Appendix for "Doctor Decision Making and Patient Outcomes"

Janet Currie[*], W Bentley MacLeod[†] and Kate Musen[‡]

January 12, 2026

[*]Princeton University and NBER
[†]Yale University and NBER
[‡]Columbia University

1 Appendix for Theory in Section 2.

This appendix lays out the detailed proofs of the model discussed in the text. The model begins with a population of patients where patient $i \in \mathscr{N}_j$ seeks treatment from doctor $j \in J$. It is assumed that neither patient or physician is sure which is the best choice. The doctor chooses between a non-intensive or an intensive treatment, denoted by $t_{ij} \in \{NI, I\}$. It is assumed that there is a best choice for the patient given by their *unobserved* state $\alpha_i \in \{L, H\}$. If $\alpha_i = L$, then the patient is low risk, and hence a non-intensive treatment is appropriate, while $\alpha_i = H$ implies that the patient is high risk, and an intensive treatment is more appropriate. This modeling strategy is based on Savage (1972 (first published 1954)'s model of Bayesian choice in which the goal of the model is not to provide a complete representation of the patient's condition, but to highlight only those aspects of a patient's state that are relevant for the decision at hand.[1]

Let the fraction of patients in $\mathscr{N}_j$ for which the doctors believe are low risk, $\alpha_i = L$, be given by $p_{Lj} \in (0, 1)$, while a fraction $p_{Hj} = 1 - p_{Lj}$ the doctors suppose are in the high risk category, $\alpha_i = H$. Doctor j cannot perfectly observe the patient's state, but after examining the patient, observes a signal:

$$T_{ij} = \begin{cases} 1 + \epsilon_i/\gamma_j, & \alpha_i = H, \\ -1 + \epsilon_i/\gamma_j, & \alpha_i = L, \end{cases} \tag{1}$$

where $\epsilon_i \sim N(0, 1)$ and γ_j is the diagnostic skill of the doctor. An increase in diagnostic skill implies a more precise assessment of a person's state. The doctor is never perfectly sure of the patient's condition since it is observed with error.

T_{ij} is increasing with α_i so it follows that the doctor's decision criterion for the treatment choice $t_{ij} \in \{NI, I\}$ takes the form:

$$t_{ij} = \begin{cases} I, & T_{ij} \geq \tau_j, \\ NI, & T_{ij} < \tau_j, \end{cases}$$

where the doctor's decision threshold is given by τ_j.

The quality of diagnosis can be measured by the likelihood that a patient is assigned to the correct treatment. There are two measures of performance corresponding to whether patients correctly or incorrectly receive the intensive treatment. Suppose a patient is in state $\alpha_i = H$ and hence should be assigned to intensive treatment. The probability that the patient correctly receives the intensive treatment, given the doctor's decision threshold, τ_j, and diagnostic skill γ_j, the *true positive rate or TPR* is given by:

$$\begin{aligned} TPR(\tau_j, \gamma_j) &\equiv \Pr[T_{ij} \geq \tau_j | \alpha_i = H], \\ &= \Pr[1 + \epsilon/\gamma_j \geq \tau_j], \\ &= F(\gamma_j (1 - \tau_j)), \end{aligned} \tag{2}$$

where $F(\cdot)$ is the Normal cumulative probability distribution.

The probability that a patient who needs non-intensive treatment ($\alpha_i = L$) receives intensive treatment

[1] See the discussion in Chapter 2 of MacLeod (2022).

is given by the *false positive rate or FPR*:

$$\begin{aligned}FPR(\tau_j, \gamma_j) &\equiv \Pr[T_{ij} \geq \tau_j | \alpha = L] \\ &= \Pr[-1 + \epsilon/\gamma_j \geq \tau_j] \\ &= F(\gamma_j(-1 - \tau_j)).\end{aligned} \tag{3}$$

The Doctor's Decision Threshold (τ_j^*)

This section derives the doctor's decision threshold, τ_j^*, given a doctor's preferences and diagnostic skill, γ_j, and the consequences for a patient getting the inappropriate treatment. It is assumed that the doctor's utility is given by the well-being of the patient plus payments that might distort this decision. In particular, the doctor would make the socially efficient solution if their preferences are given by the patient utility less the cost of treatment. Given patient type $\alpha_i \in \{H, L\}$, doctor $j's$ utility from administering treatment $t \in \{NI, I\}$ is given by:

$$U_{\alpha tj} = u_{\alpha tj} + \delta_{tj}, \tag{4}$$

where $u_{\alpha tj}$ is the expected medical benefit to a patient of type $\alpha_i \in \{L, H\}$, getting treatment $t \in \{NI, N\}$ from doctor j. For the same patient type, the outcome $u_{\alpha tj}$ can differ by doctor, a variation that we associate with a doctor's *procedural skill.* Additional factors that affect treatment, such as a payment that the doctor receives from administering the treatment, are captured by δ_{tj}. We normalize this term by setting $\delta_{Lj} = 0$ and letting $\delta_j = \delta_{IJ} \in \Re$ be the pecuniary return (that can be positive or negative) from doing the intensive procedure.

For a type $\alpha_i = L$ patient a non-intensive treatment is preferred hence $u_{LNIj} > u_{LIj}$, while for type $\alpha_i = H$ intensive treatment is preferred and hence $u_{HIj} > u_{HNIj}$.

Let $\Delta_{HIj} = \{U_{HIj} - U_{HNIj}\}$ and $\Delta_{LNIj} = \{U_{LNIj} - U_{LIj}\}$ be the increase in utility for patients who receive the appropriate treatment. Notice that:

$$\begin{aligned}\Delta_{HIj} &= \{u_{HIj} - u_{HNIj}\} + \delta_{Ij}, \\ \Delta_{LNIj} &= \{u_{LNIj} - u_{LIj}\} - \delta_{Ij}.\end{aligned}$$

Hence we have the following lemma:

Lemma 1. *Regardless of the signal T_{ij}, when $\delta_{Ij} > u_{LNIj} - u_{LIj} > 0$ then the doctor j always provides the intensive treatment, and when $\delta_{IJ} < -\{u_{HIj} - u_{HNIj}\} < 0$, then the doctor always provides the non-intensive treatment.*

Proof. The proof follows from the fact that regardless of the information received, when $\delta_{Ij} > u_{LNIj} - u_{LIj} > 0$, then $\Delta_{LNIj} < 0$ and hence the doctor would choose the intensive treatment for the low type. This condition also implies that $\Delta_{HIj} > 0$, hence regardless of type, the intensive procedure is preferred. A similar argument applies when $\delta_{IJ} < -\{u_{HIj} - u_{HNIj}\} < 0$. □

This result points out that if the pecuniary returns for choice (δ_{It}) is either very positive or very negative, then the physician will always make the same treatment choice regardless of the signal. Thus in order to observe variation in treatment choice as a function of the doctor's information T_{ij}, the absolute value of pecuniary incentives cannot be too large. In the evidence we review, insensitivity to variation in observables may be due to either lack of an effect, or excess pecuniary returns.

The doctor's *ex ante* belief regarding the appropriate treatment for a patient in this pool of potential patients is given by:

$$p_{Hj} = \Pr\left[\alpha_i = H | j\right]$$

while the belief that the probability that $\alpha_i = L$ is $p_{Lj} = 1 - p_{Hj}$.

It is worth emphasizing that p_{Hj} is the doctor's subjective belief that may not necessarily equal the true probability, p_H. In general p_{Hj} is correlated with p_H, but there can be significant variation due to a number of doctor specific factors, including poor judgment and doctor biases.

The expected utility of doctor j who chooses decision threshold τ_j for patient i is given by:

$$\begin{aligned}
u_{ij}\left(\tau_j, \gamma_j\right) &= \left(\left(u_{HIj} + \delta_j\right) \Pr\left[T_{ij} \geq \tau_j | \alpha = H\right] + u_{HNI1} \Pr\left[T_{ij} < \tau_j | \alpha = H\right]\right) \Pr\left[\alpha = H | j\right] \\
&+ \left(\left(u_{LIj} + \delta_j\right) \Pr\left[T_{ij} \geq \tau_j | \alpha = L\right] + u_{LNIj} \Pr\left[T_{ij} < \tau_j | \alpha = L\right]\right) \Pr\left[\alpha = L | j\right] \\
&= \left(u_{HNIj} + \Delta_{HIj} \Pr\left[T_{ij} \geq \tau_j | \alpha = H\right]\right) p_{Hj} \\
&+ \left(u_{LIj} - \Delta_{LNIj} \Pr\left[T_{ij} \geq \tau_j | \alpha = L\right]\right) p_{Lj}, \\
&= u_j^0 + \Delta_{HIj} TPR\left(\tau_j, \gamma_j\right) \times p_{Hj} - \Delta_{LNIj} FPR\left(\tau_j, \gamma_j\right) \times p_{Lj},
\end{aligned} \tag{5}$$

where:

$$\begin{aligned}
u_j^0 &= u_{HNIj} \Pr\left[\alpha_i = H | j\right] + u_{LIj} \Pr\left[\alpha_i = L | j\right], \\
&= u_{HNIj} \times p_{Hj} + u_{LIj} \times p_{Lj}.
\end{aligned}$$

The quantity u_j^0 is the *worst* possible medical payoff for doctor j with any of their patients. It is the outcome when all individuals with type $\alpha = H$ are given the non-intensive treatment, and all type $\alpha = L$ individuals are given the intensive treatment. The payoff to a doctor can now be written in terms of the expected gains, beliefs and expected patient outcomes.

The decision threshold for each physician is $\tau_j^* = \arg\max_{\tau \in \Re} u_{ij}\left(\tau, \gamma_j\right)$. The solution is given by the following proposition.

Proposition 1. *The doctor's decision threshold solves* $\tau_j^* = \arg\max_{\tau \in \Re} u_{ij}\left(\tau, \gamma_j\right)$. *Suppose the pecuniary return satisfies* $\delta_j \in \left(-\Delta_{HIj}, \Delta_{LNI}\right)$ *(the conditions for lemma 1 are not satisfied), then* τ_j^* *satisfies the likelihood ratio condition:*

$$L\left(\tau_j^*, \gamma_j\right) = \frac{\Delta_{LNIj}}{\Delta_{HIj}} \times \frac{p_{Lj}}{p_{Hj}}, \tag{6}$$

where the likelihood ratio is given by:

$$L\left(\tau_j^*, \gamma_j\right) = \frac{f\left(\gamma_j\left(1 - \tau_j^*\right)\right)}{f\left(\gamma_j\left(-1 - \tau_j^*\right)\right)},$$

and $f(\cdot)$ *is the Normal density function.*

Proof. The solution satisfies the first order condition:

$$
\begin{aligned}
0 &= \partial u_{ij}(\tau,\gamma_j)/\partial\tau,\\
&= (u_{HIj}+\delta_j)\,\partial TPR(\tau,\gamma_j)/\partial\tau \times p_{Hj} - \Delta_{LNIj}\partial FPR(\tau,\gamma_j)/\partial\tau \times p_{Lj},\\
&= \Delta_{HIj} f(\gamma_j(1-\tau))(-\gamma_j)\times p_{Hj} - (\Delta_{LNIj}-\delta_j)\, f\left(\gamma_j\left(-1-\tau_j^*\right)\right)(-\gamma_j)\times p_{0j}.
\end{aligned}
$$

The conditions on δ_j ensure that the ratio on the right of (6) is strictly positive. The first order condition follows from the last line. The first order conditions imply a unique decision threshold, τ_j^* satisfying:

$$
L\left(\tau_j^*,\gamma_j\right) = \frac{f\left(\gamma_j\left(1-\tau_j^*\right)\right)}{f\left(\gamma_j\left(-1-\tau_j^*\right)\right)} = \frac{\Delta_{LNIj}}{\Delta_{HIj}} \times \frac{p_{Lj}}{p_{Hj}},
$$

or:

$$
\frac{\partial TPR(\tau,\gamma_j)/\partial\tau}{\partial FPR(\tau,\gamma_j)/\partial\tau} = \frac{\Delta_{LNIj}}{\Delta_{HIj}} \times \frac{p_{Lj}}{p_{Hj}}
$$

When $\Delta_{HIj} < 0$ then $\Delta_{LNIj} > 0$ and doctor always does the non-intensive procedure. The converse holds when $\Delta_{LNIj} < 0$. □

The first order condition characterizes the global optimum, which follows from the Neyman-Pearson lemma showing that likelihood ratios are the most powerful form of hypothesis test (Neyman and Pearson (1933)).[2] When $\delta_j \in (-\Delta_{HIj}, \Delta_{LNI})$ the doctor faces uncertainty regarding choice. When this condition is not satisfied we say that the doctor is certain regarding her choice (either NI or I regardless of the test result). The model yields a closed form solution for the doctor's diagnostic rule τ_j^*, given by the following proposition:

Proposition 2. *When the doctor is uncertain, the decision threshold is given by:*

$$
\tau_j^* = b_j^*/\gamma_j^2, \tag{7}
$$

where $b_j^* \equiv \left(\ln\left(\Delta_{LNIj}/\Delta_{HIj}\right) + \ln\left(p_{Lj}/p_{Hj}\right)\right)/2$.

Proof. Observe:

$$
\begin{aligned}
\frac{f\left(\gamma_j\left(1-\tau_j^*\right)\right)}{f\left(\gamma_j\left(-1-\tau_j^*\right)\right)} &= \frac{\exp - \left\{\gamma_j\left(1-\tau_j^*\right)\right\}^2/2}{\exp - \left\{\gamma_j\left(-1-\tau_j^*\right)\right\}^2/2}\\
&= \exp\left(-\left\{\gamma_j\left(1-\tau_j^*\right)\right\}^2 + \left\{\gamma_j\left(-1-\tau_j^*\right)\right\}^2\right)/2
\end{aligned}
$$

Taking the logarithm of the first-order condition gives us:

$$
\begin{aligned}
\left(-\left\{\gamma_j\left(1-\tau_j^*\right)\right\}^2 + \left\{\gamma_j\left(-1-\tau_j^*\right)\right\}^2\right)/2 &= 2\times b_j,\\
\left(-\left\{\gamma_j^2\left(1-2\tau_j^* + \left(\tau_j^{*2}\right)^2\right)\right\} + \gamma_j^2\left(1+2\tau_j^* + \left(\tau_j^{*2}\right)^2\right)\right) &= 4b_j\\
4\gamma_{ij}^2\tau_{ij} &= 4b_j,
\end{aligned}
$$

[2] Feng et al. (2023) highlight the link between rational choice and the Neyman-Pearson lemma.

giving the desired result (7). □

Equation (7) shows that the doctor's decision threshold depends on diagnostic skill, γ_j, the relative desirability of non-intensive and intensive treatments for the two types of patients, $\Delta_{LNIj}/\Delta_{HIj}$, and the doctor's beliefs about the relative proportions of patient types, p_{Lj}/p_{Hj}, in the population. When the doctor believes that there is a higher probability that the patient needs non-intensive treatment, she adopts a higher threshold resulting in less use of the intensive treatment. Similarly, if the relative benefit from intensive treatment is higher, then this results in a lower threshold.

As diagnostic skill increases, both patient types are more likely to be allocated to the appropriate treatment. The doctor's decision rule entails patients getting the appropriate treatment with probability close to one as diagnostic skill increases. Conversely, as diagnostic skill falls, the b_j term dominates. When $b_j > 0$, treatment is biased in favor of the non-intensive treatment and the probability that patients are treated with the non-intensive procedure rises as diagnostic skill falls. When $b_j < 0$, treatment is biased in favor of intensive treatment and the probability of intensive treatment rises as diagnostic skill falls. In effect, as diagnostic skill falls, physicians choose the treatment that they believe is best for most of their patients. These observations are summarized in the following proposition:

Proposition 3. *For a doctor who is uncertain of the best course of action (b_{ij} is finite), then as diagnostic skill increases, each patient is more likely to receive treatment appropriate for their type. More precisely:*

$$\lim_{\gamma_j \to \infty} \tau_j^* = 1/2,$$

$$\lim_{\gamma_j \to \infty} u_{ij}^* = \begin{cases} u_{HIj}, & \text{if } \alpha_i = H, \\ u_{LNIj}, & \text{if } \alpha_i = L. \end{cases}$$

As diagnostic skill falls, all patients get the same treatment depending upon the sign of the decision shifter, b_j:

$$\lim_{\gamma_j \to 0} \tau_j^* = \begin{cases} \infty, & \text{if } b_j > 0, \\ 1/2, & \text{if } b_j = 0 \\ -\infty, & \text{if } b_j < 0. \end{cases}$$

$$\lim_{\gamma_j \to 0} u_{ij}^* = \begin{cases} u_{HNIj}, & \text{if } \alpha_i = H, b_j > 0, \\ u_{LNIj}, & \text{if } \alpha_i = L, b_j > 0, \\ \left(u_{HNIj} + u_{HIj}\right)/2, & \text{if } \alpha_i = H, b_j = 0, \\ \left(u_{LNIj} + u_{LIj}\right)/2, & \text{if } \alpha_i = L, b_j = 0, \\ u_{HIj}, & \text{if } \alpha_i = H, b_j < 0, \\ u_{LIj}, & \text{if } \alpha_i = L, b_j < 0. \end{cases}$$

Proof. The proof of this proposition follows from equation (7). □

1.1 Identifying the Doctor Diagnostic threshold, Diagnostic Skill, and Procedural Skill From Data

Proposition 4. *Given points* (TPR_j, FPR_j) *on an ROC curve generated by Normal errors, there is a unique solution* (τ_j, γ_j) *to:*

$$TPR_j = F\left(\gamma_j\left(1-\tau_j\right)\right),$$
$$FPR_j = F\left(\gamma_j(-1-\tau_j)\right).$$

Proof. Since $(TPR_j, FPR_j) \in (0,1)^2$, we have:

$$\gamma_j\left(1-\tau_j\right) = F^{-1}\left(TPR_j\right), \tag{8}$$
$$\gamma_j(-1-\tau_j) = F^{-1}\left(FPR_j\right). \tag{9}$$

Plugging (9) into (8) we get:

$$\begin{aligned}\gamma_j\left(1-\tau_j\right) &= \gamma_j - \gamma_j\tau_j,\\ &= 2\gamma_j + F^{-1}\left(FPR_j\right),\end{aligned}$$

and hence: -

$$\gamma_j = (F^{-1}\left(TPR_j\right) - F^{-1}\left(FPR_j\right))/2.$$

It must be the case that $\gamma_j > 0$ since from the properties of ROC curves we have $TPR_j - FPR_j > 0$ and the fact that the cumulative distribution function $F()$ is strictly increasing. Using (9) we get:

$$\tau_j = -1 - F^{-1}\left(FPR_j\right)/\gamma_j.$$

□

Abaluck et al. (2016)

The context for Abaluck et al. (2016) is ordering computerized tomography (CT) scans to test for a pulmonary embolism (PE). The use of scans is expensive, and while a pulmonary embolism is a serious condition. The goal of the paper is to ask whether or not there is excessive use of CT scans? In the context of our model, a CT scan is an intensive procedure, hence $t_{ij} = I$ if a doctor j orders a scan for patient i. The unobserved state is whether a person has a PE ($\alpha_i = H$), or does not ($\alpha_i = L$). The goal is to have a true positive rate of 1, which ensures that all individuals with a PE are tested and treated. However, the test is expensive and it is not always possible for the doctor to correctly assess the patient's condition. In general one expects to have a TPR < 1 and a FPR > 0.

