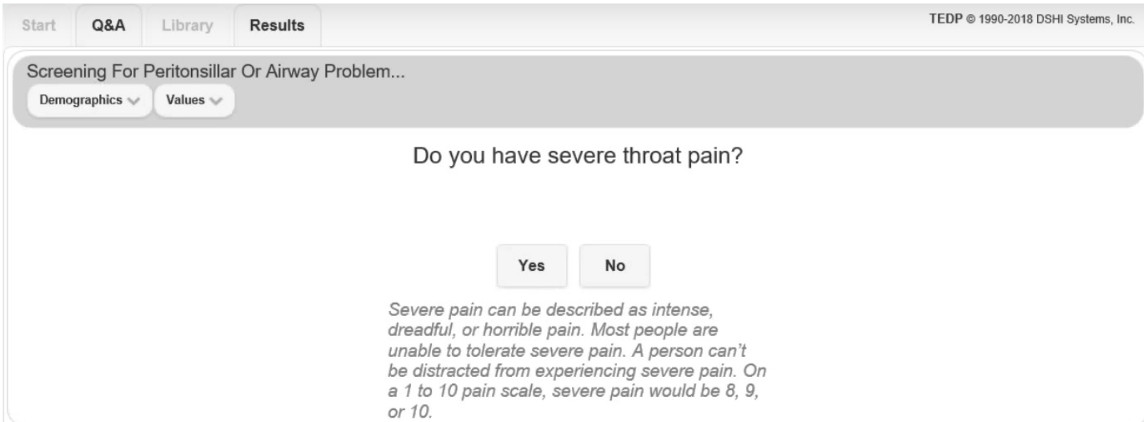


(b) Initial Inputs (Chief Complaint)



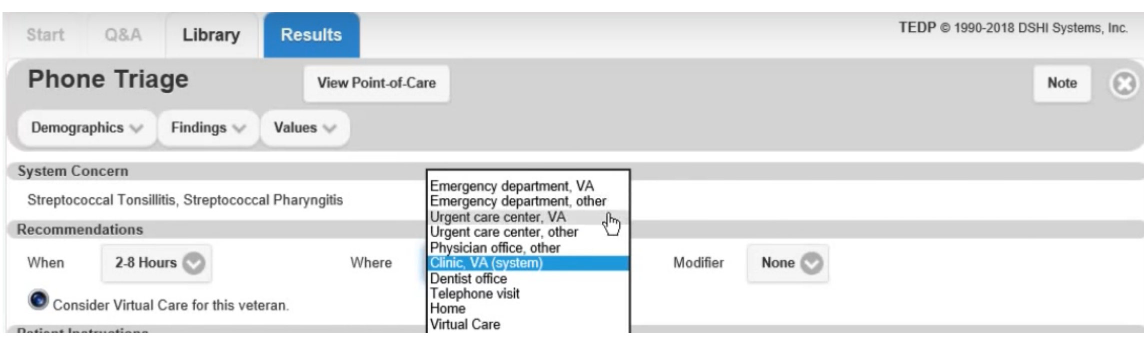
(c) Initial Inputs (Pain Scale)

Figure B1. Algorithm and Triage Process



(d) Clinical Questions

(e) Algorithm Recommendations



(f) Nurse Recommendations

Figure B1. Algorithm and Triage Process (Continued)

