NBER WORKING PAPER SERIES

PHYSICIAN PRACTICE STYLE AND PATIENT HEALTH OUTCOMES: THE CASE OF HEART ATTACKS

Janet Currie
W. Bentley MacLeod
Jessica Van Parys

Working Paper 21218
http://www.nber.org/papers/w21218

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2015

We thank the Center for Health and Well being at Princeton University for financial support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Physician Practice Style and Patient Health Outcomes: The Case of Heart Attacks
Janet Currie, W. Bentley MacLeod, and Jessica Van Parys
NBER Working Paper No. 21218
May 2015
JEL No. I11

ABSTRACT

When a patient arrives at the Emergency Room with acute myocardial infarction (AMI), doctors must quickly decide whether the patient should be treated with clot-busting drugs, or with invasive surgery. Using Florida data on all such patients from 1992-2011, we decompose physician practice style into two components: The physician's probability of conducting invasive surgery on the average patient, and the responsiveness of the physician's choice of procedure to the patient's condition. We show that practice style is persistent over time and that physicians whose responsiveness deviates significantly from the norm in teaching hospitals have significantly worse patient outcomes, including a 7% higher probability of death in hospitals among the patients who are least appropriate for the procedure. Our results suggest that a reallocation of invasive procedures from less appropriate to more appropriate patients could improve patient outcomes without increasing costs. Developing protocols to identify more and less appropriate patients could be a first step towards realizing this improvement.

Janet Currie
Princeton University
316 Wallace Hall
Princeton, NJ 08544
and NBER
jcurrie@princeton.edu

W. Bentley MacLeod
Department of Economics
Columbia University
420 West 118th Street, MC 3308
New York, NY 10027
and NBER
bentley.macleod@columbia.edu

Jessica Van Parys
Department of Economics
Hunter College
jnv2106@columbia.edu

One of the most controversial issues in medicine is whether doctors should be evaluated in terms of their adherence to simple metrics. Such metrics have become increasingly popular as a way to improve the quality of health care. For example, under the Affordable Care Act, Accountable Care Organizations are judged on the basis of criteria including: The fraction of patients who receive influenza immunization, tobacco screening, and other forms of screening, as well as whether patients with coronary artery disease are taking appropriate medications, and other such metrics.

However, doctors often argue that decisions about treatment should be tailored to the needs of individual patients, and that this type of sensitivity to patient characteristics cannot be captured through adherence to simple rules. In a recent New York Times editorial cardiologist Sandeep Jauhar argued that "… guidelines and checklists are unpopular among most American physicians. Instead of being allowed to deliver "patient-centered" care, many physicians feel they are being co-opted by regulations… Guidelines are supposed to assist and advise. But all too often, recommended care in certain situations becomes mandated care in all situations." (New York Times, Dec. 11, 2014).

In this paper, we ask whether there are differences in the extent to which doctors tailor their treatment decisions to the needs of individual patients, and whether these differences matter for costs and health outcomes. Specifically, we use a rich data set of all Florida patients arriving at the Emergency Room (ER) with acute myocardial infarction (AMI), or heart attack, between 1992 and 2011. This is a large group of patients who are being treated on an emergency basis and thus have relatively little scope for selecting a physician, at least within the hospital. When such a patient arrives at the ER, clinicians will quickly decide if the patient will be treated with drugs (thrombolysis) or with surgery (angioplasty). The goal is either a door-to-drug time of within minutes or a door-to-balloon time of within 90 minutes (Zafari and Yang, 2014).

Using a rich set of observable characteristics including the patient's age and sex, diagnosis, and comorbidities, we show that there is substantial doctor-specific variation in the extent to which observable patient characteristics affect this choice. Some doctors are much less likely to use invasive surgical procedures on the oldest and sickest patients, while others appear to pay little attention to these factors. Patients of doctors who tailor their treatment decisions to observable patient characteristics have considerably better outcomes. These patients are less likely to die in the hospital, and they incur lower costs.

If doctors fail to respond to observable patient characteristics like age because they have other, more pertinent information about the patient's condition, then we would expect doctors who disregarded the information we observe to have better rather than worse patient outcomes (because they would be relying on superior private information instead of the public information). Instead, our results suggest that there are many doctors who disregard important information about individual patients when making treatment decisions, with potentially dire health consequences.

We also ask how the quality of the doctor's decision making varies with observable characteristics of the doctor, and over time. We find that there is considerable variation and persistence in practice style. Conditional on lagged measures of practice style, experience is associated with increases in the doctor's responsiveness to the patient's condition, and reductions in aggressiveness (defined as the propensity to conduct an invasive procedure on a patient of average appropriateness). Graduates of U.S. medical schools tend to be both more responsive and more aggressive than graduates of foreign medical schools. However, we do not find any effect of graduation from a top-20 U.S. school.

The rest of the paper is laid out as follows. Section 2 presents background information. Section 3 covers data and methods. Results are presented in Section 4, and section 5 concludes.

2. Background

Health care is an important area in which we all rely on experts to make judgments about our care. Hence, it is not surprising that many studies of expertise have focused on physicians. Meehl (1954) reviewed a number of studies, mainly of clinical psychologists, and compared their forecasts to those generated by simple statistical models, including optimal linear combinations of variables that the experts also observed. He argued that predictions based on these simple models were generally more accurate than those of the experts. A more recent meta-analysis of 136 studies in clinical psychology and medicine also found that algorithms tended to either out-perform or to match the experts (Grove et al., 2000).

Kahneman and Klein (2009) argue that algorithms are most useful when we have confidence in the list of variables to be used for prediction; when we have a reliable and measurable outcome; when there is a large body of similar cases; when the cost/benefit ratio warrants the investment in developing an algorithm; and when the situation is sufficiently stable that the algorithm will not immediately become obsolete. The case of emergency treatment of AMI patients appears to satisfy all of these criteria.

In the psychological studies discussed above, the experts and the statisticians generally had access to the same data. The advantage of the algorithms arises mainly because the algorithms are more consistent than the experts. In our application, we have only relatively crude data about the patient's condition (diagnosis codes, age, gender and other demographic variables, and comorbidities) relative to what a physician might observe. But we observe the entire universe of cases over a given time period, whereas each doctor sees only his or her own cases. What we show is that even with a limited amount of information about each individual patient, the administrative data allows us to characterize physician practice style in a way that is predictive of both costs and important patient health outcomes.

Another difference between our study and many of those in psychology is that we are agnostic about the source of any "errors" in the doctor's decision making. The psychology literature is concerned about whether the errors arise from factors such as over-confidence, or other heuristic biases. We are concerned with doctors who, for a variety of possible reasons, do not make the best use of the information at their disposal in order to make the best treatment decisions.

The literature in health economics offers many possible reasons for these "mistakes" (Chandra et al., 2012). One common explanation for faulty decision making is "defensive medicine," the idea that doctors perform unnecessary procedures in order to protect themselves from lawsuits. However, Baicker et al. (2007) argue that there is little connection between malpractice liability costs and physician treatment of Medicare patients.

There is a substantial literature arguing that doctors may also be swayed by financial incentives (e.g. Gruber, Kim, Mayzlin, 1999; Gruber and Owings, 1996). However, a recent national survey of general surgeons which used hypothetical clinical scenarios suggested that the decision to operate was largely independent of malpractice concerns and financial incentives (Naimey et al., 2013).

A third possibility is that doctors are influenced by the decisions of those around them. Chandra and Staiger (2007) study the choice of surgery vs. medical management of cardiac patients. Knowledge spillovers are the main theoretical driver of small area variation in procedure use in their model. Physicians in areas that specialize in surgery are assumed to become better at surgery and worse at medical management, and vice-versa. Their model raises the possibility of mismatch between patients and physicians. All patients in high surgery areas will be more likely to have surgery, even if medical management would be more appropriate for some of them.

The most important insight from the Chandra and Staiger (2007) model may be that a reduction in the use of surgery in high use areas cannot be Pareto improving because patients who are good candidates for surgery will be harmed by a decline in the skill level of the physicians serving them. Our work builds on this insight. What matters in our application is not only whether doctors have high or low average levels of invasive procedures, but also how well they tailor their decisions to the needs of the individual patient, a factor that has not been considered earlier. We will also argue that an across-the-board cut in invasive surgeries is unlikely to be optimal. Rather, what is desirable is a reallocation of invasive surgeries away from patients who don’t need them to patients who do need them.

In an interesting recent study, Doyle et al. (2010) suggest that some doctors may just be more competent than others. Specifically, they study a setting in which patients arriving at a large medical center are randomly assigned to doctors from one of two medical groups. One of the medical groups is affiliated with a prestigious medical school while the other is not. They find that the doctors from the better medical school systematically conduct fewer tests and have lower costs. However, they found no differences in patient outcomes. We will build on this work by examining the relationship between doctor's practice style and whether they attended a top 20 medical school (using the U.S. News and World Reports rankings).

Patient preferences are often cited as a fifth potential reason for medically unnecessary procedure use. In an innovative study using vignettes from patient and physician surveys, Cutler et al. (2013) assess the hypothesis that regional variations in procedure use are driven by differences in patient demand across areas. They conclude that patient demand is a relatively unimportant determinant of regional variations and that instead the main driver is physician beliefs about appropriate treatment that are often unsupported by clinical evidence.

Finkelstein et al. (2014) address the same question using longitudinal Medicare claims data that allow them to track the same patients as they move through different health care markets. They suggest that about half of the observed variation in procedure use is due to supply-side factors, while half is due to patient-level, or demand-side. However, they conclude that much of the variation in patient demand is driven by exogenous patient health, and so probably does not simply reflect patient tastes for procedures. These findings agree with those of Cutler et al. (2013) in suggesting that patient preferences play a relatively small role in explaining variations in care.

Because we are studying heart attack patients who were admitted through the ER, it is unlikely that patient demand is driving patient choice of either physician or procedures. Nevertheless, we will show below that there is no evidence that patients with different levels of appropriateness for invasive cardiac procedures are choosing physicians with higher or lower propensities to perform them.