The goal of the paper is to assess the extent to which the decision threshold varies between doctors, and the extent to which doctors process information correctly. The challenge is that, unlike Chan, Gentzkow and Yu (2022), patients are not randomly allocated to doctors, and hence the average severity of the cases can vary by doctor. The authors address this by specifying and estimating a structural model of physician decision making. It is assumed that the signal on the condition of patient i is the expected probability that has a PE:

$$
\begin{aligned}
T_{ij} &= \Pr\left[\alpha = H | i, j\right] && (10)\\
&= \vec{x}_i \beta + a_j + \eta_{ij}, && (11)\\
&\equiv \rho_j\left(\vec{x}_i\right) + \eta_{ij} && (12)
\end{aligned}
$$

where η_{ij} is information observed by the doctor, but not the econometrician, and $\rho_j\left(\vec{x}_i\right) = \Pr\left[\alpha_i = H | \vec{x}_i, j\right]$ is the probability that the individual has PE conditional upon the observables $\vec{x}_i$ and the population of patients treated by docter j.

In this case, the decision threshold, τ_j^*, defines the cutoff probability for ordering a CT-scan. When the probability of a PE is greater than τ_{j*} then the doctor orders a CT-scan.

A key feature of this specification is the inclusion of the fixed effect a_j that captures the fact that doctors may face different distributions of patients. If patients were randomly allocated, then $a_j = a$ for some constant a for all doctors. We shall show that the challenge will be to separately estimate both a_j and the doctor's decision threshold τ_{j*}.

The authors suppose that the distribution of η_{ij} is a known *i.i.d.* distribution that is independent of patient observables $\vec{x}_i$, and with distribution $\eta_{ij} \sim H(\cdot)$, where $H(\eta) \equiv \Pr\left[\eta_{ij} \leq \eta\right]$ is the cumulative probability distribution. It is assumed $E\left\{\eta_{ij}\right\} = 0$. The online appendix of Abaluck et al. (2016) provides a parametric specification for $H(\cdot)$ (a mixture of a Uniform and Bernoulli distribution) and it is shown that it can be estimated from the data. For the current discussion, it is assumed that it is known.

Given the single index T_{ij}, Abaluck et al. (2016) and doctor practice style characterized by a threshold τ_j^*, a test is ordered whenever it is suspected that the probability of a PE is greater than τ_j^*:

$$
t_{ij} = \begin{cases} I, & T_{ij} \geq \tau_j^*, \\ NI, & T_{ij} \leq \tau_j^*, \end{cases}
$$

Thus, doctor j orders a test if and only if:

$$
\begin{aligned}
T_{ij} - \tau_j^* &\geq 0, \\
\vec{x}_i \beta + a_j - \tau_j^* + \eta_{ij} &\geq 0, \\
\vec{x}_i \beta + \mu_j + \eta_{ij} &\geq 0.
\end{aligned}
$$

Thus, the probability a test is ordered is given by:

$$
\begin{aligned}
\Pr\left[t_{ij} = I | \vec{x}_i, j\right] &= \Pr\left[T_{ij} \geq \tau_j^* | \vec{x}_i, j\right] \\
&= \Pr\left[\rho_j\left(\vec{x}_i\right) + \eta_{ij} \geq \tau_j^* | \vec{x}_i, j\right] \\
&= \Pr\left[\eta_{ij} \geq \tau_j^* - \rho_j\left(\vec{x}_i\right) | \vec{x}_i, j\right] \\
&= 1 - H\left(\vec{x}_i \beta + \mu_j\right). && (13)
\end{aligned}
$$

When estimating (13) it is not possible to separately identify τ_j^* and a_j. Rather, one can use (13) to estimate the intercept term $\mu_j \equiv a_j - \tau_j^*$ and the coefficients β and whether or not a person has PE.

To estimate τ_j^* one needs information on the probability of a PE. From the above estimate, we can define:

$$\begin{aligned} \mathrm{s}_j(\vec{x}_i) &= \rho_j(\vec{x}_i) - \tau_j^*. \\ &= \left(\vec{x}_i\beta + a_j - \tau_j^*\right) \\ &= \left(\vec{x}_i\beta + \mu_j\right) \end{aligned}$$

This function can be estimated from the data using (13), and the fact that the distribution of η_{ij} is known. The expected PE for tested individuals uses (10) to get:

$$\begin{aligned} \Pr\left[\alpha_i = H|\vec{x}_i, t_{ij} = I\right] &= \vec{x}_i\beta + a_j + E\left[\eta_{ij}|\vec{x}_i, t_{ij} = I\right] \\ &= \vec{x}_i\beta + a_j + E\left[\eta_{ij}|\eta_{ij} \geq \tau_j - \rho_j(\vec{x}_i)\right] \\ &= \tau_j^* + s_j(\vec{x}_i) + \int_{-s_j(\vec{x}_i)}^{\infty} \eta h(\eta)\, d\eta / \left(1 - H\left(\tau_j - \rho_j(\vec{x}_i)\right)\right), \\ &\equiv \tau_j^* + \lambda\left(s_j(\vec{x}_i)\right). \end{aligned} \tag{14}$$

where $h(\eta) = H'(\eta)$.[3] The key observation made by Abaluck et al. (2016) is that by construction it must be the case that $\Pr\left[\alpha_i = H|\vec{x}_i, t_{ij} = I\right] \geq \tau_j^*$, the cutoff probability. Under the hypothesis that some patients are not tested because the probability of PE is less than τ_j, implies that there exist marginal patients for which $\Pr\left[\alpha_i = H|\vec{x}_i, t_{ij} = I\right] = \tau_j^*$. The marginal patients are defined by:

$$M_j = \left\{i|\lambda\left(s(\vec{x}_i)\right) \approx 0, t_{ij} = I\right\}.$$

When the number of marginal patients is sufficiently large, then we can obtain an estimate of τ_j from:

$$\tau_j^* \simeq \frac{\sum_{i\in M_j} I_{\alpha_i = H}}{|M_i|}, \tag{15}$$

where $|M_j|$ is the number of patients in the marginal set, and $I_{\alpha_i=H} = 1$ when is $\alpha_i = H$ and zero otherwise. The implicit assumption is that the result from the CT scan is definitive and hence the true α_i is known for tested individuals. When this set M_j is large enough the authors are able to get a precise estimate of doctor's decision threshold or practice style. They show that the decision threshold does vary between doctors.

Computing the TPR and FPR

Finally, within this framework one can map the decision threshold, τ_j, into the ROC model as used by Chan, Gentzkow and Yu (2019). Here we rely upon the structural estimates for β,a_j and the distribution $H(\cdot)$. The unconditional probability a person with condition $\vec{x}_i$ has a PE is given by:

$$\rho_j(\vec{x}_i) \equiv \vec{x}_i\beta + a_j \in [0, 1].$$

[3] Abaluck et al. (2016) allows for an error term with mass point. One simply adjusts the definition of the integral to allows for such mass points, which formally is the requirement that $H(s)$ is right continuous, with jumps at the mass points.

Thus, given that for each doctor a_j is known, then we can write the probability of persons tested having a PE from (14) as a function of potential decision threshold, τ_j, as:

$$
\begin{aligned}
\Pr\left[\alpha_i = H | t_{ij} = I, \vec{x}_i, j, \tau_j\right] &= \rho_j\left(\vec{x}_i\right) + E\left[\eta_{ij} | \rho_j\left(\vec{x}_i\right) + \eta_{ij} \geq \tau_j\right] \\
&= \rho_j\left(\vec{x}_i\right) + \int_{\tau_j - \rho_j(\vec{x}_i)}^{\infty} \eta_{ij} h\left(\eta\right) ds / \left(1 - H\left(\tau_j - \rho_j\left(\vec{x}_i\right)\right)\right), \\
&= \rho_j\left(\vec{x}_i\right) + \hat{\eta}\left(\tau_j - \rho_j\left(\vec{x}_i\right)\right) / \left(1 - H\left(\tau_j - \rho_j\left(\vec{x}_i\right)\right)\right),
\end{aligned}
$$

where

$$
\hat{\eta}\left(s\right) \equiv \int_s^{\infty} \eta h\left(\eta\right) ds,
$$

is the mean value of the unobserved term, η_{ij}, greater than s. Since the mean of $\eta_{ij} = 0$ then is must be the case that $\hat{\eta}\left(s\right) \geq 0$. The support of η_{ij} must be finite in order for T_{ij} defined in (10) to be a probability, and hence $\hat{\eta}\left(s\right) = 0$ for $s > \bar{s}$ for some $\bar{s}$. From these we can compute the TPR and FPR for this model using Bayes rule:

$$
\begin{aligned}
TPR\left(\vec{x}_i, a_j, \tau_j\right) &\equiv \Pr\left[t_{ij} = I | \alpha_i = H, \vec{x}_i, a_j, \tau_j\right] \\
&= \Pr\left[\alpha_i = H | t_{ij} = I, \vec{x}_i, a_j, \tau_j\right] \times \frac{\Pr\left[t_{ij} = I | \vec{x}_j, a_j, \tau_j\right]}{\Pr\left[\alpha_i = 1 | \vec{x}_j, a_j\right]} \\
&= \left(\rho_j\left(\vec{x}_i\right) + \frac{\hat{\eta}\left(\tau_j - \rho_j\left(\vec{x}_i\right)\right)}{\left(1 - H\left(\tau_j - \rho_j\left(\vec{x}_i\right)\right)\right)}\right) \frac{\left(1 - H\left(\tau_j - \rho_j\left(\vec{x}_i\right)\right)\right)}{\rho_j\left(\vec{x}_i\right)} \\
&= \left(1 - H\left(\tau_j - \rho_j\left(\vec{x}_i\right)\right) + \frac{\hat{\eta}\left(\tau_j - \rho_j\left(\vec{x}_i\right)\right)}{\rho_j\left(\vec{x}_i\right)}\right).
\end{aligned}
$$

To compute the corresponding FPR, using Bayes rule we get:

$$
\begin{gathered}
\Pr\left[t_{ij} = I | \vec{x}_j, a_j, \tau_j\right] = \\
FPR\left(\vec{x}_i, a_j, \tau_j\right) \times \Pr\left[\alpha_i = L | \vec{x}_j, a_j\right] + TPR\left(\vec{x}_i, a_j, \tau_j\right) \times \Pr\left[\alpha_i = H | \vec{x}_j, a_j\right]
\end{gathered}
$$

From this we get:

$$
\begin{aligned}
FPR\left(\vec{x}_i, a_j, \tau_j\right) &= \frac{1 - H\left(\tau_j - \rho_j\left(\vec{x}_i\right)\right) - TPR\left(\vec{x}_i, a_j, \tau_j\right) \times \rho_j\left(\vec{x}_i\right)}{1 - \rho_j\left(\vec{x}_i\right)} \\
&= 1 - H\left(\tau_j - \rho_j\left(\vec{x}_i\right)\right) - \frac{\hat{\eta}\left(\tau_j - \rho_j\left(\vec{x}_i\right)\right)}{1 - \rho_j\left(\vec{x}_i\right)}
\end{aligned}
$$

We can see the shape of the ROC curve by looking at:

$$
\begin{aligned}
\Delta\left(\vec{x}_i, a_j, \tau_j\right) &= TPR\left(\vec{x}_i, a_j, \tau_j\right) - FPR\left(\vec{x}_i, a_j, \tau_j\right), \\
&= \hat{\eta}\left(\tau_j - \rho_j\left(\vec{x}_i\right)\right)\left(\frac{1}{\rho_j\left(\vec{x}_i\right)} + \frac{1}{1-\rho_j\left(\vec{x}_i\right)}\right), \\
&= \frac{\hat{\eta}\left(\tau_j - \rho_j\left(\vec{x}_i\right)\right)}{\rho_j\left(\vec{x}_i\right)\left(1-\rho_j\left(\vec{x}_i\right)\right)}.
\end{aligned}
$$

Hence the ROC curve can be parameterized via τ_j and given by:

$$
TPR\left(\vec{x}_i, a_j, \tau_j\right) = \frac{\hat{\eta}\left(\tau_j - \rho_j\left(\vec{x}_i\right)\right)}{\rho_j\left(\vec{x}_i\right)\left(1-\rho_j\left(\vec{x}_i\right)\right)} + FPR\left(\vec{x}_i, a_j, \tau_j\right). \tag{16}
$$

Observe that in this model all doctors have the same diagnostic skill. The ROC curve is traced out via variation in the threshold τ_j. The computation also illustrates that changes in the patient pool, via changes in the intercept term, a_j, results in changes to both the location and shape of the ROC curve via its impact on $\rho_j(\vec{x}_i)$. Thus, this model implies a single ROC for for a fixed pool of patients, a result that is inconsistent with the evidence in Chan, Gentzkow and Yu (2022).

Currie and MacLeod (2017)

This paper uses the model outlined above, where T_{ij} is a signal of patient appropriateness for an intensive procedure (a C-section). From observational data, one observes the doctor's treatment choice ($t_{ij} \in \{NI, I\}$), and some measure of patient outcomes following treatment, as well as some information on patient type that may be available in medical records. Let $\vec{x}_i$ be patient characteristics that are observable in the data. Currie and MacLeod (2017) use the vector of observed patient characteristics, $\vec{x}_i$, to estimate the probability that $\alpha_i = H$, denoted by $\rho\left(\vec{x}_i\right) = \Pr\left[\alpha_i = H|\vec{x}_i\right]$. This is estimated using the full population of patients in New Jersey, and hence it provides a measure of appropriateness that is independent of physician characteristics and practice style.

It is assumed that each physician chooses τ_j^*, as derived in the model section. This in turn determines the TPR_j and FPR_j for the doctor. Here one is implicitly assuming that the signal T_{ij} has the information contained in $\vec{x}_i$. With this definition we have:

Proposition 5. *The doctor's estimated likelihood of performing an intensive procedure is:*

$$
\Pr\left[t_{ij} = I|j, \vec{x}_i\right] = \left(TPR_j - FPR_j\right)\rho\left(\vec{x}_i\right) + FPR_j, \tag{17}
$$

where $\rho\left(\vec{x}_i\right) = \Pr\left[\alpha_i = H|\vec{x}_i\right]$ is the estimated probability that the patient needs an intensive intervention, while TPR_j and FPR_j are computed at the doctor's decision rule (proposition 2). The slope term, $\theta_j = \left(TPR_j - FPR_j\right)$ is increasing with a doctor's diagnostic skill:

$$
\frac{d\theta_j}{d\gamma_j} > 0.
$$

Finally, $\frac{d\theta_j}{db_j} > 0$ for $b_j < 0$ and $\frac{d\theta_j}{db_j} < 0$ for $b_j > 0$, namely the treatment decision is most sensitive to *the prior condition of the patient* $(\rho(\vec{x}_i))$ *when* $b_j^* = 0$.

Proof. The probability of a C-section is:

$$\begin{aligned}
\Pr[t_{ij} = I | j, \vec{x}_i] &= \Pr[t_{ij} = I | \alpha_i = H, \vec{x}_i, a_j, \tau_j] \times \Pr[\alpha_i = H | j, \vec{x}_i] \\
&\quad + \Pr[t_{ij} = I | \alpha_i = L, \vec{x}_i, a_j, \tau_j] \times \Pr[\alpha_i = L | j, \vec{x}_i] \\
&= TPR_j \times \Pr[\alpha_i = H | j, \vec{x}_i] + FPR_j \times (1 - \Pr[\alpha_i = H | j, \vec{x}_i]), \\
&= (TPR_j - FPR_j) \Pr[\alpha_i = H | j, \vec{x}_i] + FPR_j.
\end{aligned}$$

Then we have using the decision rule from proposition (1):

$$\begin{aligned}
\frac{d\theta_j}{d\gamma_j} &= \frac{dF\left(\gamma_j\left(1 - \tau_j^*\right)\right)}{d\gamma_j} - \frac{dF\left(\gamma_j\left(-1 - \tau_j^*\right)\right)}{d\gamma_j} \\
&= \frac{dF\left(\gamma_j - b_j^*/\gamma_j\right)}{d\gamma_j} - \frac{dF\left(-\gamma_j - b_j^*/\gamma_j\right)}{d\gamma_j} \\
&= \frac{b_j}{\gamma_j^2}\left(f\left(\gamma_j - b_j^*/\gamma_j\right) - f\left(-\gamma_j - b_j^*/\gamma_j\right)\right) \\
&= \frac{b_j}{\gamma_j^2} \exp\left(\gamma_j^2 + \frac{b_j^*}{\gamma_j^2}\right)\left(\exp\left(b_j^*\right) - \exp\left(-b_j^*\right)\right).
\end{aligned}$$

When $b_j > 0$ then $(\exp(b_j) - \exp(-b_j)) > 0$ and when $b_j < 0$, then $(\exp(b_j) - \exp(-b_j)) < 0$, Hence the right hand side is strictly positive when $b_j \neq 0$ and zero when $b_j = 0$, Thus the slope increases with skill.

