# The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care
# ONLINE APPENDIX

Jason Abaluck and Leila Agha
with Chris Kabrhel, Ali Raja, and Arjun Venkatesh

July 16, 2016

## 1 Physician decision tree & value of a negative CT scan

The flowchart depicted in Appendix Figure 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.

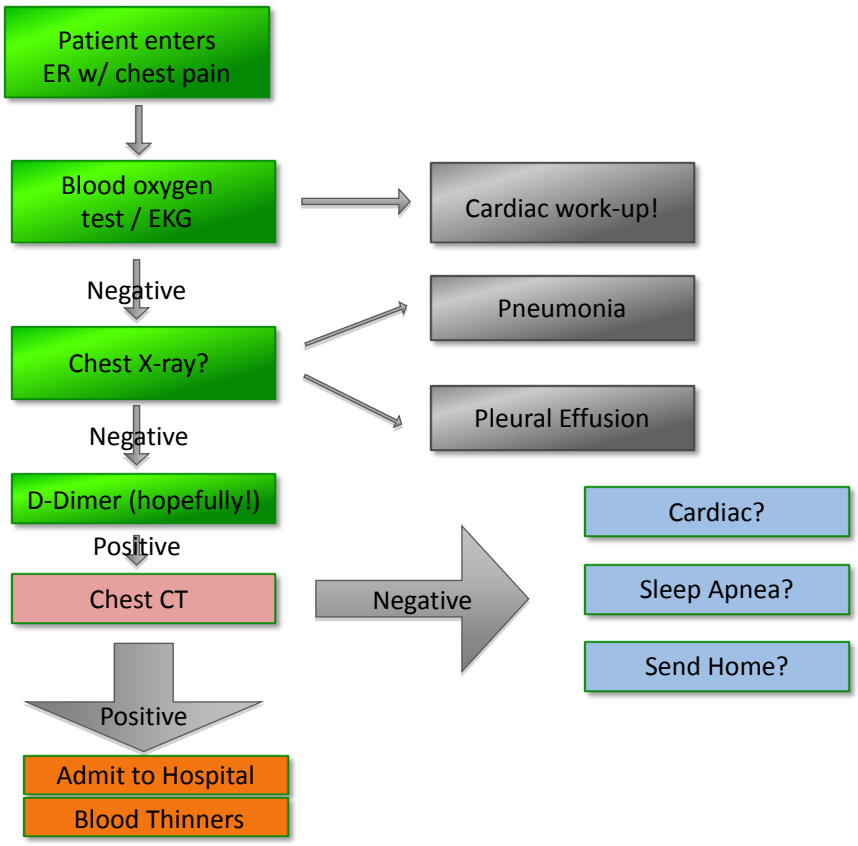
The physician will typically order a chest CT after ruling out these common causes of chest pain. A chest CT with contrast is useful for diagnosing pulmonary embolism, but otherwise adds little new information that may aid diagnosis of other possible acute conditions.[1] A positive test will typically lead to a hospital admission and treatment with blood thinners. Imaging is required

[1] In Appendix 2, we provide a detailed discussion of other conditions that can be diagnosed by chest CT and how we empirically address these possibilities.

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.

Figure 1: Clinical Assessment of Patient with Potential Pulmonary Embolism



## 2 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 percent 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 over testing, 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 percent of patients receiving chest CTs from our sample.

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 over testing.

## 3 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 percent 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 percent 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 percent 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 percent 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 percent of CT tests that we code as positive in the Medicare data, 20 percent of the patients had negative findings on their initial ED PE CT. Of the 94 percent of tests we code as negative, 1.1 percent 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 over testing in our Medicare sample.

# 4 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 3, 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 $\tau_d$, 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 3, that assuming the perceived $\alpha'_d$ equals the true $\alpha_d$ is without loss of generality). Although our baseline model assumes that $\eta_{id}$ is independently and identically distributed across doctors and patients, in Appendix 7.2 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 $\eta_{id}$ 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).

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

Assume the distribution of $\eta_{id}$ is such that $I'_{id} + v < \eta_d$ for all $I'_{id}$ and $\eta_d$ so there is no testing propensity $I'_{id}$ at which patients are always tested regardless of the value of $\eta_{id}$. 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:

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

We estimate this equation by non-linear least squares. In the heteroskedastic model, we recover: $\beta'$ (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 $\beta$), and $\hat{\theta'}_d = \frac{p_d}{2}+\frac{p_d\theta'_d+v}{2\eta_d}$. Intuitively, heteroskedasticity in $\eta_d$ 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 2 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, $\eta_d$ is identified for all doctors. In practice, for a very small number of doctors, the estimated $\eta_d$ would diverge to $\infty$ 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:

$$\begin{aligned} E(q_{id}|Test_{id}=1) &= \tau_d + I_{id} + E(\eta_{id}|\eta_{id} > -I'_{id}) \\ &= \tau_d + I_{id} + \frac{\eta_d-(I'_{id}-v)}{2} \\ &= \tau_d + (I_{id}-I'_{id}) + \frac{\eta_d+I'_{id}+v}{2} \\ &= \tau_d + \frac{\eta_d+I'_{id}+v}{2} + x_{id}(\beta-\beta') \\ &= \tau_d + \frac{\eta_d+I'_{id}+v}{2} + x_{id}(\beta-\beta') \end{aligned} \tag{3}$$

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:

$$\begin{aligned} E(Z_{id}|Test_{id}=1) &= E(q_{id}|T_{id}=1) \\ &= \tau_d + \frac{\eta_d\tilde{I'}_{id}}{p_d} + x_{id}(\beta-\beta') \end{aligned} \tag{4}$$

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 3.3, this restriction is that we effectively know $\tau_d$ 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:

$$QQ_d = \tau_d + E_{m,d}(x_{id}|Test_{id} = 1)(\beta - \beta') \tag{5}$$

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:

$$E(Z_{id}|Test_{id} = 1) - QQ_d = \frac{\eta_d \tilde{I'}_{id}}{p_d} + (x_{id} - E_{m,d}(x_{id}))(\beta - \beta') \tag{6}$$

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:

$$Y_{id} = (1 - M_d)\tau_d + \frac{\eta_d \tilde{I'}_{id}}{p_d} + X_{id}(\beta - \beta') + \epsilon_{id} \tag{7}$$

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 7. 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 7 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 4. This model is also the basis of the welfare exercises reported in Section 5.

# 5 "Empirical Bayes" Estimates of $\tau_d$

In this section, we describe how we compute the distribution of the underlying $\tau_d$ 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 $\tau_d$ 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 Appendix Table 5. (Note that the welfare results reported in Section 5 require more restrictive assumptions of the empirical Bayes procedure and do recover a posterior estimate of $\tau_d$ for each doctor. These additional restrictions are described below and in Section 5.2.)

In order to form our estimate of the true distribution of $\tau_d$, 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 Sections 4.3 and 5.2.)

We start with our estimating equation from Appendix 4, equation 7, reproduced below.

$$Y_{id} = (1 - M_d)\tau_d + \frac{\eta_d \tilde{I'}_{id}}{p_d} + X_{id}(\beta - \beta') + \epsilon_{id} \tag{8}$$

We can rewrite this equation in matrix form as:

$$Y = D\tau_{nm} + X\beta + \epsilon \tag{9}$$

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 $\tau_{nm}$ (which includes estimation error) with the underlying true variance of $\tau_{nm}$.