Our main focus is on identifying doctors who, for whatever reason, are making poor use of the observable data about their patients when making treatment decisions. We will show that patients of these doctors tend to have worse outcomes than other comparable patients. The fact that these doctors can be identified using simple models based on administrative data is relevant for policy because it suggests that it would be possible to improve patient outcomes by incorporating aids to diagnostic decision making into standard practice. Tsai et al. (2010) ask whether the treatment of AMI is consistent with guidelines from the American Council of Cardiology, and concluded that compliance is "low to moderate" suggesting that there is a great deal of room for improvement. Our results suggest that algorithms may be one way to improve care, at least for common situations like emergency treatment of heart attack.

3. Data and Methods

Our analysis is based on hospital discharge data for all heart attack patients in Florida from 1992-1994 and 1997-2011 whose physicians had a valid license number in the Florida medical practitioner database.[1] We also restrict the sample to physicians who treated at least five AMI patients per three year period, since it will be difficult to determine anything about practice style in physicians who see very few patients. Table 1 shows that this restriction reduces the sample from 1,183,458 to 1,023,821 patients, but has a much greater impact on the number of physicians, reducing it from 23,828 to 11,798. It also reduces the number of hospitals from 289 to 249.

As discussed above, we further focus our attention on patients who are admitted from the ER. Some AMI patients arrive as referrals from other physicians or as transfers from other hospitals. In these cases, the referring physician or the transferring hospital may request a specific physician. Therefore, we estimate the results using the 70% of AMI patients who are admitted to the hospital through the emergency room. We believe that the results for this sample

[1] The hospitalization data come from the Florida Agency for Healthcare Administration (FL AHCA).

of patients are less likely to be biased from selection. Hence, our analysis sample has 756,924 patients who are treated by 11,598 attending physicians in 217 hospitals. Of these patients, 304,295 have an invasive procedure.[2] Appendix Table 2 shows the distribution of patients per attending physician.

These data have information about all patient hospitalizations over the time period, including patient characteristics, admission sources, procedures, length-of-stays, charges, discharge outcomes, and physician license numbers which can be matched to Florida's physician license database to obtain additional information about the physicians in our sample. The available patient characteristics include gender, age, race, ethnicity, insurance, and up to ten diagnostic codes.

From the hospital discharge data we create several additional variables. First, we use the diagnosis codes to define patient comorbidities. Following the literature (e.g. Card et al., 2009) we pay special attention to the serious comorbidities included in the Charlson index (Charlson et al. 1987), which include cardiac arrhythmia, hypertension, congestive heart failure, peripheral vascular disease, dementia, cerebral vascular disease, coronary obstructive pulmonary disease, lupus, ulcers, liver disease, cancer, diabetes, kidney disease, and HIV. Patients with these conditions are likely to be poorer candidates for invasive procedures than healthier patients.

We measure health outcomes using information about the patient's disposition at the time of discharge.[3] Many heart attack patients die in hospital. The patient may also be discharged to

---

[2] We identify patients who received angioplasty from ICD-9-CM codes beginning with "00.6" or "00.5". Patients who received any procedure in the ICD-9-CM chapter called, "Operations on the Cardiovascular System" with the exception of section 37.2 "Diagnostic Procedures on the Heart and Pericardium" and section 38.2 "Diagnostic Procedures on the Blood Vessels," receive a value equal to 1 to indicate that a non-diagnostic, invasive treatment was prescribed. Patients who received any angioplasty also receive a value equal to 1 to show that a non-diagnostic invasive treatment was prescribed.

[3] One drawback to using visit-level hospital discharge data is that the data do not contain patient identifiers. Therefore, we can only measure health outcomes that occur on the initial visit. We cannot

home, to another hospital, or to another facility such as a skilled nursing facility or hospice. Additionally, we know whether the patient developed a hospital-acquired condition (HAC). HACs are defined by the Department of Health and Human Services and include infections like septicemia, clostridium difficile, pneumonia due to staphylococcus, catheter-associated urinary tract infection, vascular catheter infection, and surgical site infections following surgery.[4]

Our second set of outcomes includes hospital costs, length-of-stay, and the number of procedures on the hospital record. The hospital discharge data contain hospital charges, which must be converted into hospital costs. To convert charges to costs, we multiply the hospital charge by the hospital's cost-to-charge ratio (CCR) in the given year.[5] Once we convert charges to costs, we can also separate hospital costs into categories such as pharmacy costs, laboratory costs, radiology costs, costs for medical devices, cardiology, operation rooms, and all other costs.[6] We use the length of stay and the number of surgical procedures on the patient record to measure the amount of medical resources that were used to treat the patients.

From the hospital discharge data we know which hospital the patient visited and which attending physician treated the patient. The attending physician is the physician who is legally

---

measure health outcomes that occur outside of the hospital, in another hospital, or at the time of a follow-up visit.

[4] Other types of HACs include foreign body retained after surgery, air embolism, pressure ulcer stages III and IV, falls and trauma, manifestations of poor glycemic control, deep vein thrombosis and pulmonary embolism following orthopedic procedures, and lactogenic pneumothorax with venous catheterization. The HHS definitions come from a Fact Sheet report titled, "Hospital Acquired Infections (HAC) in Acute Inpatient Prospective Payment (IPPS) Hospitals," which was published in October 2012. To the HHS list we add septicemia, clostridium difficile, and pneumonia due to staphylococcus, three common infections that can be deadly and that plague patients in hospital environments.

[5] We use the group cost-to-charge ratio because some individual hospital CCRs are missing. The group CCR assigns the same ratio to similar hospitals in the same geographic area.

[6] One drawback to using hospital discharge data is that the hospital costs are not the total costs for the patient's full episode of care. The hospital costs do not include physician fees or the costs of treating the patient at another facility. Since 12% of our sample gets transferred to another hospital, and all of our patients receive treatment from at least one physician on the initial visit, the differences between the hospital costs and the total costs for the episode are care are likely to be non-negligible. Therefore, we determine how diagnostic skill and surgical preference are correlated with hospital costs, but we discuss how the results could change if we had a more comprehensive measure of total costs.

responsible for the patient's care and is in charge of making all executive decisions to treat the patient. For this reason, we focus our attention on the attending physician. To learn more about the characteristics of attending physicians in our data, we match their medical license numbers to the Florida medical license database.[7] We construct variables for physician characteristics that include the physician's specialty, experience, gender, whether the physician is a medical doctor (ME) versus osteopathic physician (OS), whether the physician attended medical school in the United States, whether the physician's medical school is ranked among the top-20 according to U.S. New and World Report research rankings, and whether the physician speaks Spanish. One important contribution of this analysis is to show which, if any, observable physician characteristics are correlated with better physician decision making and patient outcomes.

3.1 Identifying Good Candidates for Invasive Procedures

The first step in our analysis is to identify patients who appear, given their observable characteristics, to be good candidates for invasive procedures. We do this using a standard simple "machine learning" algorithm which involves estimating a logit model for the use of the invasive cardiac procedure on the observable patient characteristics.[8] Specifically, for each year of data we estimate the following model for visit *i* in quarter *t*:

$$(1)\quad Pr(Invasive_{it} = 1) = F\,(\theta_{it,diag} + \Gamma X_{it} + \delta_{it,com} + \lambda_t + \varepsilon_{ct},)$$

Where $F(\,)$ is the logit function, $Pr(Invasive_{it} = 1)$ is the probability that the patient on visit *i* in quarter *t* receives an invasive procedure, $\theta_{it,diag}$ is a vector of 20 diagnosis codes for different types of acute myocardial infarction, X_{it} is the patient's gender and a vector of age dummies (50-

[7] Link to the Florida license database: http://ww2.doh.state.fl.us/IRM00PRAES/PRASLIST.ASP

[8] See An Introduction to Statistical Learning, by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani for a discussion of machine learning, with applications to medical decision making. The logistic model is one of the simplest (and robust) methods for modelling a binary decision. We will show that it can be used to constuct an elegant two dimensional representation of physician choices.

54,…65-69, 70, 71…89, 90+), $\delta_{i\ it,com}$ is a vector of 14 comorbidities, and λ_t is a vector of quarter fixed effects.

We estimate (1) using only patients in hospitals that have accredited teaching programs, which involves assuming that these hospitals define the standard of care in each year. Note that if we construct an alternative patient appropriateness index using data from all hospitals, the correlation between this alternative index and the one that we use here is 0.964. What this implies is that if called on to rank patients from least appropriate to most appropriate for surgery, most practitioners would rank them in the same way (even though they might well choose different cutoffs for deciding who would receive surgery).

We estimate the model year by year because if, for example, surgery becomes generally safer over time, then it may make sense to use aggressive procedures on more marginal patients over time. Thus, the standard of care may evolve over time; in fact we show that there is a general trend towards using more invasive procedures in older, sicker patients over time.

These estimates are shown for several different years in Appendix Table 1. The estimates suggest that the weights that doctors place on several important determinants of appropriateness for invasive procedures have evolved over time. For example, invasive procedures are less likely to be performed on female patients, and this pattern has become stronger over time. Similarly, doctors have become less likely to perform invasive surgery on patients with congestive heart failure, dementia, cancer, and several other co-morbidities. On the other hand, doctors have become more likely to perform invasive surgeries on both relatively young patients and on relatively old patients over time. These trends show the importance of considering doctor's decisions in the context of accepted practice for the time.

We use the estimated parameters from equation (1) to construct a predicted probability of receiving an invasive procedure for each patient. In principle, this index can vary from negative infinity to positive infinity. Given this index, we can divide patients into quartiles according to their appropriateness for invasive procedures. Tables 2, 3, and 4 show means for all patients, those in the lowest quartile of appropriateness for invasive surgery (low), and those in the highest quartile of appropriateness for invasive surgery (high). Table 2 shows the mean characteristics of the patients themselves. We can see that patients who are good candidates for surgery are less likely to be female, younger (and therefore less likely to have Medicare coverage and more likely to have private health insurance coverage), and less likely to have other serious comorbid conditions.

Table 3 shows what happened to these patients. Overall, 40% have an invasive procedure, but as one might expect, this fraction rises to 67% among the high appropriateness patients and is only 18% among the least appropriate patients. The high appropriateness patients have more procedures but slightly lower lengths of stay. However, they have much higher costs in all of the categories that we can measure. In general outcomes appear somewhat better for this group in that they are less likely to have hospital acquired conditions, more likely to be discharged home, and less likely to die in the hospital.

