Treatment Effect Bounds: An Application to Swan-Ganz Catherization*

Jay Bhattacharya
School of Medicine
Stanford University

Azeem M. Shaikh
Department of Economics
Stanford University

Edward Vytlacil
Department of Economics
Stanford University

January 26, 2005

Abstract

We implement alternative bounding strategies to reanalyze data from the observational study by Connors et al. (1996) on the impact of Swan-Ganz catheterization on mortality outcomes. While the Connors et al. (1996) study assumes that there are no unobserved differences between patients who are catheterized and patients who are not, we employ a bounding strategy that allows for such unobserved differences. We implement both the traditional bounds that exploit access to an instrumental variable but impose no other assumptions (Manski, 1990) and the bounds of Shaikh and Vytlacil (2004) that impose additional relatively mild nonparametric structural assumptions. Both of these approaches require an instrumental variable that shifts the probability of catheterization, but which does not alter mortality risks. We propose and justify using indicators of weekday admission as an instrumental variable in this context.

We find that, while the traditional instrumental variable bounds are almost entirely uninformative in our application, the Shaikh and Vytlacil (2004) bounds that do impose some nonparametric structure often produce a clear answer—catheterization reduces mortality at 7 days, and increases it at 30 days and after. Applying a further nonparametric structural assumption—that doctors catheterize individuals with systematically worse latent health—further narrows these bounds and strengthens these conclusions. Our findings suggest an explanation for the fact that many ICU doctors are deeply committed to the use of the Swan-Ganz catheter. Since most ICU patients leave the ICU well before 30 days after admission have elapsed, ICU doctors never observe the increase in mortality. They do, however, observe the decline in mortality at 7 days.

*We would like to thank seminar participants at Brigham Young University, University of Chicago, Harvard/MIT, Michigan Ann-Arbor, Michigan State, and at the ZEW 2nd Conference on Evaluation Research in Mannheim. Correspondence: Landau Economics Building, 579 Serra Mall, Stanford CA 94305; Email: vytlacil@stanford.edu; Phone: 650-725-7836; Fax: 650-725-5702.

1 Introduction

In randomized settings, inferring the effect of a treatment on an individual is comparatively straightforward. Since each individual is either given or not given the treatment with equal probability, on average there are no unobserved differences between the treated and untreated groups. The researcher can therefore compare the average outcomes of the treated and untreated groups to estimate the treatment effect. Randomized studies, however, are often prohibitively expensive or suffer from ethical complications, leaving the researcher instead with more readily available observational data in which each individual or some proxy chooses whether or not to adopt the treatment. In such settings, comparing the average outcomes of the treated and untreated groups is unlikely to yield an accurate estimate of the treatment effect, as decisions regarding whether or not to adopt the treatment will be made to maximize the welfare of the recipient.

In this paper, we compare several different empirical approaches to the sample selection problem in the context of the Swan-Ganz catheter, which is a device used by intensive care unit (ICU) doctors to guide therapy in the ICU. There are two main theoretical novelties of this paper. First, we implement a new bounding approach proposed by Shaikh and Vytlacil (2004) that has heretofore not been evaluated empirically. Second, we extend the bounding approach of Shaikh and Vytlacil (2004) to exploit the assumption that doctors are performing catheterization on those patients who have the highest expected mortality rates. The Swan-Ganz example is also of interest substantively. Nearly all of the empirical work on the mortality effects of the Swan-Ganz catheter has relied on observational data, but has ignored the possibility described above that patients are catheterized non-randomly even conditional on observed covariates.[1] Even though most of this literature has concluded that the procedure increases patient mortality, the use of the Swan-Ganz catheter remains extremely common among ICU doctors. One possible reason that these studies have not let to a change in clinical practice is that they failed to account adequately for the fact that those who are catheterized are in fact the sickest patients.

This sort of sample selection problem is well known in economics,[2] and much work has been devoted to devising econometric techniques to cope with the problem (see the Handbook of Econometrics for a review). The seminal paper by Heckman (1978) proposed a strictly parametric approach to the problem—the eponymous "Heckit." Heckman frames the issue in terms of a switching regression model, where the unobservability of outcomes in one of the two branches can be addressed as a missing variable in a regressor, as long as the parametric assumptions made are correct.

[1] See Dalen (2001) and Section 2 below for a review of this literature. Recently, some randomized trials have been conducted in specialized ICU subpopulations.

[2] In part this is because randomized studies are difficult to conduct in social science settings, while observational data are much cheaper to collect. In medicine, it is rare for observational studies to worry about unobserved selection, though see McClellan et al. (1994) for an exception.

Recent work has sought to relax the reliance on unverifiable parametric assumptions.[3] Manski (1989) suggests a strategy of bounding the average treatment effect that exploits the fact that a binary outcome variable is bounded from below by zero and from above by one. The outcome variable being bounded implies that the mean of the outcome variable for all groups is also bounded. Applying this idea leads to bounds on the treatment effect with a width equal to one. Manski (1990) suggests a strategy of sharpening these bounds when an instrument is available: First, calculate the treatment effect bounds over subgroups of the data defined by the instrument; second, bound the treatment effect from below by the maximum of the resulting lower bounds and from above by the minimum of the resulting upper bounds. We will call these bounds the Manski IV bounds.

Manski (1990) does not impose any structure on the models for the outcome variable or treatment variable. Shaikh and Vytlacil (2004), in contrast, model both the treatment and the outcome variables using threshold crossing models. The assumptions underlying their bounds are therefore stronger than those imposed by Manski (1990).[4] This additional structure results in bounds on the treatment effect that are often considerably narrower than the Manski IV bounds. At the same time, Shaikh and Vytlacil (2004) still avoid imposing any of the arguably more untenable parametric or distributional assumptions underlying the analysis of Heckman (1978).

