B Physician decision tree & value of a negative CT scan

The flowchart depicted in Appendix Figure B.1 below shows a typical clinical pathway for a patient who may receive a chest CT to test for PE. The most common symptom that leads to the consideration of PE as a diagnosis is chest pain; this is a nonspecific symptom that could also indicate a cardiac problem, pneumonia, or a number of other conditions. Blood oxygen tests and an EKG are likely to be performed immediately at the bedside, and if they suggest a cardiac problem, the patient will receive a more complete cardiac workup.

If cardiac conditions are ruled out, the doctor may then be considering pneumonia, pleural effusion, and pulmonary embolism as possible diagnoses. A chest x-ray and D-dimer blood test would be the typical next steps. A chest x-ray is a low cost test with low levels of radiation exposure and little medical risk; it is highly effective at diagnosing pneumonia and pleural effusion, which are more common than PE. If the x-ray is negative, then the physician may become more concerned about the risk of PE, since other more common conditions causing chest pain have been ruled out. A chest x-ray is a commonplace and recommended antecedent to a CT scan; the popular Geneva risk scoring system for evaluating whether patient's PE risk necessitates a CT scan includes chest X-ray findings among the seven risk factors used to calculate the score.

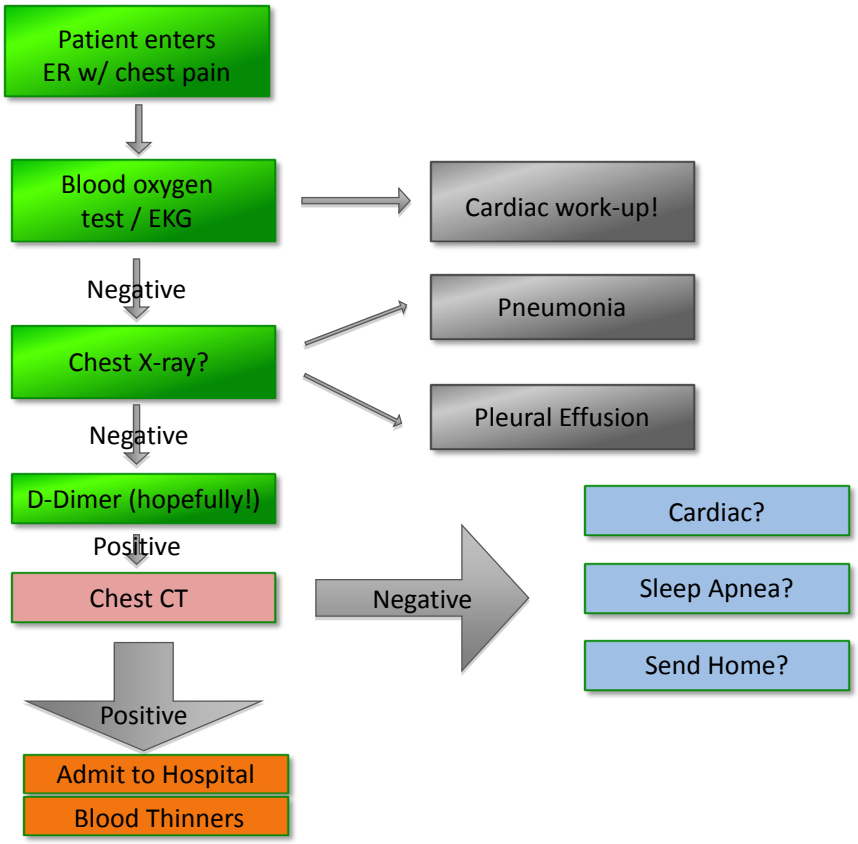
At this point, the physician may consider ordering a D-dimer, an inexpensive blood test that provides further information about a patient's risk of PE. A low-risk result on the D-dimer suggests the patient does not have a PE and the physician may forego a CT scan. A positive D-dimer result is not diagnostic of PE, but suggests an elevated probability of this condition. At this point, the physician would consider ordering a CT scan. Over our study period, the popularity of the D-dimer as an additional screening tool for PE was on the rise. Although we cannot observe the use of the D-dimer in our data, variation in D-dimer utilization is one mechanism by which physician CT ordering behavior may vary.

The physician will typically order a chest CT after ruling out these common causes of chest pain. A chest CT with contrast is useful for diagnosing pulmonary embolism, but otherwise adds little new information that may aid diagnosis of other possible acute conditions.[24] A positive test will typically lead to a hospital admission and treatment with blood thinners. Imaging is required for diagnosing PE; even high risk patients have a relatively low probability of PE and PE treatment is medically risky, so it is not a condition that would be treated presumptively without imaging.