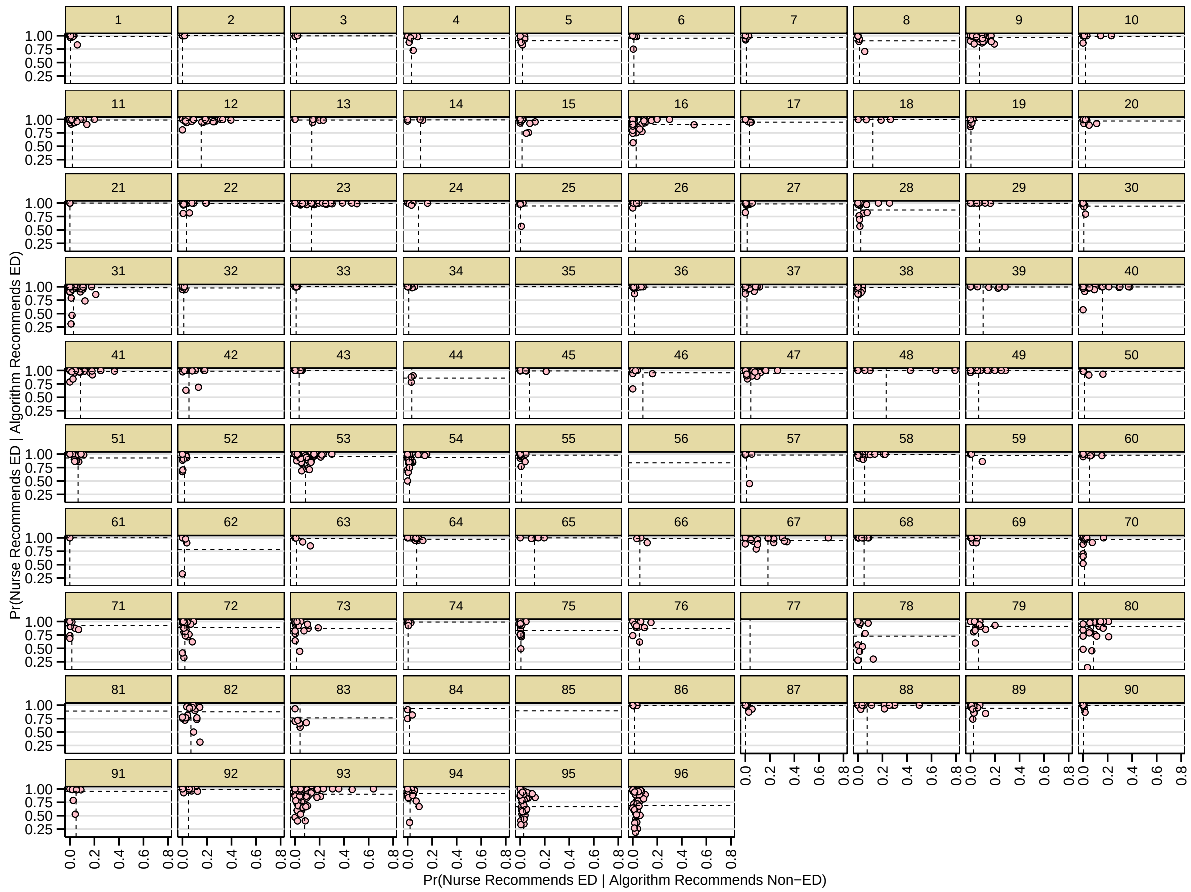


Figure B2. Triage Pattern by Call Center

Notes: This figure shows variations in the propensity to recommend ED care, conditional on algorithm disposition, between nurses within call centers. Each panel corresponds to a call center, and each circle represents a nurse within call centers. For each panel (call center), we plot each nurse's propensity to recommend ED conditional on algorithm disposition being ED care on the y-axis against the nurse's propensity to recommend ED conditional on algorithm disposition being non-ED care on the x-axis.