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

$$M_x Y = M_x D\tau_{nm} + M_x\epsilon \tag{10}$$

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

$$\hat{\tau}_{nm} = \tau_{nm} + (S'S)^{-1}S'M_x\epsilon \tag{11}$$

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 11 gives:

$$\begin{aligned} var(\hat{\tau}_{nm}) &= var(\tau_{nm}) + (S'S)^{-1}S'M_x var(\epsilon) M_x S(S'S)^{-1} \\ &= var(\tau_{nm}) + (S'S)^{-1}S'var(\epsilon)S(S'S)^{-1} \end{aligned} \tag{12}$$

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:

$$var(\tau_{nm}) = var(\hat{\tau}_{nm}) - (S'S)^{-1}\left(\sum_{i=1}^{N} e_i^2 S^{(i)} S^{(i)'}\right)(S'S)^{-1} \tag{13}$$

where $e_i$ are the residuals from equation 9. 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 $\tau_d$), we have:

$$var_d(\tau_{nm}) = var_d(\hat{\tau}_{nm}) - \frac{1}{N_{doc}} tr\left((S'S)^{-1}\left(\sum_{i=1}^{N} e_i^2 S^{(i)} S^{(i)'}\right)(S'S)^{-1}\right) \tag{14}$$

This equation allows us to recover $var_d(\tau)$, the variance of $\tau_d$ for doctors who lack marginal tested patients. In order to recover $\tau_d$ for doctors who do have marginal tested patients, we use the fact from equation 4 that:

$$E(Z_{id}|Test_{id} = 1) - x_{id}(\beta - \beta') = \tau_d \tag{15}$$

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 9, 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:

$$var_d(\tau_{marg}) = var_d(\hat{\tau}_{marg}) - \frac{1}{N_{marg}} tr\left((D'D)^{-1}\left(\sum_{i=1}^{N} e_i^2 D^{(i)} D^{(i)'}\right)(D'D)^{-1}\right) \tag{16}$$

where in this case the residuals are computed from estimation of equation 15 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 $\tau_d$, we note that $\tau_d$ 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:

$$\begin{aligned} E(\tau) &= P_m\mu_m + (1 - P_m)\mu_{nm} \\ var_d(\tau) &= P_m\sigma_m^2 + (1 - P_m)\sigma_{nm}^2 + P_m\mu_m^2 + (1 - P_m)\mu_{nm}^2 - (P_m\mu_m + (1 - P_m)\mu_{nm})^2 \end{aligned} \tag{17}$$

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 7 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 $\tau_d$ for each doctor, we redraw values of $\tau$ from the simulated distribution, order them

from least to greatest, and assign each doctor a $\tau$ from the simulated distribution which matches that doctor's rank among estimated $\tau_d$.

# 6 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 section 4.3. 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 $\tau_d$ 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 $\tau_d$ before proceeding with this counterfactual exercise. At this stage, we need to make a distributional assumption about physician testing thresholds $\tau_d$. 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 $\tau_d$).

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 10 from section 3.4 and use these estimates to simulate average test yields. Results of this simulation exercise are reported in Section 4.3 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.

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 section 3.4 equation 12). Results of this simulation exercise are reported in Section 4.3 and pictured in Figure 3.

# 7 Robustness checks

## 7.1 Stability of results to inclusion of alternate patient controls

In the spirit of Altonji et al. (2008), we explore the sensitivity of our results to the set of included variables to assess potential bias from unobservable risk factors. The rationale for this exercise is that omitting the variables $x_{id}^{omit}$ from the baseline specification could generate heteroskedasticity, if the resulting error term $\eta'_{id} = \eta_{id} + x_{id}^{omit}\beta$ is not independently and identically distributed across doctors and patients. If this heteroskedasticity substantially changes our estimates of the distribution of $\tau_d$ or the degree of misweighting for the remaining variables, this might suggest that including additional unobserved variables would change our estimates further.

Recall that we rely on comorbidities to identify the patients the doctor is just indifferent between testing and not testing, and then calculate test outcomes among that group to identify thresholds for physicians with marginal patients. In addition to testing robustness to heteroskedasticity in the error term, varying the set of included variables will also change the set of patients identified as marginal (i.e. just barely worth testing given their physician's threshold). As we remove comorbidities from the analysis, we are less able to isolate the marginal patients and may include more inframarginal patients in the group used to identify doctor's testing thresholds. To show exactly how varying the definition of marginal patients impacts the analysis separately from heteroskedasticity, we also consider explicitly varying our threshold quantile for which patients count as marginal.

The baseline model reported above included four main classes of patient level risk factors: PE specific risk factors, chronic condition warehouse comorbidities, Elixhauser comorbidities, and patient demographic variables. Because some variation in comorbidities is required to appropriately identify this model, we retain the PE specific risk factors and the chronic condition warehouse comoribidities throughout, and test the stability of our findings to excluding the Elixhauser comorbidity set and the vector of demographic variables. Results from this exercise are reported in Appendix Table 5; the empirical Bayes correction has been applied before reporting the mean and standard deviation of physician's testing thresholds.

The mean estimated value of physician's testing thresholds ranges between 5.6 percent and 6.6 percent, and shows evidence of substantial dispersion in all models. The standard deviation of $\tau_d$ ranges between 3.9 percent and 5.4 percent, depending on the set of included patient risk factors. Dropping covariates does appear to increase the value of the estimated mean $\tau_d$ although the range of values across specifications is only one quarter of the estimated across-doctor standard deviation. If including additional covariates would cause estimates of $\tau_d$ to decrease, this suggests that our results may be conservative with respect to the amount of over testing. Controlling for the full set of risk factors also appears to increase the variance in estimated testing thresholds, providing suggestive evidence that the observed variation in thresholds is not driven by the exclusion of unobserved risk factors from the model. In all of these cases, variation in testing thresholds is sufficient to imply large differences in testing probabilities for identical patients depending on which doctor they visit.

It is not surprising that the mean $\tau_d$ increases when we exclude covariates. When we exclude comorbidities from the sample, we make it more difficult to identify accurately the marginal tested patients, and may end up including more non-marginal patients in this calculation. These non-

marginal tested patients will have higher average test yields, and so will push up our estimated test thresholds. To examine more directly the sensitivity of our results to the definition of marginal patients, we explicitly vary this definition in Appendix Table 4. We include all the baseline covariates but vary the quantile of the testing propensity cutoff below which patients are defined as marginal. Less stringent definitions of marginal patients than in our baseline results recover a larger average value of the physician threshold as predicted and more stringent definitions recover a lower value, suggesting our results are conservative with respect to the degree of over testing to the extent that with more data (or more covariates) we could better identify those patients who were truly marginal.

All specifications also predict substantial misweighting of included risk factors. The average absolute value of misweighting in physicians' assessment of PE risk ranges from 0.020 to 0.023 percentage points. Perhaps unsurprisingly, the full model which includes all available risk factors as candidate sources of misweighting recovers the largest predicted amount of misweighting. In all cases, misweighting is sufficiently large that it has the potential to change testing decisions for many marginal patients. Appendix Table 4 reports that varying the definition of marginal patients also does not change the estimated misperception of PE risk.

In results reported in Appendix Table 6 we find that the specific misweighted factors identified in Table 2 and discussed in section 4.2 continue to show evidence of misweighting of similar direction and magnitude, even as we vary the set of other included comorbidities. For example, the PE risk associated with recent hospital admissions and history of PE or deep vein thrombosis appears significantly underweighted in all specifications; black patients also show evidence of being under-tested in both specifications that include demographic variables. Similarly, a consistent set of conditions shows evidence of overweighting across specifications, including ischemic heart disease, chronic obstructive pulmonary disease and atrial fibrillation. These findings are not sensitive to the choice of other included covariates.

### 7.2 Estimation with physician-specific heteroskedasticity

Even if our results are not sensitive to dropping some covariates, we might worry that PE risk factors we cannot observe from insurance claims vary systematically across doctors. Differences across doctors in the variance of $\eta_{id}$ could arise for at least three reasons. First, doctors may differ in their skill at assessing risk factors unobservable to the econometrician. A doctor with more diagnostic skill may have a higher variance in $\eta_{id}$ across his patients, since he is more discerning in his judgement of which patients should be tested on the basis of clinical presentation and symptoms. Second, doctors may differ in the variance of latent PE risk present in their patient population. A doctor with a more heterogeneous patient population may have a higher variance in $\eta_{id}$ across his patients. Finally, doctors may simply make "errors" that lead them to deviate from typical practice patterns; a doctor who frequently deviates from his peers' practice patterns in assessing PE risk may have have a higher variance in $\eta_{id}$. The model we develop in this section allows us to isolate differences in physician testing thresholds that are unrelated to possible differences in the variance of $\eta_{id}$ across physicians.