A negative CT scan will leave the physician with a broad field of possible alternative diagnoses, including a more subtle cardiac condition, sleep apnea, infection, or a false alarm, and the CT scan result will not be helpful in distinguishing between these possibilities. Ruling out a chest CT has only a modest impact on the posterior probabilities of the other conditions that may be causing a patient's symptoms, since the *ex ante* probability of PE is relatively low—even for higher risk patients. For these reasons, the informational value of a negative test is low.

[24] In Appendix C, we provide a detailed discussion of other conditions that can be diagnosed by chest CT and how we empirically address these possibilities.

Figure B.1: Clinical Assessment of Patient with Potential Pulmonary Embolism



C Testing for Multiple Conditions

An important caveat to our above analysis is that claims data is only sufficient to identify CPT codes for "chest CT with contrast"; we cannot isolate CT scans that follow the PE testing protocol specifically. Although tests for PE are the primary indication for chest CTs in the emergency room setting, there are other possibilities. Because of this limitation, some of the tests we have labeled as "negative" since the patient is not diagnosed with pulmonary embolism may be tests performed for a different indication. There are five main alternative indications for CT scans in an emergency department setting: trauma, lung or chest cancers, aortic dissection, pleural effusion, and pneumonia. We discuss our approach to each of these alternative diagnoses in turn.

We exclude from the estimation sample patients with diagnosis codes related to trauma (such as fractures, injury, motor vehicle accidents), when these codes are associated with bills on the same day as the patient's emergency department evaluation. Chest CTs for these patients are likely aiming to assess damage from a trauma rather than a pulmonary embolism. In a detailed sample of patient records from chest CT scans performed in the emergency room of a large hospital, diagnosis codes associated with the radiology bills readily distinguished traumas from other scanning indications.

Similarly, we exclude patients with a history of aortic aneurysm, aortic dissection, or other arterial dissection, in order to eliminate patients for whom chest CTs may be intended to evaluate for aortic dissection. Aortic dissections are extremely rare, with only approximately 9000 cases per year in the United States, making it over 30 times less common than pulmonary embolism (Meszaros et al. 2000).

It is unusual for a cancer diagnosis to be made for the first time in the ED, but patients with worsening symptoms as a result of tumor growth or metastasis and occasional new diagnoses may be seen. CT scanning is routinely used to diagnose and stage cancers. In our sample of detailed ED chest CT records from the academic medical center, fewer than 1% of the scans were used to diagnose or stage cancers. In the Medicare data, we exclude those patients with chest cancer indicated on their visit to the emergency room or associated inpatient visit from our preferred estimation sample.

Chest CTs can be used to guide a procedure to treat patients with pleural effusion, which is typically first diagnosed with a chest X-ray. Because a chest CT is not commonly a diagnostic test for pleural effusion but rather an input into the treatment of the disease, we can exclude patients from the sample with diagnoses of pleural effusion. Since some patients are diagnosed with both pleural effusion and pulmonary embolism, and in these patients the chest CT was likely serving a diagnostic role, we do not exclude pleural effusion patients with a diagnosis of pulmonary embolism. These sample restrictions will tend to overstate the rate of positive testing and bias us away from finding evidence of overtesting, since we may be excluding some pleural effusion patients who are being tested for pulmonary embolism but have a negative test result.

Together, these exclusions for patients with trauma, cancer, or pleural effusion remove 32% of patients receiving chest CTs from our sample. Results presented in the paper are qualitatively similar when these patients are included.

Finally, chest CTs can be used to diagnose pneumonia. Pneumonia can also be reliably diagnosed with cheaper and lower radiation technologies (David et al. 2012); the added value of a chest CT

with contrast in an ED setting for diagnosing these alternative conditions is very modest (Venkatesh et al. 2013). Technically, the value of a chest CT scan for diagnosing a condition that could otherwise be detected with an X-ray is bounded by the costs of the X-ray, which is about $30 in our sample. Accounting for a $30 additional net benefit from diagnosing pneumonia when indicated does not substantively change our results about the welfare costs of overtesting.

D Validating our approach to coding test results in claims data

We identify positive tests on the basis of Medicare Part A hospital claims that include a diagnosis code for PE among any of the diagnoses associated with the hospital stay; we assume all other CT scans failed to detect PE. We have validated our approach to identifying positive tests by using cross-referenced patient chart and hospital billing data from two large academic medical centers. The evidence from these centers suggest that we are unlikely to understate physicians' testing thresholds due to undercounting of positive test results. In particular, we may undercount positive tests in the Medicare claims data for two reasons: if patients with PE are not admitted to the hospital; or if patients with PE are admitted but their inpatient bill does not include a diagnosis of pulmonary embolism.