In the case of b_j we have:

$$\begin{aligned}
\frac{d\theta_j}{db_j} &= \frac{dF\left(\gamma_j\left(1 - \tau_j^*\right)\right)}{db_j} - \frac{dF\left(\gamma_j\left(-1 - \tau_j^*\right)\right)}{db_j} \\
&= \frac{dF\left(\gamma_j - b_j/\gamma_j\right)}{db_j} - \frac{dF\left(-\gamma_j - b_j/\gamma_j\right)}{db_j} \\
&= -\frac{1}{\gamma_j}\left(f\left(\gamma_j - b_j/\gamma_j\right) - f\left(-\gamma_j - b_j/\gamma_j\right)\right) \\
&= -\frac{1}{\gamma_j} \exp\left(\gamma_j^2 + \frac{b_j}{\gamma_j^2}\right)\left(\exp\left(b_j\right) - \exp\left(-b_j\right)\right).
\end{aligned}$$

Hence, θ_j increases with b_j if and only if $b_j < 0$. Thus θ_j is largest when $b_j = 0$. Hence, from (2) and (3) we have:

$$\theta_j = F(\gamma_j) - F(-\gamma_j)$$

□

Notice that from equation (17), as long as there is sufficient variation in the likelihood of needing intensive treatment, $\rho(\vec{x}_i)$, one can separately identify TPR_j and FPR_j in equation (17) Hence we can identify both τ_j and γ_j.

The slope term is also affected by the physician's beliefs about when invasive procedures are likely to be warranted via τ_j, and by any additional physician-specific factors that are included in δ_j. Currie and MacLeod (2017) distinguish between τ_j and γ_j by noting that in a doctor-specific regression, the constant term in Equation (17) is affected only by τ_j so given two estimated parameters and two unknowns, it is possible to identify both.

Finally, notice that patients with high *ex ante* likelihood of having a C-section ($\rho(\vec{x}_i) \approx 1$) then variation in patient outcomes is independent of both diagnostic skill and the decision threshold. Hence, we can associate variation in outcomes with procedural skill. A similar implication follows for patients with a low likelihood of a C-section ($\rho(\vec{x}_i) \approx 0$).

References

Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, and Arjun Venkatesh. 2016. "The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care." *American Economic Review*, 106(12): 3730–3764.

Chan, David C, Jr, Matthew Gentzkow, and Chuan Yu. 2019. "Selection with Variation in Diagnostic Skill: Evidence from Radiologists." National Bureau of Economic Research Working Paper 26467.

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

Currie, Janet M., and W. Bentley MacLeod. 2017. "Diagnosis and Unnecessary Procedure Use: Evidence from C-section." *Journal of Labor Economics*, 35(1): 1–42.

Feng, Kai, Han Hong, Ke Tang, and Jingyuan Wang. 2023. "Statistical Tests for Replacing Human Decision Makers with Algorithms."

MacLeod, W. Bentley. 2022. *Advanced Microenomics for Contract, Instituional and Organizational Economics.* Boston, MA:MIT Press.

Neyman, Jerzy, and Egon Sharpe Pearson. 1933. "IX. On the Problem of the Most Efficient Tests of Statistical Hypotheses." *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706): 289–337.

Savage, Leonard J. 1972 (first published 1954). *The Foundations of Statistics.* New York, N.Y.:Dover Publications.

Appendix Describing Research Papers Organized by Topic

Appendix Table 1: Health Disparities

Paper	Research Question	Data	Empirical Methods	Results	Heterogeneous Effects?
Alsan, Garrick, and Graziani (AER 2019)	How does physician race affect Black men's take up of preventative care services?	Experimental data with 1,374 recruited Black male participants, with 637 completing the study	Field experiment with random assignment to either a Black or non-Black physician in a special clinic offering preventive care. Doctor race was signaled to patients by a headshot.	Viewing the headshot did not significantly affect intended take-up of services. But patients who saw a Black patient increased demand for services ex-post by 38.79% for diabetes screening, 52.77% for cholesterol screening and 26.54% for flu shots.	No differences by income, education, or age. Effects greater for patients without a recent medical screening, with more ER visits, and with higher levels of measured medical mistrust.
Angerer, Waibel, and Stummer (AJHE 2019)	What is the effect of socioeconomic status, signaled by education level, on the probability of receiving a medical appointment and on response times?	Experimental data for April 26-June 2, 2017, with email requests for appointments sent to 1,249 Austrian specialists.	Correspondence study via email with varying email signatures to signal no degree, a doctoral degree, or a medical degree	Patients with degrees are more likely to receive an appointment, and have lower response times and lower waiting times. Whether patients are offered an appointment depends on the assistant, while response and waiting times depend on the doctor.	The effects are driven by practices that do not contract with social insurance.
Button et al. (NBER WP 2020)	How does being nonbinary or transgender interact with patient race to affect the probability of getting an appointment with a mental health care provider (MHP)?	Experimental correspondence data from 1,000 emails sent to MHPs between Jan. 28, 2020-May 15, 2020, with number of emails per zip code proportional to population.	Emails sent through an MHP appointment request website with randomly assigned content disclosing trans or nonbinary status. Names signal gender and race. Randomize whether help is sought for depression, anxiety, or "stress."	Transgender or non-binary African Americans and Hispanics are 18.7% less likely to get a positive response than cisgender whites. No evidence of differential responses by TNB status for whites.	N/A
Brekke et al. (HE 2018)	What is the relationship between SES of Type II diabetes patients and GP treatment decisions?	Norwegian administrative health data 2008- 2012; patient and GP characteristics from Statistics Norway.	GP FE models of service provision conditional on patient characteristics. Additional results using GP quits, retirements, and moves.	High ed. patients get fewer, longer visits, Less ed. patients get more medical tests and services over the course of a year. E.g. high ed. 14.79% more likely to get a visit over 20 minutes. Less ed. 3.94% more likely to get 2+ HbA1C tests.	Results are similar when disaggregated by patient age and GP sex, age, specialty, number of patients, and fixed payment vs. fee-for-service.
Cabral and Dillender (AER	How does gender concordance between	Open records request for Texas worker's	Assignment to doctors is random conditional on	Female claimants seen by a female doctor are 5.2% more likely to receive	Differences are not statistically significant but

2024)	claimants and doctors performing independent medical evaluations for workers compensation affect disability determinations?	compensation claims 2013-17, and independent medical evaluations 2005-2017; NPI registry; novel survey of 1,519 adults 30-64, 2021.	doctor's credential and the claimants' county. Estimate OLS with an interaction between female doctor and female claimant controlling for main effects, credential, and county.	benefits compared to when female claimants are seen by male doctors. Physician gender does not affect likelihood of receiving benefits for male claimants. Female claimants seen by a female doctor receive 8.6% higher benefits than female claimants seen by male doctors.	suggest larger effects for those with lower earnings, in less dangerous industries, but with worse injuries.
Chandra and Staiger (NBER WP 2010)	Are differences in the treatment of Black and female AMI patients due to physician preferences or statistical discrimination?	Clinical records for 200,000+ patients admitted for AMI in 1994 & 1995 from the Cooperative Cardiovascular Project (CCP).	Propensity score estimation; taste based discrimination implies that similar patients who receive fewer services will suffer worse outcomes.	Black and female patients receive less treatment but also receive slightly lower benefits from treatment suggesting that they are not being denied beneficial treatment due to discrimination.	N/A.
Eli, Logan, and Miloucheva (NBER WP 2019)	Use union army pension awards to examine the effect of income on mortality. Investigate differences in a board's disability evaluations by race of applicant.	Union Army and United States Colored Troops (USCT) sample from the Early Indicators Project; Rosters of Examining Surgeons from the National Archives.	Instrument pension income using leave-one-out mean of a board's pension determinations. Include board FEs. First stage shows the same boards were less generous to Black veterans.	Pension income significantly increased life expectancy. Bias against Black veterans in determining pension eligibility is substantial and accounts for much of the racial mortality gap in this population.	Bias against Black veterans is strongest for conditions where valuations may be more subjective, such as digestive diseases.
Frakes and Gruber (NBER WP 2022)	How does the availability of Black physicians on a military base affect Black Tricare patients' outcomes?	Military Health System Data Repository fiscal years 2003–2013	Mover-based ITT design exploiting differences in racial shares of physicians across bases.	1 SD increase in share of Black physicians reduces Black patients' mortality from diabetes, hypertension, high cholesterol, and cardiovascular disease by 15%. 55–69% of the effect attributed to medication adherence.	N/A.
Goyal et al. (JAMA Pediatrics 2015)	How does treatment of pain in the ED vary by race for child appendicitis patients?	National Hospital Ambulatory Medical Care Survey 2003-2010.	Multivariate logistic regression.	Black patients were less likely to receive any analgesia, adjusted OR=0.1 for moderate pain and 0.2 for severe pain. Black patients were less likely to receive opioids, adjusted OR= 0.2.	The authors test for interactions between race and sex but do not find any.

Greenwood, Carnahan, and Huang (PNAS 2018)	How does patient-attending gender concordance affect mortality from heart attacks among patients admitted to the ED? Do male doctors with more female colleagues or AMI patients have better female survival?	Census of patients admitted to hospitals in Florida 1991- 2010 from Florida's Agency for Healthcare Administration.	Assume patient assignment to physicians is conditionally random in the ED and either include physician FEs or hospital-quarter FEs. They also estimate additional specifications using matching.	In the full sample with hospital-quarter FEs, relative to male or female patients treated by female physicians, female patients treated by male doctors are 1.80% less likely to survive and male patients treated by male doctors are 0.90% less likely to survive. In the matched sample, only female patients treated by male doctors have lower survival rates.	Female survival increases when there are more female physicians in the ED, especially when they are treated by male physicians. Female patients treated by male physicians are more likely to survive as the number of female patients their doctor has treated in the prior quarter increases.
Greenwood et al. (PNAS 2020)	How does infant and maternal mortality vary as a function of patient-doctor racial concordance?	Census of patients admitted to hospitals in Florida 1992- 2015 from Florida's Agency for Healthcare Administration.	OLS with controls including physician FEs in some models.	Racial concordance between infant and physician corresponds to about a 40% reduction in gap in mortality between Black and white infants. No significant racial concordance effects are found for mothers.	Effects are more precisely estimated for infants with >=1 comorbidity and for infants in hospitals that see more Black patients. Effects are similar in % terms for pediatricians and non-pediatricians.
Hill, Jones, and Woodworth (JHE 2023)	What is the effect of physician-patient race concordance on within-hospital mortality among uninsured non-Hispanic, Black and white patients admitted through the ED?	Florida Hospital Discharge Data File from October 2011 to December 2014; Florida Physician Workforce Survey from 2008-2016.	IV measures "the lagged share of same-race physicians typically present at the indexed hospital on the weekday and shift" when patient admitted.	Physician-patient race concordance reduces mortality by 27%.	The largest effects are for subgroups of patients with high variance in number of procedures and in total charges.
Hoffman et al. (PNAS 2016)	How do false beliefs about biological racial differences among white doctors mediate racial differences in recommended for hypothetical patients?	Experimental and survey data from U.S. medical students and residents (N=222 after restricting to white, US-born, native English-speaking).	Surveys and experimental vignettes.	Participants one SD above the mean in terms of false beliefs rated the Black patient as having 0.45 less pain than the white patient on a scale of 1-10 and were less accurate in recommendations for the Black patients.	Some statistics are disaggregated by medical school year or resident status, but sample sizes are too small to draw inferences.
McDevitt and Roberts (RAND	How does the availability of female	American Medical Information's data on	Descriptive statistics and a structural model to	Counties that have one more female urologist per 100,000 residents have	

2014)	urologists relate to rates of bladder cancer death among female patients?	urologists from 2006 and 2009; Florida hospital discharge data from Jan. 2006 - June 2008; Florida Licensure Data; NCI's State Cancer Profiles; Census, BEA, ARF for each market.	explain the distribution of female urologists across counties and the lack of entry.	29.08% fewer female bladder cancer deaths per 100,000 residents. No significant associations between female urologists and male bladder cancer deaths or overall cancer deaths.	
Sabin and Greenwald (AJPH 2012)	What is the association between pediatricians' scores on an implicit bias test (IAT) and racial differences in treatment?	Survey data from 86 academic pediatricians conducted during October and September 2005.	Online survey with IAT tests plus patient vignettes describing children with pain following femur fracture, UTIs, ADHD, asthma.	Pro-white bias in the IAT is significantly correlated with not giving oxycodone to the Black vignette patient in pain after bone surgery ($p<0.05$).	N/A.
Singh and Venkataramani (NBER WP 2022)	How do racial disparities in in-hospital mortality vary with hospital capacity strain?	EHR with time stamps from 2 "highly regarded" academic hospitals serving predominantly Black patients.	OLS with rich controls; Assume that hospital capacity strain at patient arrival is conditionally independent of mortality risk.	No significant differences in conditional patient mortality by race in quintiles 1-4 of hospital capacity strain. At the fifth quintile, Black patients are 0.4 pp more likely to die on a baseline of 2%.	Effects are larger for Black women and Black patients without insurance. Effects driven by high-risk patients.
Wallis et al. (JAMA Surgery 2022)	How does surgeon-patient sex concordance affect post-operative outcomes?	Ontario Health Insurance Plan data; CIHI Discharge Abstracts and Ambulatory Care Reporting Services System; Registered Persons Data; Corporate Provider Database.	Population-based, retrospective cohort study.	Sex discordance was associated with increased likelihood of death (adjusted odds ratio 1.07) and complications (adjusted odds ratio 1.09), but not readmission.	They disaggregate by patient sex and find that effects are driven by male surgeons treating female patients. They also find stronger effects for cardiothoracic surgery.

Appendix Table 2: Effect of Experience and Training on Doctor Skills

Paper	Research Question	Data	Empirical Methods	Results	Heterogeneous Effects?
Chan and Chen (NBER WP, 2023)	How do NPs compare to doctors with respect to patient outcomes and resource use in the ED? How does variation in provider skill vary across and within professions?	Administrative health records from the VHA for ED visits between 01/2017 and 01/2020 (1.1 million cases, 44 EDs) linked to death records.	Use number of NPs on duty as IV for assignment to an NP vs. a doctor on arrival at the ED.	Assignment to an NP increases patient length of stay by 11%, increases cost of care by 7%, and increases 30-day preventable hospitalizations by 20%. Productivity variation is greater within than between each profession.	The NP-physician performance gap is smaller for experienced providers and larger for patients with complex or severe conditions. Many NPs are more skilled than some doctors.
Currie and Zhang (ReStat, 2023)	Are some physicians more effective in promoting patient health? Correlation in effectiveness across domains of patient care? Do effective providers have lower/higher costs?	EHR data from the Veterans Health Administration's Corporate Data Warehouse for 2004 to Feb. 2020, VHA Vital Status files, CDC National Death Index Plus files.	Quasi-random assignment of veterans to PCP teams in the VHA system; value-added measure of provider effectiveness.	PCPs with 1 SD higher mental health effectiveness, circulatory condition effectiveness, or ACSC effectiveness have a 27-44% reduction in adverse outcomes. Effectiveness measures positively correlated. Assignment to a PCP with a 1 SD higher effectiveness reduces mortality 3.6-4.2 % and reduces patient costs 2.5-5.4% over the next three years.	Provider effectiveness increases with provider age and number of patients seen.
Doyle, Ewer, and Wagner (JHE, 2010)	Do residents from highly ranked programs do better than residents from lower ranked programs re: costs and health outcomes?	Veteran's Administration inpatient data 1993-2006; 2000 Census zip code level data.	Residency teams randomly assigned to patients based on the last digit of the SSN.	Patients assigned residents from lower ranked program had 11.96% longer stays and 13.31% higher costs. No differences in health outcomes.	Differences in costs were higher for more serious conditions.
Doyle (NBER WP, 2020)	Does having cardiologists in the ER affect treatment and outcomes for patients with heart failure? Does additional experience with heart failure patients affect outcomes?	Medicare claims data (1998-2002) linked to mortality data; AMA's Masterfile for physician characteristics.	Estimate the effect of the share of physicians of different types in the ER, conditional on hospital*quarter *day-of-week FE.	Controlling for number of physicians available, 1-year mortality falls by 1.10% with each additional cardiologist. Additional cardiologists increase intensity of care. A doctor seeing 10 more heart failure patients yearly reduces mortality 1.2%.	Mortality point estimates larger for patients with higher predicted mortality, in high-volume hospitals, and for patients seen on slow days but differences imprecisely estimated.
Epstein,	Compare effect of initial	Florida and New York	Initial skill defined as	Without hospital FE, initial skill	Privately insured patients

Nicholson, and Asch (AJHE 2016)	skill to the effect of experience in predicting obstetrician performance?	all-payer discharge databases (1992 to 2012); AMA Physician Masterfile; AMA FREIDA identifiers of hospitals with OB residency training.	physician's normalized, risk-adjusted maternal complication rate in the 1st year.	explains much of the variance in performance. After 16 years, it explains 39-75% of performance. With hospital FEs initial skill explains only 1-9%, suggesting better doctors go to better hospitals. Experience explains little.	respond to recent measures of physician skill. Robustness checks with physician "stayers" only show similar results.
Facchini (Health Econ, 2022)	Does the recent volume of C-sections performed affect the outcomes of a surgeon performing a nonelective C-section?	Birth certificates from a large public hospital in Tuscany, Italy (2011 to 2014)	Patients cannot select their surgeon though more skilled surgeons may get harder cases. Include surgeon FEs.	Recent experience defined as #C-sections in the last 4 weeks. A one SD increase in experience reduces NICU admission 13.86% and reduces low APGAR 13.19%.	N/A.
Gowrisankaran, Joiner, and Léger (Management Science 2023)	How are measures of physician practice style and of physician skill correlated in the context of patients visiting the ED?	La Régie de l'as-surance maladie du Québec (RAMQ) data on Montreal patients who visited an ED between April and Dec. 2006.	Identification relies on conditional random assignment of patients within an ED. Physician practice style and skill estimated from physician FEs.	Physicians with more intensive practice style have worse outcomes on average. Practice intensity correlated across conditions, as is skill.	Negative correlation intensive practice style and patient outcomes strongest for appendicitis, weakest for transient ischemic attacks.
Schnell and Currie (AJHE, 2018)	How does a doctor's medical school rank affect their propensity to prescribe opioids? How does this relationship vary over time and between specialties with different levels of training in pain relief?	QuintilesIMS opioid prescription data 2006-2014; US News and World Reports; CMS provider utilization and payment data; ACS data; Mortality data.	FE models (specialty, county of practice, practice address).	Physicians from the lowest ranked medical school are 121% more likely to prescribe any opioids and prescribe 160% more than physicians trained at the top school.	Rank doesn't matter for specialties with pain medicine training. Rank matters less for more recent cohorts. Foreign physicians from low prescribing areas have low prescription rates.
Simeonova, Skipper, and Thingholm (JHR, 2024)	Do health management skills (HMS) of primary care physicians affect medication adherence and hospitalizations for cardiovascular (CV) disease, and CV hospital costs of patients on statins? Do skills change with age?	Danish registry data on population of statin users and their PCPs (01/2004-06/2008). However, cannot observe PCP for 54% of clinics.	Leave-one-out adherence rates for each physician adjusted for patient and physician observables. Event studies after changes in PCP induced by clinic closures or patient moves.	A one SD increase in PCP HMS is associated with a 1.10% increase in medication adherence and 1.47% fall in CV hospitalization. CV hospital expenditures fall by 0.298%. Skill declines with physician age.	N/A.