Table 4 shows the characteristics of doctors who treat these patients. The unit of observation is still the patient. The table shows that there is little difference between “high” and “low” appropriateness patients in terms of the type of medical school attended or gender of the physician. However, “high” patients are more likely to see cardiologists, as well as doctors who have slightly less experience. Recall that the high appropriateness patients tend to be younger, less likely to have co-morbidities, and more likely to have private insurance.   Thus these differences do not necessarily indicate that patients with high and low appropriateness for surgery are seeing doctors with different practice styles (since differences in practice style have to do with treating patients with the same characteristics differently rather than just tending to see different types of patients). In order to investigate this issue, it is first necessary to try to measure practice style.

3.2 Measuring Practice Style

The next step in our analysis involves the estimation of doctor-specific regressions which show how the doctor's decision to use invasive cardiac procedures varies with the index we constructed above, which summarizes all of the information that we observe about the patient's medical condition. These regressions tell us whether the doctor in question is more or less likely to use invasive procedures relative to the state-wide standard defined in equation (1). Rather than assuming that a doctor's behavior is constant over time, we allow it to evolve with physician experience, measured as the number of years since the physician completed his or her residency. We create 3-year experience level bins, where k =1 if the physician has less than 3 years of experience, k=2 indicates 3-6 years of experience, and so on, up to more than 30 years of experience (k=11). Specifically, we estimate a logit model for each physician j in experience level k,

$$Pr(Invasive_{ijk} = 1) = F(\alpha_{jk} + \beta_{jk}*Patient\_Index_i + \varepsilon_{ijk}) \qquad \text{for each } j*k=1,\ldots, JK \quad (2)$$

Where again, $F(\ )$ indicates the logit function. Here α_{jk} captures physician j's propensity to perform an invasive treatment on the median patient given experience level k, and β_{jk} captures the relative weight that physician j places on the index summarizing the patient i's appropriateness for invasive treatment. Because we estimate separate equations for the same physician at different levels of experience, α_{jk} and β_{jk} vary within physicians over time. This allows us to test whether physician are more or less likely to consider observable patient characteristics as they gain experience. In what follows we will refer to these two dimensions of doctor behavior as aggressiveness and responsiveness.

The parameters α_{jk} and β_{jk} can take values from negative infinity to positive infinity, but there are some special cases that illustrate the intuition behind our model. When $\beta_{jk}=0$, physicians ignore patient characteristics and have the same probability of performing an invasive treatment on all patients. One way to characterize this behavior is that the doctor has a particular preference for invasive procedures that is independent of patient characteristics. When $\alpha_{jk}=0$, the doctor's behavior depends only on the patient's appropriateness for invasive procedures. If in addition $\beta_{jk}=1$, then physicians behave in exactly the way predicted by our equation (1) model. The coefficient α_{jk} can also be characterized as the probability that a doctor will perform an invasive procedure on a patient with an index of patient appropriateness equal to zero, that is, on a patient of median appropriateness.

We are not the first to try to measure physician practice style (see for example, Epstein and Nicholson, 2009). However, practice style is usually modeled as a physician-specific fixed effect. Instead, we allow there to be two dimensions to practice style. Moreover, we allow practice style to evolve over time. We will show that this richer model is useful in conceptualizing practice style.

Table 5a shows the distribution of the estimated parameters α and β in our sample. This table does not take the precision of the coefficient estimates into account. Therefore, Table 5b divides the data into cells according to whether the physician's behavior shows a statistically significant departure from the norm. That is, we ask whether α is significantly different from zero, and whether β is significantly different than one.   Not surprisingly then, the largest cell in Table 5b is the middle cell where $\alpha=0$ and $\beta=1$.   These are physicians whose behavior is consistent with the state-wide norm established by the accredited teaching hospitals for that year. However, Table 5b shows that 21.6% of patients who have a physician who is less responsive to patient characteristics than the norm (i.e. β<1), and 42.3% of patients have a physician who is less aggressive than the norm established in the teaching hospitals (i.e. α<0). Only 7.94% of patients have a doctor who is significantly more aggressive than that norm (i.e. α>0), and very few (1.4%) have a physician that is more responsive to patient characteristics than the norm (i.e. have β>1)

We can now return to the question of whether patients whose medical conditions make them good candidates for invasive procedures go to doctors with systematically different practice styles than patients who are poor candidates for such surgery. Table 5c shows the results of taking the estimated α's and β's and regressing them on year and hospital effects as well as the patient's gender, age, comorbidities, and an indicator equal to one if the patient had a previous AMI.[9] Table 5c shows that there are no differences in the mean residuals between the high appropriateness and low appropriateness patients, which suggests that patients are not being selectively matched to physicians with different practice styles.

3.3. Models of Costs, Patient Outcomes, and Correlates of Physician Practice Style

Given the data we have constructed, we can now ask how the variations in physician treatment style affect costs and patient outcomes, and also what characteristics of physicians are associated with differences in practice style. We have defined a standard of care using data from teaching hospitals. This standard of care has two dimensions which can be thought of as the

[9] As discussed above, we cannot follow patients over time, but the diagnosis for each AMI patient indicates whether their has been a previous AMI.

average level of use of invasive procedures (α) and the extent to which the physician responds to the observable condition of the patient (given largely by their age and comorbidities) when deciding whether to perform an invasive procedure (β). If the standard is useful then we will find that patients of physicians who deviate from the standard will have poorer outcomes. For example, physicians who are aggressive will generate worse health outcomes for patients who are inappropriate candidates for invasive procedures. The reason is that aggressive physicians will prescribe invasive treatments even for patients who are poor candidates for these treatments. As a result, these patients may be more likely to have complications from surgery resulting in outcomes such as hospital-acquired conditions or in-hospital mortality.

In order to measure the effect of physician characteristics on outcomes, we estimate models of the form:

$$Y_{ijkht} = \phi_1 * LowResponsiveness_{ijkt} + \phi_2 * LowAggressiveness_{ijkt} + \phi_3 * HighAggressiveness_{ijkt}$$

$$+ \Pi Z_i + \Omega X_i + \delta_h + \lambda_t + \varepsilon_{ijkt}, \qquad (3)$$

where *LowResponsiveness* corresponds to an estimated physician β that is significantly less than one; *LowAggressiveness* corresponds to an estimated physician α that is significantly less than zero, and *HighAggressiveness* corresponds to an estimated physician α that is significantly greater than zero. For the time being, we ignore the possibility that β is significantly greater than one since it is so rare in our data. However, below we present models that also include an indicator for this possibility. These measures are specific to patient *i* of physician *j* with experience level *k* in year-quarter *t*.

The outcomes Y_{ijkht} include health outcomes, total hospital costs, itemized costs, length-of-stay, and the number of surgical procedures. Health outcomes include the probability of developing a hospital-acquired condition (HAC), the probability of being discharged home (Home), and the probability of dying in the hospital (Died). The subscript *h* indicates that outcomes may also vary at the hospital level. The cost variables are in logs, while the number of invasive procedures and the length of stay variables are in levels.

The vector Z_i includes other observable characteristics of physicians, including experience, specialty, gender, whether he or she attended medical school in the US or abroad, whether the physician attended a top-20 medical school, and whether the physician is a medical doctor versus osteopathic physician. The vector X_i includes gender, age, the patient's propensity index, and the patient's comorbidities. We also estimate alternative models adding in patient race, ethnicity, and type of health insurance which may have independent effects on treatment choices and on outcomes. We control for hospital fixed effects (δ_h) in order to be sure that we are capturing physician-level differences, and for a separate indicator for each combination of year and quarter in the sample (λ_t, i.e. fall 2000) in order to capture things like technological improvements that may improve survival over time.

Finally, we examine the way that the estimated parameters of physician practice style vary with physician characteristics, and with their experience. In order to conduct this examination we estimate:

$$\alpha_{jkt} = \phi_1 * \alpha_{jkt-1} + \Pi Z_{it} + \delta_h + \lambda_t + \varepsilon_{jkt}, \quad (4a)$$

$$\beta_{jkt} = \phi_2 * \beta_{jkt-1} + \Pi Z_{it} + \delta_h + \lambda_t + \varepsilon_{jkt.} \quad (4b)$$

where the estimated parameters at time t depend on physician characteristics including experience as well as their lagged values, hospital effects, and year-quarter effects.

4. Estimation Results

Estimates of equation (3) are shown in Table 6. Table 6 suggests that patients of physicians who are not responsive to observable patient characteristics (here patient age and comorbidities) tend to have worse outcomes. Among patients who are not good candidates for invasive surgery, low responsiveness predicts more hospital acquired infections, a higher probability of dying in hospital, and a lower probability of being discharged home. For example, the point estimate of 0.013 on dying in hospital corresponds to an increase of 8.7% in the probability of this outcome in this group of patients. Among patients who are good candidates for invasive surgery, low responsiveness predicts a significantly lower probability of being discharged home: The point estimate of -0.046 corresponds to a reduction of 7% in this probability.

Turning to the aggressiveness of the physicians, Table 6 shows that patients of physicians who are significantly less aggressive than the standard tend to be less likely to die in hospital on the index visit, but also less likely to be discharged home. In fact, these patients are more likely to be transferred to another facility. Patients who need invasive procedures are also less likely to be discharged home because they are being transferred to other facilities. In contrast, patients who are good candidates for aggressive procedures are more likely to be discharged home when they have a highly aggressive physician (i.e. one who is more likely than the average to perform invasive procedures regardless of patient condition).

Appendix Table 3 shows all of the coefficient estimates on the included covariates. These estimates largely show the expected patterns. Older patients and those with comorbidities are less likely to have positive outcomes. It is also notable (given that so many people are treated by non-cardiologists) that cardiologists tend to have better outcomes across the board.