At the two academic medical centers, we found that 90% of patients who test positive for PE in the emergency department were admitted within 1 day. Patients with very small PEs may occasionally be discharged after brief observation and treated with blood thinning agents as outpatients if the PE appeared small on the scan and the patient has no other complicating health conditions; this likely accounts for most of the cases where a test is coded as positive on the basis of patient chart data but no inpatient admission is recorded. Note that this suggests that we are undercounting positive tests precisely for the patient group for whom the benefits of treatment are the lowest.

Among patients with positive PE CT scans recorded in chart data who are subsequently admitted to the hospital, 87% have a diagnosis of pulmonary embolism recorded on the bill for their inpatient hospital stay. PE may not be recorded on the bill for two main reasons: the patient may have other medical conditions that are treated during the hospital stay and are reimbursed at a higher rate, such that there is no billing incentive to include PE among the inpatient diagnoses; or, the bill may simply be incorrectly coded. In total, 21% of patients diagnosed with PE in the emergency department (ED) do not have an inpatient claim with a PE diagnosis.

Of patients with a negative PE CT scan recorded in their emergency department chart, 1.5% have a diagnosis of pulmonary embolism recorded on the bill for an ensuing hospital stay. In the claims data, we would mistakenly attribute this diagnosis to the ED workup. This error could occur if the patient develops a PE later in his hospital course and receives a subsequent positive CT test, a plausible mechanism given that the immobilization frequently associated with hospital stays is a risk factor for PEs; alternatively, these PE diagnosis codes could indicate billing errors.

Taken together, these data suggest that of the 6% of CT tests that we code as positive in the Medicare data, 20% of the patients had negative findings on their initial ED PE CT. Of the 94% of tests we code as negative, 1.1% of the patients had positive ED PE CTs. The overall rate of positive tests is almost exactly equal to what it would be if no such coding mistakes were made, since

these two types of coding errors offset each other. This suggests that the limitations of this coding algorithm should not contribute to overstatements of the degree of overtesting in our Medicare sample.

E Derivation and estimation of structural model

In this section, we describe the derivation and estimation of our structural model in more detail. This section is meant to complement the discussion in Section 4, by filling in additional algebraic steps needed to complete the estimation. We begin by outlining our parametric assumptions and describe the testing equation. Second, we derive the test outcome equation which is used to estimate the distribution of τ_d, the degree of misweighting, and a scaling factor which relates the testing and test outcome equations.

Recall our assumption that doctor d's ex ante belief about the probability of a positive test for patient i is given by $q'_{id} = x_{id}\beta' + \alpha_d + \eta_{id}$ (noting, as in Section 4, that assuming the perceived α'_d equals the true α_d is without loss of generality). Although our baseline model assumes that η_{id} is independently and identically distributed across doctors and patients, in Section 6.2 we extend the model to allow for physician-specific heteroskedasticity. The motivation and results of this extension are discussed in more detail in that section. Because the heteroskedastic estimation procedure is a straightforward generalization of our baseline model, we use notation below that allows for heteroskedasticity and thus covers both the baseline model and its heteroskedastic extension.

We assume that the distribution of η_{id} follows a particular functional form, which is a mixture of a Uniform and a Bernoulli distribution; in particular, $\eta_{id} \sim U(-\eta_d, \eta_d)$ with probability $1 - p_d$ and $\eta_{id} \sim U[v - \eta_d, v + \eta_d]$ with probability p_d. The baseline model in the text assumes homoskedasticity, so that $p_d = p$ and $\eta_d = \eta$ and we note below how this affects the estimation procedure.

Assume that doctors test a patient if and only if the patient's perceived probability of a positive test exceeds a physician-specific threshold, i.e. $q'_{id} > \tau_d$. Let $I'_{id} \equiv x_{id}\beta' + \theta'_d$ where $\theta'_d = \alpha_d - \tau_d$. Also as in the text, $q_{id} = x_{id}\beta + \alpha_d + \eta_{id}$ gives the actual ex ante likelihood of a positive test. Let $I_{id} \equiv x_{id}\beta + \theta_d$ denote the unprimed version of the propensity to test (i.e. the testing propensity we would observe if physicians correctly weighted observable comorbidities to maximize test yields).

$$\begin{aligned} Pr(Test_{id} = 1) &= Pr(q'_{id} > \tau_d) \\ &= Pr(I'_{id} + \eta_{id} > 0) \\ &= 1 - Pr(\eta_{id} < -I'_{id}) \end{aligned} \tag{20}$$