Van Parys (PLOS One, 2016)	How are variations in ED physicians' treatment of minor injuries related to physician characteristics including experience? Does practice style explain persistence as an ED physician?	All Florida ED visits for minor injuries 2005-2011 matched to Florida Healthcare Practitioner Database; HCUP databases.	OLS assuming little systematic matching of physicians and patients conditional on observables.	Physicians with <2 years of experience spend 4.60% more and perform 3.46% more procedures than physicians with 7+ years. High-cost physicians are 3% less likely to work in a Florida ED 2 years after start.	Differences in care intensity fall with experience after 2-7 years of experience.

Appendix Table 3: Time Pressure and Fatigue

Paper	Research Question	Data	Empirical Methods	Results	Heterogeneous Effects?
Chan (2018) Econometrica	How does ER physician decision- making change over the course of a shift?	Data on physician shifts from the ER in a large, U.S. academic, tertiary-care center 06/2005-12/2012.	Exploits randomness and pre-determination of shifts and overlap in shifts. Counter-factual simulations of patient assignments.	8.70% shorter visits in the 4th to last hour before shift ends, 44.40% shorter in last hour. Patients arriving in last hour have 10.44% more tests/treatments, a 5.7 pp (21.19%) higher likelihood of admission, and 23.12% higher total costs. No significant effects beyond the last hour. No effects found with respect to 30-day mortality or 14-day bounce back.	The effects on workload-adjusted length-of-stay are greater in the daytime and disappear if the index physician has enough time to offload cases to the incoming physician.
Chu et al. (2024) Working Paper	How does cognitive load affect how a physician takes notes, orders tests, and treats patients?	High frequency "click stream" data from EHRs, for patients over 18 at the UCSF ED (2017-2019)	Cognitive load proxied by complexity of patient caseloads. Predict physician orders from past orders; measure deviations in actual orders as a function of load.	When load is high, physicians reduce note editing by 7-14% and increase diagnostic orders by 2-5%, with higher entropy in diagnostic tests. For every 1 SD from expected orders induced by cognitive load, probability of admission increases 3.4 p.p. (14%).	N/A.
Costa-Ramón et al. (JHE 2018)	How does time of delivery affect unscheduled C-sections, and infant health.	6163 births in 4 Spanish public hospitals 2014-2016. Scheduled and breech deliveries excluded.	IV estimation using an indicator for births between 11 p.m. and 4 a.m.	Unplanned C-sections increase by 53.21% between 11 p.m. and 4 a.m. There is a negative effect on 1-minute and 5-minute APGAR (-0.992 and -0.936).	N/A
Freedman et al. (JHE 2021)	Unexpected scheduling changes and decisions of PCPs.	EMR data on all visits to 31 primary care centers in a health system 2005-2015.	Physician FE models with unexpected schedule changes in minutes as the independent variable.	10-minute increase in waiting time reduces total/new (0.19%/0.14%), referrals (0.32%), opioid Rx (0.33%), pap tests (0.39%). Increases scheduled/unscheduled follow ups (0.80%/0.50%), inpatient visits within 14/30 days (1.15%/1.85%), and hospital care within 30 days (0.17%). No effect on ER visits, imaging, antibiotic Rx, diabetes management.	Effects with respect to PT referrals and opioid Rx among opioid-naïve patients are not significant in the baseline specification.
Gruber, Hoe, and Stoye (ReStat 2021)	Studies an English policy limiting ER wait times to 4 hours for 95% of	Records of all visits to public hospitals at the visit level linked	Bunching estimator using the four-hour target. Assumes that only patients around	Wait times fell 8% in patients with wait times of 180-400 minutes, and by 59 minutes for patients moved from the post-threshold period to the pre-threshold period. Increased 30-day	Larger wait time effects and mortality for sicker patients. No significant difference in probability of hospital

	patients at public hospitals.	to vital statistics mortality records for 4/2011-03/2013.	the four-hour mark are affected.	total costs (4.9%); hospital admissions (12.2%); tests in the ER (4.6%); Decreased 30/90-day mortality (13.8%/7.9%); discharge probability (7%); referrals (8.9%). No effect on 1-year mortality, length of stay or number of inpatient procedures.	admission. Most mortality reduction driven by circulatory, respiratory, and digestive problem deaths.
Linder et al. (JAMA IM 2014)	How does time in shift affect the decision to prescribe antibiotics?	Billing and EMRs for visits to 23 Partners HealthCare-affiliated PCPs 05/2011-09/2012.	Logistic regression.	Relative to the first hour of a shift, the adjusted odds ratios of antibiotic prescribing in the 2nd, 3rd, and 4th hours were 1.01 (95% CI, 0.91-1.13), 1.14 (95% CI, 1.02-1.27), and1.26 (95% CI, 1.13-1.41). 44.46% of the sample was prescribed antibiotics.	N/A.
Neprash et al. (JAMA HF 2023)	What is the association between primary care visit length and inappropriate prescribing?	Claims and EHR data from AthenaHealth Inc., 2017.	Descriptive; linear probability models with physician FEs and patient covariates.	An additional minute of visit duration decreases inappropriate antibiotic prescribing 0.11 pp (0.2%), opioid and benzodiazepine co-prescribing for pain 0.01 pp (0.3%), and a prescribing of medications from the Beers List to older adults 0.004 pp (0.4%).	For patients with an anxiety and pain , each additional minute of visit duration decreased dangerous opioid and benzodiazepine co-prescribing 0.05 pp.
Shurtz et al. (RAND, 2022)	Do PCPs increase treatment intensity and screening in response to time pressure caused by absent colleagues?	Administrative data from the largest HMO in Israel covering all primary care visits in Jerusalem 2011-2014.	Event studies at physician-day level. IV for visit length is %caseload missing physicians. (Alt. IV= any doctors missing). Nonparametric methods to bound the ATE.	A 1 minute longer visit increases use of any diagnostic input 4.50% and referrals 7.93%. No significant effects on imaging, pain killer Rx, antibiotic Rx, additional visits.	Effects on use of diagnostic tools bigger for older patients (>60 years) and patients with higher predicted utilization of primary care.
Persson et al. (HE 2019)	How are orthopedic surgeons' decisions affected by the number of patients already seen in a shift?	848 Swedish orthopedic clinic visits spanning 133 work shifts by eight surgeons between 10/2015-12/2015.	Logits with surgeon fixed-effects, assuming patient allocation to time slots is exogenous conditional on observables.	Every additional patient already seen decreases the odds an operation is scheduled by 10.5% (OR = 0.895, CI 0.842 to 0.951). Patients seen in the afternoon are 1.955x more likely to be scheduled for surgery (CI 1.110 to 3.486). Surgery prescribed in 32% of cases.	N/A.
Tai-Seale and McGuire (HE 2012)	Do physicians have a target time per patient?	385 video-taped visits 1998-2000 with 35 PCPs; patient surveys.	Logits on the probability of a topic being the last of the visit.	Topics in the 1st 5 minutes=reference group. Probability of a topic being last increases by 16.8 pp, 26.8 pp, and 35.7 pp for topics raised at 5-10, 10-15, 15+ minutes.	Academic medical centers demonstrated sharpest increase in the shadow price of time.

Appendix Table 4: Peer Effects and Team Dynamics

Paper	Research Question	Data	Empirical Methods	Results	Heterogeneous Effects?
Agha and Molitor (ReStat 2018)	Does proximity to lead investigators in new cancer drug trials increase the propensity to prescribe new drugs?	Medicare Part B claims 1998-2008; Dartmouth Atlas data; FDA drug application data.	DiD, patient location IV (secondary analysis).	Cancer patients in lead investigator's HHR 4.04 pp (36%) more likely to get new cancer drug, with convergence after 4 years. No effect in other authors' HHRs. IV estimates smaller.	Effects bigger in areas with slower drug adoption. Convergence suggests lead investigators are not in areas with higher latent demand for the cancer drug.
Chan (JPE, 2016)	Is doctor shirking (i.e. working slowly to avoid work) reduced when doctors vs. nurse schedulers do patient assignments?	6 years of ED data from an academic medical center. ED had 2 pods of doctors.	Natural experiment in which a nurse-managed pod became doctor-managed, as the other pod was.	The doctor-managed system reduced patient wait times by 13.67% with no significant effects on quality, cost, or utilization.	Patient assignment is more negatively correlated with a physician's number of patients in doctor-managed system (consistent with it being a stronger signal of true workload).
Chan (AEJ: EP, 2021)	How much influence do senior residents have on team decisions? How do junior resident's decisions vary with experience?	Five years of data from the internal medicine residency program of a large teaching hospital.	RE model exploiting discontinuity caused by promotion of junior residents to senior.	There is a jump in the SD of log costs after promotion. Senior residents are responsible for almost all of the variance in decision making within a team of residents.	The jump in practice variation is highest for diagnostic spending (vs. medication, blood work, or nursing). No differences by patient characteristics.
Chen (AER, 2021)	How does the length of time that PCI/CABG surgeons and other hospital physicians have worked together affect patient outcomes?	20% of Medicare claims 2008-2016 linked to Vital Statistics, MD-PPAS 2008–2016, Physician Compare 2014–2017.	1.Restrict to admissions through ED and include FEs for proceduralists. 2.TWFE model with FEs for proceduralists and PCPs.	1 SD increase in shared work experience reduces 30-day mortality by 10 to 14%. Shared work experience decreases use of medical resources and length of stay.	Effect of shared work experience declines with individual physicians' experience, but this decline is small. The effect is larger for more complex cases.
Molitor (AEJ: EP 2018)	How are cardiologists affected when they move to areas with different practice styles?	Medicare fee-for-service claims 1998–2012; AMA Masterfile;	"Movers" design follows cardiologists moves across HRRs; event study and difference-in-	A 1pp increase in cardiac catheterization in the new HRR increases the physician's own rate 0.628 pp (1.36%). A 1pp increase in the rate at the physician's	Effects of moving larger for moves from low to high-intensity areas. Effects similar for moving earlier vs. later in their careers. Effects of moving are larger for

			differences.	hospital leads to a 0.796 pp (1.72%) increase in the physician's own rate.	more marginally appropriate patients.
Silver (ReStud 2021)	How do peer-groups affect speed and outcomes in the ED?	All ED visits from New York (2005-2013). Linked to state physician license register, public physician profiles, and vital statistics mortality data.	Peers vary across shifts. Decompose variation in outcomes attributable to physicians and physician-peer matches. Use peer group as IV for outcomes.	First-Stage: A 10% increase in the speed of a physician's peers increases own speed 1.47% with controls. 2SLS: A peer group that increases a physician's speed by 10% decreases charges by 2.17% with no significant effect on the 30-day mortality of discharged patients.	Physicians work faster in smaller groups and when all of their peers are male. 2SLS: In at-risk patients, peer groups that increase physician speed by 10% decrease charges 2.55% and increase 30-day mortality in discharged patients by 0.2121 pp (5.65%) .

Appendix Table 5: U.S. Financial Incentives

Paper	Research Question	Data	Empirical Methods	Results	Elasticity	Het. Effects?
Papers with Defined Price Elasticities						
Allen, Fichera, and Sutton (HE, 2016)	Examined an English policy that increased payments 24% for outpatient cholecystectomies while inpatient reimbursement were unchanged.	Hospital Episode Statistics from the NHS Information Centre for Health and Social Care from 12/2007-03/2011.	D-in-D using a set of control procedures with similar recommended outpatient rates that were not affected.	Planned outpatient surgeries increased by 27% of baseline mean. Reversion from laparoscopic to open surgery decreased. No effect on deaths or readmissions.	Elasticity of outpatient surgery supply w.r.t. payment: 1.21	N/A.
Alexander and Schnell (AEJ:AE, 2024)	What was the impact of increasing Medicaid PCP payments in 2013 and 2014 to comply with the ACA?	State-level Medicaid reimbursement rates; NHIS (2009–2015); NAEP (2009, 2011, 2013).	D-in-D and event studies exploiting variation in the effect of ACA rule given pre-ACA reimbursement rates.	A $10 rise in payments (a 13.2% rise) decreases prob. doctors decline new Medicaid patients by 0.71pp or 11.5%. Also decreases prob. that parents have trouble finding a doctor for child 25%. Increased payments increased doctor visits, improve reported health, and reduce school absences.	Elasticity of getting and appointment w.r.t. payment: 11.5/13.2 =0.87	Effects on school absences are larger and more precisely estimated for younger students.
Bisgaier and Rhodes (NEJM, 2011)	How does public vs. private insurance affect the probability that specialists will accept new pediatric patients, and wait times?	Experiment with 546 paired calls to 273 specialty clinics. Private insurance pays 60% more.	Audit study. One call with public insurance and one a month later with private insurance.	Private insurance accepted 89.4% of the time, public ins. accepted 34.4% of the time. Medicaid-CHIP callers were 6.2 times more likely to be denied an appointment. Conditional on getting an appointment, Medicaid-CHIP callers waited 22 days longer.	Elasticity of getting an appointment w.r.t. payment: [(89.4-34.4)/34.4]/60 =2.66.	N/A.
Cabral, Carey, and Miller (NBER Working Paper,	How did increased payments to providers of evaluation & management services to dual-eligible beneficiaries under the ACA affect care provision?	20% random sample of Medicare beneficiaries from Master Beneficiary Summary File and medical claims files	DiD and triple differences using non-duals and non-qualifying providers as control groups.	Increased payments increased evaluation & management services for dual-eligible beneficiaries by 6.3% and reduced fraction with no evaluation & management visits by 8.7%.	Elasticity of evaluation & management services/appointments w.r.t. payment: 1.2	Larger effects for younger/white beneficiaries, and beneficiaries not living in HPSAs.

2024)		(2010–2014); Medicaid Analytic Extract (2011–2013)				
Chen and Lakdawalla (JHE, 2019)	Do physician responses to changes in Medicare reimbursement vary with patient income?	Medicare Current Beneficiary Survey (MCBS) 1993- to 2002; Federal Registers from 1993 to 2002.	2SLS: Instruments are changes in fees from 1997 consolidation of Medicare areas and 1999 changes in estimation of expenses.	A 10% increase in patient income increases price elasticity for services 0.051 (53% of the mean). Different physician responses wrt patient income explain 53% of the increase in the gap in services received by high-income vs. low-income patients.	Mean elasticity= 0.095.	0.05 at 10th percentile of patient income. 0.15 at 90th percentile of patient income.
Clemens and Gottlieb (AER, 2014)	How do changes to Medicare physician payment rates affect provision of care, technology adoption, and patient health?	Medicare Part B claims 1993-2005.	Natural experiment: 1997 consolidation of Medicare geographic areas. Event study with nearest-neighbor matching on counties.	Higher fees increase elective procedures and RVUs per physician. Imprecise effects on MRIs by non-radiologists. Increases in hospitalization for AMI within 1 year, but no effect on 4-year mortality. A "1 percent change in reimbursement rates thus translates, on average, into a 2.5 percent change in the physician's net wage."	Elasticities for RVUs per patient w.r.t. payment: Short run =0.82 Medium run =2.01 Long run= 1.46.	Heterogenous effects by patient age and state-level intensity of care. Higher care elasticities for older patients and patients from states with more intense care.
Coudin, Pla, and Samson (HE, 2015)	How did a French reform that increased the proportion of GPs subject to price regulation, affect the provision of health services?	Administrative INSEE-CNAMTS-DGFiP File on physicians for 2005-2008.	Fuzzy RD using increase in the requirements for GPs to "bill freely" in their contracts with public health insurance.	Price regulation increased the supply of medical care by 66.53% and the number of procedures by 84.23%.	Provision of total medical procedures wrt payment= 1.61	Male GPs increase labor supply more and also increase home visits and prescriptions.
Fortin et al. (JAE, 2021)	Compare FFS contracts vs. contracts that pay a per diem plus a smaller amount per service. Effects on care rendered by pediatricians?	Doctor time-use survey linked to records from Health Insurance Organization of Quebec (1996–2002).	Structural discrete choice model with variation from a reform introducing an optional per diem plus payment contract.	Small changes in time spent with patients, but services rendered under mixed remuneration contract decrease by 5-12%.	Elasticity of hours wrt wages ~0. Elasticity of services: -0.124.	Female doctors and younger doctors are more likely to switch to the per diem contract.