In this paper, we develop an extended case study of these two different approaches for bounding the treatment effect by reanalyzing the data from a well known observational study by Connors et al. (1996), which examines the impact of Swan-Ganz catheterization on mortality outcomes among a population of patients admitted to the intensive care unit (ICU) at five prominent hospitals. All five of the hospitals are members of the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT) group—Duke University Medical Center, Durham, NC; MetroHealth Medical Center, Cleveland, OH; St. Joseph's Hospital, Marshfield, WI; and University of California Medical Center, Los Angeles, CA. Analyzing observational data from these ICUs, Connors et al. (1996) reach the controversial conclusion that patients who receive Swan-Ganz catheterization during their first day in the ICU are 1.27 times *more* likely to die within 180 days of their admission. Even at 7 days after ICU admission, Connors et al. (1996) find that catheterization increases mortality. This conclusion is surprising to ICU doctors because these catheters are often placed in

[3]Another possible approach is to use linear two stage least squares (TSLS). Under the conditions considered in this paper, TSLS will not recover the average effect of catheterization on mortality. However, Angrist et al. (1996) suggest redefining the object of interest to be the "Local Average Treatment Effect" (LATE) instead of the usual average treatment effect. As long as an instrument that satisfies their conditions is available, and as long as there are no covariates or the model is fully saturated, they show that TSLS will yield consistent estimates of the LATE parameter. See Heckman and Vytlacil (2000) for the relationship between the LATE parameter and other mean treatment parameters including the average treatment effect.

[4]Note that Shaikh and Vytlacil also impose more structure than the bounding analysis of Heckman and Vytlacil (1999), who impose a threshold crossing model on the treatment variable but not on the outcome variable. See Heckman and Vytlacil (2001) for the relationship between the Heckman and Vytlacil (1999) bounds and both the Manski (1990) IV bounds and the Balke and Pearl (1997) IV bounds.

severely ill ICU patients to inform better therapy aimed at reducing mortality. The finding that catheterization increases mortality even by 7 days is especially puzzling since one might expect ICU doctors to detect such an effect.[5] The statistical strategy Connors et al. (1996) use to reach this conclusion—the propensity score matching method—assumes away the possibility of unobserved differences between patients who are catheterized and patients who are not.

Our analysis, by comparison, permits the possibility of unobserved differences between these groups. We rely, instead, on the existence of an instrument for Swan-Ganz placement, and we use this instrument to compute both of the bounds on the treatment effect described above. We use the day of the week that the patient was admitted to the ICU as an instrument for Swan-Ganz catheterization.[6] We argue that this variable meets the two crucial requirements for an instrument's validity. First, it is strongly correlated with the application of the treatment. On weekends, patients are less likely to receive any medical procedure during hospitalization, including Swan-Ganz catheterization, as fewer medical staff are available on these days. Second, we argue that within observable risk classes, day-of-the-week is uncorrelated with the outcome of treatment; that is, the particular day of the week that a patient is admitted to the ICU has little to do with mortality rates, and more to do with the arc of the patient's medical condition. We test whether weekday admission monotonically increases the odds of catheterization.

The main theme of our empirical analysis is to quantify in an important empirical setting the trade-off between precision and verifiability. Accepting structural assumptions, like a threshold crossing model for catheterization, increases the former at the cost of the latter. On the other hand, estimators that are stubbornly agnostic altogether about unverifiable assumptions yield imprecise estimates that provide little guidance to ICU doctors deciding whether or not to catheterize a patient.

2 Background on Swan-Ganz Catheterization

A Swan-Ganz catheter is a long slender tube outfitted with sensors designed to measure hemodynamic pressures in the right side of the heart and in the pulmonary artery. The catheter also has a small balloon tip that allows a doctor to wedge the catheter into a pulmonary artery, providing a measure of left atrial pressure. Under sterile conditions, an ICU doctor will typically insert the catheter into the left subclavian vein (underneath the clavicle) or right internal jugular vein, and advance the catheter through the superior vena cava and into right side of the heart. Once in place, the catheter is often left in place for days, so it can continuously provide information to ICU doctors about the hemodynamic status of the patient.

[5] ICU doctors would typically not be expected to detect increased mortality if it occurs after the patient has left the ICU. At 7 days, a substantial portion of the ICU population may still be in the ICU.

[6] This is not the first study in health economics to use this variable as an instrument for treatment decisions. In a study of the effect of queuing time on mortality in a Canadian population undergoing hip-fracture surgery, Hamilton et al. (2000) use day of the week of admission as an instrument for queuing time. Their study does not examine the ICU population or the effect of Swan-Ganz catheterization.

A Swan-Ganz catheter is primarily used to measure pressure at different locations (depending upon where the tip of the catheter is) in the right side of the heart: central venous pressure, right atrial pressure, pulmonary artery pressure, and pulmonary artery wedge pressure. These pressure measurements can provide important diagnostic information, such as whether the patient's heart valves are working, or whether the patient has pulmonary hypertension. Information gleaned from Swan-Ganz measurements is often used by ICU doctors to make decisions about treatment, such as whether to give the patient medications that affect the functioning of the heart.

The placement of Swan-Ganz catheters is an extremely common procedure for ICU patients—over 2 million patients in North America are catheterized each year (Paunovic and Sharma, 2003). Nevertheless, there are a number of potentially life-threatening complications that can occur as a result of the procedure. During insertion, the doctor might miss the subclavian vein and puncture the lung underneath it instead. The placement of any foreign body such as a catheter can cause blood clots to form leading to a deadly pulmonary embolism.[7] Left in place for a while, the catheters can cause infections at the insertion point which can spread into the blood stream and into the heart itself.

Even if the placement of the catheter is successful, the information generated by the pressure measurements may itself cause harm. No test is perfect, Swan-Ganz catheterization included; false positive diagnoses of heart failure, for example, or other misdiagnoses may lead doctors to administer inappropriate medications and treatments that lead to bad outcomes. In light of these potential adverse consequences from catheter placement, the Connors et al. (1996) assertion that Swan-Ganz catheters increase mortality, while surprising, is not *prima facie* implausible and needs to be taken seriously.