Assume the distribution of η_{id} is such that $I'_{id} + v < \eta_d$ for all I'_{id} and η_d so there is no testing propensity I'_{id} at which patients are always tested regardless of the value of η_{id}. Assume further that patients are never tested if the v shock is not realized. For example, the v shock could represent symptoms that would lead the physician to suspect PE, such as chest pain and shortness of breath. Then, given our distributional assumptions: $Pr(\eta_{id} < -I'_{id}) = 1 - p_d + p_d \cdot \min\left\{1, \frac{\eta_d - (I'_{id} + v)}{2\eta_d}\right\}$. Thus:

$$
\begin{aligned}
Pr(Test_{id}=1) &= p\left[1-\min\left\{1,\frac{1}{2}-\frac{I'_{id}+v}{2\eta_d}\right\}\right] \\
&= \max\left\{0,\frac{p_d}{2}+\frac{p_d(I'_{id}+v)}{2\eta_d}\right\} \qquad (21)
\end{aligned}
$$

We estimate this equation by non-linear least squares. In the heteroskedastic model, we recover: β' (up to a scaling normalization), $\hat{\eta}_d = C\frac{p_d}{2\eta_d}$ (where the value of the constant C depends on the normalization of β), and $\hat{\theta'}_d = \frac{p_d}{2}+\frac{p_d\theta'_d+v}{2\eta_d}$. Intuitively, heteroskedasticity in η_d is identified by the fact that observables are less predictive of testing behavior for doctors with more private information. In the homoskedastic model where $p_d = p$ and $\eta_d = \eta$, this simplifies so that we are estimating $\hat{\beta}' = \frac{p\beta'}{2\eta}$ and $\hat{\theta'}_d = \frac{p}{2}+\frac{p(\theta'_d+v)}{2\eta}$.

In either the homoskedastic or heteroskedastic case, we can use the predicted values from estimation of equation 21 to construct an estimate of $\tilde{I'}_{id} = \frac{p_d}{2}+\frac{p_d(I'_{id}+v)}{2\eta_d}$. Estimating the heteroskedastic model requires an additional sample restriction at this stage. In theory, η_d is identified for all doctors. In practice, for a very small number of doctors, the estimated η_d would diverge to ∞ because patients with larger $x_{id}\beta'$ are less likely to be tested, due to random variation in a limited per-doctor sample. These doctors are excluded from the final sample for estimation when we turn to the heteroskedastic model.

Returning to the testing outcomes equation, our distributional assumptions imply that: $E(\eta_{id}|\eta_{id} > -I'_{id}) = \frac{\eta_d-(I'_{id}-v)}{2}$. Thus:

$$
\begin{aligned}
E(q_{id}|Test_{id}=1) &= \tau_d + I_{id} + E(\eta_{id}|\eta_{id} > -I'_{id}) \\
&= \tau_d + I_{id} + \frac{\eta_d-(I'_{id}-v)}{2} \\
&= \tau_d + (I_{id}-I'_{id}) + \frac{\eta_d+I'_{id}+v}{2} \\
&= \tau_d + \frac{\eta_d+I'_{id}+v}{2} + x_{id}(\beta-\beta') \\
&= \tau_d + \frac{\eta_d+I'_{id}+v}{2} + x_{id}(\beta-\beta') \qquad (22)
\end{aligned}
$$

From our definition of $\tilde{I'}_{id}$ above, it follows that $\frac{\eta_d+I'_{id}+v}{2} = \frac{\eta_d\tilde{I'}_{id}}{p_d}$ and so:

$$
\begin{aligned}
E(Z_{id}|Test_{id}=1) &= E(q_{id}|T_{id}=1) \\
&= \tau_d + \frac{\eta_d\tilde{I'}_{id}}{p_d} + x_{id}(\beta-\beta') \qquad (23)
\end{aligned}
$$

where $\tilde{I'}_{id}$ is the propensity estimated from the testing equation, and Z_{id} is the realized testing outcome (1 for a positive test, 0 for a negative test).

We can estimate this model by non-linear least squares but we need an additional exclusion restriction so that the coefficient on $\tilde{I'}_{id}$ is identified by more than just functional form. As discussed in Section 4.3, this restriction is that we effectively know τ_d for high volume doctors who test

marginal patients—i.e. patients who are very unlikely to be tested based on observables but are nonetheless tested—because we observe test outcomes among those patients. In practice, we also need to be careful about the misweighting term. If we average observed test outcomes Z_{id} among tested marginal patients (i.e. patients with $\tilde{I'}_{id} = 0$) for doctors who have such patients, then for each of those doctors we obtain an estimate of:

$$QQ_d \quad = \quad \tau_d + E_{m,d}(x_{id}|Test_{id} = 1)(\beta - \beta') \tag{24}$$

where $E_{m,d}(x_{id}|Test_{id} = 1)$ gives the mean of x_{id} among only tested marginal patients for a given doctor. For doctors with marginal patients, we have:

$$E(Z_{id}|Test_{id} = 1) - QQ_d = \frac{\eta_d \tilde{I'}_{id}}{p_d} + (x_{id} - E_{m,d}(x_{id}))(\beta - \beta') \tag{25}$$