Recall the assumption we made in Section 3 that $\eta_{id}$ followed a mixture of a Bernoulli and uniform distribution. We maintain the basic shape of the distribution but now allow both the

Bernoulli probability and the variance of the uniform distribution to vary across doctors, so that $\eta_{id} \sim U(-\eta_d, \eta_d)$ with probability $1 - p_d$ and $\eta_{id} \sim U[v - \eta_d, v + \eta_d]$ with probability $p_d$.

Following the derivation in Appendix 5, the more flexible distributional assumption implies the testing equation takes this form:

$$Pr(Test_{id} = 1) \quad = \quad \max\left\{0, \frac{p_d}{2} + \frac{p_d(I'_{id} + v)}{2\eta_d}\right\} \tag{18}$$

From the testing equation above, we can see that heteroskedasticity in $\eta_{id}$ is identified by the fact that observables are less predictive of testing behavior for doctors with a high variance in $\eta_{id}$, i.e. a smaller value of $\frac{p_d}{\eta_d}$. As described in the appendix, the testing equation can be used to estimate $C\frac{p_d}{2\eta_d}$, where $C$ is an unknown scaling constant. For computational tractability given the demands of this more flexible estimation strategy, we randomly exclude half of the physicians from our sample to reduce sample size, and drop the Elixhauser comorbidities and demographic risk factors from our list of included covariates.

With the introduction of heteroskedasticity, the conditional probability of a positive test is given by:

$$E(q_{id}|Test_{id} = 1) \quad = \quad \tau_d + \frac{C}{2}\frac{\tilde{I'}_{id}}{\hat{\eta}_d} + (x_{id} - E_d(x_{id}))(\beta - \beta') \tag{19}$$

where $\hat{\eta_d} = C\frac{p_d}{2\eta_d}$ are the variances estimated in the testing equation. Further details of the estimation strategy are provided in Appendix 5.

Appendix Table 5 reports the results of this analysis in panel 4, which can be compared to results from the baseline model with the same excluded comorbidity set, as reported in panel 3. The mean value of physicians' test thresholds $\tau_d$ is slightly higher at 7.0 percent in the model allowing for heteroskedasticity compared to 6.6 percent in the baseline model with the same covariates. Estimates of the standard deviation of $\tau_d$ are are also higher at 5.1 percentage points in the heteroskedastic model compared to 3.9 percentage points in the homoskedastic model. Thus, the cross-physician variation in testing behavior is not explained by differences in the variance of $\eta_{id}$ across doctors. This provides reassuring evidence that the assumption of homoskedasticity in the baseline model was not leading us to overstate differences across physicians in testing thresholds. Finally, the degree of misweighting remains very similar to the original estimates, with the average absolute value of misweighting estimated at 0.021 in the heteroskedastic model compared to 0.020 in the baseline model.

The role of physician diagnostic judgment in driving testing behavior and outcomes was previously explored by Doyle et al. (2010). In a natural experiment, they find that physicians from more prestigious residency programs achieve similar patient outcomes at 10-25 percent lower cost compared to their less skilled peers. One potential explanation for this phenomenon is that physicians from less prestigious schools prefer to administer more low-value care and could achieve the same outcomes at lower cost if they cut back some services. In the language of our model, these less skilled physicians might have lower testing thresholds, i.e. smaller $\tau_d$. A second explanation is that these less skilled physicians just need to use more medical resources to achieve the same quality of

care, because they are less accurate in their assessments of ex ante patient risk. In the language of our model, this decreased diagnostic accuracy would correspond to a lower variance of $\eta_{id}$, since these less skilled physicians would be failing to incorporate clinical information about patient risk to improve test targeting. Our results suggest that the heterogeneity in measured $\tau_d$ across physicians persists even after allowing for heterogeneous variance of $\eta_{id}$ acoss doctors. This finding raises the possibility that cost variance across physicians is driven in part by lower marginal value services provided by doctors with lower expected benefit thresholds.

### 7.3 Estimation of a semiparametric selection model

Next we test whether our results are sensitive to the shape of the distribution assumed for the unobserved component of patient PE risk, $\eta_{id}$. We previously imposed a strict distributional assumption, requiring $\eta_{id}$ to be distributed according to a mixture of Bernoulli and Uniform distributions. Now, we relax this assumption by estimating equation 10 from section 3.4 as a semiparametric binary choice model, using the Klein and Spady (1993) binary choice estimator. This robustness exercise will ensure that differences in testing thresholds observed in the previous sections are not driven solely by the strong distributional assumptions which restricted the functional form of the testing equation and the shape of the selection correction function $\lambda(\cdot)$. To implement the semiparametric model, we return to our original, strong version of the ignorability assumption that $\eta_{id}$ is i.i.d. across physicians and patients.

Estimation of the semiparametric model proceeds as follows. Let $g$ denote the probability that patient $i$ is tested given index $I'_{id} = x_{id}\beta' + \theta'_d$. The log likelihood is given by:

$$L(\beta, g) = \sum_i [Test_{id} \ln g(x_{id}\beta' + \theta'_d) + (1 - Test_{id})(1 - \ln g(x_{id}\beta' + \theta'_d))] \tag{20}$$

The idea of the Klein-Spady estimator is to approximate $g$ using a "leave-one-out" estimator which predicts the probability of testing for a particular patient, giving more weight to patients with nearby indices $I'_{id}$. Specifically, we substitute for $g$ using the following function:

$$\hat{g}_{-i,d} = \frac{\sum_{j \neq i} k\left(\frac{I'_{jd} - I'_{id}}{h}\right) Test_j}{\sum_{j \neq i} k\left(\frac{I'_{jd} - I'_{id}}{h}\right)} \tag{21}$$

We use a 4th-order Gaussian Kernel, $k(\cdot)$, and empirically select for the smallest bandwidth $h$ such that $\hat{g}$ is a monotonic function of the index $I'_{id}$.

Given the propensity to test index $I'_{id}$ from estimating equation 10 from section 3.4 by the Klein-Spady procedure, the next step is to estimate the testing outcome equation. Echoing the derivation in Section 3.2, the probability of a positive test among tested patients is given by:

$$E(Z_{id}|Test_{id} = 1) = \tau_d + x_{id}(\beta - \beta') + \lambda(I'_{id}) \tag{22}$$

where $\lambda(I'_{id}) = I'_{id} + h(I'_{id})$. Because we no longer assume a particular distribution of $\eta_{id}$, we now fit the function $\lambda(\cdot)$ flexibly, reporting results with $\lambda(\cdot)$ as a linear function and as a cubic polynomial, and estimate the net benefit equation by OLS.

Note that the Klein-Spady estimator only recovers $I'_{id}$ up to a location and scale normalization. The scale normalization is embedded in the function $\lambda(\cdot)$. We impose the appropriate location normalization so that at the smallest value of $I'_{id}$ among tested patients, $\underline{I}$, we have $\lambda(\underline{I}) = 0$ as shown in Section 3.3.[2]

Estimation of the semiparametric model is quite computationally intensive, and as a result, we maintain the restricted sample size and covariate set also used in the estimation of the heteroskedastic model in the previous section. Each time we construct the likelihood function, we need to construct a jackknife estimate for each observation which is a weighted average across all other observations given our kernel and bandwidth. This is nested within an optimization problem in which we estimate the parameters of our model for a given bandwidth. We then iterate the entire procedure, searching over for the smallest bandwidth that gives a monotonic result.