Connors et al. (1996) is not the first observational study to conclude that Swan-Ganz catheterization may kill ICU patients. In the specialty journal, *Chest*, Gore et al. (1985) report an autopsy study of the effects of Swan-Ganz catheterization on patients who had suffered an acute myocardial infarction (heart attack). Zion et al. (1990) conduct a case-control study of the effects of Swan-Ganz catheterization on a different set of heart attack patients. Both groups of authors conclude that compared with patients who had catheters placed, patients without catheters were less likely to die within a fixed period after the hospitalization for the heart attack. The major criticism that both studies faced was that they did not sufficiently allow for clinically important differences between the patients who had catheters placed and those who did not. For example, Dalen (2001), commenting on the two reports, says:

> The studies reported by Gore et al. (1985) and Zion et al. (1990) were not randomized trials; rather, they were observational studies. Therefore, it is possible that physicians accurately assessed patients at risk for increased risk of an adverse outcome and selected patients who were more seriously ill for PAC use.

[7] Swan-Ganz catheters are often coated with an anti-coagulant to prevent blood clots from forming, but this expedient is not 100% effective.

The Connors et al. (1996) study was conceived in part as a response to this criticism. The SUPPORT team included a dizzying array of clinical variables designed to control as exhaustively as possible for observed differences between the patients with catheters placed and patients without catheters. In addition, Connors et al. (1996) expanded the set of ICU patients beyond just heart attack patients to all ICU patients. Weil (1998) argues that because Connors et al. (1996) expanded the set of patients considered, they fail to consider important clinical variables in their statistical work:

> With respect to... disorder groups, it is not possible to join diagnoses regardless of expected outcome. Let me cite aseptic meningitis. It has a very different course than meningococcal meningitis, and high-output failure of septic shock has a different course than low-output failure of ischemic heart disease... None of these were discretely classified in the Connors et al. (1996) study.

Despite substantial criticism, the publication of the Connors et al. (1996) study was seminal in the Swan-Ganz catheterization literature in the sense that it has guided the arc of subsequent research in two concrete ways. Subsequent studies have focused on expanding the set of ICU patients considered in the analysis, and on minimizing the possibility of selection bias.

Polanczyk et al. (2001) conduct another observational study of Swan-Ganz catheterization aimed at expanding the set of ICU patients considered to those undergoing elective non-cardiac surgery. Swan-Ganz catheters are often placed in patients undergoing serious non-cardiac surgery to help manage cardiac complications arising from the surgery. Like Connors et al. (1996), Palanczyk and colleagues use a large set of clinical control variables along with propensity score matching methods to limit the possibility of selection bias. Unlike Connors et al. (1996), their major endpoints are cardiac events like heart attack, unstable angina, congestive heart failure, and ventricular fibrillation. In their propensity score-matched analysis, they find that the placement of a catheter in this group of patients doubles the odds of congestive heart failure.

Hirano and Imbens (2001) reanalyze the Connors et al. (1996) data, using a modified version of the propensity score matching method. Their key modification is aimed at thinning the set of clinical control variables that are used to construct the propensity score. To do this, they initially regress an indicator variable that equals one if a patient had a catheter placed on each potential clinical control variable in turn, one at a time. Their final propensity score model includes all control variables that show up as statistically significant (at some preset level) in the univariate regression. Their main finding is that the Connors et al. (1996) conclusion that catheterization increases mortality risk is robust to their model selection exercise.

Prior to Connors et al. (1996), attempts to organize a randomized trial failed because doctors refused to recruit patients into the control group—the belief in the efficacy of catheterization was so strong that doctors believed it unethical to deny this procedure to patients on

the basis of chance.[8]

Since Connors et al. (1996), there have been two randomized trials on specialized ICU populations. Unlike the observational studies, these studies do not find that catheterization kills. Sandham et al. (2003) focus on high risk surgical patients aged 60 years and older who were admitted to ICUs. While patients in the catheterized group were statistically more likely to suffer a pulmonary embolism, there were no differences between the groups in survival in hospital or at 6 and 12 months post-ICU admission. Richard et al. (2003) conduct a randomized trial on 676 French ICU patients admitted because of shock, acute respiratory distress syndrome (ARDS) or both. They find no statistically significant differences between the catheterized and non-catheterized groups in 14, 28, and 90 day mortality; nor did they find any differences in the likelihood of organ failure or other measures of serious morbidity. These two studies are just the start of this research arc: there are a number of other trials currently in preparation that examine the effect of catheterization on patients with a wide variety of disease conditions (Dalen, 2001).

While it would be appealing to compare our results directly with the outcomes from these two randomized trials, differences in the populations studied preclude anything but a tentative comparison. Unlike the Sandham et al. (2003) trial, the Connors et al. (1996) SUPPORT data that we reanalyze includes adult patients of all ages, not just elderly patients. Our sample also includes non-surgical ICU patients, not just surgical patients. Like Sandham et al. (2003), however, patients in our sample are severely ill, even by the standards of the typical ICU patient. In principle, we could analyze the surgical ICU patients in our sample who are over 60 and with the same minimum level of disease severity (as measured by predicted mortality rate at the time of ICU admission)—this would provide a more direct comparison. Unfortunately, there are too few such patients in our sample to analyze fruitfully. The Richard et al. (2003) trial includes both severely ill and less severely ill ICU patients. While Richard et al. (2003) admit only patients with ARDS or shock or both to their trial, our data excludes no patients on the basis of their diagnosis. One similarity between the Richard et al. (2003) sample and ours is that, like them and unlike Sandham et al. (2003), we exclude no adults on the basis of their age from the sample. We report some results for ARDS patients alone that can be compared with the Richard et al. (2003) finding, though it is unclear whether our results for ARDS patients would pertain if we were to analyze less severely ill patients, as had Richard et al. (2003).