Johnson and Rehavi (AEJ:EP, 2016)	How is the probability of C-section affected if the patient is a physician? Is there an interaction with financial incentives?	Confidential CA Vital Statistics data, 1996-2005; CA physician licensure data; TX birth data 1996–2003 and 2005–2007.	Comparison group is educated mothers. Nearest neighbor matching regressions for CA. Hospital fixed effects.	California physicians are 1.17 pp (6.13%) less likely to have an unscheduled C-section at non-HMO hospitals. In Texas physicians are 2.09 pp (6.39%) less likely to receive a C-section. Financial incentives affect C-section rates only among non-physicians.	Elasticity~0 for physician-mothers. Non-zero for other mothers but not computable from paper.	Effects greater for physician parents who specialize in areas related to childbirth.
Papers about Capitation/Managed Care Organizations.						
Dickstein (WP 2017)	Are there differences in how physicians in capitated plans prescribe for depression compare to physicians in non-capitated plans?	MarketScan: 2003-2005 Commercial Claims & Benefit Plan Design Data; County-level IRS Income; National Ambulatory Medical Care Survey.	Structural model, instrumenting drug price with sum of price changes within an insurer's plan for all other drugs.	Prescribers in capitated plans are more likely to choose generic Rx. Patients have higher adherence and less medication switching but also higher relapse rates.	Lower drug switching may promote adherence but has negative effects on patients at highest risk of relapse.	
Ding and Liu (JHE, 2021)	How does capitation affect treatment of lower back pain?	MarketScan Commercial Claims 2003- 2006.	Plan history FEs and physician FEs.	Providers with capitation use 12.2% fewer medical resources to evaluate and treat lower back pain with no effect on relapse probabilities.	Effects are biggest for physical therapy and diagnostic testing. But do capitated providers report all procedures?	
Chorniy, Currie, and Sonchak (JHE, 2018)	How does switching from FFS to MMC affect children's treatment of asthma and ADHD?	60% random sample of all South Carolina (SC) Medicaid enrollees < 17, 2005-2015; Vital Statistics	Staggered roll out of MMC contracts with higher capitated payments for children with chronic conditions; child FEs.	Switching to MMC increased ADHD caseloads by 11.6% and asthma caseloads by 8.2%. No significant effects on hospitalization and increases in ER use.	N/A.	
Physician Detailing						
Agha and Zeltzer (AEJ: EP, 2022)	How do pharma payments affect the prescribing of physicians who only share patients with physicians who receive payments?	Medicare Part D (2014–2016); Open Payments database (2013–2016); CMS Referral Patterns;	Event studies; DiD-style regressions with doctor-drug and drug-quarter-specialty FEs	Peers of physicians who receive payments for speaking, consulting, etc., increase prescribing of the promoted drug 1.8%. Spillovers account for ¼ of increased prescribing	Effects are larger for peer physicians with more shared patients with the physician receiving payments.	

		Physician Compare.		from payments.	
Carey, Daly, and Li (NBER WP, 2024)	How do pharma payments affect the prescribing of physician-administered cancer drugs in Medicare?	Open Payments database; claims from 20% sample of Medicare FFS (2014–2018).	D-in-D and event study models with physician-drug and time-drug FEs.	Payments increase Rx the marketed drug by 4% in the year after payment. No improvement in patient mortality. No elasticity because payment value not reported.	Targeted doctors increase treatment of patients with lower expected mortality.
Carey, Lieber, and Miller (JPubE, 2021)	How does detailing affect physician prescribing behavior in terms of drug efficacy, and use of generics?	20% Medicare Part D 2013-2015; Open Payments database; hand-collected data on drug efficacy.	Event studies with physician by drug FEs	Prescribing of the detailed drug increases by 2.2% in the 6 months following payment. No significant effects on efficacy or transitions to generics.	Results are similar when restricting sample to physicians who receive small payments.
Newham and Valente (JHE, 2024)	How do gifts to doctors from pharmaceutical companies affect antidiabetic drug prescribing patterns and costs?	Open Payments database; Medicare Part D data (2014–2017); demographic and health data from ACS and CDC.	Compare physicians with similar propensities to receive payments and use random timing. Residuals from outcome models regressed on residuals from payment models.	An increase in payments by the average yearly payment of $65 increases Rx of branded antidiabetic drugs by 4.8%, increasing costs of Rx drugs.	Effects are higher for doctors in areas with a higher proportion of patients receiving subsidies for out-of-pocket drug costs for low-income individuals.
Shapiro (MS, 2018)	Compare effect of new information from clinical trials and detailing on PCP prescribing behavior for Seroquel.	AlphaImpactRx monthly panel of 1,762 PCPs 2002-2006 (links self-reported detailing, patient treatment).	Two clinical trials over sample period, plus record of detailing. Examine effects in models with physician and month FEs.	No effect of the clinical trial information. Detailing increased after both trials. Detailing increased Seroquel Rx 26% in the month of the visit.	One third of the increase in prescribing occurred in off-label uses.
Other Papers without Defined Elasticities					
Alexander (JPE, 2020)	When hospitals offer incentives to physicians to lower costs, does it affect (1) who is admitted (2) which hospital they are	New Jersey Uniform Billing Records (2006-2013); AHA annual survey; Medicare cost-to-	D-in-D with doctor FEs using the New Jersey Gainsharing Demonstration as a policy experiment.	The policy doesn't reduce costs or change procedure choice. But lower predicted cost patients are sorted towards participating hospitals.	Effects are less precisely estimated for surgical patients, where there is less opportunity for gaming.

	admitted to, and (3) how intensely they are treated?	charge ratio series.			
Alexander and Currie (EHB, 2017)	What is the effect of private vs. public insurance on propensity to be admitted to hospital from ED? Are effects moderated by capacity constraints?	New Jersey Uniform Billing Records 2006- 2012.	Exogenous variation in hospital bed supply due to local flu conditions; hospital FEs.	In high flu weeks, publicly insured children are .3 p.p. (6.4%) less likely to be admitted for non-flu conditions compared to privately insured children. Outcomes are no worse for marginal children.	Effects are larger when restricting to diagnoses with mid-range admissions rates.
Brekke et al. (JHE, 2019)	How does GP compensation and relationship with patients affect their propensities to issue sick-leave certificates patients need to claim benefits?	Norwegian administrative data 2006–2014 linking health, national insurance, and labor market data.	Physicians see patients both in their own practices and in EDs where they do not face reputational effects. Models with physician and patient FEs.	GPs with a FFS contract are 34.63% more likely to issue sickness certificates for own patients vs. ED patients. For GPs with fixed salaries the gap is 24.15%.	GPs with new practices have similar effects with FFS but not for fixed salary. The effect for fixed salary is driven by relationships with patients. Effects larger in areas with more GPs per capita and where GPs have more openings.
Chernew et al. (JHE, 2021)	How much of the variation in prices for lower-limb MRIs is explained by physician referral patterns vs. patient characteristics?	2013 insurance claims from a large national insurer; data from the company's online price comparison tool; SK&A physician-level dataset.	Restrict to lower-limb MRIs without contrast since these are "shoppable, homogeneous MRI scans." Estimate models with referrer FEs.	Referrer FEs explain 52% of the variance in patient spending on lower-limb MRIs. Patient cost-sharing and characteristics explain less than 1%. Patient HHR FEs explain 2%. Going to the cheapest provider within the same driving distance would reduce spending 35.83%.	The mean vertically- integrated physician refers 52% of patients to a hospital-based MRI provider compared to 19% for non-vertically-integrated physicians.
Clemens et al. (NBER WP, 2024)	How do measures of provider preferences for treatment intensity relate to utilization and spending for commercially insured patients? How do financial incentives mediate these relationships?	Health Care Cost Institute Commercial Claims Database; survey data from Cutler, Skinner, Stern, and Wennberg (2019)	Descriptive analysis following Cutler et al. (2019) with additional covariates to represent different financial incentives in commercial insurance.	Provider preference measures (share Cowboy, Comforter High Follow-Up, Low Follow-Up) are weakly related to utilization and spending, in contrast to Cutler et al. (2019). Private insurance offers lower prices in areas with a higher share of Cowboys/High Follow-Up, offsetting provider preferences.	Relationship between provider preference measures and non-price utilization measures are weaker than relationship between provider preference measures and payments.
Frakes	Does physician behavior	National Hospital	Focus on AMI and C-	After adoption of a national-standard	Disaggregates by whether states

(AER, 2013)	converge towards national averages when states change malpractice laws to consider national rather than local norms?	Discharge Survey (1977-2005), Natality Data (1978-2004); Mortality Data (1977-2004).	section. Event study exploiting variation in states adoption of national-standard rules.	rule, the deviations between state and national C-section rates fall by 4.87 pp (48.31%). Estimates for AMI are noisier. No convergence in outcomes.	have rates that are initially higher or lower rates than the national rate. Convergence occurs in subsamples.
Gupta (AER, 2021)	Effects of the Hospital Readmissions Reduction Program (HRRP) on care quality and admissions for patients with heart attacks, heart failure, and pneumonia?	Medicare fee-for-service claims 07/2006-07/2006; 20% sample of all Medicare beneficiaries.	D-in-D, IV using baseline predicted readmission rate.	HRRP reduced 30-day readmissions by 10.5% and 30-day returns to the hospital by 6.92%. Little effect on admission decisions or upcoding. Increases in procedures for AMI patients and 8.87% fall in 1-year mortality.	Readmission rates lower for patients initially admitted to index hospital, not for those originally seen elsewhere. Government hospitals respond less. Higher volume hospitals and at-risk systems respond more.
Howard and McCarthy (JHE, 2021)	Did a DOJ investigation of Medicare fraud re: implantable cardiac defibrillators (ICDs) change practice?	All-payer data from Florida; ED data from Florida's Agency for Healthcare Administration.	D-in-D using ICD procedures not subject to the investigation as a control.	The investigation plus new checklists that were part of the settlement caused a 22% decline in unnecessary ICD implantations.	The decline in ICDs was stronger for hospitals involved in the lawsuit. Decline for Medicare patients smaller in percent but larger in absolute terms compared to patients with other insurance.
Johnson et al. (NBER 2016)	Are OBs more/less likely to do unscheduled C-sections on own patients? Effects recent patients' laceration rates?	EMR and billing databases for three practice groups.	They use rotating call schedules of OB groups as a plausibly exogenous source of OB assignments.	OBs are 4 pp (25.97%) more likely to perform a C-section and 2.5 pp (25.0%) less likely to use vacuum or forceps on their own patients vs. another OB's.	Higher rates of recent lacerations increase the probability of C-section for an OB's own patients but not for other patients.
Wilding et al. (JHE, 2022)	How did increased stringency of blood pressure targets for patients <80 affect English GPs' treatment and testing decisions for hypertensive patients?	EHRs from Clinical Practice Research Datalink (04/2010-03/2017); Health Survey for England.	D-in-D comparing patients over and under 80; bunching estimators.	Stricter targets did not increase diagnoses of hypertension in new patients but increased antihypertensive Rx 1.2 pp. Doctors did multiple tests when patients failed, reported more patients as exempt from reporting, and increased reports of patients exactly meeting targets.	Lower-performing practices increased reporting of patients as exempt more than higher-performing practices, but other effects were similar. No data on health outcomes.

Note: One could compute detailing elasticities for some of the papers above, but these measures are difficult to interpret because detailing involves more than payment. Carey, Lieber, and Miller (JPubE, 2021) find that effect sizes are very similar when restricting to small payments, suggesting that direct remuneration is not the main reason that detailing affects physician decision making.

Appendix Table 6: Doctor Responses to New Information

Paper	Research Question	Data	Methods	Results	Heterogeneous Effects?
Avdic et al. (JHE, 2024)	New stents were first thought to reduce complications and then to increase them. How did cardiologists respond to new information and guidelines?	Swedish Coronary Angiography and Angioplasty Registry 2002-2011.	Separate models for periods after positive info, after negative info, and after guidelines allow physician-specific intercepts and trends.	Doctors responded more quickly to negative information than to the initial positive information.	Doctors slow to take up new stents were more likely to use the appropriate stent and had better patient outcomes. No heterogeneity within hospitals. Slow responders more likely to practice in teaching hospitals.
Ahomaki, Pitkanen, Soppi, and Saastamoinen (JHE, 2020)	Experiment with letters sent to Finnish doctors who prescribed 100+ paracetamol-codeine pills to a new patient.	National Prescription Register including all purchases, merged to Nordic Product Number and physician characteristics.	D-in-D using new patients where non-targeted physicians are the control. "Treatment" is intent-to-treat.	Significant 6.13 tablet decrease in number of pills purchased by new patients of treated doctors relative to patients of untreated doctors (12.8% of treatment group baseline).	Treatment effects larger for high prescribers. Top 5 specialties have Similar effect size. The decrease in large purchases was greatest in urban areas and not significant in rural areas.
Bradford & Kleit (HE, 2015)	The effect of the 2005 Blackbox warning on NSAID prescriptions, and how it was mediated by advertising, media coverage, and patient characteristics.	EMRs from the Primary Care Practices Research Network; media data from Competitive Media Reporting, Inc. and Lexis/Nexis; NSAID sample dispensation data from IMS health.	Probit models on having active prescription for non-COX-2 inhibitor NSAIDs, COX-2 inhibitor NSAIDs, opioids, and other analgesics.	Blackbox warnings resulted in a 2.8pp (54.90%) decrease in prescriptions for COX-2-inhibitors and 2.8pp (23.14%) increase in prescriptions for a non-COX-2-inhibitor ($p<.001$).	Patients with cardiovascular disease had a similar decrease in prescription of COX-2-inhibitors, but no significant increase in non-COX-2-inhibitors. These patients substituted toward opioids and other analgesics.
Currie and Musen (Working Paper, 2025)	Effect of prior authorization policies on prescribing of antipsychotics to kids on Medicaid.	New hand-collected data on Medicaid prior authorization policies (2005–2020); IQVIA LRx database of psychotropic Rx (2006–2019).	Staggered DiD using state-level rollout of prior authorization policies.	Comprehensive pediatric prior authorization policies reduced providers' prescribing of antipsychotics to children ages 3-5 on Medicaid by 30%.	No spillovers to older children or children on private insurance, suggesting hassle costs instead of information as the primary mechanism behind main findings.
DeCicca, Isabelle, and Malak (HE Letters, 2024)	Effect of Term Breech Trial and its subsequent overturning on C-sections for breech births.	U.S. Birth Certificate Records 1995–2010.	D-in-D using complication-free births as control group.	No effect of original Term Breech Trial on C-section rates. Reversal of trial findings reduced C-sections for breech babies by 15–23%.	Reductions in C-sections greater in counties with younger physicians and more IMGs and among non-white, less educated patients.
Doctor, Nguyen,	Effect of notification of	Opioid dispensing from	RCT with intent-to-	Milligram morphine	N/A

Lev, Lucas, Knight, Zhao, and Menchine (Science, 2018)	patient death by overdose on future opioid prescribing.	California's Prescription Drug Monitoring Program database.	treat analysis. Letters from the Chief Medical Examiner of CA.	equivalents prescribed down 9.7% in treatment vs. control 3 months after intervention.	
Dubois and Tuncel (JHE, 2021)	How did French physicians respond to the 2004 information that SSRIs increase suicidal thinking in children?	Cegedim proprietary longitudinal patient data covering all prescriptions by 386 GPs. Includes doctor and patient demographics, and visit-level information.	D-in-D estimation, older patients are control. Random coefficient discrete choice logit examines choice across drug categories.	Child SSRI prescriptions fell 9.9 pp (19.8%). The baseline effect for adults was -2.8 pp (5.6%). Many physicians decreased prescription of other classes of anti-depressants but substituted to off-label use of other drugs.	25% of the physicians prescribe an SSRI for depression <20% of the time before the warning, and 25% prescribe an SSRI >73% of the time. Over 25% of physicians never prescribe SSRIs to children after the warning.
Howard, David, and Hockenberry (JEMS, 2016)	Variation in surgeon responses to the information that arthroscopic knee surgery is ineffective by whether it is a hospital or a free-standing surgery.	Outpatient claims data from Florida's State Ambulatory Surgery Database, 1998-2000. Surgeons cannot be linked over time. Analysis at facility level.	Triple D-in-D, alternative model using differential trends in the ratio of knee to shoulder surgeries (preferred specification).	Preferred specification: if free-standing centers responded like hospitals the number of surgeries would be reduced 6.27-11.37% on a baseline of 34,000 each year.	Disaggregating by procedure type, the differential decline between free-standing centers and hospital centers is driven by meniscectomies, which have received more insurance company scrutiny.
Howard and Hockenberry (HSR, 2019)	How is physician age related to the response to new information that episiotomies are ineffective?	Pennsylvania Inpatient Hospital Discharge Data (1994–2010)	Descriptive. LPM with hospital FEs.	Physicians who started delivering babies 10 years earlier are 6 pp (19.5%) more likely to perform an episiotomy.	The relationship between physician age and episiotomy rate has decreased over time and is weaker in teaching hospitals, which promote evidence-based medicine.
Kolstad (AER, 2013)	Effects of quality "report cards" for Coronary Artery Bypass Graft (CABG) surgeries. Is provider response profit motivated?	Pennsylvania Health Care Cost Containment Council data for 89,406 CABG surgeries 1994-1995, 2000, and 2002-2003 merged with surgeon tenure. Focus is on the surgeons' mortality rate before report cards less the report card risk-adjusted rate.	Reduced form responses to differences between own mortality rates and other doctors'. Structural model of consumer demand separates "intrinsic" and "extrinsic" motivations.	Counterfactuals indicate that "extrinsic" incentives induced a 3.5% decline in predicted risk-adjusted mortality whereas "intrinsic" incentives induced a 13% decline in predicted risk-adjusted mortality.	The response is larger for surgeons who are worse than other surgeons in their own hospital compared to surgeons who are just worse than expected.