Because we observe only a small number of marginal patients for each doctor, we can construct: $\widehat{QQ}_d = QQ_d + e_d$, a noisy estimate of QQ_d. Thus, let $Y_{id} = Z_{id}$ for doctors with no marginal tested patients and $Y_{id} = Z_{id} - \widehat{QQ}_d$ for doctors with marginal tested patients. Further, let $X_{id} = (x_{id} - E_{m,d}(x_{id}))$ for doctors with marginal tested patients and $X_{id} = x_{id}$ for doctors with no marginal tested patients. Finally, let M_d denote an indicator for whether a doctor has marginal tested patients. This gives the estimating equation:

$$Y_{id} \quad = \quad (1 - M_d)\tau_d + \frac{\eta_d \tilde{I'}_{id}}{p_d} + X_{id}(\beta - \beta') + \epsilon_{id} \tag{26}$$

where $\epsilon_{id} = M_d e_d + u_{id}$ includes both the noise in the estimation of QQ_d and the prediction error in $Z_{id} = E(q_{id}|Test_{id} = 1) + u_{id}$. This model can be estimated by least squares.

In the homoskedastic case, $\frac{\eta_d}{p_d}$ is a constant which we recover from least squares estimation of equation 26. In the heteroskedastic model, we estimated $\hat{\eta}_d = C\frac{p_d}{2\eta_d}$ in the testing equation, so the 2nd term in equation 26 is replaced by $\frac{\tilde{I'}_{id}}{\hat{\eta}_d}$ and the recovered coefficient tells us $\frac{C}{2}$, which is sufficient given $\hat{\eta}_d$ to recover $\frac{p_d}{\eta_d}$.

Following this procedure, we estimate the model and analyze the results described in Section 5. This model is also the basis of the welfare exercises reported in Section 7.

F "Empirical Bayes" Estimates of τ_d

In this section, we describe how we compute the distribution of the underlying τ_d from the observed distribution of $\hat{\tau}_d$ which includes both the underlying true variation and sampling error. We call this an "empirical Bayes" estimate because of the intuition that we are recovering the true underlying distribution of τ_d from noisy estimates, but our specific model does not recover a posterior mean estimate of the parameter for each doctor. Results of this procedure are reported in Table 5. (Note that the welfare results reported in Section 7 require more restrictive assumptions of the empirical Bayes procedure and do recover a posterior estimate of τ_d for each doctor. These additional restrictions are described below and in Section 7.2.)

In order to form our estimate of the true distribution of τ_d, we will proceed as follows:

1. Estimate the mean and variance of this distribution for doctors with no marginal tested patients.

2. Estimate the mean and variance of this distribution for doctors who do have marginal tested patients.

3. Apply the law of total variance to compute the mean and variance of the mixture distribution which combines the distributions for doctors with and without marginal tested patients.

4. Make a parametric assumption so that the mean and variance uniquely pin down the posterior distribution. (Required only for welfare simulations reported in Section 7.2.)

We start with our estimating equation from Appendix E, equation 26, reproduced below.

$$Y_{id} = (1 - M_d)\tau_d + \frac{\eta_d \tilde{I'}_{id}}{p_d} + X_{id}(\beta - \beta') + \epsilon_{id} \tag{27}$$

We can rewrite this equation in matrix form as:

$$Y = D\tau_{nm} + X\beta + \epsilon \tag{28}$$

where D includes the doctor fixed effects for all doctors who lack marginal tested patients (as indicated by the nm subscript) and $X\beta$ includes the constant terms, the $\tilde{I'}_{id}$ terms and the misweighting terms.

Our goal econometrically will be to relate the observed across doctor variance of τ_{nm} (which includes estimation error) with the underlying true variance of τ_{nm}.

Let $M_x = I_n - X(X'X)^{-1}X'$ where I_n is the identity matrix. Partialing out gives:

$$M_x Y = M_x D\tau_{nm} + M_x\epsilon \tag{29}$$

Let $S = M_x D$. Then our estimator of τ is given by:

$$\hat{\tau}_{nm} = \tau_{nm} + (S'S)^{-1}S'M_x\epsilon \tag{30}$$

For a vector x, define $var(x) = E(xx') - E(x)E(x')$. Define $var_d(x) = E(x'x) - E_d(x)^2$, i.e. the scalar generated by taking the variance across the observations in the vector. Taking the "outer product" variance of both sides of equation 30 gives:

$$\begin{aligned} var(\hat{\tau}_{nm}) &= var(\tau_{nm}) + (S'S)^{-1}S'M_x var(\epsilon) M_x S(S'S)^{-1} \\ &= var(\tau_{nm}) + (S'S)^{-1}S'var(\epsilon)S(S'S)^{-1} \end{aligned} \tag{31}$$

where the second line uses the fact that $M_x M_x = M_x$. Let $S^{(i)'}$ denote the ith row of S. Assuming $var(\epsilon)$ is a diagonal matrix, $S_0 = \frac{1}{N}\sum_{i=1}^{N} e_i^2 S^{(i)}S^{(i)'} \to_p \frac{1}{N}\sum_{i=1}^{N}\epsilon_i^2 S^{(i)}S^{(i)'} = \frac{1}{N}S'var(\epsilon)S$. This is

asymptotically equivalent to:

$$var(\tau_{nm}) = var(\hat{\tau}_{nm}) - (S'S)^{-1}\left(\sum_{i=1}^{N} e_i^2 S^{(i)} S^{(i)'}\right)(S'S)^{-1} \tag{32}$$

where e_i are the residuals from equation 28. Finally, using the fact that $var_d(\tau_{nm}) = \frac{1}{N_{doc}} tr(var(\tau_{nm}))$ where N_{doc} is the number of doctors with no marginal tested patients (i.e. the docs for whom we are currently estimating τ_d), we have:

$$var_d(\tau_{nm}) = var_d(\hat{\tau}_{nm}) - \frac{1}{N_{doc}} tr\left((S'S)^{-1}\left(\sum_{i=1}^{N} e_i^2 S^{(i)} S^{(i)'}\right)(S'S)^{-1}\right) \tag{33}$$

This equation allows us to recover $var_d(\tau)$, the variance of τ_d for doctors who lack marginal tested patients. In order to recover τ_d for doctors who do have marginal tested patients, we use the fact from equation 23 that:

$$E(Z_{id}|Test_{id} = 1) - x_{id}(\beta - \beta') = \tau_d \tag{34}$$

if we restrict to marginal tested patients of those doctors (meaning that $\tilde{I}'_{id} = 0$). This equation can be written as a special case of equation 28, with $Y_{id} = Z_{id} - x_{id}(\beta - \beta')$. Note that D now denotes the matrix of doctor fixed effects for doctors *with* marginal tested patients, N_{marg} denotes the number of doctors with marginal tested patients, and $X = 0$. This simplification means that $S = D$ and we have:

$$var_d(\tau_{marg}) = var_d(\hat{\tau}_{marg}) - \frac{1}{N_{marg}} tr\left((D'D)^{-1}\left(\sum_{i=1}^{N} e_i^2 D^{(i)} D^{(i)'}\right)(D'D)^{-1}\right) \tag{35}$$

where in this case the residuals are computed from estimation of equation 34 by OLS on the sample of physicians with marginal tested patients and only those marginal tested patients included in the estimation.

To combine these distributions into a single distribution of τ_d, we note that τ_d is a random variable whose mean and variance are $\mu_m = E(\tau_{marg})$ and $\sigma_m^2 = Var_d(\tau_{marg})$ with probability P_m (the fraction of doctors who have some marginal tested patients) and $\mu_{nm} = E(\tau_{nm})$ and $\sigma_{nm}^2 = Var_d(\tau_{nm})$ respectively with probability $1 - P_m$. This implies:

$$\begin{aligned} E(\tau) &= P_m\mu_m + (1 - P_m)\mu_{nm} \\ var_d(\tau) &= P_m\sigma_m^2 + (1 - P_m)\sigma_{nm}^2 + P_m\mu_m^2 + (1 - P_m)\mu_{nm}^2 - (P_m\mu_m + (1 - P_m)\mu_{nm})^2 \end{aligned} \tag{36}$$

where the second equation follows from the law of total variance.

For simulations and welfare analyses, we further assume that $\tau_d + M$ is log-normally distributed with mean $E(\tau)$, variance $var_d(\tau)$ and minimum possible value $M = fp$. fp is the value we would estimate for patients in equation 26 if there were no PE incidence so that the only positive tests were false positives (implying $E(Z_{id}|Test_{id} = 1) = fp$, the rate of false positives). In order to recover an estimate of τ_d for each doctor, we redraw values of τ from the simulated distribution, order them

from least to greatest, and assign each doctor a τ from the simulated distribution which matches that doctor's rank among estimated τ_d.

G Simulations of testing behavior and test yields

This section describes how we apply our structural model to simulate the relationships plotted in Figure 3 and discussed in sections 5.1 and 5.3. The first exercise illustrates the hypothetical relationship between average physician testing propensities and positive test rates, if all doctors were to have the same testing threshold. We simulate testing decisions and test outcomes under a counterfactual where τ_d is held constant across doctors, at the estimated average value $E(\tau_d) = 0.056$.