3 Notation and Assumptions

We now define notation and assumptions that we will use for our bounding analysis. Let Y denote mortality, with $Y = 1$ if the individual dies within the given number of days of admission into the ICU unit and $Y = 0$ otherwise. Let D denote the treatment, with $D = 1$ if the

[8] See Fowler and Cook (2003) and Guyatt (1991).

individual is catheterized and $D = 0$ otherwise. Denote by X the observed individual characteristics determining mortality and by Z the observed individual characteristics determining catheterization. We suppose that both Y and D are determined by threshold crossing models:

$$\begin{aligned} Y^* &= r(X, D) - \epsilon \\ Y &= \mathbf{1}[Y^* \geq 0] \end{aligned} \tag{1}$$

$$\begin{aligned} D^* &= s(Z) - \nu \\ D &= \mathbf{1}[D^* \geq 0] \end{aligned} \tag{2}$$

where $\mathbf{1}[\cdot]$ is the indicator function taking the value 1 if its argument is true and the value 0 otherwise. In these equations, ϵ and ν are unobserved random variables, where ϵ and ν are possibly dependent, giving rise to selection bias. Y^* and D^* are latent indices, with Y^* interpreted as a latent measure of health status and D^* interpreted as a latent measure of desire by hospital staff to conduct the catherization. We do not impose any functional form on r or s, and we do not impose any parametric distributional assumption on (ϵ, ν).

We will assume that $(X, Z) \perp\!\!\!\perp (\epsilon, \nu)$. Thus, we are allowing catheterization to be endogenous, reflecting the possible correlation between ϵ and ν, but we are assuming that all other regressors are exogenous. We also assume that (ϵ, ν) has a strictly positive density with respect to Lebesgue measure on $\Re^2$. This assumption eases the exposition, but it is not essential to any of our arguments. We will also assume that there is at least one variable in Z that is not in X. In other words, there is one variable that affects the decision to perform catheterization but does not directly affect mortality. Such a variable is often referred to as an instrument or an excluded variable, as it is excluded from the Y^* equation. In our application to catherization, we assume that admission into the hospital on a weekend as opposed to a weekday affects the hospital staff's desire to provide catherization but does not directly affect mortality.

Using potential outcome notation, let Y_1 denote the outcome that would be observed if the individual receives treatment, and let Y_0 denote the outcome that would be observed if the individual does not receive treatment. The effect of catheterization on mortality for a given individual is $Y_1 - Y_0$, and the average effect of the catheterization on mortality given covariates is then defined as

$$E(Y_1 - Y_0 \mid X) = \Pr[Y_1 = 1 \mid X] - \Pr[Y_0 = 1 \mid X].$$

The identification problem is that Y_1 is only observed for individuals who receive catheterization, and Y_0 is only observed for individuals who did not receive catheterization, and thus without further assumptions it is not possible to identify the average treatment effect directly.

Given our model and assumptions, the potential outcomes are given by

$$\begin{aligned} Y_1 &= \mathbf{1}[r(X, 1) - \epsilon \geq 0] \\ Y_0 &= \mathbf{1}[r(X, 0) - \epsilon \geq 0], \end{aligned}$$

and the average treatment effect conditional on X is

$$E(Y_1 - Y_0 \mid X) = \Pr[Y_1 = 1 \mid X] - \Pr[Y_0 = 1 \mid X] = F_\epsilon(r(X,1)) - F_\epsilon(r(X,0)).$$

It is worthwhile to note that a special case of our model is the standard bivariate probit model, which imposes linear index and bivariate normality assumptions: $r(X,D) = X\beta + DX\alpha$, $s(Z) = Z\gamma$, and (ϵ, ν) distributed bivariate normal with the normalizations that $E(\epsilon) = E(\nu) = 0$ and $\text{Var}(\epsilon) = \text{Var}(\nu) = 1$. In this case,

$$\begin{aligned} Y_1 &= \mathbf{1}[X(\beta + \alpha) - \epsilon \geq 0] \\ Y_0 &= \mathbf{1}[X\beta - \epsilon \geq 0] \end{aligned}$$

and

$$E(Y_1 - Y_0 \mid X) = \Phi(X(\beta + \alpha)) - \Phi(X\beta)$$

where $\Phi(\cdot)$ is the standard normal cdf. Our model nests the bivariate probit model as a special case but does not impose any of these parametric assumptions.

4 IV Bounds

For ease of exposition, suppose that there are no X covariates and that Z is a binary random variable. We will later relax these assumptions to allow for X covariates and to allow Z to be a general random vector. For ease of exposition, order Z so that $\Pr[D = 1|Z = 1] > \Pr[D = 1|Z = 0]$; $Z = 1$ is therefore associated with the higher rate of catherization than $Z = 0$. Thus, in our application, $Z = 1$ corresponds to a admission into an ICU on a weekday while $Z = 0$ corresponds to admission on a weekend. Assume the researcher has available an i.i.d. sample of (Z_i, D_i, Y_i) observations. We now consider two alternative bounding strategies: the Manski (1990) IV bounds and the Shaikh and Vytlacil (2004) bounds.[9,10] We also compare the Shaikh and Vytlacil bounds to the Manski and Pepper (2000) bounds that impose monotone treatment response and instrumental variable assumptions, and we also consider an extension of the Shaikh-Vytlacil bounds than strengthen their assumptions by imposing that doctors are more likely to catheterize patients with the worst latent health.

First consider the Manski (1990) instrumental variables bounds. Manksi assumes that Y_1 and Y_0 are independent of Z, $\Pr[Y_0 = 1 \mid Z] = \Pr[Y_0 = 1]$ and $\Pr[Y_1 = 1 \mid Z] = \Pr[Y_1 = 1]$, but does not impose any assumptions on the model for either the treatment or outcome.[11]

[9] Since our outcome variable is binary, we are specializing the Manski (1990) bounds to the special case of a binary outcome variable. His bounds do not require the outcome to be binary.