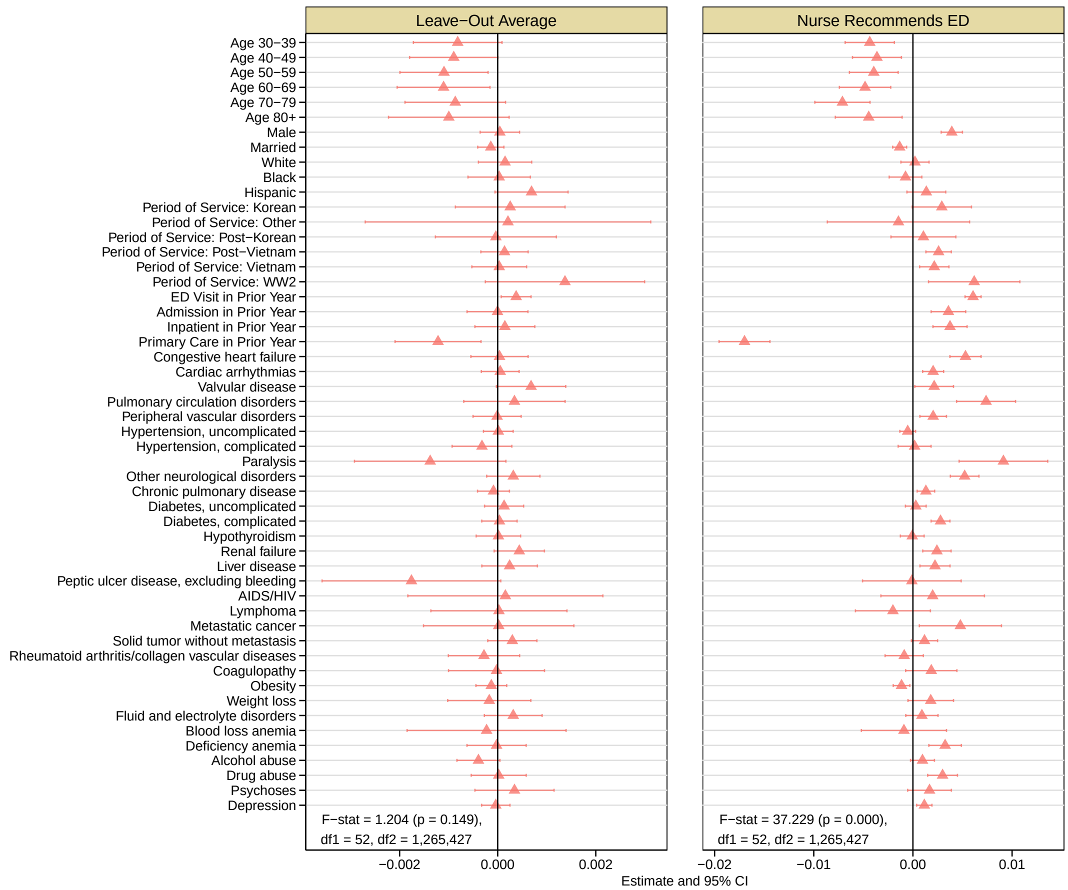


Figure B3. Balance Check (F-Test) (ED - UC)

Notes: This figure tests our assumption that the assigned nurse's ED recommendation propensity is independent of the patient's characteristics, conditional on the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county) described in Appendix Table B3. The left panel plots the estimated coefficients of a multivariate regression of the leave-one-patient-out measure Z_i^{ED} on the hold-out variables H_i, conditional on the baseline controls X_i^0. For this balance check, the leave-out ED tendency only residualizes for the baseline controls X_i^0 in equation (5) and does not partial out the hold-out variables H_i. The right panel plots the estimated coefficients of a multivariate regression of the (actual) ED recommendation indicator R_i^{ED} on the hold-out variables H_i, conditional on the baseline controls X_i^0. The F-statistic and p-value are reported in the figure to test for the joint significance of the hold-out variables. The F-test degrees of freedom are 52 and 1,265,427.