Results of the semiparametric estimation are reported in Appendix Table 5, panels 5 and 6. This semiparametric estimation approach estimates the mean value of $\tau_d$ at 6.7 percent (linear) or 6.6 percent (cubic), similar to the parametric model estimate of 6.6 percent in the sample with identical comorbidities. We continue to find a large amount of cross-doctor dispersion in estimated testing thresholds. The standard deviation of $\tau_d$ is 5.4 percent across doctors, compared to 3.9 percent in the parametric model with the same covariates (but interestingly nearly identical to the parametric model with the full set of covariates included). Our assessment of misweighting continues to be highly consistent across models, with an average absolute value of the error due to misweighting at 2.1 percent in the semiparametric model, compared to 2.0 percent in the parametric model.

Taken together, these robustness checks, including varying the set of included covariates, allowing for physician-specific heteroskedasticity, and estimating a semiparametric selection model, all suggest that our findings on the dispersion in testing thresholds and amount of misweighting are very stable across alternative modeling assumptions. We find substantial variance in testing thresholds of similar magnitude in all specifications, suggesting that much of the observed variation in testing behavior may be driven by differences in practice styles. Further, doctors are misassessing patient PE risk by similar amounts in percentage point terms across all models.

# 8 Computing the welfare costs of over testing and misweighting

In order to calculate the welfare costs of over testing 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. Let $fp$ denote the likelihood of a false positive, $s$ the sensitivity of the test (one minus the probability of a false negative), $MB$ the medical benefits of treating a PE, $MC$ the medical costs and $CT$ the financial costs of treatment. In this section, we show that allowing for false positives and false negatives results in a model which is isomorphic to the one above with $NU$ replaced by $\hat{NU} = \frac{s}{s-fp}MB - MC - CT$ and $c$ replaced by $\hat{c} = c + \frac{s \cdot fp}{s-fp}MB$.

We begin by calculating the net utility of treatment, given that there are both false positive and

[2]This normalization can be implemented by omitting the constant term from the polynomial $\lambda(\cdot)$ and subtracting a constant $\underline{I}$ from $\hat{I'}_{id}$; thus the resulting polynomial $\lambda(I'_{id} - \underline{I})$ will equal 0 for $I'_{id} = \underline{I}$. To avoid sensitivity to outliers, we normalize $I'_{id}$ so that $\lambda(\underline{I}) = 0$ for $I'_{id}$ in the 10th percentile amongst tested patients, which agrees with our definition of marginal patients in Section 3.3.

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:

$$NU_{id} = Pr(PE_{id}|Z_{id} = 1)MB - MC - CT \tag{23}$$

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:

$$NU_{id} = \frac{s(q_{id} - fp)}{q_{id}(s - fp)}MB - MC - CT \tag{24}$$

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:

$$\begin{aligned} B_{id} &= q_{id}NU_{id} - c \\ &= \frac{s(q_{id} - fp)}{(s - fp)}MB - q_{id}MC - q_{id}CT - c \end{aligned} \tag{25}$$

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:

$$B_{id} = q_{id}\hat{NU} - \hat{c} \tag{26}$$

The optimal testing threshold $\tau^*$ 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 3 and Appendix 4, to compute the welfare cost of over testing 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 $\tau_d$ 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 $\tau^*$, 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:

$$\begin{aligned} MB(\tau_d, \Delta\beta) &= \sum_i Pr(Test_i = 1) \cdot Pr(PE_{id}|Test_i = 1)MB_{id} \quad (27) \\ &= \sum_i \hat{t}_{id}(\tau_d, \Delta\beta)\frac{s(\hat{Z}_{id}(\tau_d, \Delta\beta) - fp)}{(s - fp)}MB_{id} \\ MC(\tau_d, \Delta\beta) &= \sum_i Pr(Test_i = 1)Pr(Z_{id} = 1|Test_i = 1)MC_{id} \quad (28) \\ &= \sum_i \hat{t}_{id}(\tau_d, \Delta\beta)\hat{Z}_{id}(\tau_d, \Delta\beta)MC_{id} \\ FC(\tau_d, \Delta\beta) &= \sum_i Pr(Test_i = 1)(c + P(Z_{id} = 1|Test_i = 1)CT_{id}) \quad (29) \\ &= \sum_i \hat{t}_{id}(\tau_d, \Delta\beta)(c + \hat{Z}_{id}(\tau_d, \Delta\beta)CT_{id}) \\ NB(\tau_d, \Delta\beta) &= MB(\tau_d, \Delta\beta) - MC(\tau_d, \Delta\beta) - FC(\tau_d, \Delta\beta) \quad (30) \end{aligned}$$

where $MB$ denote the medical benefits of testing, $MC$ 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 over testing 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 $\tau_d$ which we recover using the methods outlined in Appendix 5 above.

# 9 Appendix Bibliography

# References


Altonji, J. G., T. E. Elder, and C. R. Taber (2008). Using selection on observed variables to assess bias from unobservables when evaluating swan-ganz catheterization. *The American Economic Review 98*(2), pp. 345–350.

David, S., P. Beddy, J. Babar, and A. Devaraj (2012, Feb). Evolution of ct pulmonary angiography: referral patterns and diagnostic yield in 2009 compared with 2006. *Acta Radiologica 53*(1), 36–43.

Doyle, J. J., S. M. Ewer, and T. H. Wagner (2010). Returns to physician human capital: Evidence from patients randomized to physician teams. *Journal of health economics 29*(6), 866–882.

Klein, R. W. and R. H. Spady (1993). An efficient semiparametric estimator for binary response models. *Econometrica: Journal of the Econometric Society*, 387–421.

Meszaros, I., J. Morocz, J. Szlavi, J. Schmidt, L. Tornoci, L. Nagy, and L. Szep (2000, May). Epidemiology and clinicopathology of aortic dissection. *Chest 117*(5), 1271–1278.

Venkatesh, A., J. A. Kline, and C. Kabrhel (2013, Jan. 28). Computed tomography in the emergency department setting–reply. *Journal of the American Medical Association Internal Medicine 173*(2), 167–168.

# 10 Appendix Tables

Appendix Table 1: Summary statistics

| | *A. Untested patients* | *B. Patients with negative tests* | *C. Patients with positive tests* |
|---|---|---|---|
| *Patient characteristics* | | | |
| Age | 77.6 | 76.8 | 76.9 |
| Female | 0.586 | 0.602 | 0.600 |
| Black | 0.082 | 0.066 | 0.083 |
| History of PE | 0.003 | 0.006 | 0.017 |
| *Doctor, hospital and region characteristics* | | | |
| Doctor experience | 16.5 | 16.4 | 16.8 |
| | (8.3) | (8.4) | (8.5) |
| Top 50 research med. school | 0.28 | 0.29 | 0.30 |
| Top 50 primary med. school | 0.26 | 0.27 | 0.28 |
| Academic hospital | 0.33 | 0.34 | 0.356 |
| For profit hospital | 0.12 | 0.13 | 0.120 |
| HRR avg spending (in $) | 8,198 | 8,173 | 8,089 |
| | (959) | (972) | (936) |
| Average income in region | 22,771 | 23,005 | 23,039 |
| | (5521) | (5490) | (5710) |
| Joint and several liability | 0.69 | 0.70 | 0.692 |
| Malpractice damage caps | 0.70 | 0.76 | 0.747 |
| Number of observations | 1,819,015 | 66,677 | 4,968 |

Notes: Table reports means and standard deviations (in parentheses). Data is from the Medicare claims 2000-2009, the American Hospital Association annual survey, the American Medical Association Masterfile, the Dartmouth Atlas, and the Avraham Database of State Tort Law Reform. Results reported separately for patients who do not receive a CT scan (column A), patients who receive a negative test (column B), and patients with a positive test (column C). We observe the testing behavior of over 6600 physicians, with an average of 284 ED patients per physician.