[10] Other IV bounds not considered in this paper include the contaminated IV bounds of Hotz et al. (1997) and the IV bounds of Balke and Pearl (1997) and Heckman and Vytlacil (1999, 2001).

[11] In the general case with Y potentially not binary, the Manski IV bounds impose only a mean independence assumption and not a full independence assumption.

Thus, Manski's assumptions are strictly weaker than the assumptions considered in this paper. Consider $\Pr[Y_1 = 1 \mid Z = z]$ and note that

$$\Pr[Y_1 = 1 \mid Z = z] = \Pr[D = 1, Y_1 = 1 \mid Z = z] + \Pr[D = 0, Y_1 = 1 \mid Z = z].$$

The event $(D = 1, Y_1 = 1)$ is the same as $(D = 1, Y = 1)$ since $Y = Y_1$ when $D = 1$. Thus, the term $\Pr[D = 1, Y_1 = 1 \mid Z = z] = \Pr[D = 1, Y = 1 \mid Z = z]$ is immediately identified from the data. The term $\Pr[D = 0, Y_1 = 1 \mid Z = z] = \Pr[D = 0 \mid Z = z] \Pr[Y_1 = 1 \mid D = 0, Z = z]$, on the other hand, is not identified from the data since we never know the outcome with treatment when the individual did not receive treatment. However, any probability is bounded by zero and one, so $0 \leq \Pr[Y_1 = 1 \mid D = 0, Z = z] \leq 1$, and thus

$$\Pr[D = 1, Y = 1|Z = z] \leq \Pr[Y_1 = 1|Z = z] \leq \Pr[D = 1, Y = 1|Z = z] + \Pr[D = 0|Z = z].$$

Since $\Pr[Y_1 = 1] = \Pr[Y_1 = 1 \mid Z = z]$ by assumption, we have

$$\Pr[D = 1, Y = 1|Z = z] \leq \Pr[Y_1 = 1] \leq \Pr[D = 1, Y = 1|Z = z] + \Pr[D = 0|Z = z].$$

Since we can evaluate the bounds at any value of z, we can take the maximum of the lower bounds and the minimum of the upper bounds over possible values of z. Following the parallel analysis for $\Pr[Y_0 = 1]$, we have the following bounds:

$$B_M^L \leq E(Y_1 - Y_0) \leq B_M^U$$

where

$$B_M^L = \max_z \{\Pr[D = 1, Y = 1|Z = z]\} - \min_z \{\Pr[D = 0, Y = 1|Z = z] + \Pr[D = 1|Z = z]\},$$

$$B_M^U = \min_z \{\Pr[D = 1, Y = 1|Z = z] + \Pr[D = 0|Z = z]\} - \max_z \{\Pr[D = 0, Y = 1|Z = z]\}.$$

These bounds can be consistently estimated by substituting the conditional sample means for the conditional population means. These bounds are sharp given Manski's mean independence assumption, which leaves open the question of whether the bounds can be improved upon if imposing the threshold crossing, binary choice model on the treatment variable and the outcome variable.

Now consider the Shaikh-Vytlacil bounds. The analysis of Shaikh and Vytlacil (2004) imposes all of our assumptions: the threshold crossing, binary choice model for both D and the outcome variable Y, as well as independence of (ϵ, ν) with Z. They do not, however, assume any parametric distributional assumption on (ϵ, ν), and do not require a linear index assumption. Thus, their assumptions are stronger than the assumptions imposed in the Manski IV bounds but are much weaker than the assumptions required for a bivariate probit. Using the model and independence assumptions,

$$\begin{aligned} \Pr[Y = 1 \mid Z] &= \Pr[D = 1, Y = 1 \mid Z] + \Pr[D = 0, Y = 1 \mid Z] \\ &= \Pr[D = 1, Y_1 = 1 \mid Z] + \Pr[D = 0, Y_0 = 1 \mid Z] \\ &= \Pr\left[\nu \leq s(Z), \epsilon \leq r(1)\right] + \Pr\left[\nu > s(Z), \epsilon \leq r(0)\right]. \end{aligned}$$

Recall that we have ordered Z so that $\Pr[D = 1 \mid Z = 1] > \Pr[D = 1 \mid Z = 0]$. Given our model and independence assumptions, $\Pr[D = 1 \mid Z = 1] > \Pr[D = 1 \mid Z = 0]$ implies $s(1) > s(0)$. Thus, if $r(1) > r(0)$,

$$\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0] = \Pr\left[s(0) < \nu \leq s(1), r(0) < \epsilon \leq r(1)\right];$$

and if $r(1) < r(0)$ then

$$\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0] = -\Pr\left[s(0) < \nu \leq s(1), r(1) < \epsilon \leq r(0)\right].$$

This analysis is graphically illustrated in Figure 1 for the case where $r(1) > r(0)$. Thus $\Pr[Y = 1 \mid Z = 1] \geq \Pr[Y = 1 \mid Z = 0]$ indicates that $r(1) \geq r(0)$. Likewise, $\Pr[Y = 1 \mid Z = 1] \leq \Pr[Y = 1 \mid Z = 0]$ indicates that $r(1) \leq r(0)$. From the structure of the threshold crossing model on Y, $r(1) \geq r(0)$ implies that $Y_1 \geq Y_0$ and $r(1) \leq r(0)$ implies that $Y_1 \leq Y_0$. We can thus improve upon the Manski results. If, for example, $\Pr[Y = 1 \mid Z = 1] \geq \Pr[Y = 1 \mid Z = 0]$, then $\Pr[Y = 1 \mid D = 1, Z]$ can be used as an upper bound on the unidentified term $\Pr[Y_0 = 1 \mid D = 1, Z]$ instead of 1; and $\Pr[Y = 1 \mid D = 0, Z]$ can be used as a lower bound on the unidentified term $\Pr[Y_1 = 1 \mid D = 0, Z]$ instead of 0. The resulting bounds are:

$$B^L_{SV} \leq E(Y_1 - Y_0) \leq B^U_{SV}$$

where, if $\Pr[Y = 1 \mid Z = 1] > \Pr[Y = 1 \mid Z = 0]$, then

$$\begin{array}{rcl} B^L_{SV} & = & \Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0] \\ B^U_{SV} & = & \Pr[D = 1, Y = 1 \mid Z = 1] + \Pr[D = 0 \mid Z = 1] - \Pr[D = 0, Y = 1 \mid Z = 0]; \end{array} \tag{3}$$

and if $\Pr[Y = 1 \mid Z = 1] < \Pr[Y = 1 \mid Z = 0]$, then

$$\begin{array}{rcl} B^L_{SV} & = & \Pr[D = 1, Y = 1 \mid Z = 1] - \Pr[D = 0, Y = 1 \mid Z = 0] - \Pr[D = 1 \mid Z = 0] \\ B^U_{SV} & = & \Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0]; \end{array} \tag{4}$$

and $B^L_{SV} = B^U_{SV} = 0$ if $\Pr[Y = 1 \mid Z = 1] = \Pr[Y = 1 \mid Z = 0]$.

Consider the difference between these bounds and the Manski IV bounds. In general, the Manski IV bounds do not simplify to a form that makes comparison with the Shaikh-Vytlacil bounds convenient. However, given our assumption that the treatment is given by a threshold crossing model, we have by the analysis of Heckman and Vytlacil (2001) that the Manski IV bounds simplify to the following form:

$$B^L_M = \Pr[D = 1, Y = 1 | Z = 1] - \Pr[D = 0, Y = 1 | Z = 0] - \Pr[D = 1 \mid Z = 0]$$

$$B^U_M = \Pr[D = 1, Y = 1 | Z = 1] + \Pr[D = 0 \mid Z = 1] - \Pr[D = 0, Y = 1 | Z = 0].$$

Suppose that $\Pr[Y = 1 \mid Z = 1] \geq \Pr[Y = 1 \mid Z = 0]$. In this case, $B^U_{SV} = B^U_M$. In other words, the upper bound on average treatment effect has not changed. The lower bound, however, has now changed to $\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0]$. Notice that $B^L_{SV} - B^L_M = \Pr[D =$

$0, Y = 1 \mid Z = 1] - \Pr[D = 1, Y = 1 \mid Z = 0] + \Pr[D = 1 \mid Z = 0] = \Pr[D = 0, Y = 1 \mid Z = 1] + \Pr[D = 1, Y = 0 \mid Z = 0] \geq 0$, so the Shaikh and Vytlacil lower bound is larger than the Manski lower bound. If, on the other hand, $\Pr[Y = 1 \mid Z = 1] \leq \Pr[Y = 1 \mid Z = 0]$, the lower bound is the same, while the Shaikh and Vytlacil upper bound is now smaller than the Manski upper bound.

Notice further that if $\Pr[Y = 1 \mid Z = 1] > \Pr[Y = 1 \mid Z = 0]$, then the lower bound on the average treatment effect is $\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0] > 0$; conversely, if $\Pr[Y = 1 \mid Z = 1] < \Pr[Y = 1 \mid Z = 0]$, then the upper bound on the average treatment effect is $\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0] < 0$. Thus, the bounds always lie on one side of zero, unless $\Pr[Y = 1 \mid Z = 1] = \Pr[Y = 1 \mid Z = 0]$, in which case the average treatment effect is point identified to be zero.

Unlike the Manksi IV bounds, the Shaikh and Vytlacil (2004) bounds cannot be consistently estimated simply by substituting the conditional sample means for the conditional population means. The reason is that their bounds are a highly discontinuous function of $\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0]$ at the origin and thus implementing the empirical analog of the population Shaikh and Vytlacil bounds will result in an estimator that is inconsistent if $\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0] = 0$.[12]

It is, however, possible to circumvent this difficulty and construct a consistent estimator of the bounds in the following way. Let $\epsilon_n \searrow 0$ be any sequence of constants such that $\sqrt{n}\epsilon_n \nearrow \infty$. If the sample analogue of $\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0]$ is larger than ϵ_n, then use the sample analogue of (3) as the estimate of the bounds; if the sample analogue of $\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0]$ is less than $-\epsilon_n$, then use the sample analogue of (4) as the estimate of the bounds; otherwise, use $\{0\}$ as the estimate of the bounds. If $\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0] > 0$, then the rate restriction on the ϵ_n sequence implies that with probability approaching 1 the sample analogue will eventually be larger than ϵ_n. Similarly, when $\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0] < 0$ or $= 0$, we have that with probability approaching 1 the sample analogue of $\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0]$ will be less than $-\epsilon_n$ or fall in $(-\epsilon_n, \epsilon_n)$, respectively. Therefore, the estimate of the bounds constructed in this way will be consistent. Since this argument only imposes a rate restriction on the ϵ_n sequence, it will be implicit in our later analysis that we have chosen the level of the ϵ_n sequence in a way so that the sample analogue of $|\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0]|$ is always larger than ϵ_n.