McKibbin (JHE, 2023)	How do physicians change prescribing of off-label cancer drugs in response to new information from RCTs?	Data on FDA approvals and RCT results, 100% Outpatient and 20% Carrier Claims files for Medicare part B, 1999-2013.	Event studies comparing drug-cancer pairs with and without newly presented RCT evidence from academic conferences.	8 quarters after a conference, prescriptions of drugs with confirmed efficacy up 192%. Prescribing falls by 33% over 8 quarters with negative information.	Responses discontinuous around p-value 0.05. When the abstract describing the RCT has no mention of improvements in quality of life or side effects, adoption and de-adoption rates are less asymmetric.
Olson and Yin (HE, 2021)	Physician responses to changes in drug labeling from the FDA's 1997 Pediatric Exclusivity provision (provides 6 months of exclusivity in return for conducting Pediatric trials).	Prescription data from NAMC; Label changes and exclusivity from FDA; journal publication data from Benjamin et al. (2006) and PubMed; IMS health data on drug promotions; disease prevalence from MEPS.	D-in-D with treatment group defined as children <18 years old and controls as adults >35 (using a zero-inflated negative binomial model).	In their preferred specification, the marginal effect of a pediatric label change is 2.09 fewer prescriptions (12.67 %) for children.	Negative information added to the label reduces prescribing more than positive information. Magnitudes are larger for physicians in solo practice. No clear pattern by child age group. Estimates somewhat sensitive to included controls.
Persson et al. (NBER WP, 2021)	Do doctors consider the diagnosis of an older sibling when evaluating children for ADHD?	Swedish population register 1990-2018, (2016 for HS records); prescription drug claims July 2005-Dec. 2017; birth records data from NHBW, 1996-2016.	Birthday cut-off RD using older sib or cousin's birth date and school eligibility cutoffs to use "young for grade" sib's higher prob. of ADHD diagnosis.	An older sibling born after the school entry cutoff decreases the probability of ADHD diagnosis by 0.59 pp (12.04%) and decreases the probability of ADHD drug claims by 0.55 pp (9.82%). Smaller results for cousins.	Effects on younger siblings are greater before older siblings graduate from HS. Spillovers greater in cities with more funding for special needs children. Cousin spillover effects are greater when cousins are in the same municipality.
Sacarny, Yokum, Finkelstein, and Agrawal (HA 2016)	Effect of letters from Medicare to outlier prescribers of controlled substances on future opioid prescriptions.	CMS Integrated Data Repository-- records for prescription drugs covered by Medicare Part D with prescriber ID.	RCT with analysis of intent-to-treat.	Statistically insignificant increase of 0.8% relative to the control mean after 90 days, 95% CI (-1.38%, 2.91%).	No evidence of heterogeneity by prescriber specialty, geographic region, prescribing pre-treatment, and whether the physician had been investigated for fraud.
Sacarny, Barnett, Le, Tetkoski, Yokum, and Agrawal (JAMA Psych, 2018).	Effect of three letters sent by Medicare to outlier prescribers of quetiapine on future quetiapine prescriptions.	100% Medicare claims data 2013-2017; enrollment data 2015-2017; risk-adjustment data 2013-2014.	RCT with analysis of intent-to-treat.	11.1% fewer days over 9 months vs. control mean (11.99% of the sample mean). Effects lasted 2+ years. No negative effects on patients.	The reduction in prescribing was larger for patients with low-value indications and smaller for guideline-concordant patients.

Wu and David (JHE, 2022)	How did relative procedural skill affect the prob. that doctors abandoned laparoscopic hysterectomy after a negative info shock about the safety of the procedure?	All hospital inpatient and outpatient visit data for patients receiving hysterectomies in Florida (January 2012 – Sept. 2015).	Leave-one-out IV for physician skill at laparotomy/ laparoscopic hysterectomy; DiD event study estimates before/after 2014 FDA announcement.	A 1 SD increased in relative skill in laparoscopic hysterectomy decreased prob. of abandoning the procedure by 4.6–4.9 p.p. (6.2–6.5% reduction from pre-period mean). Only top laparotomy doctors increased laparotomies.	Patients with characteristics that indicate less appropriateness for the laparoscopic procedure had greater reductions in likelihood of receiving a laparoscopic procedure after the announcement.

Appendix Table 7: Heuristics and Guidelines

Paper	Research Question	Data	Empirical Methods	Results	Heterogeneous Effects?
Abaluck et al. (NBER WP 2021)	How does the proportion of physicians following guidelines for anticoagulants for atrial fibrillation patients change after 2006 guidelines? Is lack of implementation due to awareness or nonadherence?	Text mining of EMRs from the VA for patients newly diagnosed with atrial fibrillation between Oct. 2002-Dec. 2013; Patient-level data for 8 clinical trials of anticoagulants.	Causal-forest model to estimate heterogenous treatment effects using data from eight RCTs; Chernozhukov et al. (2018) approach to calculating best linear predictions of conditional average treatment effects.	After 1st mention of guidelines, physicians become more compliant. Stricter adherence could prevent 24% more strokes.	Most departures from guidelines are not justified by measurable treatment effect heterogeneity (though RCTs were not originally randomized on the observables analyzed).
Almond et al. (QJE, 2010)	Does the care of newborns change discretely at the threshold for being classified "very low birthweight" and does this affect mortality?	NCHS linked birth/infant death files (1983-1991 and 1995-2002); linked birth, death, hospital discharge data from California (1991 -2002); HCUP for AZ, NJ, MD, NY.	RD centered around threshold of 1,500 grams.	Relative to the means just above the threshold, VLBW classification has an 11.11% effect on spending and a 5.93% effect on length of hospital stay.	Effects are greater for non-NICU and Level 0/1/2 NICU hospitals than for Level 3A-3D NICU hospitals.
Coussens (Working Paper 2022)	Do doctors use simple heuristics in patient age to make treatment decisions for ischemic heart disease (IHD)?	Truven Commercial Claims and Encounters database 2005-2013; ED records from a large Boston-area hospital 01/2010-05/2015.	Regression discontinuity centered at age 40	Turning 40 increases the probability of being tested, diagnosed, or admitted for IHD by 0.887pp, 0.131pp, and 0.068pp, respectively. Relative changes compared to intercepts are 9.51%, 19.29%, and 17.80%, respectively.	Effects are larger for women and patients presenting without chest pain. Effects are also stronger when the ED is less busy and in the 1st half of a physician's shift.
Cuddy and Currie (PNAS, 2020)	What is the probability that adolescents with private insurance receive appropriate care following an initial diagnosis of mental illness? What factors are related to the type of care received?	Claims data for a large national insurer. Children covered for at least a year between 2012 and 2018 who were ever diagnosed with a mental health condition.	Observational study using linear probability models. Define "red-flag" treatment as prescribing that falls outside accepted guidelines.	Only 75% of adolescents receive follow-up care within 3 months. Of those receiving drugs, 44.85% receive "red flag" drugs. Composition of clinicians affects treatment: More psychiatrists→more drug use vs. more therapists →more therapy.	Any treatment, drug treatment, red-flag drugs increase with age. Girls more likely to be treated, to get therapy, and to get be red-flag drugs. Variation *across* zip codes explains less than half of overall treatment variation.
Cuddy and Currie (JPE, forthcoming)	Would adherence to guidelines improve outcomes? Is there a	Claims data for a large national insurer. Children diagnosed with depression	Instrument individual prescriptions with area-level practice style	Outcomes for red-flag vs. grey-area vs. FDA approved drug treatment after 24	P(drug treatment) is higher for girls, older children, and children whose 1st visit

	difference between "grey-area" prescribing sanctioned by professional societies but not by FDA, and "red-flag" prescribing not sanctioned by either?	or anxiety for the first time 2012-2018. Measures of local practice style computed from IQVIA and from the claims data.	measures interacted with patient characteristics (use Lasso to choose instrument set).	months: P(self-harm): 5.8%; 4.9%; 3.8%. P(ED or hosp.): 33.6%; 18.6%; 26.8%. Total costs: \$9557; \$1745; \$9658. Red-flag has highest costs and worst outcomes.	resulted in hospitalization.
Currie and MacLeod (Econometrica 2020)	Would adherence to professional guidelines improve outcomes? Does the answer to this question vary with the physician's skill?	Claims data for a large national insurer. Adults ever diagnosed with depression 2013-2016; NPPES; Experimental propensity is measured using prescription dispersion across drugs in IQVIA Xponent prescription data base.	Patient FE models of effects of having more experimental doctors and of violations of guidelines. Simulations measure benefits of experimentation for different skill groups. (Psychiatrists assumed more skilled than GPs).	Violations of professional guidelines are associated with worse subsequent outcomes (spending, hospitalizations, ED visits) for all patients.	Among patients seeing psychiatrists, switching to a more experimental doctor improves outcomes (a 0.25 increase reduces P(ED visit or hospitalization) by 10.2%). No effect of experimentation with less skilled doctors.
Geiger et al. (JAMA HF, 2021)	What is the effect of a designation of "advanced maternal age" (AMA) on prenatal care and birth outcomes?	Claims and monthly enrollment data from a large, nationwide commercial insurer 2008-2009; zip-code level public ACS data.	Focus on discontinuities in care for mothers 35+ on expected delivery date. Donut RD excluding women with due dates within 7 days of their 35th birthday.	AMA increases screening, specialty visits; decreases perinatal mortality by 0.39pp or 42.39% of sample mean. No effects on severe maternal morbidity, preterm birth, or low birth weight.	As a percentage of baseline the effects on prenatal care services and perinatal mortality are much greater for low-risk pregnancies than for the full sample.
Kowalski (ReStud, 2023)	Are women who are more likely to receive mammograms different from women who are less likely? How does the probability of being "over-diagnosed" vary with the propensity to receive mammograms?	RCT data from the Canadian National Breast Cancer Screening Study (CNBSS) linked to cancer registries and the mortality data. Allows long-term follow up to see cancers that are detected but would not have caused symptoms.	Extension of Imbens and Angrist (1994) framework in the context of an RCT (which provides identifying variation).	In women who are treated compliers w.r.t. screening guidelines, 14% of breast cancers are "over-diagnosed". For always takers, over 36% of breast cancers are over-diagnosed. Results suggest current guidelines should be revised to reduce mammography.	Women who are more likely to receive mammograms are healthier and of higher socioeconomic status on average.
Ly (Annals of Emergency Medicine, 2021)	Are physicians more likely to test for pulmonary embolism (PE) in the ED when they recently treated a patient with PE?	National EHR data from the VA Corporate Data Warehouse (2011–2018)	Linear probability model with time and physician FEs and clinical and demographic covariates	In the first 10 days after treating a patient with PE, physicians increase testing for PE by 15%. No change in testing behavior in the 50 days after the first 10 days.	N/A.

Ly, Shekelle, and Song (JAMA Internal Medicine, 2023)	Do physicians delay testing for pulmonary embolism (PE) in patients with congestive heart failure presenting in the ED with shortness of breath when congestive heart failure is documented in triage?	National EHR data from the VA Corporate Data Warehouse (2011–2018)	Linear probability model with time and physician FEs and clinical and demographic covariates	The mention of congestive heart failure in triage reduced testing in the ED by 4.6 p.p. (34.8%) and delayed testing in the ED by 15.5 minutes (20.5% increase). Patients were 0.15 p.p. (65.2%) less likely to be diagnosed with PE in the ED but no difference in diagnosis of PE w/in 30 days.	N/A.
Olenski et al. (NEJM, 2020)	Do physicians use simple heuristics in patient age to make treatment decisions for Coronary Artery Bypass Graft Surgery (CABG)?	Medicare data from 2006 to 2012.	Regression discontinuity at age 80.	Patients admitted in the 2 weeks after their 80th birthday were 1.7pp (28.05%) less likely to get CABG than patients admitted 2 weeks before their birthday.	N/A.
Singh (Science, 2021)	Do physicians switch delivery mode after a complication with their previous patient?	EHR (2000–2020) from the obstetric wards of two academic hospitals.	Linear probability model with time, physician, and hospital FEs and clinical and demographic covariates	After a complication with a C-section, physicians are 3.4% more likely to use a vaginal delivery with the next patient. After a complication with a vaginal delivery, physicians are 3.6% more likely to use a C-section with the next patient.	Effects are larger for more experienced physicians.

Appendix Table 8: Technology

Paper	Research Question	Data	Empirical Methods	Results	Heterogeneous Effects?
Agarwal et al. (NBER WP 2024)	How do radiologists use AI predictions and clinical histories in diagnosis? What is optimal use of AI?	Patient cases from Stanford University healthcare; data from an experiment on radiologist decisions and decision time.	2x2 experiment with radiologists. Add AI prediction, clinical history from referring doctor, or both; random forest regression.	AI does not improve performance. Access to clinical history reduces deviation from diagnostic standards by 4%. Optimal to have AI decide cases when confident and radiologists decide all other cases w/o AI.	When the AI tool has high confidence, AI improves radiologist diagnosis. When the tool has low confidence, AI worsens radiologist diagnostic accuracy.
Agha (JHE 2014)	Impact of EMRs plus clinical decision supports on quality and cost of care.	20% sample of Medicare claims, 1998- 2005; Health Information and Management System Survey.	Exploits differential timing of Health Information Technology (HIT) adoption at hospital level w FE.	HIT adoption increases spending 1.3%. No effect on 1-year patient mortality, length of stay, #physicians seen within a year of admission, intensity of care, 30-day readmissions, complications, or an index of care quality.	No evidence of higher returns to more comprehensive HIT systems. Do not see larger effects in larger hospitals.
Alpert, Dystra, and Jacobson (AEJ:EP, 2024)	How much does information versus hassle costs from MA-PDMPs affect opioid prescribing?	Claims data from Optum's Clinformatics Data Mart (2006–2016).	DiD and event studies using policy change in Kentucky. Triple differences comparing opioid naïve and non-naïve patients.	Hassle and information explain 69% and 31% of fall in opioid Rx respectively. MA-PDMPs reduce opioid Rx 6.8% for opioid naïve patients, 10.6% for non-naïve patients, and 16% for patients with opioid-inappropriate conditions.	Declines in prescribing to opioid non-naïve patients occur for patients with history of doctor shopping or high dose/quantity of opioid use.
Arrow, Bilir, and Sorenson (AEJ: AE 2020)	Does access to an electronic database for pharmaceuticals affect doctors' prescribing of cholesterol drugs?	IMS Health Xponent database 2000-2010; data from the firm that owns the studied electronic reference database.	Models with zip-code-month FEs, physician FEs, and physician-specific time trend; IV doctor's access using share of area doctors using database.	Database increases prescribing of generic Rx in its 1st year by 1.3 pp (3.7%). No effect on new branded Rx. New and old generic Rx increase; Old branded Rx decrease. Providers prescribe 0.7% more unique Rx.	In zip codes with more pharmaceutical patenting, database has less effect on drug adoption. Effects stronger for providers who access the database more frequently upon adoption.
Buchmueller and Carey (AEJ: Economic Policy, 2018)	How do MA-PDMPs versus PDMPS without must-access provisions affect opioid use in Medicare?	PDMP info from Prescription Drug Abuse Policy System; 5% Medicare beneficiaries in Part D and FFS in any year 2007–2013.	DiD and event study models using variation in state-level policy.	Without must-access provisions PDMPs have no effect on opioid utilization. MA-PDMPs reduce doctor shopping by 8% and pharmacy shopping by 15%. Neither PDMP significantly affects opioid poisoning rates.	Effect sizes are larger must access provisions are broader.

Buchmueller, Carey, and Meille (Health Economics 2020)	Effect of Kentucky's must-access PDMP program on opioid prescribing.	Kentucky (2006-2016) and Indiana (2012-2016) PDMPs; CDC data on opioid prescriptions; ARCOS 2006-2016.	DiD comparing Kentucky (treated) to Indiana (control).	Quarterly morphine equivalents per capita fell 11–13% in KY vs. IN. Providers prescribing any opioids fell by 3.8 pp (5%). The number of patients prescribed fell 16% among providers prescribing any opioids.	Providers who initially prescribed fewer opioids were more likely to stop prescribing. Reductions in prescribing greater for patients who used opioids multiple times and doctor-shoppers.
Dahlstrand (Working Paper, 2021 updated 2024)	How much could patient outcomes be improved by using an algorithm to match patients and GPs?	Data from Sweden's largest digital healthcare platform (2016–2018) matched to Swedish registry health data.	Physician skill estimated using leave-one-out measures with shrinkage. Match effects exploit the platform's conditional random assignment of patients.	Using an algorithm with positive assortative matching could reduce avoidable hospitalizations by 8%, all hospitalizations by 3%, and counter-guideline antibiotic Rx by 3%.	Effects are smaller for patients seeing a doctor within the day/hour. In urban areas, similar improvements are possible by restricting matches to doctors patients can travel to see in person.
Ellyson, Grooms, and Ortega (Health Economics 2022)	Do the effects of must-access PDMPs vary by specialty?	CMS Part D public use files 2010–2017; AMA Physician Masterfile; PDMP start dates from Prescription Drug Abuse Policy System.	DiD and event study.	Primary care doctors decrease opioid prescribing by 4% after MA-PDMP implementation. No significant effect for providers in IM, EM, surgery, palliative care, oncology, and pain medicine.	Primary care and IM providers with initially low prescribing stop prescribing opioids after MA-PDMP.
Goetz (International Journal of Industrial Organization 2023)	How does an increase in competition on a telehealth platform for talk therapy affect providers' pricing and exit decisions?	Therapist data collected from Psychology Today in 2020; controls from Canadian government sources and Facebook's Movement Range maps.	Propensity score matched DiD exploiting change in how platform shows providers to patients. For areas with <20 providers, platform made providers outside area visible.	Increased competition caused by the platform displaying more providers decreases the likelihood that affected providers provide sliding scale discounts by 8.9%.	Providers with more training respond to competition by stopping sliding scale offers; providers with less training exit the platform. Bigger effects on late adopters of teletherapy.
Horwitz et al. (NBER Working Paper 2024)	How do Certificate of Need (CON) laws affect imaging? How does this vary by the value of imaging?	Hand-coded laws; AHA's Annual Survey of Hospitals 2018; accreditor data on free-standing CT/MRIs; 20% sample Medicare FFS claims 2009–2014.	RDD at state borders where one state has a CON law and the other does not.	The prob. of receiving an MRI is 2% lower on the CON side of the state border, compared to the mean on the non-CON side. Overall, no effect on prob. of a CT.	The prob. of receiving a high-value MRI does not change at border, the prob. of receiving a high-value CT on the CON side falls by 6% of non-CON mean. Low-value imaging falls 20–26%.