To calculate the new values of the testing propensities under this counterfactual where $\tau_d = E(\tau_d)$ for all doctors, we start by considering the estimated testing propensity: $\tilde{I}'_{id} = \frac{p}{2} + \frac{p(x_{id}\beta' + \theta'_d + v)}{2\eta}$. To simulate the testing propensity under the counterfactual where testing thresholds are held constant at their mean, $\tilde{I}'^{\tau_d = E(\tau_d)}_{id}$, we need to add our estimate of $(\widehat{\tau}_d - E(\tau_d))\frac{p}{2\eta}$ back to our original estimate of $\tilde{I}'_{id}$.

Because the estimated $\widehat{\tau}_d$ are noisy and overstate the true variance in the distribution, we calculate a posterior, shrunk estimate of each τ_d before proceeding with this counterfactual exercise. At this stage, we need to make a distributional assumption about physician testing thresholds τ_d. We assume they follow a log-normal distribution with mean and variance determined by the empirical Bayes estimates described above, and the same relative rank as in the raw estimated distribution (i.e. the doctor with the 20th largest estimated $\widehat{\tau}_d$ will also have the 20th largest posterior τ_d).

Plugging in our new, simulated estimates of $\tilde{I}'^{\tau_d = E(\tau_d)}_{id}$ and setting $\tau_d = E(\tau_d)$, we calculate $E(Z_{id}|Test_{id} = 1)$ for each patient following equation 13 and use these estimates to simulate average test yields. Results of this simulation exercise are reported in Section 5.1 and pictured in Figure 3.

The second simulation exercise considers the role of misweighting in determining the relationship between testing propensities and test yield. We simulate the counterfactual relationship between physicians' average testing propensities and test yields that would be observed if there were no heterogeneity in testing thresholds *and* no misweighting of observable risk factors. Eliminating misweighting should increase the test yield for all values of the testing propensity by improving the targeting of PE CT tests to the highest risk patients.

First we simulate how testing propensities $\tilde{I}'^{\tau_d = E(\tau_d)}_{id}$ would change if there were also no misweighting of patient risk factors. In particular, we add a correction factor $x\frac{\beta - \beta'}{2(\eta/p)}$ to $\tilde{I}'^{\tau_d = E(\tau_d)}_{id}$ to calculate new simulated testing propensities $\tilde{I}^{sim}_{id}$ under the counterfactual with no misweighting. Based on these new values of $\tilde{I}^{sim}_{id}$, we calculate the expected test yield according to the formula $E(Z^{sim}_{id}|Test_{id} = 1) = E(\tau_d) + \frac{\eta}{p}\tilde{I}^{sim}_{id}$ (from equation 13). Results of this simulation exercise are reported in Section 5.3 and pictured in Figure 3.

H Computing the welfare costs of overtesting and misweighting

In order to calculate the welfare costs of overtesting and misweighting, we must first understand how false positive and false negative test results will affect the costs and benefits of testing, and the

calibrated optimal physician testing threshold. We begin by calculating the net utility of treatment, given that there are both false positive and false negative test results. Let PE_{id} denote the event that patient i truly has a PE. As before, Z_{id} is an indicator which is 1 if a test is positive. MB denotes the medical benefits of treatment if the patient has a PE, MC denotes the medical costs of treatment and CT denotes the financial cost of treatment. Then the net utility of a positive test is given by:

$$NU_{id} = Pr(PE_{id}|Z_{id} = 1)MB - MC - CT \tag{37}$$

The medical benefits of treatment accrue only if the positive test result is a "true positive," i.e. the patient actually has a PE. If there are more false positives, the medical benefits of any observed positive test will be smaller. In contrast, the medical risks and financial costs of treatment are incurred for any treated patient regardless of whether he actually has a PE.

Let s denote the sensitivity of the test (one minus the probability of a false negative) and fp denote the probability of a false positive. Applying Bayes' Rule and the law of total probability, we can rewrite net utility as:

$$NU_{id} = \frac{s(q_{id} - fp)}{q_{id}(s - fp)}MB - MC - CT \tag{38}$$

Given the net utility associated with treating a patient with a positive test, the net benefits of testing also depend on the probability of a positive test, q_{id} and the costs of testing c. We can therefore write the net benefits of testing as:

$$\begin{aligned} B_{id} &= q_{id}NU_{id} - c \\ &= \frac{s(q_{id} - fp)}{(s - fp)}MB - q_{id}MC - q_{id}CT - c \end{aligned} \tag{39}$$

Let $\hat{NU} = \frac{s}{s-fp}MB - MC - CT$ and $\hat{c} = c + \frac{s \cdot fp}{s-fp}MB$. Then we can rewrite the net benefits of testing as:

$$B_{id} = q_{id}\hat{NU} - \hat{c} \tag{40}$$

The optimal testing threshold τ^* will be the threshold at which the expected net benefits of testing are zero, or $\tau^*\hat{NU} = \hat{c}$.