The Shaikh and Vytlacil analysis is related to the analysis of Manski and Pepper (2000). Manski and Pepper consider combining a weakened instrumental variable assumption ("monotone instrumental variables", MIV) with a "monotone treatment response" (MTR) assumption. The MTR assumption is that one knows *a priori* that $Y_1 \geq Y_0$ for all individuals or one knows *a*

[12] If one is willing to believe *a priori* that $E(Y_1 - Y_0) \neq 0$, then $\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0] \neq 0$ and the Shaikh-Vytlacil bounds can be consistently estimated simply by substituting conditional sample means for the conditional population means.

priori that $Y_0 \geq Y_1$ for all individuals. In the present context of the effect of catheterization on mortality where much of the debate focuses on whether the average effect of catheterization is positive, negative, or zero, imposing the MTR is not attractive since it would involve imposing the answer to the question of interest. In comparison, the Shaikh and Vytlacil bounds identify the sign of the average treatment effect from the data and do not impose it *a priori*. Shaikh and Vytlacil (2004) compare their bounds to the Manski and Pepper (2000) bounds that would result from imposing MTR and the Manski IV assumption. They show that if the treatment effect is positive, then the Shaikh and Vytlacil bounds coincide with the Manski and Pepper bounds that impose *a priori* a positive effect and impose the Manski IV assumption. If the treatment effect is negative, then the Shaikh and Vytlacil bounds coincide with the Manski and Pepper bounds that impose *a priori* a negative effect and impose the Manski IV assumption. Thus, there is a tight link between the Manski and Pepper bounds and the Shaikh and Vytlacil bounds, with the tradeoff that the Manski and Pepper analysis requires that one knows *a priori* the sign of the treatment effect but does not impose the threshold crossing structure imposed by Shaikh and Vytlacil.

Finally, we consider an extension to Shaikh and Vytlacil analogous to the idea of monotone treatment selection in Manski and Pepper (2000). Consider adding the additional assumption that doctors catheterize individuals with systematically worse latent health. In particular, assume that individuals with unobserved (by the econometrician) characteristics that make them more likely to be catheterized (have a low value of ν) are individuals with unobserved (by the econometrician) characteristics that make them more likely to suffer mortality (have a low values of ϵ). More formally, assume that ϵ and ν are positive quadrant dependent so that[13]

PQD: $\Pr[\epsilon \leq t_0 \mid \nu \leq t_1] \geq \Pr[\epsilon \leq t_0]$ for all t_0, t_1.

In other words, the probability that ϵ is small conditional on ν being small is at least as large as the unconditional probability of ϵ being small. One can easily show that PQD is equivalent to the assumption that $\Pr[\epsilon \leq t_0 \mid \nu \leq t_1] \geq \Pr[\epsilon \leq t_0 \mid \nu > t_1]$ for all t_0, t_1, i.e., equivalent to the statement that individuals receiving catheterization are at least as likely to have poor latent health as individuals who do not receive catheterization. Using the PQD assumption, we then have that

$$\begin{aligned} \Pr[Y_1 = 1|D = 1, Z] &= \Pr\left[\epsilon \leq r(1)|\nu \leq s(Z)\right] \\ &\geq \Pr\left[\epsilon \leq r(1)|\nu > s(Z)\right] \\ &= \Pr[Y_1 = 1|D = 0, Z] \end{aligned}$$

where the inequality is exploiting the PQD assumption. Thus, we can bound $\Pr[Y_1 = 1|D = 0, Z]$ from above by $\Pr[Y = 1|D = 1, Z]$. Following symmetric reasoning, we can bound

[13]Positive quadrant dependence is a relatively weak measure of positive dependence between two random variables. See Joe (1997) for the relationship between positive quadrant dependence and other concepts of positive dependence.

$\Pr[Y_0 = 1|D = 1, Z]$ from below by $\Pr[Y = 1|D = 0, Z]$. Combining this analysis with the Shaikh and Vytlacil analysis, we have that, if $\Pr[Y = 1 \mid Z = 1] \geq \Pr[Y = 1 \mid Z = 0]$, then

$$\begin{aligned}
\Pr[Y = 1 \mid D = 1, Z] \geq &\quad \Pr[Y_1 = 1 \mid D = 0, Z] &&\geq \Pr[Y = 1 \mid D = 0, Z] \\
\Pr[Y = 1 \mid D = 1, Z] \geq &\quad \Pr[Y_0 = 1 \mid D = 1, Z] &&\geq \Pr[Y = 1 \mid D = 0, Z];
\end{aligned}$$

if $\Pr[Y = 1 \mid Z = 1] \leq \Pr[Y = 1 \mid Z = 0]$, then

$$\begin{aligned}
\min\{\Pr[Y = 1 \mid D = 1, Z], \Pr[Y = 1 \mid D = 0, Z]\} \geq &\quad \Pr[Y_1 = 1 \mid D = 0, Z] &&\geq 0 \\
\max\{\Pr[Y = 1 \mid D = 1, Z], \Pr[Y = 1 \mid D = 0, Z]\} \leq &\quad \Pr[Y_0 = 1 \mid D = 1, Z] &&\leq 1.
\end{aligned}$$

We can now bound $\Pr[Y_0 = 1]$ and $\Pr[Y_1 = 1]$. For example, if $\Pr[Y = 1 \mid Z = 1] > \Pr[Y = 1 \mid Z = 0]$ then $\Pr[Y_1 = 1]$ can be bounded from above by the following argument:

$$\begin{aligned}
\Pr[Y_1 = 1] &= \Pr[Y_1 = 1 \mid Z = z] \\
&= \Pr[D = 1 \mid Z = z] \Pr[Y_1 = 1 \mid D = 1, Z = z] \\
&\quad + \Pr[D = 0 \mid Z = z] \Pr[Y_1 = 1 \mid D = 0, Z = z] \\
&\leq \Pr[Y = 1 \mid D = 1, Z = z].
\end{aligned}$$