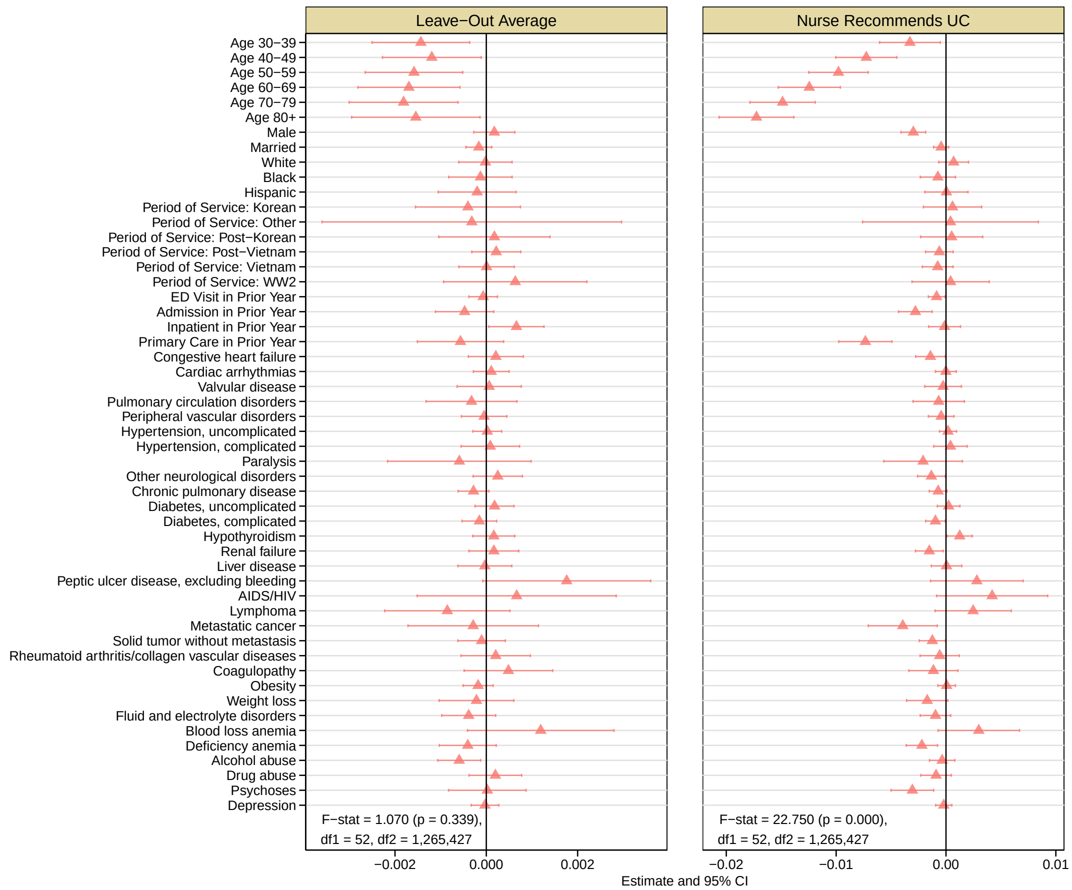


Figure B4. Balance Check (F-Test) (UC - PC)

Notes: This figure tests our assumption that the assigned nurse's UC recommendation propensity is independent of the patient's characteristics, conditional on the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county) described in Appendix Table B3. The left panel plots the estimated coefficients of a multivariate regression of the leave-one-patient-out measure Z_i^{UC} on the hold-out variables H_i, conditional on the baseline controls X_i^0. For this balance check, the leave-out UC tendency only residualizes for the baseline controls X_i^0 in equation (5) and does not partial out the hold-out variables H_i. The right panel plots the estimated coefficients of a multivariate regression of the (actual) UC recommendation indicator R_i^{UC} on the hold-out variables H_i, conditional on the baseline controls X_i^0. The F-statistic and p-value are reported in the figure to test for the joint significance of the hold-out variables. The F-test degrees of freedom are 52 and 1,265,427.