Appendix Table 2: Regressions of testing threshold on physician characteristics and practice environment

| | ***Dependent variable: Physician testing threshold*** | | | |
|---|---|---|---|---|
| | OLS | FGLS | OLS | FGLS |
| *Independent variables:* | (1) | (2) | (3) | (4) |
| Doctor experience | 0.0007 | 0.0007 | 0.0007 | 0.0008 |
| | (0.0001) | (0.0001) | (0.0002) | (0.0001) |
| Top 50 research medical school | 0.0047 | 0.0050 | 0.0053 | 0.0032 |
| | (0.0038) | (0.0031) | (0.0047) | (0.0037) |
| Top 50 primary care medical school | -0.0062 | -0.0042 | -0.0077 | -0.0030 |
| | (0.0039) | (0.0032) | (0.0048) | (0.0037) |
| Academic hospital | 0.0006 | 0.0007 | | |
| | (0.0026) | (0.0022) | | |
| For profit hospital | -0.0004 | -0.0018 | | |
| | (0.0041) | (0.0032) | | |
| Log(HRR avg Medicare spending) | -0.0391 | -0.0474 | | |
| | (0.0109) | (0.0093) | | |
| Average income in region (in $10k) | 0.0000 | 0.0000 | | |
| | (0.0025) | (0.0019) | | |
| Joint and several liability | 0.0001 | 0.0003 | | |
| | (0.0027) | (0.0023) | | |
| Malpractice damage caps | -0.0029 | -0.0053 | | |
| | (0.0028) | (0.0023) | | |
| Hospital Fixed Effects | No | No | Yes | Yes |

Notes: Each column reports results from a regression of estimated physician testing thresholds $\tau_d$ on characteristics of the physician's training and practice environment. Even numbered columns report FGLS estimates which account for estimation error in $\tau_d$. Columns 3 and 4 include hospital fixed effects. An observation is an individual doctor; there are 6636 observations.
Source: Data are from the Medicare claims 2000-2009, the American Hospital Association annual survey, the Avraham database of state tort law reforms, the Dartmouth Atlas, and US News and World Report.

Appendix Table 3: Comorbidities with no significant misweighting:
Impact of comorbidity on testing decisions and estimated misassessment of PE risk (continued)

| | Marginal effect from testing eqn (1) | Misassessment of PE risk (2) | Std. error of misassessment (3) | T statistic of misassessment (4) |
|---|---|---|---|---|
| *Other comorbidities* | | | | |
| History of hip fracture (CCW) | -0.0035 | 0.0192 | 0.0116 | 1.6552 |
| Alzheimer's related dementias (CCW) | -0.0060 | 0.0077 | 0.0047 | 1.6383 |
| Anemia (CCW) | -0.0023 | 0.0038 | 0.0024 | 1.5833 |
| Depression (CCW) | -0.0008 | 0.0042 | 0.0031 | 1.3548 |
| Hypertension (CCW) | 0.0008 | 0.0033 | 0.0025 | 1.3200 |
| Solid tumor w/o metastisis (Elixhauser) | -0.0066 | 0.0145 | 0.0112 | 1.2946 |
| Benign prostatic hyperplasia (CCW) | -0.0014 | 0.0046 | 0.0038 | 1.2105 |
| Hypothyroidism (Elixhauser) | -0.0009 | 0.0068 | 0.0060 | 1.1333 |
| Liver disease (Elixhauser) | -0.0066 | 0.0219 | 0.0195 | 1.1231 |
| Prior surgery within 1 year | 0.0136 | 0.0239 | 0.0215 | 1.1116 |
| Blood loss anemia (Elixhauser) | -0.0044 | 0.0126 | 0.0118 | 1.0678 |
| Breast cancer (CCW) | 0.0066 | 0.0046 | 0.0049 | 0.9388 |
| Stroke / Transient ischemic attack (CCW) | -0.0099 | 0.0035 | 0.0046 | 0.7609 |
| Chronic kidney disease (CCW) | -0.0091 | 0.0024 | 0.0042 | 0.5714 |
| Psychoses (Elixhauser) | -0.0057 | 0.0046 | 0.0126 | 0.3651 |
| Congestive heart failure (Elixhauser) | -0.0022 | 0.0018 | 0.0056 | 0.3214 |
| Congestive heart failure (CCW) | -0.0006 | 0.0008 | 0.0028 | 0.2857 |
| Drug abuse (Elixhauser) | 0.0059 | 0.0060 | 0.0304 | 0.1974 |
| Alcohol abuse (Elixhauser) | 0.0008 | 0.0020 | 0.0149 | 0.1342 |
| Pulmonary circulation disease (Elixhauser) | -0.0035 | 0.0009 | 0.0107 | 0.0841 |
| Acute myocardial infarction (CCW) | -0.0058 | 0.0002 | 0.0090 | 0.0222 |
| Lymphoma (Elixhauser) | -0.0174 | -0.0005 | 0.0220 | -0.0227 |
| Coagulation deficiency (Elixhauser) | -0.0001 | -0.0006 | 0.0109 | -0.0550 |
| Weight loss (Elixhauser) | -0.0054 | -0.0021 | 0.0119 | -0.1765 |
| Prior surgery within 30 days | 0.0151 | -0.0047 | 0.0191 | -0.2461 |
| Arthritis (Elixhauser) | 0.0044 | -0.0032 | 0.0096 | -0.3333 |
| Fluid & electrolyte disorders (Elixhasuer) | -0.0013 | -0.0022 | 0.0047 | -0.4681 |
| Acquired hypothyroidism (CCW) | 0.0022 | -0.0020 | 0.0035 | -0.5714 |
| Hyperlipidemia (CCW) | 0.0054 | -0.0017 | 0.0024 | -0.7083 |
| Hypertension (CCW) | 0.0012 | -0.0051 | 0.0040 | -1.2750 |
| Diabetes w/chronic complications (Elixhauser) | -0.0080 | -0.0176 | 0.0115 | -1.5304 |
| Glaucoma (CCW) | -0.0003 | -0.0047 | 0.0029 | -1.6207 |
| Diabetes w/o chronic complications (Elixhauser) | -0.0023 | -0.0085 | 0.0051 | -1.6667 |
| Lung cancer (CCW) | -0.0142 | -0.0198 | 0.0113 | -1.7522 |
| Cataracts (CCW) | -0.0010 | -0.0037 | 0.0021 | -1.7619 |
| Valvular disease (Elixhauser) | -0.0031 | -0.0116 | 0.0060 | -1.9333 |

Notes: Results continued from Table 2; this table includes only covariates without significant evidence of misweighting. Column 1 reports marginal effects from coefficient estimates of the testing equation (i.e. equation 2); for example, patients with hip fracture history are 0.35 percentage points less likely to be tested, after controlling for included PE risk factors and physicians' testing thresholds. Column 2 reports estimates of physicians' misweighting of these PE risk factors estimated from equation 14; for example, physicians' observed testing patterns suggest they are overestimating the PE risk associated with hip fracture history by 1.92 percentage points. Column 3 reports standard errors for coefficients in column 3. Column 4 reports t-statistics. Variables are sorted by statistical significance.