Table 7 shows the way that these differences in practice style affect the probability of having invasive procedures, the number of invasive procedures, costs, and length of stay. The estimates in the first row suggest that when physicians are not responsive to patient conditions, they are more likely to perform invasive procedures on patients who are not appropriate for such procedures. They also perform more procedures. Both tendencies increase costs and length of stay for these patients. For example, these less responsive physicians have total costs that are 11% higher. Conversely, these "low responsiveness" physicians are less likely to perform invasive procedures on the highly appropriate patients, so their patients receive fewer procedures, have lower costs, and shorter lengths of stay. "Low aggressiveness" corresponds to fewer procedures and costs across the board, while "high aggressiveness" implies the reverse. Table 8 shows that the same patterns that apply to total costs are reflected in virtually every category of costs.

As discussed above, we only observe hospital costs. In most cases, there will also be costs associated with both the treating physicians and anesthetists. Since these costs are likely to rise with the number of procedures, the comparisons above are likely to understate the extent to which additional procedures drive up costs.

4.1 Correlates of Practice Style

Table 9 shows how the two dimensions of practice style that we have identified (responsiveness and aggressiveness) vary with other physician characteristics. Because we allow practice style to evolve over time, we can also ask how it varies with experience (measured as years since residency). The first and fourth columns show models without lagged practice style measures. In these models, physician specialty is the main determinant of practice style with cardiologists being significantly less responsive and more aggressive than other doctors.

The remaining columns do control for lagged practice style. The estimates on the lags suggest that practice style is quite persistent over time, although it does evolve. This is an interesting observation given past work asking whether physicians are "punished" for aggressive practice styles (Dranove et al., 2011). Apparently they are not punished enough to change rapidly.

Conditional on lagged practice style, cardiologists are somewhat more responsive and more aggressive than other physicians. Also, greater experience is associated with more responsiveness and less aggressiveness. Graduating from a U.S. medical school is associated with both more responsiveness and more aggressiveness, but there is no significant effect of having graduated from a "top 20" medical school.

4.2 Robustness

Appendix Table 4 shows estimates of a model that controls for combinations of aggressiveness and responsiveness. Table 5b suggested that very few patients saw doctors with an estimated β significantly greater than one. Moreover, relatively few patients had doctors with β significantly less than one and α significantly greater than one. Therefore we collapse this category together with β less than one and α equal to zero and consider only four alternatives to the baseline case in which β=1 and α=0. These alternatives are: α<0 and β>=1, α>0 and β>=1, α<0 and β<1, and α>=0 and β<1.

Appendix Table 4 shows that the two low responsiveness conditions are associated with a higher probability of hospital acquired infections in the "low appropriateness" group. Patients of doctors with typical aggressiveness and low responsiveness also have a higher probability of death in the hospital in the low appropriateness group of patients.   The two low responsiveness conditions are associated with lower probabilities of being discharged home for *both* high and low appropriateness patients. Finally, low aggressiveness with typical responsiveness is associated with a lower probability of dying in hospital among the low appropriateness group, but also a lower probability of being discharged home for both groups of patients considered reflecting the fact that the patients are being transferred elsewhere. Overall then, deviations from the norm established in the teaching hospitals are associated with worse outcomes, especially among the subgroup of patients who are the least appropriate for invasive procedures. These results are consistent with those of Table 6.

Appendix Table 5 shows models similar to those in Table 6 except that they also include the patients' race, ethnicity, and type of insurance. These are variables that might be predictive of patient outcomes, and might also be correlated with the type of treatment they receive. However, including these variables has virtually no effect on the estimated coefficients on the measures of practice style.

5. Conclusions

In this paper we use the behavior of doctors in accredited teaching hospitals to predict which heart attack patients arriving in the ER are more or less appropriate candidates for invasive surgery.   This prediction is made using the patient's diagnosis, demographic information, and co-morbidities. Given these predictions we can identify doctors whose behavior deviates significantly from the norm defined in the teaching hospitals. We also show that practice style is quite persistent and relatively insensitive to measures of "quality" such as the ranking of the medical school attended.

We show that patients of doctors who are less responsive than the norm to patient characteristics have significantly worse outcomes. In particular, patients who are not good candidates for invasive surgery are more likely to get hospital-acquired conditions and to die in hospital if they are treated by a less responsive physician.

The effects on costs are more ambiguous. A physician who is not responsive to the patient's condition is more likely to perform invasive treatments when it is inappropriate (thereby incurring higher costs for the "low appropriateness" patients) but is also less likely to perform such surgery when it is appropriate (thereby incurring lower costs for "high appropriateness" patients). These results suggest that it would be possible to improve outcomes holding cost constant by reallocating procedures across patients. Thus, they contribute to the large literature demonstrating inefficiencies in the health sector (Garber and Skinner, 2008).

Our results suggest that simple machine learning protocols can indeed be used to identify doctors whose behavior deviates systematically from accepted norms, and that using such protocols to change physician behavior could have positive health consequences. Moreover, our protocol uses only the limited information available to us in standard hospital discharge data. It is highly likely that more sensitive protocols could be developed using data that is routinely collected (such as patient histories) but which we could not access. Hence, the development of treatment protocols that are more sophisticated than the simple checklists now in vogue (because they account for multiple dimensions of patient condition) offers a potential way to improve the cost-effectiveness of medical care and patient outcomes.

Finally, we do not mean to suggest that protocols are a substitute for the diagnostic capabilities of a skilled physician. Providers that have successfully implemented protocols often use them to indicate reasonable treatment choices, and allow physicians to over-rule the protocols when they have reason to do so. This is a model that has been successfully demonstrated in individual hospitals such as those operated by the Intermountain system in Utah (James and Savitz, 2011; Morris et al. 2008). Our results suggest that it is a model that deserves wider consideration.

References

Card, David, Carlos Dobkin and Nicole Maestas. “Does Medicare Save Lives?” Quarterly Journal of Economics, 124(2): 597-636, 2009.

Charlson, Mary E.; Pompei, Peter; Ales, Kathy L.; MacKenzie, C. Ronald. "A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation". Journal of Chronic Diseases 40 (5): 373–83, 1987.

Doyle, Joseph J., Steven M. Ewer, Todd H. Wagner. “Returns to physician human capital: Evidence from patients randomized to physician teams,” Journal of Health Economics, 29 2010, 866-882

Baicker, K., E. S. Fisher, and A. Chandra (2007, MAY-JUN). Malpractice liability costs and the practice of medicine in the medicare program. Health Affairs 26 (3), 841–852.

Chandra, A., D. Cutler, and Z. Song (2012). Chapter six - who ordered that? The economics of treatment choices in medical care. In T. G. M. Mark V. Pauly and P. P. Barros (Eds.), Handbook of Health Economics, volume 2 of Handbook of Health Economics, pp. 397 – 432. Elsevier.

Chandra, A. and D. O. Staiger (2007). Productivity spillovers in health care: Evidence from the treatment of heart attacks. Journal of Political Economy 115 (1), pp. 103–140.

Currie, J. and W. B. MacLeod (2008, May). First do no harm? tort reform and birth outcomes. Quarterly Journal of Economics 123 (2), 795–830.

Cutler, D., J. Skinner, A. D. Stern, and D. Wennberg (2013, August). Physician beliefs and patient preferences: A new look at regional variation in health care spending. Technical Report 19320, NBER.

Dranove, D., S. Ramanarayanan, and A. Sfekas (2011). Does the market punish aggressive experts? Evidence from cesarean sections. B E Journal oOf Economic Analysis & Policy 11 (2).

Epstein, A. J. and S. Nicholson (2009). The formation and evolution of physician treatment styles: An application to cesarean sections. Journal of Health Economics 28, 1126–1140.

Finkelstein, A., M. Gentzkow, and H.Williams (2014, July). Sources of geographic variation in health care: Evidence from patient migration. Technical report, University of Chicago.

Garber, A. M. and J. Skinner (2008). Is american health care uniquely inefficient? The Journal of Economic Perspectives 22 (4), pp. 27–50.

Gawande, A. (2009). The Checklist Manifesto: Getting Things Right. New York: Picador.

Gaynor, M., J. B. Rebitzer, and L. J. Taylor (2004, Aug). Physician incentives in health maintenance organizations. Journal of Political Economy 112 (4), 915–931. English.
30

Grove, W., D. H. Zald, B. S. Lebow, B. E. Snitz, and C. Nelson (2000). Clinical versus mechanical prediction: A meta-analysis. Psychological Assessment 12, 19–30.

Gruber, J., J. Kim, and D. Mayzlin (1999, AUG). Physician fees and procedure intensity: the case of cesarean delivery. Journal of Health Economics 18 (4), 473–490.

Gruber, J. and M. Owings (1996). Physician financial incentives and cesarean section delivery. The RAND Journal of Economics 27 (1), pp. 99–123.

James, Brent, Savitz, Lucy. "How Intermountain Trimmed Health Care Costs Through Robust Quality Improvement Efforts," Health Affairs, June 2011, vol. 30 no. 6 1185-1191.

James, Gareth, Witten, Daniela,Hastie, Trevor, Tibshirani, Robert. An Introduction to Statistical Learning with Applications in R (New York: Springer Science) 2013.

Kahneman, D. and G. Klein (2009). Conditions for intuitive expertise a failure to disagree. American Psychologist 64 (6), 515–526.

Kessler, D. and M. McClellan (1996). Do doctors practice defensive medicine? Quarterly Journal of Economics 111 (2), 353–90.

Meehl, P. E. (1954). Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence. University of Minnesota Press.

Morris, Alan, Orme, James, Rocha, Beatriz, Holmen, John, et al. "An Electronic Protocol for Translation of Research Results to Clinical Practice: A Preliminary Report," Journal of Diabetes Science and Technology," September 2008 vol. 2 no. 5 802-808.

Rathore SS, Curtis JP, Chen J, Wang Y, Nallamothu BK, Epstein AJ, et al. "Association of door-to-balloon time and mortality in patients admitted to hospital with ST elevation myocardial infarction: national cohort study," British Medical Journal. May 19 2009;338:b1807.

Tsai CL, Magid DJ, Sullivan AF, Gordon JA, Kaushal R, Michael Ho P, et al. Quality of care for acute myocardial infarction in 58 U.S. emergency departments. Academic Emergency Medicine. Sep 2010;17(9):940-50.