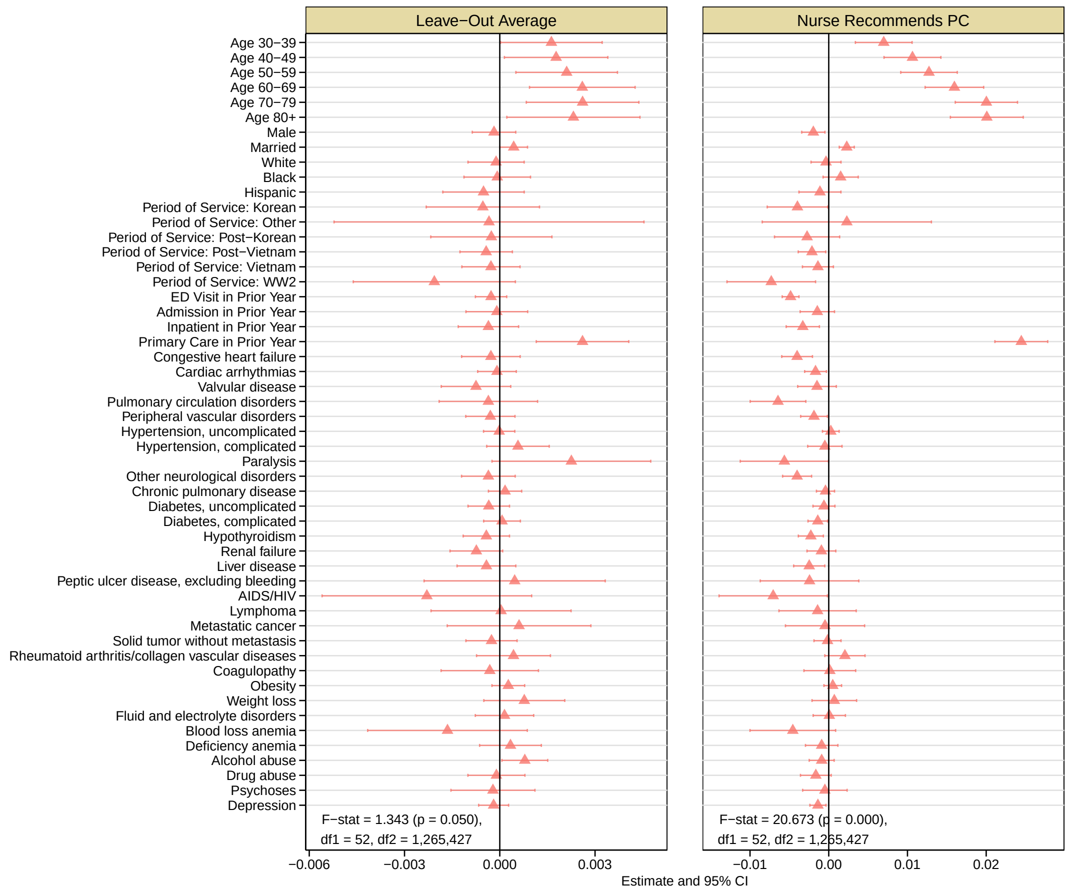


Figure B5. Balance Check (F-Test) (PC - Home)

Notes: This figure tests our assumption that the assigned nurse's UC recommendation propensity is independent of the patient's characteristics, conditional on the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county) described in Appendix Table B3. The left panel plots the estimated coefficients of a multivariate regression of the leave-one-patient-out measure Z_i^{PC} on the hold-out variables H_i, conditional on the baseline controls X_i^0. For this balance check, the leave-out PC tendency only residualizes for the baseline controls X_i^0 in equation (5) and does not partial out the hold-out variables H_i. The right panel plots the estimated coefficients of a multivariate regression of the (actual) PC recommendation indicator R_i^{PC} on the hold-out variables H_i, conditional on the baseline controls X_i^0. The F-statistic and p-value are reported in the figure to test for the joint significance of the hold-out variables. The F-test degrees of freedom are 52 and 1,265,427.