Appendix Table 4: Distribution of test thresholds and misweighting:
Robustness to varying definition of marginal tested patient

| | **Strict definition** | **Baseline definition** | **Lax definition** |
|---|---|---|---|
| | (1) | (2) | (3) |
| Fraction of tested patients defined as marginal | 5% | 10% | 15% |
| Mean of $\tau_d$ | 0.0492 | 0.0563 | 0.0620 |
| Standard deviation of $\tau_d$ | 0.0497 | 0.0540 | 0.0582 |
| Average absolute value of PE misassessment | 0.0228 | 0.0226 | 0.0227 |
| Standard deviation of PE misassessment | 0.0344 | 0.0347 | 0.0353 |
| Number of observations | 1,890,660 | 1,890,660 | 1,890,660 |

Notes: Recall that the test yield among each doctor's marginal tested patients (those just barely worth testing) is used to estimate the doctor's test thresholds and form exclusion restrictions that identify the model. Each column of this table reports estimation results under an alternative definition of the marginal tested patient. The baseline results, reported in column 2 for easy comparison, define a patient as marginal if they are in the bottom 10 percent of tested patients on the basis of their estimated testing propensity index. Column 1 employs a stricter definition, allowing only the bottom 5 percent of tested patients to be counted as marginal; column 3 employs a weaker definition, allowing the bottom 15 percent of testd patients to be counted as marginal. Each column reports the estimated posterior mean and standard deviation of physician testing thresholds $\tau_d$ from the model, after applying the Bayesian shrinkage described in Appendix 5. Recall that $\tau_d$ is the threshold probability of a positive test at which a physician determines it is worthwhile to test a patient. The average absolute value of PE risk misassessment calculates the absolute value of the difference between physicians' assessment of the patient's PE probability and the estimated risk associated with the patient's comorbidities, and then averages this value across all patients. The standard deviation of PE misassessment describes how the total amount of misweighting varies across patients.

Appendix Table 5: Distribution of testing thresholds and misweighting under alternative estimation strategies

| | Baseline parametric model, all comorbidities | Parametric model, Elixhauser comorbidities excluded | Parametric model, Elixhauser comorbidities & demographics excluded |
|---|---|---|---|
| | (1) | (2) | (3) |
| Mean of $\tau_d$ | 0.0563 | 0.0623 | 0.0662 |
| Standard deviation of $\tau_d$ | 0.0540 | 0.0396 | 0.0394 |
| Average absolute value of PE misassessment | 0.0226 | 0.0214 | 0.0200 |
| Standard deviation of PE misassessment | 0.0347 | 0.0336 | 0.0329 |
| Number of observations | 1,890,660 | 1,890,660 | 1,890,660 |
| | Heteroskedastic parametric model | Semiparametric model, linear polynomial | Semiparametric model, cubic polynomial |
| | (4) | (5) | (6) |
| Mean of $\tau_d$ | 0.0703 | 0.0672 | 0.0661 |
| Standard Deviation of $\tau_d$ | 0.0514 | 0.0539 | 0.0541 |
| Average absolute value of PE misassessment | 0.0212 | 0.0207 | 0.0208 |
| Standard deviation of PE misassessment | 0.0361 | 0.0357 | 0.0364 |
| Number of observations | 861,707 | 861,707 | 861,707 |

Notes: Panel 1 reports the estimated posterior mean and standard deviation of physician testing thresholds $\tau_d$ from our baseline parametric model, after applying the Bayesian shrinkage described in Appendix 6. Recall that $\tau_d$ is the threshold probability of a positive test at which a physician determines it is worthwhile to test a patient. The average absolute value of misweighting calculates the absolute value of the difference between physicians' assessment of the patient's PE probability and the estimated risk associated with the patient's comorbidities, and then averages this value across all patients. The standard deviation of misweighting describes how the amount of misweighting varies across patients. Panel 2 reports results from the parametric model that excludes all Elixhauser comorbidities. Panel 3 reports results from the parametric model that excludes both Elixhauser comorbidities and demographic variables. Panel 4 reports results from the heteroskedastic model described in Appendix 7.2, which allows the variance of $\eta_{id}$ to differ across physicians. Panels 5 and 6 report results from the semiparametric model described in Appendix 7.3, where Panel 5 fits the function $\lambda(\cdot)$ with a linear function and Panel 6 applies a cubic polynomial. Models estimated in Panels 4, 5, and 6 exclude Elixhauser comorbidities and demographic variables and are estimated on a random subsample of half of the physicians for computational tractability.
Source: Data are from the Medicare claims 2000-2009.

Appendix Table 6: Part 1: Assessment of misweighting with varying included covariates

| | *All comorbidities* | | *Excluding Elixhauser comorbidities* | | *Excluding Elixhauser comorbidities and demographics* | |
|---|---|---|---|---|---|---|
| | Misassessment of PE risk | Standard error | Misassessment of PE risk | Standard error | Misassessment of PE risk | Standard error |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Underweighted risk factors* | | | | | | |
| Prior hospital visit w/in 30 days | 0.1070 | 0.0121 | 0.1025 | 0.0125 | 0.1045 | 0.0125 |
| Prior hospital visit w/in 7 days | 0.1128 | 0.0130 | 0.1091 | 0.0133 | 0.1105 | 0.0133 |
| Prostate cancer (CCW) | 0.0298 | 0.0048 | 0.0311 | 0.0048 | 0.0318 | 0.0046 |
| Cancer metastisis (Elixhauser) | 0.0726 | 0.0128 | 0.0843 | 0.0134 | 0.0892 | 0.0134 |
| History of deep vein thrombosis | 0.0571 | 0.0114 | 0.0560 | 0.0113 | 0.0570 | 0.0113 |
| History of pulmonary embolism | 0.0666 | 0.0145 | 0.0800 | 0.0142 | 0.0827 | 0.0141 |
| Rhumatoid arthritis, osteoarthritis (CCW) | 0.0091 | 0.0024 | 0.0097 | 0.0025 | 0.0108 | 0.0024 |
| Endometrial cancer (CCW) | 0.0547 | 0.0153 | 0.0438 | 0.0154 | 0.0405 | 0.0153 |
| Obesity (Elixhauser) | 0.0218 | 0.0076 | | | | |
| Paralysis (Elixhauser) | 0.0331 | 0.0117 | | | | |
| Other neurological conditions (Elixhauser) | 0.0194 | 0.0075 | | | | |
| Any prior admission history | 0.0102 | 0.0041 | 0.0033 | 0.0029 | 0.0028 | 0.0029 |
| Alzheimer's disease (CCW) | 0.0152 | 0.0064 | 0.0158 | 0.0065 | -0.0036 | 0.0092 |
| Colorectal cancer (CCW) | 0.0136 | 0.0067 | 0.0166 | 0.0067 | 0.0163 | 0.0067 |
| *Overweighted risk factors* | | | | | | |
| Ischemic heart disease (CCW) | -0.0226 | 0.0023 | -0.0233 | 0.0023 | -0.0226 | 0.0023 |
| Chronic obstructive pulmonary disease (CCW) | -0.0182 | 0.0036 | -0.0158 | 0.0037 | -0.0159 | 0.0037 |
| Atrial fibrillation (CCW) | -0.0156 | 0.0036 | -0.0172 | 0.0036 | -0.0175 | 0.0036 |
| Depression (Elixhauser) | -0.0208 | 0.0069 | | | | |
| Peripheral vascular disease (Elixhauser) | -0.0214 | 0.0071 | | | | |
| Diabetes (CCW) | -0.0087 | 0.0029 | -0.0115 | 0.0028 | -0.0105 | 0.0028 |
| Osteoperosis (CCW) | -0.0087 | 0.0033 | -0.0079 | 0.0033 | -0.0075 | 0.0032 |
| Deficiency anemias (Elixhauser) | -0.0142 | 0.0056 | | | | |
| Asthma (CCW) | -0.0088 | 0.0040 | -0.0086 | 0.0040 | -0.0072 | 0.0040 |
| Chronic pulmonary disease (Elixhauser) | -0.0094 | 0.0048 | | | | |
| *Demographic factors* | | | | | | |
| Black | 0.0257 | 0.0044 | 0.0189 | 0.0045 | | |
| Asian | -0.0386 | 0.0118 | -0.0392 | 0.0118 | | |
| Hispanic | -0.0168 | 0.0097 | -0.0142 | 0.0100 | | |
| Female | 0.0000 | 0.0024 | 0.0000 | 0.0024 | | |
| Age 65-69 | 0.0119 | 0.0037 | 0.0103 | 0.0037 | | |
| Age 70-74 | 0.0129 | 0.0052 | 0.0092 | 0.0053 | | |
| Age 75-79 | 0.0140 | 0.0038 | 0.0122 | 0.0038 | | |
| Age 80-84 | 0.0166 | 0.0039 | 0.0133 | 0.0039 | | |
| Age 85-89 | 0.0208 | 0.0042 | 0.0181 | 0.0042 | | |
| Age 90-94 | 0.0132 | 0.0078 | 0.0075 | 0.0081 | | |

Notes: Table continued on next page. Column 1 reports estimates of physicians' misweighting of these PE risk factors estimated from equation 14 under the baseline specification with full set of included covariates. Column 2 reports standard errors on these misweighting terms. (Columns 1 and 2 replicate results reported in Table 2 for purposes of comparison.) Columns 3 and 4 also report misweighting terms and standard errors, now from the model that excludes the Elixhauser comorbidity set. Columns 5 and 6 report results from the model that excludes both Elixhauser comoribidites and demographic factors.