Wilson, Niamey P., Francis Wilson, Mark Neuman, Adnrew Epstein, et al. "Determinants of Surgical Decision Making: A National Survey," The American Journal of Surgery, 206, 2013, 970-978.

Zafari, A. Maziar, Yang, Eric. "Myocardial Infarction Treatment and Management," Dec. 1, 2014. http://emedicine.medscape.com/article/155919-treatment#aw2aab6b6b3

Table 1: Derivation of Analysis Sample

Diagnosis Name & ICD-9-CM Code	# Patients	# Attendings	# Hospitals
All Patients			
AMI Codes beginning with "410"	1,183,458	23,828	289
Angioplasty or other surgery	549,420	18,509	258
Restricting to physicians who treat >= 5 AMI patients in a 3 year period			
All Patients	1,023,821	11,798	249
Angioplasty or other surgery	481,935	11,234	235
Further Restricting to patients admitted from the ER (Analysis Sample)			
Patients Admitted from ER with "410"	756,924	11,598	217
Admitted from ER with surgery	304,295	10,805	209

Note: We restrict the sample to physicians whose license numbers on the patient records match the license numbers in Florida's medical practicioner database. We also restrict the sample to physicians who treat at least 5 AMI patients per 3-year period.

Table 2: Mean Patient Characteristics

Appropriateness for Surgery	All	Low	High
Female	0.42	0.59	0.25
Age	70.90	83.71	61.24
White	0.79	0.85	0.75
Black	0.08	0.05	0.10
Hispanic	0.10	0.09	0.11
Medicaid	0.04	0.01	0.06
Medicare	0.67	0.90	0.43
Private Insurance	0.20	0.07	0.34
Self Pay or Other	0.08	0.02	0.16
Morbidity Index	0.29	-0.98	1.52
Subsequent AMI	0.05	0.09	0.01
#Diagnoses	8.09	8.56	7.75
Arrhythmia	0.27	0.34	0.20
Hypertension	0.44	0.44	0.41
Congestive Heart Failure	0.34	0.58	0.11
Peripheral Vascular Disease	0.05	0.05	0.03
Dementia	0.04	0.14	0.00
Cerebral Vascular Disease	0.07	0.10	0.04
COPD	0.18	0.27	0.06
Lupus	0.02	0.04	0.01
Ulcer	0.01	0.01	0.00
Liver Disease	0.02	0.01	0.02
Cancer	0.07	0.12	0.02
Diabetes	0.22	0.24	0.15
Hemiplegia	0.00	0.00	0.00
Kidney Disease	0.15	0.09	0.23
N	756,924	205,419	186,687

Table 3: Mean Procedure and Outcome Rates

Appropriateness for Surgery:	All	Low	High
Any invasive procedure	0.40	0.18	0.67
#Invasive Procedures	1.53	0.66	2.81
Length of Stay	6.51	6.67	6.22
Total Costs	14610	9439	20920
Pharmacy Costs	2355	1505	3336
Laboratory Costs	2051	1674	2465
Radiology Costs	867	731	949
Medical Devices Costs	1923	949	3224
Cardiology Costs	2402	867	4558
Operating Room Costs	645	229	1098
Other Costs	4367	3482	5290
Hospital-Acquired Conditions	0.15	0.19	0.12
Discharged to Home	0.55	0.45	0.66
Died in the Hospital	0.11	0.15	0.09
N	756,924	205,419	186,687

Table 4: Physician Characteristics

Appropriateness for Surgery:	All	Low	High
Cardiologist	0.24	0.17	0.30
Internal Medicine	0.59	0.63	0.56
Family Practice or Other	0.17	0.19	0.14
Experience (yrs since residency)	12.46	13.20	12.24
Medical Doctor (MD vs. OS)	0.93	0.93	0.93
US Medical School	0.45	0.46	0.44
Top-20 Medical School	0.06	0.06	0.06
Female Physician	0.12	0.12	0.13
Spanish-Speaking Physician	0.27	0.27	0.26
N	756,924	205,419	186,687

Table 5a: Distributions of Physician Responsiveness (Beta) and Aggressiveness (Alpha) Across Patients

Patient Percentile	Responsiveness (Beta)	Aggressiveness (Alpha)
1%	-1.28	-3.47
5%	-0.47	-2.49
10%	-0.17	-2.04
25%	0.30	-1.26
50%	0.71	-0.55
75%	1.06	0.02
90%	1.47	0.59
95%	1.81	1.08
99%	3.01	3.44
N	756924	756924

Note: For each physician, we estimate a model for every three years of our sample period and obtain an estimate of alpha and beta.

Table 5b: Fraction of Estimated Physician Coefficients that are Significantly Different than Beta=1 and Alpha=0

	Beta<1	Beta=1	Beta>1
Alpha<0	0.129	0.286	0.008
Alpha=0	0.068	0.424	0.006
Alpha>0	0.019	0.060	0.0004
Total	0.216	0.77	0.0144

N=756924 patients.

Table 5c: Physician Practice Style Conditional on Hospital, Year and Patient Gender, Age, Comorbidities, and Previous AMI

Appropriateness:	All	Low	High
Mean Residual Beta*	8.96E-13	0.0016	0.0018
Mean Residual Alpha*	-1.15E-11	-0.0221	0.0538
N	756,924	205,419	186,687

Table 6: Outcomes Associated with Physician Responsiveness and Aggressiveness for Patients with High and Low Appropriateness for Invasive Procedures.

	(1)	(2)	(3)	(4)	(5)	(6)
Appropriateness for Invasive Procedure:	**Low**	**High**	**Low**	**High**	**Low**	**High**
Outcome:	**Hosp. Aquired Infection**	**Hosp. Aquired Infection**	**Died in Hospital**	**Died in Hospital**	**Discharged to Home**	**Discharged to Home**
Low Responsiveness	0.018***	-0.002	0.013***	0.003	-0.020***	-0.046***
(Beta<1)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
Low Aggressiveness	-0.003	-0.001	-0.012***	-0.003	-0.010***	-0.050***
(Alpha<0)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
High Aggressiveness	-0.000	0.002	-0.001	-0.004	0.009	0.015**
(Alpha>0)	(0.005)	(0.003)	(0.005)	(0.002)	(0.007)	(0.005)
Quarter since sample start	Y	Y	Y	Y	Y	Y
Hospital FE	Y	Y	Y	Y	Y	Y
Patient Appropriateness Index	Y	Y	Y	Y	Y	Y
Patient Age Categories & Gende	Y	Y	Y	Y	Y	Y
Patient Comorbidities	Y	Y	Y	Y	Y	Y
Physician Characteristics	Y	Y	Y	Y	Y	Y
N	205419	186687	205419	186687	205419	186687
R^2	0.06	0.13	0.05	0.12	0.10	0.31

Notes: Standard errors are clustered at the physician level and shown in parentheses. * indicates p<0.05, ** indicates p< 0.01, *** indicates p<0.001. Alphas and Betas vary with each 3 years of physician experience.

"Low appropriateness" indicates patient is below the 25th percentile of our index of appropriateness for invasive procedures. "High appropriateness" indicates patient is above the 75th percentile.

Table 7: Physician Responsiveness and Aggressiveness, Procedure Use, Log(Costs), and Length of Stay for Patients with High and Low Appropriateness for Invasive Procedures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Appropriateness for Invasive Procedure:	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**
Outcome:	**Any Invasive Procedure**	**Any Invasive Procedure**	**# Invasive Procedures**	**# Invasive Procedures**	**Total Costs**	**Total Costs**	**Length of Stay**	**Length of Stay**
Low Responsiveness	0.10***	-0.09***	0.24***	-0.46***	0.11***	-0.13***	0.52***	-0.36***
(Beta<0)	(0.00)	(0.00)	(0.01)	(0.02)	(0.01)	(0.01)	(0.04)	(0.07)
Low Aggressiveness	-0.08***	-0.07***	-0.21***	-0.31***	-0.10***	-0.10***	-0.43***	-0.31***
(Alpha<0)	(0.00)	(0.00)	(0.01)	(0.02)	(0.00)	(0.01)	(0.04)	(0.06)
High Aggressiveness	0.20***	0.07***	0.73***	0.47***	0.32***	0.20***	1.00***	0.99***
(Alpha>0)	(0.01)	(0.00)	(0.05)	(0.03)	(0.02)	(0.02)	(0.14)	(0.15)
Quarter since sample start	Y	Y	Y	Y	Y	Y	Y	Y
Hospital FE	Y	Y	Y	Y	Y	Y	Y	Y
Patient Appropriateness Index	Y	Y	Y	Y	Y	Y	Y	Y
Patient Age Categories & Gender	Y	Y	Y	Y	Y	Y	Y	Y
Patient Comorbidities	Y	Y	Y	Y	Y	Y	Y	Y
Physician Characteristics	Y	Y	Y	Y	Y	Y	Y	Y
N	205419	186687	205419	186687	205419	186687	205419	186687
R^2	0.09	0.21	0.11	0.32	0.27	0.34	0.09	0.14

Notes: Standard errors are clustered at the physician level and shown in parentheses. * indicates p<0.05, ** indicates p< 0.01, *** indicates p<0.001. Alphas and Betas vary with each 3 years of physician experience. "Low appropriateness" indicates patient is below the 25th percentile of our index of appropriateness for invasive procedures. "High appropriateness" indicates patient is above the 75th percentile.