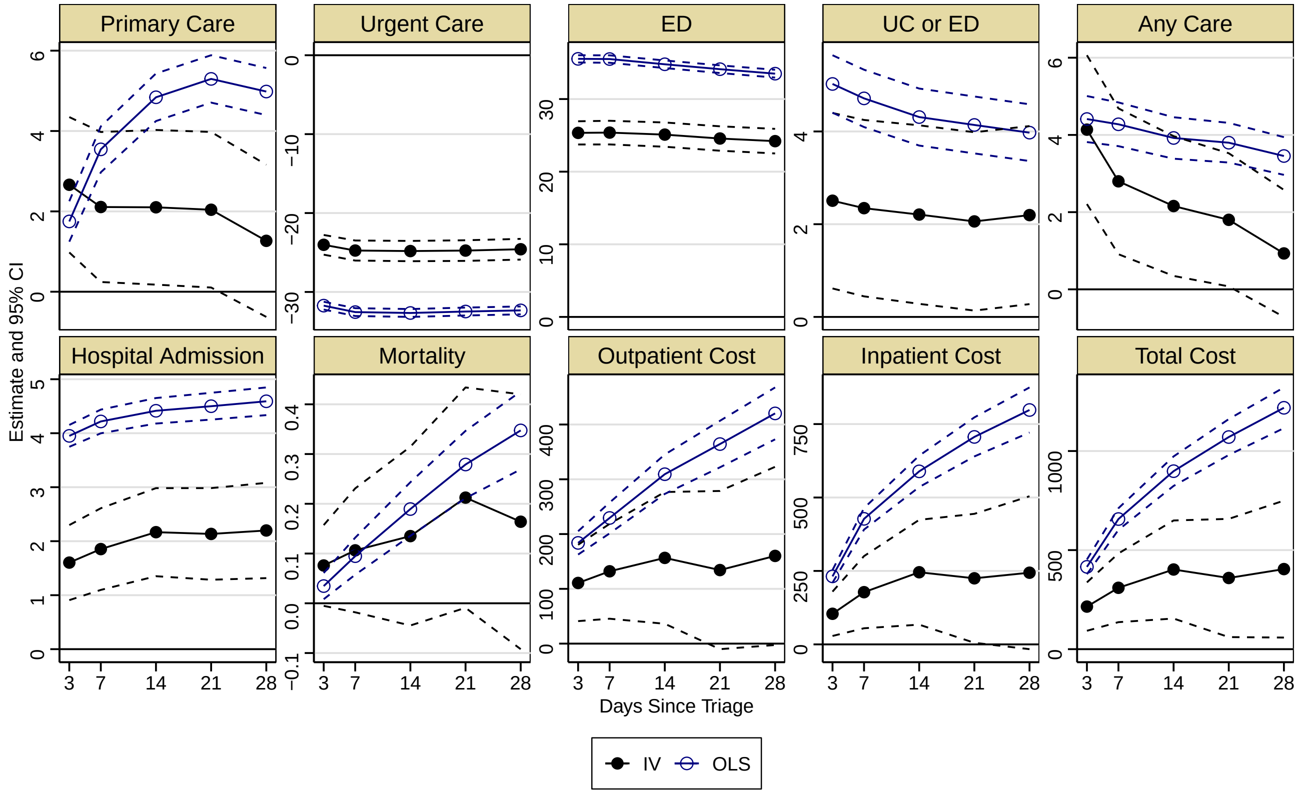


Figure B6. IV Results: Effect of Nurse ED Recommendation (Relative to UC Recommendation)

Notes: This figure reports TSLS (black solid circles) and OLS (navy hollow circles) estimates of the effect of nurse's ED recommendation. All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. The dashed lines show the 95% confidence intervals based on standard errors clustered at the call center-by-call time level. The y-axis scales vary across panels.