McCullough et al. (Health Affairs 2010)	How is quality of care related to EMR adoption 2004-2007?	AHA's annual survey; Health Information and Management Systems Society Analytics database.	OLS with hospital and year fixed effects, coefficient of interest is on the one-year lag of EMR adoption.	Pneumococcal vaccination rates up 2.1pp (3.2%); use most appropriate antibiotic for pneumonia up 1.3pp (1.6%). No effect on other quality of care measures studied.	The relationship between quality measures and EMR adoption is stronger in academic vs. non-academic hospitals.
Miller and Tucker (JPE 2011)	Does EMR adoption lower neonatal mortality.	Linked birth and infant death data 1995–2006; AHA surveys; BEA Regional Accounts; CBP; HIMS Analytics Data; Georgetown Health Privacy Project; Lexis-Nexis.	Construct balanced county-level panel over 12 years. OLS w county and year FEs; IV for EMR adoption using state medical privacy laws.	A 10% increase in EMR adoption reduces neonatal mortality by 3%. Reductions are due to prematurity and complications not to accidents, SIDS, or congenital defects.	Larger effects when EMRs combined with digital storage, and obstetric-specific/decision support technologies. Larger gains for mothers who are Black, Hispanic, unmarried, or have < high school education.
Neumark and Savych (American Journal of Health Economics, 2023)	How do MA-PDMPs and laws that limit initial opioid Rx length for patients with work-related injuries?	Workers Compensation Research Institute claims for workers injured Oct. 2009 – March 2018.	DiD using state-level variation in laws.	Laws that limit opioid Rx length have no effect on opioid Rx (w/pre-trend w/o state trends). MA-PDMPs reduce opioid Rx on intensive but not extensive margin. For neuro spine pain, non-opioid pain Rx increase 14%.	Effects of MA-PDMPs are larger for neurologic spine pain, spine sprains and strains, and other sprains and strains cases.
Obermeyer et al. (Science 2019)	Is there racial bias in algorithms used to target care for high-risk patients? Do doctors correct for algorithmic biases?	Data from all primary care patients enrolled in risk-based contracts at a large academic medical center, 2013-2015.	Descriptive statistics and simulations.	Conditional on chronic condition, Black patients get less recommended care. Black patients have 26% more chronic conditions at the 97th percentile of the risk score. Simulations suggest that physicians do not counteract bias in the algorithms.	Algorithm was trained on spending. Conditional on diagnosis, Black patients have lower spending and algorithm reproduces this bias. Changing algorithm to target health outcomes could potentially resolve the problem.
Mullainathan and Obermeyer (QJE 2022)	Ask how the actual decision to test for heart attacks differs from algorithmically predicted risks and explore health implications.	"Large urban hospital's" HER from Jan. 2010 to May 2015 linked to Social Security Death Index; 20% sample Medicare FFS claims Jan. 2009 to June 2013.	Descriptive comparisons of output from risk model and actual physician decisions; shift-to-shift variation in average testing rates associated with triage team.	Physicians over test low-risk patients and under test high-risk patients because they focus on salient and representative symptoms, ignoring more complicated predictors of risk. High risk patients who arrive at the ED during high-testing shifts have 32% lower 1-year mortality.	Stress testing is more overused than catheterization. More experienced physicians test less but more accurately target tests toward high-risk patients.
Sacks et al. (JHE,	What are the	Commercial claims	DiD using state-level	MA-PDMPs decrease hazard of a	Increases in new opioid Rx in

2021)	effects of MA-PDMPs and laws that limit initial opioid Rx length on opioid-naïve patients?	from "large, national insurer" (20% sample and 100% sample for patients w/opioid Rx) Jan. 2007–Apr. 2018.	variation in laws.	new opioid Rx by 4.7%. Laws that limit initial Rx length increase hazard of new opioid Rx by 8.7%—reductions in Rx for >7 days are more than offset by increase in Rx for <7 days.	response to laws that limit initial opioid Rx length are stronger for PCPs, providing evidence that these laws may inadvertently signal that short prescriptions are safe.
Van Parys and Brown (NBER WP 2023)	Did broadband access improve the outcome of joint replacement outcomes?	Federal Communication Commission data on broadband roll-out; Medicare Current Beneficiary Survey; TM Claims 1999–2014.	DiD exploiting staggered rollout of broadband; discrete choice model	Broadband access explains 16% of the improvement in joint replacement outcomes between 1999-2008. 10% stems from patients seeking better providers and 6% stems from improvements in care conditional on patient demand.	Improvements in outcomes due to hospital access to broadband are driven by hospitals in markets with less competition.
Zeltzer et al. (JHE 2023)	How does the adoption of a digital device to assist with telehealth visits affect health care?	EHR data from Isreali Clait Health Services (an HMO covering ~half the Israeli population) from 2018–2022.	Matched DiD and event study.	Device-assisted telemedicine increases primary care visits 12%, increases antibiotic use 15.6%, and decreases urgent care/ED/inpatient visits 11–24% compared to baseline mean.	Adults have a smaller increase in primary care use and a larger decrease in urgent care/ER/impatient visits than pediatric patients.
Zeltzer et al. (JEEA 2024)	Impact of increased access to telemedicine during COVID-19 after lockdowns lifted were in May–June 2020.	EHR data from Israeli Clait Health Services from January 2019 to June 2020.	DiD at the patient level. Treatment is a patients' physicians' propensity to use telemedicine during the initial March–May 2020 lockdown.	Having a PCP who was a high user of telemedicine increased the prob. of a primary care visit by 3.6% but reduced visit costs by 5.7% (of the pre-lockdown mean). Visits had fewer Rx and referrals. No evidence of more missed diagnoses for patients of high adopters.	Effects measured in % changes with respect to baseline are similar across patient age, gender, and SES. Reduction in Rx larger for providers who prescribed more in the pre-period.

Glossary of Table Terms

AHA – American Hospital Association
AKM– Abowd, Kramarz, and Margolis (1999)
AMA – American Medical Association
AMI/MI –Acute myocardial infarction
ATE—Average Treatment Effect
CCI—Charlson Comorbidity Index
CDC – Center for Disease Control and Prevention
CMS –Centers for Medicare and Medicaid Services
CPOE – Computerized provider order entry
DEA – Drug Enforcement Authority
D-in-D – Difference in differences
DO – Doctor of Osteopathic Medicine
ED/ER – emergency department
EMR/EHR – Electronic medical/health record
FDA— Food and Drug Administration (United States)
FE – Fixed effects
FFS—fee-for-service
GP—General Practitioner
HCUP – Health care utilization project
HIT – Health information technology
HRR – Hospital referral regions (from the Dartmouth Atlas)
IV –Instrumental variable
MA-PDMP – Must-Access Prescription Drug Monitoring Program
MD – Medical Doctor
MMC—Medicaid managed care
NCHS -- National Center for Health Statistics
NHS—National Health Service (U.K., Norway)
NPI – National Provider Identifier
OR – Odds ratio
PCP –Primary care provider
PDMP – Prescription drug monitoring program
pp – percentage point
PSI – Patient safety indicator
RCT – Randomized controlled trial

RD—Regression discontinuity
Rx—Prescription
SES – Socioeconomic status
SSRI—Selective Serotonin Reuptake Inhibitor
VHA—Veterans Health Administration (United States)

Bibliography for Table Citations

Abaluck, Jason, Leila Agha, Jr Chan David C., Daniel Singer, and Diana Zhu. 2021. "Fixing Misallocation with Guidelines: Awareness vs. Adherence." Working Paper 27467. National Bureau of Economic Research. https://www.nber.org/papers/w27467.

Agarwal, Nikhil, Alex Moehring, Pranav Rajpurkar, and Tobias Salz. 2024. "Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology." Working Paper 31422. National Bureau of Economic Research. https://doi.org/10.3386/w31422.

Agha, Leila. 2014. "The Effects of Health Information Technology on the Costs and Quality of Medical Care." *Journal of Health Economics* 34 (March):19–30. https://doi.org/10.1016/j.jhealeco.2013.12.005.

Agha, Leila, and David Molitor. 2018. "The Local Influence of Pioneer Investigators On Technology Adoption: Evidence From New Cancer Drugs." *The Review of Economics and Statistics* 100 (1): 29–44. https://doi.org/10.1162/REST_a_00670.

Agha, Leila, and Dan Zeltzer. 2022. "Drug Diffusion through Peer Networks: The Influence of Industry Payments." *American Economic Journal: Economic Policy* 14 (2): 1–33. https://doi.org/10.1257/pol.20200044.

Ahomäki, Iiro, Visa Pitkänen, Aarni Soppi, and Leena Saastamoinen. 2020. "Impact of a Physician-Targeted Letter on Opioid Prescribing." *Journal of Health Economics* 72 (July):1–22. https://doi.org/10.1016/j.jhealeco.2020.102344.

Alexander, Diane. 2020. "How Do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs." *Journal of Political Economy* 128 (11): 4046–96. https://doi.org/10.1086/710334.

Alexander, Diane, and Janet Currie. 2017. "Are Publicly Insured Children Less Likely to Be Admitted to Hospital than the Privately Insured (and Does It Matter)?" *Economics & Human Biology*, In Honor of Nobel Laureate Angus Deaton: Health Economics in Developed and Developing Countries, 25 (May):33–51. https://doi.org/10.1016/j.ehb.2016.10.005.

Alexander, Diane, and Molly Schnell. 2024. "The Impacts of Physician Payments on Patient Access, Use, and Health." *American Economic Journal: Applied Economics*. https://doi.org/10.1257/app.20210227.

Allen, Thomas, Eleonora Fichera, and Matt Sutton. 2016. "Can Payers Use Prices to Improve Quality? Evidence from English Hospitals." *Health Economics* 25 (1): 56–70. https://doi.org/10.1002/hec.3121.

Almond, Douglas, Joseph J. Doyle Jr., Amanda E. Kowalski, and Heidi Williams. 2010. "Estimating Marginal Returns to Medical Care: Evidence from At-Risk Newborns." *The Quarterly Journal of Economics* 125 (2): 591–634. https://doi.org/10.1162/qjec.2010.125.2.591.

Alpert, Abby, Sarah Dykstra, and Mireille Jacobson. 2024. "Hassle Costs versus Information: How Do Prescription Drug Monitoring Programs Reduce Opioid Prescribing?" *American Economic Journal: Economic Policy* 16 (1): 87–123. https://doi.org/10.1257/pol.20200579.

Alsan, Marcella, Owen Garrick, and Grant Graziani. 2019. "Does Diversity Matter for Health? Experimental Evidence from Oakland." *American Economic Review* 109 (12): 4071–4111. https://doi.org/10.1257/aer.20181446.

Angerer, Silvia, Christian Waibel, and Harald Stummer. 2019. "Discrimination in Health Care: A Field Experiment on the Impact of Patients' Socioeconomic Status on Access to Care." *American Journal of Health Economics* 5 (4): 407–27. https://doi.org/10.1162/ajhe_a_00124.

Arrow, Kenneth J., L. Kamran Bilir, and Alan Sorensen. 2020. "The Impact of Information Technology on the Diffusion of New Pharmaceuticals." *American Economic Journal: Applied Economics* 12 (3): 1–39. https://doi.org/10.1257/app.20170647.

Avdic, Daniel, Stephanie von Hinke, Bo Lagerqvist, Carol Propper, and Johan Vikström. 2024. "Do Responses to News Matter? Evidence from Interventional Cardiology." *Journal of Health Economics* 94 (March). https://doi.org/10.1016/j.jhealeco.2023.102846.

Bisgaier, Joanna, and Karin V. Rhodes. 2011. "Auditing Access to Specialty Care for Children with Public Insurance." *New England Journal of Medicine* 364 (24): 2324–33. https://doi.org/10.1056/NEJMsa1013285.

Bradford, W. David, and Andrew N. Kleit. 2015. “Impact of FDA Actions, DTCA, and Public Information on the Market for Pain Medication.” *Health Economics* 24 (7): 859–75. https://doi.org/10.1002/hec.3067.

Brekke, Kurt R., Tor Helge Holmås, Karin Monstad, and Odd Rune Straume. 2018. “Socio-Economic Status and Physicians’ Treatment Decisions.” *Health Economics* 27 (3): e77–89. https://doi.org/10.1002/hec.3621.

———. 2019. “Competition and Physician Behaviour: Does the Competitive Environment Affect the Propensity to Issue Sickness Certificates?” *Journal of Health Economics* 66 (July):117–35. https://doi.org/10.1016/j.jhealeco.2019.05.007.

Buchmueller, Thomas C., and Colleen Carey. 2018. “The Effect of Prescription Drug Monitoring Programs on Opioid Utilization in Medicare.” *American Economic Journal: Economic Policy* 10 (1): 77–112. https://doi.org/10.1257/pol.20160094.

Buchmueller, Thomas C., Colleen M. Carey, and Giacomo Meille. 2020. “How Well Do Doctors Know Their Patients? Evidence from a Mandatory Access Prescription Drug Monitoring Program.” *Health Economics* 29 (9): 957–74. https://doi.org/10.1002/hec.4020.

Button, Patrick, Eva Dils, Benjamin Harrell, Luca Fumarco, and David Schwegman. 2020. “Gender Identity, Race, and Ethnicity Discrimination in Access to Mental Health Care: Preliminary Evidence from a Multi-Wave Audit Field Experiment.” Working Paper 28164. National Bureau of Economic Research. https://doi.org/10.3386/w28164.

Cabral, Marika, Colleen Carey, and Sarah Miller. 2021. “The Impact of Provider Payments on Health Care Utilization of Low-Income Individuals: Evidence from Medicare and Medicaid.” Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w29471.

Cabral, Marika, and Marcus Dillender. 2024. “Gender Differences in Medical Evaluations: Evidence from Randomly Assigned Doctors.” *American Economic Review* 114 (2): 462–99. https://doi.org/10.1257/aer.20220349.

Carey, Colleen, Michael Daly, and Jing Li. 2024. “Nothing for Something: Marketing Cancer Drugs to Physicians Increases Prescribing Without Improving Mortality.” Working Paper. National Bureau of Economic Research. https://doi.org/10.3386/w32336.

Carey, Colleen, Ethan M. J. Lieber, and Sarah Miller. 2021. “Drug Firms’ Payments and Physicians’ Prescribing Behavior in Medicare Part D.” *Journal of Public Economics* 197 (May):104402. https://doi.org/10.1016/j.jpubeco.2021.104402.

Chan, David C. 2016. “Teamwork and Moral Hazard: Evidence from the Emergency Department.” *Journal of Political Economy* 124 (3): 734–70. https://doi.org/10.1086/685910.

———. 2018. “The Efficiency of Slacking off: Evidence from the Emergency Department.” *Econometrica* 86 (3): 997–1030. https://doi.org/10.3982/ECTA13565.

———. 2021. “Influence and Information in Team Decisions: Evidence from Medical Residency.” *American Economic Journal: Economic Policy* 13 (1): 106–37. https://doi.org/10.1257/pol.20180501.

Chan, Jr, David C., and Yiqun Chen. 2022. “The Productivity of Professions: Evidence from the Emergency Department.” Working Paper 30608. National Bureau of Economic Research. https://doi.org/10.3386/w30608.

Chandra, Amitabh, and Douglas O. Staiger. 2010. “Identifying Provider Prejudice in Healthcare.” Working Paper 16382. National Bureau of Economic Research. https://doi.org/10.3386/w16382.

Chen, Alice, and Darius N. Lakdawalla. 2019. “Healing the Poor: The Influence of Patient Socioeconomic Status on Physician Supply Responses.” *Journal of Health Economics* 64 (March):43–54. https://doi.org/10.1016/j.jhealeco.2019.02.001.

Chen, Yiqun. 2021. “Team-Specific Human Capital and Team Performance: Evidence from Doctors.” *American Economic Review* 111 (12): 3923–62. https://doi.org/10.1257/aer.20201238.

Chernew, Michael, Zack Cooper, Eugene Larsen Hallock, and Fiona Scott Morton. 2021. “Physician Agency, Consumerism, and the Consumption of Lower-Limb MRI Scans.” *Journal of Health Economics* 76 (March):102427. https://doi.org/10.1016/j.jhealeco.2021.102427.

Chorniy, Anna, Janet Currie, and Lyudmyla Sonchak. 2018. “Exploding Asthma and ADHD Caseloads: The Role of Medicaid Managed Care.” *Journal of Health Economics* 60 (July):1–15. https://doi.org/10.1016/j.jhealeco.2018.04.002.
Chu, Bryan, Ben Handel, Jonathan Kolstad, Jonas Knecht, Ulrike Malmendier, and Filip Matejka. 2024. “Cognitive Capacity, Fatigue and Decision Making: Evidence from the Practice of Medicine.” UC Berkeley.
Clemens, Jeffrey, and Joshua D. Gottlieb. 2014. “Do Physicians’ Financial Incentives Affect Medical Treatment and Patient Health?” *American Economic Review* 104 (4): 1320–49. https://doi.org/10.1257/aer.104.4.1320.
Clemens, Jeffrey, Pierre-Thomas Léger, Yashna Nandan, and Robert Town. 2024. “Physician Practice Preferences and Healthcare Expenditures: Evidence from Commercial Payers.” Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w33090.
Costa-Ramón, Ana María, Ana Rodríguez-González, Miquel Serra-Burriel, and Carlos Campillo-Artero. 2018. “It’s about Time: Cesarean Sections and Neonatal Health.” *Journal of Health Economics* 59 (May):46–59. https://doi.org/10.1016/j.jhealeco.2018.03.004.
Coudin, Elise, Anne Pla, and Anne-Laure Samson. 2015. “GP Responses to Price Regulation: Evidence from a French Nationwide Reform.” *Health Economics* 24 (9): 1118–30. https://doi.org/10.1002/hec.3216.
Coussens, Stephen. 2022. “Behaving Discretely: Heuristic Thinking in the Emergency Department.” Columbia University. https://www.dropbox.com/s/6sdlwnniq2unlnk/behaving_discretely_web.pdf?dl=0&unfurl=1.
Cuddy, Emily, and Janet Currie. 2020. “Treatment of Mental Illness in American Adolescents Varies Widely within and across Areas.” *Proceedings of the National Academy of Sciences* 117 (39): 24039–46. https://doi.org/10.1073/pnas.2007484117.
———. 2024. “Rules vs. Discretion: Treatment of Mental Illness in U.S. Adolescents.” *Journal of Political Economy*. https://doi.org/10.3386/w27890.
Currie, Janet M, and W Bentley MacLeod. 2020. “Understanding Doctor Decision Making: The Case of Depression Treatment.” *Econometrica: Journal of the Econometric Society* 88 (3): 847–78. https://doi.org/10.3982/ECTA16591.
Currie, Janet, and Kate Musen. 2025. “Effects of Medicaid Prior Authorization on Child Antipsychotic Prescribing: Spillovers, Information, and Hassle Costs.” ASSA 2025 Annual Meeting. San Fransisco.
Currie, Janet, and Jonathan Zhang. 2023. “Doing More with Less: Predicting Primary Care Provider Effectiveness.” *The Review of Economics and Statistics*, February, 1–45. https://doi.org/10.1162/rest_a_01290.
Dahlstrand, Amanda. 2024. “Defying Distance? The Provision of Services in the Digital Age.”
DeCicca, Philip, Maripier Isabelle, and Natalie Malak. 2024. “How Do Physicians Respond to New Medical Research?” *Health Economics* 33 (10): 2206–28. https://doi.org/10.1002/hec.4879.
Dickstein. 2017. “Physician vs. Patient Incentives in Prescription Drug Choice.” http://www.michaeljdickstein.com/posts/2015/9/24/patient-vs-physician-incentives-in-prescription-drug-choice.
Ding, Yu, and Chenyuan Liu. 2021. “Alternative Payment Models and Physician Treatment Decisions: Evidence from Lower Back Pain.” *Journal of Health Economics* 80 (December). https://doi.org/10.1016/j.jhealeco.2021.102548.
Doctor, Jason N., Andy Nguyen, Roneet Lev, Jonathan Lucas, Tara Knight, Henu Zhao, and Michael Menchine. 2018. “Opioid Prescribing Decreases after Learning of a Patient’s Fatal Overdose.” *Science* 361 (6402): 588–90. https://doi.org/10.1126/science.aat4595.
Doyle, Joseph J., Steven M. Ewer, and Todd H. Wagner. 2010. “Returns to Physician Human Capital: Evidence from Patients Randomized to Physician Teams.” *Journal of Health Economics* 29 (6): 866–82. https://doi.org/10.1016/j.jhealeco.2010.08.004.
Doyle, Jr., Joseph J. 2020. “Physician Characteristics and Patient Survival: Evidence from Physician Availability.” Working Paper 27458. National Bureau of Economic Research. https://doi.org/10.3386/w27458.