Table 8: Physician Responsiveness and Aggressiveness, Procedure Use and Log(Costs) for Patients with High and Low Appropriateness for Invasive Procedures - Breakdown of Costs by Type

Appropriateness for Invasive Procedure:	(1) Low	(2) High	(3) Low	(4) High	(5) Low	(6) High
Type of Cost:	**Pharmacy**	**Pharmacy**	**Laboratory**	**Laboratory**	**Radiology**	**Radiology**
Low Responsiveness	0.14***	-0.10***	0.06***	-0.06***	0.09***	-0.04***
(Beta<0)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Low Aggressiveness	-0.10***	-0.07***	-0.06***	-0.04***	-0.06***	-0.01
(Alpha<0)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
High Aggressiveness	0.32***	0.20***	0.16***	0.16***	0.16***	0.15***
(Alpha>0)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)
R^2	0.20	0.25	0.42	0.44	0.23	0.31
Type of Cost:	**Medical Devices**	**Medical Devices**	**Cardiology**	**Cardiology**	**Operating Room**	**Operating Room**
Low Responsiveness	0.20***	-0.31***	0.06***	-0.16***	0.08***	-0.24***
(Beta<0)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
Low Aggressiveness	-0.18***	-0.24***	-0.11***	-0.13***	-0.19***	-0.06*
(Alpha<0)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
High Aggressiveness	0.59***	0.32***	0.43***	0.09***	0.69***	0.76***
(Alpha>0)	(0.04)	(0.02)	(0.02)	(0.01)	(0.06)	(0.06)
R^2	0.16	0.32	0.27	0.47	0.23	0.21
Quarter since sample start	Y	Y	Y	Y	Y	Y
Hospital FE	Y	Y	Y	Y	Y	Y
Patient Appropriateness Index	Y	Y	Y	Y	Y	Y
Patient Age Categories & Gende	Y	Y	Y	Y	Y	Y
Patient Comorbidities	Y	Y	Y	Y	Y	Y
Physician Characteristics	Y	Y	Y	Y	Y	Y

Notes: Standard errors are clustered at the physician level and shown in parentheses. * indicates p<0.05, ** indicates p< 0.01, *** indicates p<0.001. Alphas and Betas vary with each 3 years of physician experience.
"Low appropriateness" indicates patient is below the 25th percentile of our index of appropriateness for invasive procedures. "High appropriateness" indicates patient is above the 75th percentile.

Table 9: Relationship between Physician Aggressiveness and Responsiveness and Other Physician Characteristics

	(1)	(2)	(3)	(5)	(6)	(7)
Physician Characteristics	**Responsiveness**	**Responsiveness**	**Responsiveness**	**Aggressiveness**	**Aggressiveness**	**Aggressiveness**
Responsiveness (t-1)		0.7190***	0.7205***			-0.0555***
		(0.0063)	(0.0062)			(0.0043)
Aggressiveness (t-1)			-0.1088***		0.8530***	0.8541***
			(0.0108)		(0.0098)	(0.0098)
Experience	-0.0072***	0.0074***	0.0043**	-0.0228***	0.0046***	0.0034***
	(0.0021)	(0.0014)	(0.0015)	(0.0033)	(0.0010)	(0.0010)
Cardiologist	0.0007	0.0062	0.0693***	0.5921***	0.1148***	0.1142***
	(0.0161)	(0.0086)	(0.0095)	(0.0301)	(0.0070)	(0.0070)
Family Practice	-0.0297	0.0347***	0.0238*	-0.0780***	0.0093	0.0043
	(0.0172)	(0.0101)	(0.0100)	(0.0175)	(0.0066)	(0.0066)
Other Specialty	-0.0817***	-0.0036	0.0203	0.1722***	0.0071	0.0003
	(0.0243)	(0.0150)	(0.0176)	(0.0450)	(0.0101)	(0.0098)
Medical Doctor	0.0199	0.0116	0.0316*	0.1656***	0.0188*	0.0190*
	(0.0214)	(0.0128)	(0.0127)	(0.0266)	(0.0089)	(0.0088)
US Medical School	0.0164	0.0263***	0.0402***	0.1339***	0.0328***	0.0318***
	(0.0134)	(0.0075)	(0.0076)	(0.0208)	(0.0057)	(0.0057)
Top 20 Medical School	0.0135	-0.0040	-0.0034	-0.0040	0.0073	0.0074
	(0.0243)	(0.0143)	(0.0160)	(0.0515)	(0.0128)	(0.0121)
Female Physician	-0.0188	-0.0057	-0.0157	-0.0788***	-0.0001	-0.0008
	(0.0171)	(0.0085)	(0.0085)	(0.0181)	(0.0066)	(0.0066)
Spanish-Speaking	-0.0200	0.0002	0.0046	0.0461*	0.0138*	0.0122*
	(0.0142)	(0.0081)	(0.0087)	(0.0233)	(0.0062)	(0.0062)
Quarter since sample start	Y	Y	Y	Y	Y	Y
Patient Propensity Score	Y	Y	Y	Y	Y	Y
Patient Characteristics	Y	Y	Y	Y	Y	Y
Patient Comorbidities	Y	Y	Y	Y	Y	Y
Hospital FEs	Y	Y	Y	Y	Y	Y
N	756924	723136	723136	756924	723136	723136
R^2	0.12	0.55	0.56	0.40	0.82	0.82

Notes: Standard errors are clustered at the physician level and shown in parentheses. * indicates p<0.05, ** indicates p< 0.01, *** indicates p<.001.

Alphas & Betas vary with each 3 years of physician experience. "Low appropriateness" indicates patient is below the 25th percentile of appropriateness for invasive procedures. "High appropriateness" indicates patient is above the 75th percentile.

Appendix Table 1: Modeling the Probability of Invasive Procedures

Year:	(1) 1992	(2) 1997	(3) 2002	(4) 2007	(5) 2011
female	0.0194	-0.159***	-0.269***	-0.267***	-0.319***
	(0.28)	(-3.31)	(-5.87)	(-5.14)	(-5.99)
Co-morbidities					
arrhythmia	-0.156*	0.168***	0.0951	0.0682	-0.164*
	(-2.23)	(3.31)	(1.90)	(1.04)	(-2.18)
hypertension	-0.0212	-0.0965*	-0.0704	0.0817	-0.0860
	(-0.32)	(-2.07)	(-1.55)	(1.46)	(-1.52)
congestive heart	-0.314***	-0.371***	-0.344***	-0.309***	-0.580***
failure	(-4.54)	(-7.27)	(-6.78)	(-5.30)	(-8.92)
peripheral disease	0.0264	0.0585	-0.0706	0.728**	-0.116
	(0.19)	(0.60)	(-0.81)	(3.17)	(-0.60)
dementia	-0.903**	-1.133***	-1.098***	-1.604***	-1.526***
	(-2.59)	(-6.43)	(-7.80)	(-5.72)	(-5.49)
cere disease	0.0150	0.202*	-0.0790	-0.126	-0.150
	(0.12)	(2.32)	(-1.01)	(-1.13)	(-1.48)
chronic obstructive	-0.0781	-0.188**	-0.0561	-0.358***	-0.614***
pulmonary disease	(-0.90)	(-3.06)	(-0.99)	(-4.91)	(-6.03)
lupus	0.0370	-0.325*	-0.0607	-0.371	-0.675**
	(0.12)	(-2.05)	(-0.46)	(-1.93)	(-3.10)
ulcer	-0.0940	-0.293	-0.767**	-0.247	-1.192**
	(-0.26)	(-1.09)	(-2.75)	(-0.70)	(-3.14)
liver disease	-0.491	-0.0568	-0.452**	-0.0986	0.169
	(-1.46)	(-0.24)	(-3.04)	(-0.37)	(0.98)
cancer	-0.259	-0.308***	-0.460***	-0.843***	-0.833***
	(-1.70)	(-3.47)	(-5.58)	(-6.68)	(-6.48)
diabetes	-0.239**	-0.0909	-0.133**	-0.207**	-0.109
	(-3.07)	(-1.71)	(-2.67)	(-2.75)	(-1.40)
kidney disease	0.242	0.536***	0.533***	0.0956	0.0889
	(1.79)	(6.27)	(7.24)	(1.47)	(1.44)
hiv		-0.794	-0.0776	-0.850**	-0.331
		(-1.90)	(-0.28)	(-2.86)	(-1.18)
Age Group					
age 50-54	0.261	0.337**	0.144	0.430**	0.516**
	(1.63)	(2.58)	(1.17)	(2.98)	(3.26)
age 55-59	0.0986	0.306*	0.349**	0.353**	0.0809
	(0.63)	(2.57)	(3.02)	(2.67)	(0.60)
age 60-64	0.314*	0.275*	0.194	0.406**	0.347**
	(2.08)	(2.35)	(1.75)	(3.20)	(2.60)
age 65-69	0.227	0.377***	0.176	0.123	0.355**

	(1.59)	(3.32)	(1.60)	(1.01)	(2.77)
age 70	0.157	0.0807	0.182	0.110	0.109
	(1.11)	(0.75)	(1.66)	(0.90)	(0.85)
age 71	0.0948	0.140	0.0838	0.114	0.547*
	(0.47)	(0.90)	(0.53)	(0.59)	(2.58)
age 72	-0.121	-0.0823	0.0121	0.0657	0.147
	(-0.58)	(-0.53)	(0.08)	(0.33)	(0.78)
age73	0.0346	-0.132	-0.117	-0.0669	0.246
	(0.17)	(-0.86)	(-0.75)	(-0.35)	(1.22)
age74	0.286	-0.286	-0.165	0.00794	0.410
	(1.39)	(-1.84)	(-1.10)	(0.04)	(1.84)
age75	0.0968	-0.0889	-0.133	0.152	-0.132
	(0.47)	(-0.56)	(-0.86)	(0.84)	(-0.63)
age76	-0.184	-0.363*	-0.328*	-0.0985	-0.118
	(-0.82)	(-2.33)	(-2.09)	(-0.57)	(-0.63)
age77	0.281	-0.280	0.00598	-0.0891	-0.0989
	(1.23)	(-1.77)	(0.04)	(-0.51)	(-0.53)
age78	-0.0216	-0.226	-0.0907	-0.260	0.0647
	(-0.10)	(-1.36)	(-0.57)	(-1.36)	(0.34)
age79	-0.475*	-0.582***	-0.151	-0.246	-0.298
	(-1.97)	(-3.55)	(-0.99)	(-1.35)	(-1.53)
age 80	-0.130	-0.345*	-0.375*	-0.386*	-0.283
	(-0.49)	(-2.03)	(-2.37)	(-2.20)	(-1.48)
age 81	-0.638*	-0.650***	-0.431*	-0.468*	-0.00360
	(-2.36)	(-3.78)	(-2.56)	(-2.39)	(-0.02)
age 82	-0.805**	-0.446*	-0.410*	0.0939	-0.511**
	(-3.12)	(-2.53)	(-2.56)	(0.48)	(-2.60)
age 83	-0.661*	-0.553**	-0.441**	-0.514**	-0.366
	(-2.36)	(-3.07)	(-2.68)	(-2.73)	(-1.91)
age 84	-0.957**	-0.906***	-0.802***	-0.765***	-0.221
	(-3.23)	(-4.83)	(-4.69)	(-4.01)	(-1.08)
age 85	-1.318**	-1.203***	-0.899***	-0.924***	-0.376
	(-3.25)	(-6.04)	(-4.69)	(-4.75)	(-1.79)
age 86	-1.619***	-1.306***	-1.177***	-0.839***	-0.458*
	(-3.58)	(-6.10)	(-5.38)	(-4.35)	(-2.23)
age 87	-1.447**	-1.107***	-1.206***	-0.809***	-0.836***
	(-3.10)	(-4.56)	(-6.02)	(-3.84)	(-3.77)
age 88	-1.118*	-0.620**	-1.498***	-1.160***	-1.113***
	(-2.39)	(-2.81)	(-6.62)	(-4.67)	(-5.04)
age 89	-1.057*	-1.482***	-1.358***	-1.024***	-0.668**
	(-2.22)	(-5.18)	(-5.97)	(-4.17)	(-3.06)
age 90	-2.530*	-2.250***	-1.654***	-1.330***	-0.815**
	(-2.41)	(-5.66)	(-5.92)	(-5.52)	(-3.27)