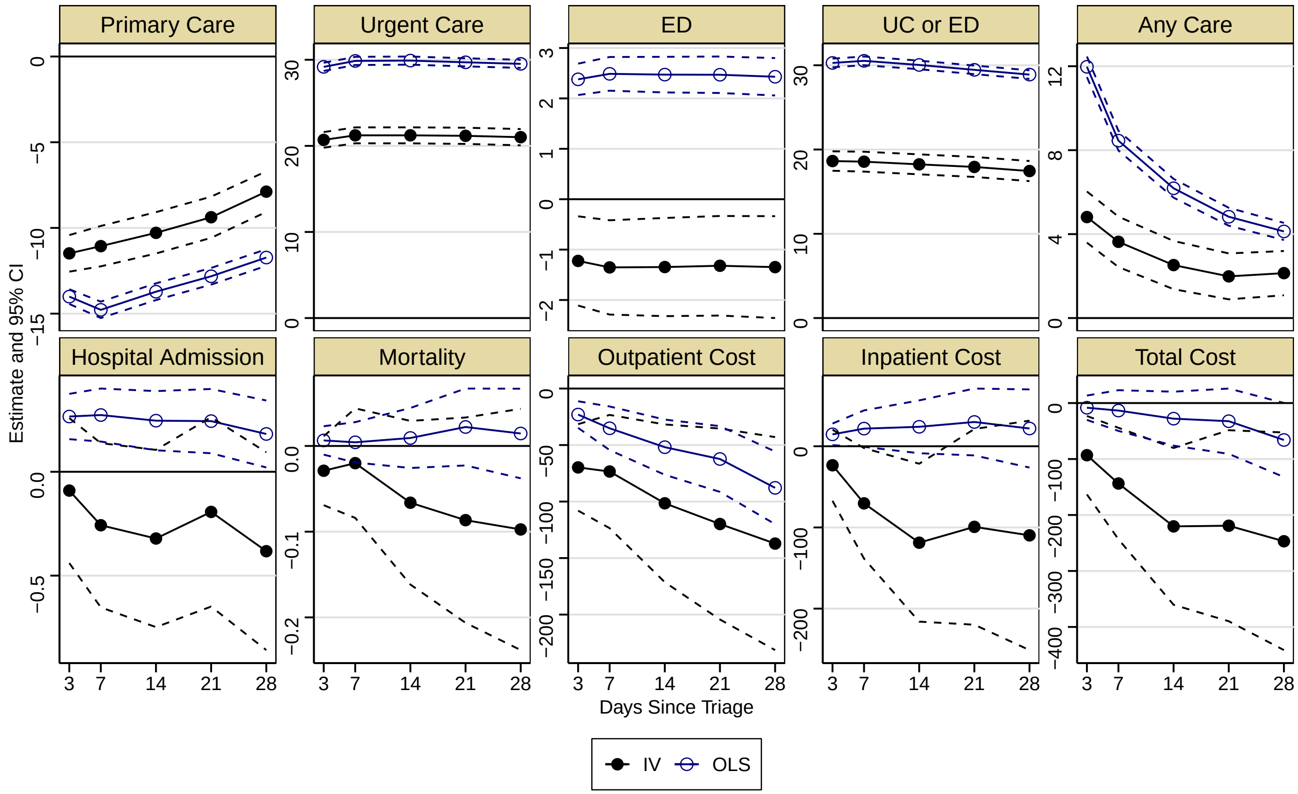


Figure B7. IV Results: Effect of Nurse UC Recommendation (Relative to PC Recommendation)

Notes: This figure reports TSLS (black solid circles) and OLS (navy hollow circles) estimates of the effect of nurse's UC recommendation. All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. The dashed lines show the 95% confidence intervals based on standard errors clustered at the call center-by-call time level. The y-axis scales vary across panels.