Once we have recovered the optimal testing threshold, we can apply the structural model described in Section 4 and Appendix E, to compute the welfare cost of overtesting as follows. Let $\hat{t}_{id}(\tau_d, \Delta\beta)$ denote the probability that consumer i is tested by doctor d as a function of τ_d and the vector of weighting errors physicians make in assessing PE risk. The vector of misweighting errors is labeled as $\Delta\beta = \beta - \beta'$. Let $\hat{Z}_{id}(\tau_d, \Delta\beta)$ denote the probability of a positive test conditional on testing.

To compute testing behavior under the counterfactual where all doctors utilize the optimal testing threshold τ^*, we estimate $\hat{t}_{id}(\tau^*, \Delta\beta)$ using the fact that $I(\tau^*, \Delta\beta) = I(\tau_d, \Delta\beta) + (\tau_d - \tau^*)$ which implies $\tilde{I}'(\tau^*, \Delta\beta) = \tilde{I}'(\tau_d, \Delta\beta) + \frac{p(\tau_d - \tau^*)}{2\eta}$. Having adjusted the testing propensities, we can now calculate the expected probability of a positive test $\hat{Z}_{id}(\tau^*, \Delta\beta) = \frac{\eta\tilde{I}'_{id}(\tau^*, \Delta\beta)}{p} + x_{id}(\beta - \beta')$.

Welfare simulations to evaluate the costs of misweighting parallel the derivation above. In

particular, to compute the propensity to test with no misweighting, $\hat{t}_{id}(\tau_d, 0)$, we use the fact that $I(\tau_d, 0) = I(\tau_d, \Delta\beta) + x_{id}\Delta\beta$ which implies $\tilde{I}'(\tau_d, 0) = \tilde{I}'(\tau_d, \Delta\beta) + \frac{px_{id}\Delta\beta}{2\eta}$. Given this adjustment to the testing propensities, we can calculate expected test outcomes according to the following formula: $\hat{Z}_{id}(\tau_d, 0) = \tau_d + \frac{\eta \tilde{I}'_{id}(\tau_d, 0)}{p}$.

To complete the welfare calculations, we must apply assumptions about the expected medical benefits, medical costs and financial costs associated with treatment of positive tests. Following the notation above, we have:

$$\begin{aligned}
MB(\tau_d, \Delta\beta) &= \sum_i Pr(Test_i = 1) \cdot Pr(PE_{id}|Test_i = 1) MB_{id} && (41)\\
&= \sum_i \hat{t}_{id}(\tau_d, \Delta\beta) \frac{s(\hat{Z}_{id}(\tau_d, \Delta\beta) - fp)}{(s - fp)} MB_{id} \\
MC(\tau_d, \Delta\beta) &= \sum_i Pr(Test_i = 1) Pr(Z_{id} = 1|Test_i = 1) MC_{id} && (42)\\
&= \sum_i \hat{t}_{id}(\tau_d, \Delta\beta) \hat{Z}_{id}(\tau_d, \Delta\beta) MC_{id} \\
FC(\tau_d, \Delta\beta) &= \sum_i Pr(Test_i = 1)(c + P(Z_{id} = 1|Test_i = 1) CT_{id}) && (43)\\
&= \sum_i \hat{t}_{id}(\tau_d, \Delta\beta)(c + \hat{Z}_{id}(\tau_d, \Delta\beta) CT_{id}) \\
NB(\tau_d, \Delta\beta) &= MB(\tau_d, \Delta\beta) - MC(\tau_d, \Delta\beta) - FC(\tau_d, \Delta\beta) && (44)
\end{aligned}$$

where MB denote the medical benefits of testing (derived in Section 7.1), MC denotes the medical costs of testing, FC denotes the financial costs of testing and NB denotes the net benefits of testing as a function of these objects. The test sensitivity is given by s, and fp is the false positive rate. We define the welfare cost of overtesting as $NB(\tau^*, \Delta\beta) - NB(\hat{\tau}_d, \Delta\beta)$ and the welfare cost from misweighting as $NB(\hat{\tau}_d, 0) - NB(\hat{\tau}_d, \Delta\beta)$ where $\hat{\tau}_d$ is drawn from the estimated underlying distribution of τ_d which we recover using the methods outlined in Appendix F above.