age > 90	-2.611***	-1.911***	-1.615***	-1.720***	-1.535***
	(-4.91)	(-9.30)	(-9.93)	(-10.89)	(-9.53)
Heart diagnoses (ICD codes)					
410.00	0.462	-0.601			
	(0.46)	(-0.48)			
410.01	0.0340	0.478*	1.119***	2.100***	2.657***
	(0.14)	(2.46)	(5.52)	(6.92)	(8.10)
410.02	0.0589	-0.673	-0.271	-0.782	
	(0.08)	(-1.29)	(-0.51)	(-0.77)	
410.10		0.0853		1.500	
		(0.11)		(1.31)	
410.11	0.413*	0.650***	1.404***	2.306***	2.786***
	(2.08)	(3.91)	(8.30)	(9.46)	(10.66)
410.12	1.044***	-0.000854	-0.195	-0.317	0.161
	(3.38)	(-0.00)	(-0.80)	(-0.84)	(0.41)
410.20					
410.21	0.0117	0.508*	1.274***	1.928***	2.644***
	(0.04)	(2.40)	(5.45)	(5.73)	(7.45)
410.22	0.751	1.204	0.579	1.837	
	(1.19)	(1.89)	(1.10)	(1.88)	
410.30					
410.31	0.742*	0.614*	1.685***	2.721***	3.480***
	(2.37)	(2.57)	(5.66)	(5.45)	(5.26)
410.32	0.464	-0.521	-0.919		
	(0.47)	(-0.83)	(-1.09)		
410.40	0.305	1.483	1.030	0.856	0.367
	(0.32)	(1.17)	(0.98)	(0.69)	(0.26)
410.41	0.361	0.602***	1.462***	2.386***	2.865***
	(1.87)	(3.70)	(8.80)	(9.83)	(10.93)
410.42	0.870**	-0.155	-0.0703	0.390	0.622
	(3.26)	(-0.72)	(-0.27)	(1.07)	(1.39)
410.50		-0.275			
		(-0.23)			
410.51	-0.317	0.232	1.241***	2.114***	3.054***
	(-1.19)	(1.02)	(4.23)	(5.04)	(5.11)
410.52	-0.221	-0.515	-0.220	-0.551	
	(-0.32)	(-0.87)	(-0.31)	(-0.44)	
410.60					
410.61	0.371	0.186	1.363**	2.701*	1.399*
	(0.99)	(0.50)	(2.91)	(2.35)	(2.06)

410.62	1.874	-1.327	0.763		
	(1.46)	(-0.84)	(0.88)		
410.70		1.068	-0.661	1.035	1.302*
		(1.53)	(-1.02)	(1.30)	(2.52)
410.71	-0.234	-0.0700	0.631***	0.908***	1.130***
	(-1.20)	(-0.44)	(4.14)	(4.33)	(5.22)
410.72	0.847**	-0.366	-0.0170	-0.125	0.0282
	(3.04)	(-1.88)	(-0.09)	(-0.53)	(0.12)
410.80		-0.0472	-0.793	0.394	
		(-0.05)	(-0.73)	(0.32)	
410.81	0.0692	0.265	0.583**	1.079***	1.341***
	(0.29)	(1.32)	(2.65)	(3.91)	(4.12)
410.82	0.231	0.172	0.345	0.401	0.960
	(0.46)	(0.43)	(0.41)	(0.51)	(1.25)
410.90	-1.268*	-0.945*	-0.132	0.0326	0.260
	(-2.11)	(-1.99)	(-0.14)	(0.05)	(0.48)
410.91	-0.224	-0.336	0.329	0.927***	1.190***
	(-1.03)	(-1.88)	(1.94)	(4.08)	(5.03)
Quarter 1	0.00265	-0.0513	-0.0575	0.0509	-0.0838
	(0.03)	(-0.82)	(-0.95)	(0.73)	(-1.15)
Quarter 2	0.104	-0.0434	-0.0359	0.0928	-0.134
	(1.19)	(-0.68)	(-0.58)	(1.28)	(-1.80)
Quarter 3	-0.0331	-0.0935	-0.0113	0.000612	-0.179*
	(-0.37)	(-1.45)	(-0.18)	(0.01)	(-2.38)
Constant	-0.117	0.407*	0.105	-0.0553	-0.174
	(-0.53)	(2.19)	(0.59)	(-0.24)	(-0.72)
Dep variable mean	0.33	0.41	0.46	0.53	0.56
N	4714	9209	10224	8734	8077
Pseudo R^2	0.0800	0.1044	0.1158	0.1388	0.1483

Notes: These models are estimated using only patients in programs accredited to teach internal medicine. These are coefficients from the logit model (not marginal effects).

Appendix Table 2: Distribution of Patients Per Physician Before and After Sample Restriction

Percentile	Before	After	Per 3-years After
1%	3	8	5
5%	15	24	8
10%	30	41	11
25%	79	90	20
50%	181	192	45
75%	372	378	90
90%	651	649	156
95%	890	870	223
99%	1561	1547	425
N	1,183,548	1,023,821	1,023,821

Note on sample restrictions: We restrict the sample to physicians whose license numbers on the patient records match the license numbers in Florida's medical practicioner database. We also restrict the sample to physicians who treat at least 5 AMI patients per 3-year period.

Appendix Table 3: Main Results Showing all Covariates (See Table 5)

	(1)	(2)	(3)	(4)	(5)	(6)
Appropriateness for Invasive Procedure:	**Low**	**High**	**Low**	**High**	**Low**	**High**
Outcome:	**Hosp. Aquired Infection**	**Hosp. Aquired Infection**	**Died in Hospital**	**Died in Hospital**	**Discharged Home**	**Discharged Home**
Low Responsiveness	0.018***	-0.002	0.013***	0.003	-0.020***	-0.046***
(Beta<1)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
Low Aggressiveness	-0.003	-0.001	-0.012***	-0.003	-0.010***	-0.050***
(Alpha<0)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
High Aggressiveness	-0.000	0.002	-0.001	-0.004	0.009	0.015**
(Alpha>0)	(0.005)	(0.003)	(0.005)	(0.002)	(0.007)	(0.005)
<u>Patient Characteristics</u>						
Index of Patient	0.012***	-0.025***	-0.003	0.009***	-0.011***	0.021***
Appropriateness for	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)	(0.002)
Invasive Procedures						
Patient had subsequent AMI	-0.053***	-0.052***	-0.067***	-0.050***	0.110***	0.110***
	(0.003)	(0.007)	(0.002)	(0.006)	(0.004)	(0.011)
Female	0.080***	0.070***	-0.004*	0.020***	-0.038***	-0.022***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
age50 or less	0.020	0.007**	-0.005	-0.001	-0.006	-0.008*
	(0.014)	(0.003)	(0.012)	(0.002)	(0.023)	(0.004)
age55	0.013	0.019***	0.020	0.005*	-0.024	-0.023***
	(0.012)	(0.003)	(0.011)	(0.002)	(0.020)	(0.004)
age60	0.038***	0.029***	0.029**	0.012***	-0.000	-0.035***
	(0.011)	(0.003)	(0.010)	(0.002)	(0.017)	(0.004)
age65	0.039***	0.044***	0.047***	0.023***	-0.024	-0.056***
	(0.010)	(0.003)	(0.009)	(0.002)	(0.016)	(0.004)
age70	0.060***	0.056***	0.049***	0.038***	-0.060***	-0.089***
	(0.009)	(0.003)	(0.008)	(0.003)	(0.015)	(0.004)
age71	0.060***	0.065***	0.051***	0.046***	-0.079***	-0.102***