Dubois, Pierre, and Tuba Tunçel. 2021. "Identifying the Effects of Scientific Information and Recommendations on Physicians' Prescribing Behavior." *Journal of Health Economics* 78 (July):102461. https://doi.org/10.1016/j.jhealeco.2021.102461.
Eli, Shari, Trevon D. Logan, and Boriana Miloucheva. 2019. "Physician Bias and Racial Disparities in Health: Evidence from Veterans' Pensions." Working Paper 25846. National Bureau of Economic Research. https://doi.org/10.3386/w25846.
Ellyson, Alice M, Jevay Grooms, and Alberto Ortega. 2022. "Flipping the Script: The Effects of Opioid Prescription Monitoring on Specialty-Specific Provider Behavior." *Health Economics* 31 (2): 297–341. https://doi.org/10.1002/hec.4446.
Epstein, Andrew J., Sean Nicholson, and David A. Asch. 2016. "The Production of and Market for New Physicians' Skill." *American Journal of Health Economics* 2 (1): 41–65. https://doi.org/10.1162/AJHE_a_00033.
Facchini, Gabriel. 2022. "Forgetting-by-Not-Doing: The Case of Surgeons and Cesarean Sections." *Health Economics* 31 (3): 481–95. https://doi.org/10.1002/hec.4460.
Fortin, Bernard, Nicolas Jacquemet, and Bruce Shearer. 2021. "Labour Supply, Service Intensity, and Contracts: Theory and Evidence on Physicians." *Journal of Applied Econometrics* 36 (6): 686–702. https://doi.org/10.1002/jae.2840.
Frakes, Michael. 2013. "The Impact of Medical Liability Standards on Regional Variations in Physician Behavior: Evidence from the Adoption of National-Standard Rules." *American Economic Review* 103 (1): 257–76. https://doi.org/10.1257/aer.103.1.257.
Frakes, Michael D., and Jonathan Gruber. 2022. "Racial Concordance and the Quality of Medical Care: Evidence from the Military." Working Paper 30767. National Bureau of Economic Research. https://doi.org/10.3386/w30767.
Freedman, Seth, Ezra Golberstein, Tsan-Yao Huang, David J. Satin, and Laura Barrie Smith. 2021. "Docs with Their Eyes on the Clock? The Effect of Time Pressures on Primary Care Productivity." *Journal of Health Economics* 77 (May):102442. https://doi.org/10.1016/j.jhealeco.2021.102442.
Geiger, Caroline K., Mark A. Clapp, and Jessica L. Cohen. 2021. "Association of Prenatal Care Services, Maternal Morbidity, and Perinatal Mortality With the Advanced Maternal Age Cutoff of 35 Years." *JAMA Health Forum* 2 (12): e214044. https://doi.org/10.1001/jamahealthforum.2021.4044.
Goetz, Daniel. 2023. "Telemedicine Competition, Pricing, and Technology Adoption: Evidence from Talk Therapists." *International Journal of Industrial Organization* 89 (July):102956. https://doi.org/10.1016/j.ijindorg.2023.102956.
Gowrisankaran, Gautam, Keith Joiner, and Pierre Thomas Léger. 2022. "Physician Practice Style and Healthcare Costs: Evidence from Emergency Departments." *Management Science*, October. https://doi.org/10.1287/mnsc.2022.4544.
Goyal, Monika K., Nathan Kuppermann, Sean D. Cleary, Stephen J. Teach, and James M. Chamberlain. 2015. "Racial Disparities in Pain Management of Children with Appendicitis in Emergency Departments." *JAMA Pediatrics* 169 (11): 996–1002. https://doi.org/10.1001/jamapediatrics.2015.1915.
Greenwood, Brad N., Seth Carnahan, and Laura Huang. 2018. "Patient–Physician Gender Concordance and Increased Mortality among Female Heart Attack Patients." *Proceedings of the National Academy of Sciences* 115 (34): 8569–74. https://doi.org/10.1073/pnas.1800097115.
Greenwood, Brad N., Rachel R. Hardeman, Laura Huang, and Aaron Sojourner. 2020. "Physician–Patient Racial Concordance and Disparities in Birthing Mortality for Newborns." *Proceedings of the National Academy of Sciences* 117 (35): 21194–200. https://doi.org/10.1073/pnas.1913405117.
Gruber, Jonathan, Thomas P. Hoe, and George Stoye. 2021. "Saving Lives by Tying Hands: The Unexpected Effects of Constraining Health Care Providers." *The Review of Economics and Statistics*, May, 1–45. https://doi.org/10.1162/rest_a_01044.
Gupta, Atul. 2021. "Impacts of Performance Pay for Hospitals: The Readmissions Reduction Program." *American Economic Review* 111 (4): 1241–83. https://doi.org/10.1257/aer.20171825.

Hill, Andrew J., Daniel B. Jones, and Lindsey Woodworth. 2023. "Physician-Patient Race-Match Reduces Patient Mortality." *Journal of Health Economics* 92 (December). https://doi.org/10.1016/j.jhealeco.2023.102821.

Hoffman, Kelly M., Sophie Trawalter, Jordan R. Axt, and M. Norman Oliver. 2016. "Racial Bias in Pain Assessment and Treatment Recommendations, and False Beliefs about Biological Differences between Blacks and Whites." *Proceedings of the National Academy of Sciences* 113 (16): 4296–4301. https://doi.org/10.1073/pnas.1516047113.

Horwitz, Jill, Austin Nichols, Carrie H. Colla, and David M. Cutler. 2024. "Technology Regulation Reconsidered: The Effects of Certificate of Need Policies on the Quantity and Quality of Diagnostic Imaging." Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w32143.

Howard, David H., Guy David, and Jason Hockenberry. 2017. "Selective Hearing: Physician-Ownership and Physicians' Response to New Evidence." *Journal of Economics & Management Strategy* 26 (1): 152–68. https://doi.org/10.1111/jems.12178.

Howard, David H., and Jason Hockenberry. 2019. "Physician Age and the Abandonment of Episiotomy." *Health Services Research* 54 (3): 650–57. https://doi.org/10.1111/1475-6773.13132.

Howard, David H., and Ian McCarthy. 2021. "Deterrence Effects of Antifraud and Abuse Enforcement in Health Care." *Journal of Health Economics* 75 (January):102405. https://doi.org/10.1016/j.jhealeco.2020.102405.

Johnson, Erin M., and M. Marit Rehavi. 2016. "Physicians Treating Physicians: Information and Incentives in Childbirth." *American Economic Journal: Economic Policy* 8 (1): 115–41. https://doi.org/10.1257/pol.20140160.

Johnson, Erin, M. Marit Rehavi, Jr Chan David C., and Daniela Carusi. 2016. "A Doctor Will See You Now: Physician-Patient Relationships and Clinical Decisions." Working Paper 22666. National Bureau of Economic Research. https://doi.org/10.3386/w22666.

Kolstad, Jonathan T. 2013. "Information and Quality When Motivation Is Intrinsic: Evidence from Surgeon Report Cards." *American Economic Review* 103 (7): 2875–2910. https://doi.org/10.1257/aer.103.7.2875.

Kowalski, Amanda E. 2023. "Behaviour within a Clinical Trial and Implications for Mammography Guidelines." *The Review of Economic Studies* 90 (1): 432–62. https://doi.org/10.1093/restud/rdac022.

Linder, Jeffrey A., Jason N. Doctor, Mark W. Friedberg, Harry Reyes Nieva, Caroline Birks, Daniella Meeker, and Craig R. Fox. 2014. "Time of Day and the Decision to Prescribe Antibiotics." *JAMA Internal Medicine* 174 (12): 2029–31. https://doi.org/10.1001/jamainternmed.2014.5225.

Ly, Dan P. 2021. "The Influence of the Availability Heuristic on Physicians in the Emergency Department." *Annals of Emergency Medicine* 78 (5): 650–57. https://doi.org/10.1016/j.annemergmed.2021.06.012.

Ly, Dan P., Paul G. Shekelle, and Zirui Song. 2023. "Evidence for Anchoring Bias During Physician Decision-Making." *JAMA Internal Medicine* 183 (8): 818–23. https://doi.org/10.1001/jamainternmed.2023.2366.

McCullough, Jeffrey S., Michelle Casey, Ira Moscovice, and Shailendra Prasad. 2010. "The Effect Of Health Information Technology On Quality In U.S. Hospitals." *Health Affairs* 29 (4): 647–54. https://doi.org/10.1377/hlthaff.2010.0155.

McDevitt, Ryan C., and James W. Roberts. 2014. "Market Structure and Gender Disparity in Health Care: Preferences, Competition, and Quality of Care." *The RAND Journal of Economics* 45 (1): 116–39. https://doi.org/10.1111/1756-2171.12044.

McKibbin, Rebecca. 2023. "The Effect of RCTs on Drug Demand: Evidence from off-Label Cancer Drugs." *Journal of Health Economics* 90 (July):102779. https://doi.org/10.1016/j.jhealeco.2023.102779.

Miller, Amalia R., and Catherine E. Tucker. 2011. "Can Health Care Information Technology Save Babies?" *Journal of Political Economy* 119 (2): 289–324. https://doi.org/10.1086/660083.

Molitor, David. 2018. "The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration." *American Economic Journal: Economic Policy* 10 (1): 326–56. https://doi.org/10.1257/pol.20160319.

Mullainathan, Sendhil, and Ziad Obermeyer. 2022. "Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care." *The Quarterly Journal of Economics* 137 (2): 679–727. https://doi.org/10.1093/qje/qjab046.

Neprash, Hannah T., John F. Mulcahy, Dori A. Cross, Joseph E. Gaugler, Ezra Golberstein, and Ishani Ganguli. 2023. "Association of Primary Care Visit Length with Potentially Inappropriate Prescribing." *JAMA Health Forum* 4 (3): e230052. https://doi.org/10.1001/jamahealthforum.2023.0052.

Neumark, David, and Bogdan Savych. 2023. "Effects of Opioid-Related Policies on Opioid Utilization, Nature of Medical Care, and Duration of Disability." *American Journal of Health Economics* 9 (3): 331–73. https://doi.org/10.1086/722981.

Newham, Melissa, and Marica Valente. 2024. "The Cost of Influence: How Gifts to Physicians Shape Prescriptions and Drug Costs." *Journal of Health Economics* 95 (May):102887. https://doi.org/10.1016/j.jhealeco.2024.102887.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366 (6464): 447–53. https://doi.org/10.1126/science.aax2342.

Olenski, Andrew R., André Zimerman, Stephen Coussens, and Anupam B. Jena. 2020. "Behavioral Heuristics in Coronary-Artery Bypass Graft Surgery." *New England Journal of Medicine* 382 (8): 778–79. https://doi.org/10.1056/NEJMc1911289.

Olson, Mary K., and Nina Yin. 2021. "New Clinical Information and Physician Prescribing: How Do Pediatric Labeling Changes Affect Prescribing to Children?" *Health Economics* 30 (1): 144–64. https://doi.org/10.1002/hec.4182.

Persson, Emil, Kinga Barrafrem, Andreas Meunier, and Gustav Tinghög. 2019. "The effect of decision fatigue on surgeons' clinical decision making." *Health Economics* 28 (10): 1194–1203. https://doi.org/10.1002/hec.3933.

Persson, Petra, Xinyao Qiu, and Maya Rossin-Slater. 2021. "Family Spillover Effects of Marginal Diagnoses: The Case of ADHD." Working Paper 28334. National Bureau of Economic Research. https://doi.org/10.3386/w28334.

Sabin, Janice A., and Anthony G. Greenwald. 2012. "The Influence of Implicit Bias on Treatment Recommendations for 4 Common Pediatric Conditions: Pain, Urinary Tract Infection, Attention Deficit Hyperactivity Disorder, and Asthma." *American Journal of Public Health* 102 (5): 988–95. https://doi.org/10.2105/AJPH.2011.300621.

Sacarny, Adam, Michael L. Barnett, Jackson Le, Frank Tetkoski, David Yokum, and Shantanu Agrawal. 2018. "Effect of Peer Comparison Letters for High-Volume Primary Care Prescribers of Quetiapine in Older and Disabled Adults: A Randomized Clinical Trial." *JAMA Psychiatry* 75 (10): 1003–11. https://doi.org/10.1001/jamapsychiatry.2018.1867.

Sacarny, Adam, David Yokum, Amy Finkelstein, and Shantanu Agrawal. 2016. "Medicare Letters To Curb Overprescribing Of Controlled Substances Had No Detectable Effect On Providers." *Health Affairs* 35 (3): 471–79. https://doi.org/10.1377/hlthaff.2015.1025.

Sacks, Daniel W., Alex Hollingsworth, Thuy Nguyen, and Kosali Simon. 2021. "Can Policy Affect Initiation of Addictive Substance Use? Evidence from Opioid Prescribing." *Journal of Health Economics* 76 (March):102397. https://doi.org/10.1016/j.jhealeco.2020.102397.

Schnell, Molly, and Janet Currie. 2018. "Addressing the Opioid Epidemic: Is There a Role for Physician Education?" *American Journal of Health Economics* 4 (3): 383–410. https://doi.org/10.1162/ajhe_a_00113.

Shapiro, Bradley T. 2018. "Informational Shocks, Off-Label Prescribing, and the Effects of Physician Detailing." *Management Science* 64 (12): 5925–45. https://doi.org/10.1287/mnsc.2017.2899.

Shurtz, Ity, Alon Eizenberg, Adi Alkalay, and Amnon Lahad. 2022. "Physician Workload and Treatment Choice: The Case of Primary Care." *The RAND Journal of Economics* 53 (4): 763–91. https://doi.org/10.1111/1756-2171.12425.

Silver, David. 2021. "Haste or Waste? Peer Pressure and Productivity in the Emergency Department." *The Review of Economic Studies* 88 (3): 1385–1417. https://doi.org/10.1093/restud/rdaa054.

Simeonova, Emilia, Niels Skipper, and Peter Rønø Thingholm. 2024. "Physician Health Management Skills and Patient Outcomes." *Journal of Human Resources* 59 (3): 777–809. https://doi.org/10.3368/jhr.0420-10833R1.

Singh, Manasvini. 2021. “Heuristics in the Delivery Room.” *Science* 374 (6565): 324–29. https://doi.org/10.1126/science.abc9818.
Singh, Manasvini, and Atheendar Venkataramani. 2022. “Rationing by Race.” Working Paper 30380. National Bureau of Economic Research. https://doi.org/10.3386/w30380.
Tai-Seale, Ming, and Thomas McGuire. 2012. “Time Is up: Increasing Shadow Price of Time in Primary-Care Office Visits.” *Health Economics* 21 (4): 457–76. https://doi.org/10.1002/hec.1726.
Van Parys, Jessica. 2016. “Variation in Physician Practice Styles within and across Emergency Departments.” *PLOS ONE* 11 (8). https://doi.org/10.1371/journal.pone.0159882.
Van Parys, Jessica, and Zach Y. Brown. 2023. “Broadband Internet Access and Health Outcomes: Patient and Provider Responses in Medicare.” Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w31579.
Wallis, Christopher J. D., Angela Jerath, Natalie Coburn, Zachary Klaassen, Amy N. Luckenbaugh, Diana E. Magee, Amanda E. Hird, et al. 2022. “Association of Surgeon-Patient Sex Concordance with Postoperative Outcomes.” *JAMA Surgery* 157 (2): 146–56. https://doi.org/10.1001/jamasurg.2021.6339.
Wilding, Anna, Luke Munford, Bruce Guthrie, Evangelos Kontopantelis, and Matt Sutton. 2022. “Family Doctor Responses to Changes in Target Stringency under Financial Incentives.” *Journal of Health Economics* 85 (September). https://doi.org/10.1016/j.jhealeco.2022.102651.
Wu, Bingxiao, and Guy David. 2022. “Information, Relative Skill, and Technology Abandonment.” *Journal of Health Economics* 83 (May):102596. https://doi.org/10.1016/j.jhealeco.2022.102596.
Zeltzer, Dan, Liran Einav, Joseph Rashba, and Ran D Balicer. 2024. “The Impact of Increased Access to Telemedicine.” *Journal of the European Economic Association* 22 (2): 712–50. https://doi.org/10.1093/jeea/jvad035.
Zeltzer, Dan, Liran Einav, Joseph Rashba, Yehezkel Waisman, Motti Haimi, and Ran D. Balicer. 2023. “Adoption and Utilization of Device-Assisted Telemedicine.” *Journal of Health Economics* 90 (July):102780. https://doi.org/10.1016/j.jhealeco.2023.102780.