Appendix Table 6 Part 2: Assessment of misweighting with varying included covariates

| | *All comorbidities* | | *Excluding Elixhauser comorbidities* | | *Excluding Elixhauser comorbidities and demographics* | |
|---|---|---|---|---|---|---|
| | Misassessment of PE risk | Standard error | Misassessment of PE risk | Standard error | Misassessment of PE risk | Standard error |
| *Other comorbidities* | (1) | (2) | (3) | (4) | (5) | (6) |
| History of hip fracture (CCW) | 0.0192 | 0.0116 | 0.0025 | 0.0118 | 0.0042 | 0.0117 |
| Alzheimer's related dementias (CCW) | 0.0077 | 0.0047 | 0.0070 | 0.0048 | 0.0070 | 0.0049 |
| Anemia (CCW) | 0.0038 | 0.0024 | 0.0014 | 0.0024 | 0.0024 | 0.0024 |
| Depression (CCW) | 0.0042 | 0.0031 | -0.0006 | 0.0029 | -0.0010 | 0.0029 |
| Hypertension (CCW) | 0.0033 | 0.0025 | 0.0042 | 0.0024 | 0.0052 | 0.0024 |
| Solid tumor w/o metastisis (Elixhauser) | 0.0145 | 0.0112 | | | | |
| Benign prostatic hyperplasia (CCW) | 0.0046 | 0.0038 | 0.0062 | 0.0038 | 0.0070 | 0.0035 |
| Hypothyroidism (Elixhauser) | 0.0068 | 0.0060 | | | | |
| Liver disease (Elixhauser) | 0.0219 | 0.0195 | | | | |
| Prior surgery within 1 year | 0.0239 | 0.0215 | 0.0352 | 0.0217 | 0.0293 | 0.0218 |
| Blood loss anemia (Elixhauser) | 0.0126 | 0.0118 | | | | |
| Breast cancer (CCW) | 0.0046 | 0.0049 | 0.0089 | 0.0049 | 0.0095 | 0.0049 |
| Stroke / Transient ischemic attack (CCW) | 0.0035 | 0.0046 | 0.0027 | 0.0047 | 0.0050 | 0.0047 |
| Chronic kidney disease (CCW) | 0.0024 | 0.0042 | 0.0031 | 0.0044 | 0.0014 | 0.0044 |
| Psychoses (Elixhauser) | 0.0046 | 0.0126 | | | | |
| Congestive heart failure (Elixhauser) | 0.0018 | 0.0056 | -0.0053 | 0.0056 | -0.0055 | 0.0056 |
| Congestive heart failure (CCW) | 0.0008 | 0.0028 | 0.0007 | 0.0028 | 0.0020 | 0.0028 |
| Drug abuse (Elixhauser) | 0.0060 | 0.0304 | | | | |
| Alcohol abuse (Elixhauser) | 0.0020 | 0.0149 | | | | |
| Pulmonary circulation disease (Elixhauser) | 0.0009 | 0.0107 | | | | |
| Acute myocardial infarction (CCW) | 0.0002 | 0.0090 | -0.0026 | 0.0092 | 0.0153 | 0.0066 |
| Lymphoma (Elixhauser) | -0.0005 | 0.0220 | | | | |
| Coagulation deficiency (Elixhauser) | -0.0006 | 0.0109 | | | | |
| Weight loss (Elixhauser) | -0.0021 | 0.0119 | | | | |
| Prior surgery within 30 days | -0.0047 | 0.0191 | -0.0066 | 0.0192 | -0.0031 | 0.0192 |
| Arthritis (Elixhauser) | -0.0032 | 0.0096 | | | | |
| Fluid & electrolyte disorders (Elixhasuer) | -0.0022 | 0.0047 | | | | |
| Acquired hypothyroidism (CCW) | -0.0020 | 0.0035 | 0.0007 | 0.0030 | 0.0013 | 0.0030 |
| Hyperlipidemia (CCW) | -0.0017 | 0.0024 | -0.0005 | 0.0025 | -0.0013 | 0.0025 |
| Hypertension (CCW) | -0.0051 | 0.0040 | | | | |
| Diabetes w/complications (Elixhauser) | -0.0176 | 0.0115 | | | | |
| Glaucoma (CCW) | -0.0047 | 0.0029 | -0.0043 | 0.0029 | -0.0023 | 0.0029 |
| Diabetes w/o complications (Elixhauser) | -0.0085 | 0.0051 | | | | |
| Lung cancer (CCW) | -0.0198 | 0.0113 | -0.0219 | 0.0117 | -0.0266 | 0.0116 |
| Cataracts (CCW) | -0.0037 | 0.0021 | -0.0029 | 0.0021 | -0.0017 | 0.0020 |
| Valvular disease (Elixhauser) | -0.0116 | 0.0060 | | | | |

Notes: Table continued from previous page. Column 1 reports estimates of physicians' misweighting of these PE risk factors estimated from equation 14 under the baseline specification with full set of included covariates. Column 2 reports standard errors on these misweighting terms. (Columns 1 and 2 replicate results reported in Table 2 for purposes of comparison.) Columns 3 and 4 also report misweighting terms and standard errors, now from the model that excludes the Elixhauser comorbidity set. Columns 5 and 6 report results from the model that excludes both Elixhauser comorbidities and demographic factors.

Appendix Table 7: Part 1: Assessing the costs of misweighting by variable

| | *Net Benefits* | *Change in net benefits* |
|---|---|---|
| **Original** | **13.279** | |
| Age 65-69 | 12.323 | -0.956 |
| Age 70-74 | 12.078 | -0.245 |
| Age 75-79 | 11.580 | -0.498 |
| Age 80-84 | 11.988 | 0.408 |
| Age 85-89 | 13.560 | 1.572 |
| Age 90-94 | 13.695 | 0.135 |
| Black | 15.486 | 1.791 |
| Asian | 15.707 | 0.221 |
| Hispanic | 15.802 | 0.095 |
| Acute myocardial infarction (CCW) | 15.802 | 0.000 |
| Alzheimer's disease (CCW) | 16.712 | 0.910 |
| Chronic obstructive pulmonary disease (CCW) | 18.879 | 2.167 |
| Congestive heart failure (CCW) | 18.815 | -0.064 |
| History of hip fracture (CCW) | 18.980 | 0.165 |
| Anemia (CCW) | 19.164 | 0.184 |
| Asthma (CCW) | 19.343 | 0.179 |
| Hyperlipidemia (CCW) | 19.516 | 0.173 |
| Benign prostatic hyperplasia (CCW) | 19.591 | 0.075 |
| Hypertension (CCW) | 19.432 | -0.159 |
| Acquired hypothyroidism (CCW) | 19.426 | -0.006 |
| Alzheimer's related dementias (CCW) | 19.644 | 0.218 |
| Atrial fibrillation (CCW) | 20.498 | 0.854 |
| Cataracts (CCW) | 20.625 | 0.127 |
| Chronic kidney disease (CCW) | 20.611 | -0.014 |
| Diabetes (CCW) | 21.392 | 0.781 |
| Glaucoma (CCW) | 21.484 | 0.092 |
| Ischemic heart disease (CCW) | 23.516 | 2.032 |
| Depression (CCW) | 23.616 | 0.100 |
| Osteoperosis (CCW) | 23.677 | 0.061 |
| Rhumatoid arthritis, osteoarthritis (CCW) | 24.503 | 0.826 |
| Stroke / Transient ischemic attack (CCW) | 24.603 | 0.100 |
| Breast cancer (CCW) | 24.664 | 0.061 |
| Colorectal cancer (CCW) | 25.079 | 0.415 |
| Prostate cancer (CCW) | 26.588 | 1.509 |
| Lung cancer (CCW) | 26.541 | -0.047 |
| Endometrial cancer (CCW) | 27.117 | 0.576 |