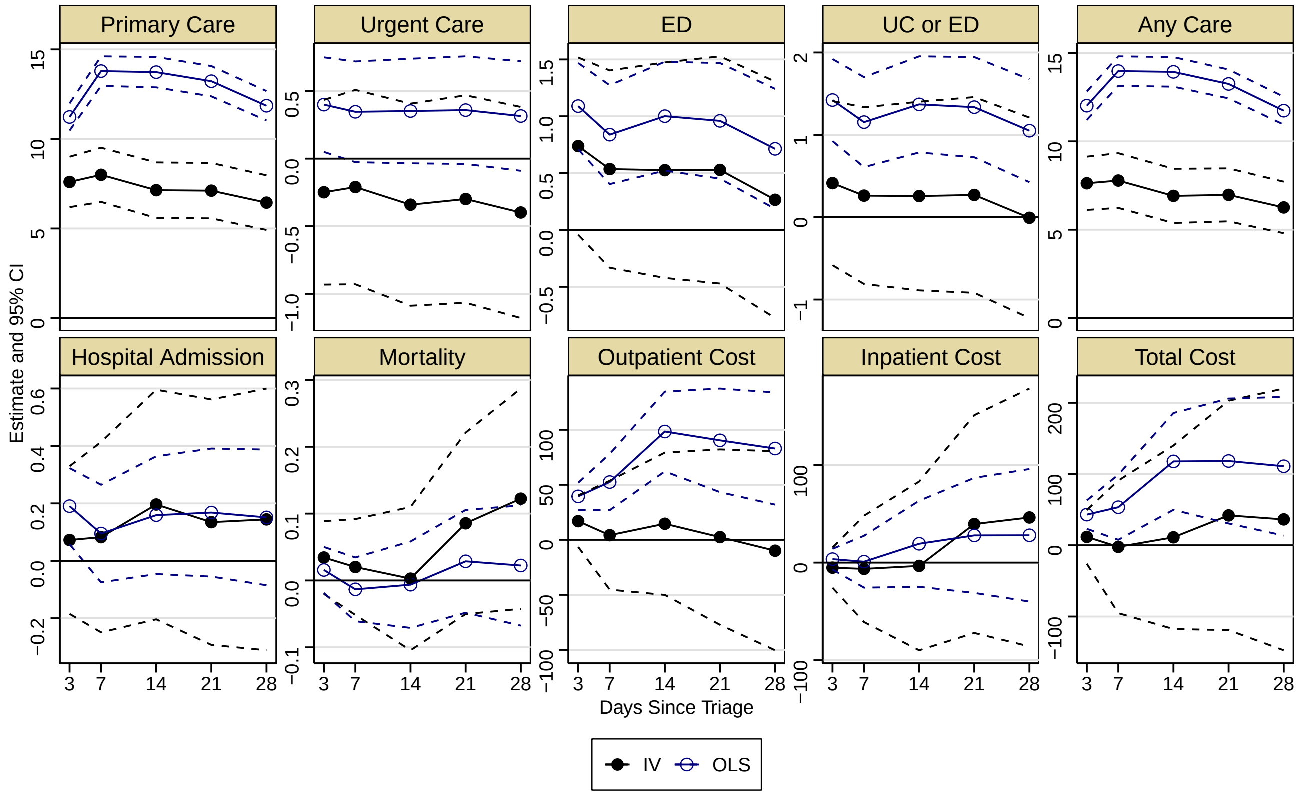


Figure B8. IV Results: Effect of Nurse PC Recommendation (Relative to Home Recommendation)

Notes: This figure reports TSLS (black solid circles) and OLS (navy hollow circles) estimates of the effect of nurse's PC recommendation. All binary outcomes (primary care - mortality) are multiplied by 100 so that estimates can be interpreted as percentage points. Costs are measured in US dollars. In addition to the baseline controls X_i^0 (call center-by-call time, call center-by-algorithm disposition, and patient county), all regressions control for the hold-out variables (age, sex, marital status, race/ethnicity, period of service, prior healthcare utilization, and prior diagnoses) for precision. The baseline and hold-out controls, X_i^0 and H_i, are described in Appendix Tables B3 and B4, respectively. The dashed lines show the 95% confidence intervals based on standard errors clustered at the call center-by-call time level. The y-axis scales vary across panels.