	(0.010)	(0.006)	(0.010)	(0.005)	(0.017)	(0.007)
age72	0.063***	0.069***	0.049***	0.057***	-0.077***	-0.118***
	(0.010)	(0.006)	(0.009)	(0.006)	(0.017)	(0.008)
age73	0.073***	0.075***	0.049***	0.055***	-0.074***	-0.128***
	(0.010)	(0.006)	(0.009)	(0.005)	(0.016)	(0.008)
age74	0.067***	0.064***	0.067***	0.073***	-0.086***	-0.142***
	(0.010)	(0.006)	(0.009)	(0.006)	(0.016)	(0.008)
age75	0.077***	0.069***	0.064***	0.062***	-0.083***	-0.131***
	(0.010)	(0.006)	(0.009)	(0.006)	(0.016)	(0.008)
age76	0.081***	0.096***	0.065***	0.091***	-0.090***	-0.168***
	(0.010)	(0.007)	(0.009)	(0.007)	(0.015)	(0.009)
age77	0.090***	0.077***	0.065***	0.088***	-0.096***	-0.179***
	(0.010)	(0.007)	(0.009)	(0.007)	(0.015)	(0.009)
age78	0.086***	0.085***	0.068***	0.084***	-0.115***	-0.176***
	(0.009)	(0.008)	(0.009)	(0.007)	(0.015)	(0.009)
age79	0.095***	0.070***	0.070***	0.099***	-0.113***	-0.191***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.015)	(0.010)
age80	0.089***	0.060***	0.073***	0.094***	-0.108***	-0.172***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.015)	(0.011)
age81	0.095***	0.081***	0.082***	0.110***	-0.123***	-0.204***
	(0.009)	(0.009)	(0.008)	(0.009)	(0.015)	(0.011)
age82	0.098***	0.094***	0.087***	0.110***	-0.134***	-0.218***
	(0.009)	(0.011)	(0.008)	(0.010)	(0.015)	(0.012)
age83	0.100***	0.082***	0.084***	0.125***	-0.129***	-0.259***
	(0.009)	(0.011)	(0.008)	(0.011)	(0.015)	(0.013)
age84	0.113***	0.079***	0.089***	0.123***	-0.144***	-0.278***
	(0.009)	(0.013)	(0.008)	(0.013)	(0.015)	(0.015)
age85	0.107***	0.061***	0.094***	0.108***	-0.140***	-0.269***
	(0.009)	(0.013)	(0.008)	(0.013)	(0.015)	(0.017)
age86	0.114***	0.093***	0.094***	0.148***	-0.153***	-0.265***
	(0.009)	(0.015)	(0.008)	(0.016)	(0.015)	(0.019)
age87	0.124***	0.104***	0.099***	0.124***	-0.161***	-0.276***

	(0.009)	(0.018)	(0.008)	(0.017)	(0.015)	(0.022)
age88	0.127***	0.081***	0.108***	0.143***	-0.176***	-0.304***
	(0.009)	(0.020)	(0.008)	(0.020)	(0.015)	(0.023)
age89	0.138***	0.091***	0.110***	0.134***	-0.186***	-0.254***
	(0.010)	(0.024)	(0.008)	(0.022)	(0.015)	(0.027)
age90	0.142***	0.120***	0.119***	0.146***	-0.201***	-0.317***
	(0.010)	(0.033)	(0.009)	(0.031)	(0.015)	(0.038)
age95 plus	0.166***	0.054*	0.130***	0.218***	-0.240***	-0.386***
	(0.009)	(0.024)	(0.008)	(0.027)	(0.015)	(0.028)
Patient Co-mobidities						
arrhythmia	-0.026***	-0.022***	0.009***	0.017***	0.002	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
hypertension	-0.100***	-0.098***	-0.069***	-0.060***	0.090***	0.095***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
chf	-0.015***	-0.001	0.002	0.023***	-0.019***	-0.031***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.004)
peripheral_disease	-0.046***	-0.040***	-0.015***	-0.010**	0.033***	0.039***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.005)	(0.005)
dementia	0.017***	-0.030	-0.009**	-0.004	-0.129***	-0.096**
	(0.003)	(0.025)	(0.003)	(0.023)	(0.004)	(0.035)
cere_disease	0.017***	0.045***	0.043***	0.087***	-0.116***	-0.174***
	(0.003)	(0.005)	(0.003)	(0.005)	(0.004)	(0.006)
copd	-0.019***	-0.017***	-0.001	0.007*	-0.005	0.013**
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.004)
lupus	-0.013**	-0.026***	-0.038***	-0.001	-0.012*	0.036***
	(0.004)	(0.007)	(0.004)	(0.007)	(0.006)	(0.009)
ulcer	-0.053***	-0.042**	-0.041***	-0.044***	0.058***	0.039*
	(0.007)	(0.014)	(0.006)	(0.010)	(0.010)	(0.019)
liver_disease	0.019*	0.041***	0.101***	0.178***	-0.079***	-0.146***
	(0.007)	(0.007)	(0.008)	(0.007)	(0.009)	(0.008)
cancer	-0.043***	-0.052***	-0.004	0.022***	0.022***	0.017*
	(0.003)	(0.004)	(0.003)	(0.005)	(0.004)	(0.007)

diabetes	-0.009***	-0.008***	-0.021***	-0.027***	0.019***	0.046***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
hemiplegia	0.753***		-0.101***		-0.458***	
	(0.013)		(0.011)		(0.016)	
kidney_disease	0.069***	0.165***	0.125***	0.136***	-0.116***	-0.162***
	(0.004)	(0.003)	(0.004)	(0.002)	(0.004)	(0.003)
HIV	0.011	0.045***	0.049*	0.026**	-0.008	-0.015
	(0.019)	(0.011)	(0.021)	(0.009)	(0.029)	(0.013)
<u>Other Physician Characteristics</u>						
Experience Category	-0.003***	-0.000	0.002***	0.001*	0.002***	0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Cardiologist	-0.044***	-0.049***	-0.022***	-0.027***	0.071***	0.055***
	(0.003)	(0.002)	(0.003)	(0.002)	(0.005)	(0.004)
Family Practice	0.002	-0.007*	-0.000	-0.002	-0.005	0.001
	(0.004)	(0.003)	(0.003)	(0.002)	(0.004)	(0.005)
Other Specialty	0.025***	0.008	0.006	0.012*	-0.029***	-0.020*
	(0.005)	(0.005)	(0.004)	(0.005)	(0.006)	(0.008)
Medical Doctor (MD vs. OS)	-0.010*	0.001	-0.004	-0.008*	0.025***	0.012*
	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)
U.S. Medical School	-0.006*	-0.002	-0.001	-0.003	0.014***	0.000
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Top 20 Medical School	-0.005	-0.007*	-0.005	0.001	-0.004	0.001
	(0.005)	(0.004)	(0.004)	(0.003)	(0.007)	(0.006)
Female Physician	0.005	0.003	-0.002	-0.006*	-0.001	0.003
	(0.003)	(0.003)	(0.003)	(0.002)	(0.004)	(0.004)
Spanish Speaking	-0.001	-0.002	-0.002	-0.004*	0.001	0.002
	(0.003)	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)
Constant	0.106***	0.126***	-0.039***	0.022***	0.622***	0.771***
	(0.013)	(0.007)	(0.011)	(0.006)	(0.019)	(0.010)
Quarter since sample start	Y	Y	Y	Y	Y	Y
Hospital FE	Y	Y	Y	Y	Y	Y
N	205419	186687	205419	186687	205419	186687

R^2	0.06	0.13	0.05	0.12	0.10	0.31

Notes: See Table 6. Experience category is a discrete variable where 1 corresponds to 1-3 years of experience, 2 corresponds to 4-6 years of experience and so on.

Appendix Table 4: Robustness of main results to an alternative categorization of alpha and beta.

	(1)	(2)	(3)	(4)	(5)	(6)
Appropriateness for Invasive Procedure:	**Low**	**High**	**Low**	**High**	**Low**	**High**
Outcome:	**Hospital Aquired Infection**	**Hospital Aquired Infection**	**Died in Hospital**	**Died in Hospital**	**Discharged Home**	**Discharged Home**
Low Aggressiveness, Typical Responsiveness (Alpha<0, Beta>=1)	-0.002	0.000	-0.010***	-0.003	-0.012***	-0.044***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
High Aggressiveness, Typical Responsiveness (Alpha>0, Beta>=1)	0.003	0.003	-0.001	-0.006*	0.008	0.009
	(0.005)	(0.003)	(0.005)	(0.003)	(0.008)	(0.005)
Low Aggressiveness, Low Responsiveness (Alpha<0, Beta<1)	0.015***	-0.006	-0.001	-0.001	-0.029***	-0.110***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.006)
Typical Aggressiveness, Low Responsiveness (Alpha>=0, Beta<1)	0.019***	0.002	0.020***	0.003	-0.023***	-0.029***
	(0.004)	(0.003)	(0.003)	(0.002)	(0.005)	(0.004)
Quarter since sample start	Y	Y	Y	Y	Y	Y
Hospital FE	Y	Y	Y	Y	Y	Y
Patient Appropriateness Inde	Y	Y	Y	Y	Y	Y
Patient Age & Gender	Y	Y	Y	Y	Y	Y
Patient Comorbidities	Y	Y	Y	Y	Y	Y
Physician Characteristics	Y	Y	Y	Y	Y	Y
N	205419	186687	205419	186687	205419	186687
R^2	0.06	0.13	0.05	0.12	0.10	0.31

Notes: See Table 5.

Appendix Table 5: Robustness of Main Results to Inclusion of Race, Ethnicity, and Insurance Status

	(1)	(2)	(3)	(4)	(5)	(6)
Appropriateness for Invasive Procedure:	**Low**	**High**	**Low**	**High**	**Low**	**High**
Outcome:	**Hosp. Aquired Infection**	**Hosp. Aquired Infection**	**Died in Hospital**	**Died in Hospital**	**Discharged to Home**	**Discharged to Home**
Low Responsiveness	0.018***	-0.002	0.013***	0.003	-0.020***	-0.046***
(Beta<0)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
Low Aggressiveness	-0.002	-0.002	-0.012***	-0.003	-0.011***	-0.050***
(Alpha<0)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
High Aggressiveness	0.000	0.002	-0.001	-0.004	0.009	0.015**
(Alpha>0)	(0.005)	(0.003)	(0.005)	(0.002)	(0.007)	(0.005)
Quarter since sample start	Y	Y	Y	Y	Y	Y
Hospital FE	Y	Y	Y	Y	Y	Y
Patient Appropriateness Index	Y	Y	Y	Y	Y	Y
Patient Age & Gender	Y	Y	Y	Y	Y	Y
Patient Comorbidities	Y	Y	Y	Y	Y	Y
Physician Characteristics	Y	Y	Y	Y	Y	Y
Race, Ethnicity, Insurance	Y	Y	Y	Y	Y	Y
N	205419	186687	205419	186687	205419	186687
R^2	0.06	0.14	0.05	0.12	0.10	0.31

Notes: Standard errors are clustered at the physician level and shown in parentheses. * indicates $p<0.05$, ** indicates $p< 0.01$, *** indicates $p<0.001$. Alphas and Betas vary with each 3 years of physician experience. "Low appropriateness" indicates patient is below the 25th percentile of our index of appropriateness for invasive procedures. "High appropriateness" indicates patient is above the 75th percentile.