Notes: This table is continued on the next page. This table reports results of a series of simulation exercises where we test the welfare impact of correcting for physician misweighting of observed risk factors, one variable at a time. This exercise allows us to assess which specific risk factors are the biggest contributors to the welfare costs associated with misweighting. We proceed in the order listed in the table and show how the total net benefits of testing (in $ millions) change from their observed value of 13.279 to the final value 49.132 in the absence of any misweighting, by correcting one additional variable in each row. Note that because we continue to allow physician thresholds to vary and do not correct for all risk factors at once, correcting a single additional risk factor occasionally leads to a small decline in net benefits. The results of this exercise may be sensitive to the order in which risk factors are corrected.

Appendix Table 7 Part 2: Assessing the costs of misweighting by variable

| | *Net Benefits* | *Change in net benefits* |
|---|---|---|
| Prior surgery within 30 days | 26.311 | -0.806 |
| Prior surgery within 1 year | 30.794 | 4.483 |
| Any prior admission history | 32.632 | 1.838 |
| Valvular disease (Elixhauser) | 32.534 | -0.098 |
| Pulmonary circulation disease (Elixhauser) | 32.546 | 0.012 |
| Peripheral vascular disease (Elixhauser) | 32.496 | -0.050 |
| Paralysis (Elixhauser) | 32.927 | 0.431 |
| Other neurological conditions (Elixhauser) | 33.271 | 0.344 |
| Diabetes w/o chronic complications (Elixhauser) | 33.100 | -0.171 |
| Diabetes w/chronic complications (Elixhauser) | 33.058 | -0.042 |
| Hypothyroidism (Elixhauser) | 33.195 | 0.137 |
| Liver disease (Elixhauser) | 33.287 | 0.092 |
| Lymphoma (Elixhauser) | 33.286 | -0.001 |
| Solid tumor w/o metastisis (Elixhauser) | 33.518 | 0.232 |
| Arthritis (Elixhauser) | 33.509 | -0.009 |
| Coagulation deficiency (Elixhauser) | 33.504 | -0.005 |
| Obesity (Elixhauser) | 33.840 | 0.336 |
| Weight loss (Elixhauser) | 33.825 | -0.015 |
| Fluid & electrolyte disorders (Elixhasuer) | 33.770 | -0.055 |
| Blood loss anemia (Elixhauser) | 33.866 | 0.096 |
| Deficiency anemias (Elixhauser) | 33.668 | -0.198 |
| Alcohol abuse (Elixhauser) | 33.673 | 0.005 |
| Drug abuse (Elixhauser) | 33.675 | 0.002 |
| Psychoses (Elixhauser) | 33.687 | 0.012 |
| Depression (Elixhauser) | 33.706 | 0.019 |
| Hypertension (Elixhauser) | 33.176 | -0.530 |
| History of deep vein thrombosis | 34.174 | 0.998 |
| History of pulmonary embolism | 35.186 | 1.012 |
| Prior hospital visit w/in 30 days | 43.135 | 7.949 |
| Prior hospital visit w/in 7 days | 47.871 | 4.736 |
| Female | 47.871 | 0.000 |
| Chronic pulmonary disease (Elixhauser) | 47.903 | 0.032 |
| Congestive heart failure (Elixhauser) | 47.914 | 0.011 |
| Cancer metastisis (Elixhauser) | 49.132 | 1.218 |

Notes: This table is continued from the previous page. This table reports results of a series of simulation exercises where we test the welfare impact of correcting for physician misweighting of observed risk factors, one variable at a time. This exercise allows us to assess which specific risk factors are the biggest contributors to the welfare costs associated with misweighting. We proceed in the order listed in the table and show how the total net benefits of testing (in $ millions) change from their observed value of 13.279 to the final value 49.132 in the absence of any misweighting. Note that because we continue to allow physician thresholds to vary and do not correct for all risk factors at once, correcting a single additional risk factor occasionally leads to a small decline in net benefits. The results of this exercise may also be sensitive to the order in which risk factors are corrected.

Appendix Table 8: Sensitivity of welfare simulations to calibration parameters

| | A. Counterfactual with no overtesting | | |
|---|---|---|---|
| | *Percent tested* | *Test yield* | *Change in net benefits* |
| *False positive rate* | | | |
| 0.00 | 0.037 | 0.071 | 0.093 |
| 0.03 | 0.026 | 0.083 | 3.802 |
| **0.04** | 0.019 | 0.090 | 8.144 |
| *Value of a statistical life* | | | |
| $500,000 | 0.005 | 0.137 | 15.748 |
| **$1,000,000** | 0.019 | 0.090 | 8.144 |
| $1,500,000 | 0.025 | 0.081 | 5.249 |
| *Test sensitivity* | | | |
| 0.75 | 0.019 | 0.090 | 8.080 |
| **0.83** | 0.019 | 0.090 | 8.144 |
| 0.90 | 0.018 | 0.090 | 8.191 |
| *Financial cost of testing* | | | |
| $0 | 0.033 | 0.075 | 0.725 |
| **$300** | 0.019 | 0.090 | 8.144 |
| $500 | 0.012 | 0.104 | 16.872 |
| | **B. Counterfactual with no misweighting** | | |
| | *Percent tested* | *Test yield* | *Change in net benefits* |
| *False positive rate* | | | |
| 0.00 | 0.043 | 0.090 | 44.134 |
| 0.03 | 0.043 | 0.090 | 38.094 |
| **0.04** | 0.043 | 0.090 | 35.853 |
| *Value of a statistical life* | | | |
| $500,000 | 0.043 | 0.090 | 13.184 |
| **$1,000,000** | 0.043 | 0.090 | 35.853 |
| $1,500,000 | 0.043 | 0.090 | 58.522 |
| *Test sensitivity* | | | |
| 0.75 | 0.043 | 0.090 | 36.120 |
| **0.83** | 0.043 | 0.090 | 35.853 |
| 0.90 | 0.043 | 0.090 | 35.660 |
| *Financial cost of testing* | | | |
| $0 | 0.043 | 0.090 | 38.882 |
| **$300** | 0.043 | 0.090 | 35.853 |
| $500 | 0.043 | 0.090 | 33.834 |

Notes: This table supplements Tables 4 and 5 and displays the simulated welfare benefits of changing physician practice patterns under a range of calibration parameters. Each row represents a separate simulation exercise; bold rows indicate the baseline parameter values used for our main welfare analysis. The changes in net benefits (column 3) are reported in millions of dollars, compared to welfare under observed testing thresholds and misweighting. In any given row, all parameters aside from the one in question are kept constant at the values listed in Table 3. Panel A displays testing behavior and the improvement in social welfare under simulations assuming all physicians with thresholds below the calibrated optimum are reassigned to the optimal testing threshold of $\tau_d = \tau^*$ (but maintaining the observed degree of misweighting). Panel B displays testing behavior and the improvement in social welfare under simulations assuming that physicians target testing to patients with the highest expected probability of a positive test based on observable demographics and comorbidities (but maintaining the observed degree of over testing).