Since this upper bound hold for $z = 0, 1$, we have

$$\Pr[Y_1 = 1] \leq \min_z \{\Pr[Y = 1 \mid D = 1, Z = z]\}.$$

Following an argument analogous to those given in Shaikh and Vytlacil (2004), one can show that, given our assumptions,

$$\min_z \{\Pr[Y = 1 \mid D = 1, Z = z]\} = \Pr[Y = 1 \mid D = 1, Z = 1].$$

Following this line of reasoning, one can show the following bounds on $E(Y_1 - Y_0)$:

$$B^L_{PQD} \leq E(Y_1 - Y_0) \leq B^U_{PQD}$$

where, if $\Pr[Y = 1 \mid Z = 1] > \Pr[Y = 1 \mid Z = 0]$, then

$$\begin{aligned}
B^L_{PQD} &= \Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0] \\
B^U_{PQD} &= \Pr[Y = 1 \mid D = 1, Z = 1] - \Pr[Y = 1 \mid D = 0, Z = 0],
\end{aligned}$$

and if $\Pr[Y = 1 \mid Z = 1] < \Pr[Y = 1 \mid Z = 0]$, then

$$B^L_{PQD} = \Pr[D = 1, Y = 1 \mid Z = 1] - \Pr[D = 0, Y = 1 \mid Z = 0] - \Pr[D = 1 \mid Z = 0]$$

$$\begin{aligned}
B^U_{PQD} = &\Pr[D = 1, Y = 1 \mid Z = 1] \\
&+ (\Pr[D = 0 \mid Z = 1] \times \min\{\Pr[Y = 1 \mid D = 1, Z = 1], \Pr[Y = 1 \mid D = 0, Z = 1]\}) \\
&\qquad - \Pr[D = 0, Y = 1 \mid Z = 0] \\
&\quad - (\Pr[D = 1 \mid Z = 0] \times \max\{\Pr[Y \mid D = 1, Z = 0], \Pr[Y = 1 \mid D = 0, Z = 0]\}),
\end{aligned}$$

and $B^L_{PQD} = B^U_{PQD} = 0$ if $\Pr[Y = 1 \mid Z = 1] = \Pr[Y = 1 \mid Z = 0]$. Notice imposing our PQD assumption decreases the upper bound on $E(Y_1 - Y_0)$ compared to the Shaikh-Vytlacil bounds while leaving the lower bound unchanged. Notice that the reduction of the upper bound is nontrivial if $\Pr[Y = 1 \mid Z = 1] > \Pr[Y = 1 \mid Z = 0]$. As with the Shaikh-Vytlacil bounds, these bounds are a discontinuous function of the quantity $\Pr[Y = 1 \mid Z = 1] - \Pr[Y = 1 \mid Z = 0]$, which prohibits consistent estimation by simply substituting conditional sample means for their population counterparts, but consistent estimation of these bounds can be achieved by following a procedure similar to the one described above for the Shaikh-Vytlacil bounds.

Thus far, we have assumed that there are no X covariates and that Z is binary. Allowing for X covariates only changes the analysis in a trivial fashion – the same analysis now holds conditional on X. For estimation, if X is discrete or can be discretized, one can simply continue to use conditional means in place of conditional expectations.[14] If X is continuous, one needs to use a nonparametric smoothing estimator such as a kernel regression estimator to estimate the relevant conditional expectations. Allowing Z to take more than two values is again straightforward. Suppose that z_1 maximizes $P(z)$ and that z_0 minimizes $P(z)$. Then all of our analysis continues to hold simply by substituting the z_1 evaluation point for the 1 evaluation point and the z_0 evaluation point for the 0 evaluation point. For estimation, if Z is discrete, one can examine the sample means of D conditional on Z to determine which z value maximizes and minimizes $P(z)$. If Z is continuous, then, as in the case when X is continuous, one must resort to a more sophisticated nonparametric estimator.

5 Data

The Connors et al. (1996) data comes from five medical centers (the SUPPORT group), and were collected to study medical decision-making for severely ill adult patients admitted to hospitals and ICUs. To be admitted to the study, patients had to have a predicted six-month mortality rate of 50% or more at the time of admission, and have one of nine disease conditions: acute respiratory failure, chronic obstructive pulmonary disease, congestive heart failure, cirrhosis, nontraumatic coma, metastatic colon cancer, late-stage non-small cell lung cancer, and multiorgan system failure with malignancy or sepsis. Murphy and Cluff (1990) provide a detailed description of patient recruitment procedures, including a detailed list of exclusion criteria. Most prominent among these criteria are trauma, pregnancy, psychiatric disease, and AIDS.

The SUPPORT group's main research aim in collecting these data was to measure the effect of placing a Swan-Ganz catheter early in the course of hospitalization. They count a patient as having a catheter placed if the procedure was performed within 24 hours of

[14] Shaikh and Vytlacil (2004) consider how imposing that there is a covariate in X that is not contained in Z can be exploited to narrow the bounds on the average treatment effect. See also Vytlacil and Yildiz (2004). However, in our application there are no variables contained in X that are not contained in Z.

entering the hospital. The main reason for this focus and this definition of catheterization is their hypothesis that the biggest harm from catheterization comes from enabling ICU doctors to act too aggressively in prescribing medications and treatments with severe side-effects, inadvertently harming patients. The data includes 5,735 patients, all of whom were admitted to or transferred to the ICU within 24 hours of entering the hospital. By including only people who were admitted to the ICU, they allow the longest possible time for iatrogenic injury to occur. This empirical strategy has the additional advantage of reducing unobserved clinical heterogeneity among study patients. All patients in the sample share a common feature in their clinical history—they were sick enough to end up in an ICU soon after hospital admission.

Connors et al. (1996) collected a large amount of information about each patient via standardized medical chart abstraction methods and interviews with patients and patient surrogates. Chart abstractors focused mainly on a detailed set of laboratory and clinical variables, measured on the first through fourth weeks of the hospital stay. We use the information from the first week of the hospital stay, as it is collected for every patient in the sample. The interviews were used to collect both basic demographic information, as well as information about functional status (that is, whether the patient is unable to perform some basic tasks like bathing or dressing oneself).

Tables 1 - 3 compare patients who were catheterized during their first day of admission to the ICU with those who were not. These tables present mean values of each variable on which we make these comparisons, along with standard deviations of those variables, and the p-value from a test of the hypothesis that the means of the variables are equal.

From these statistics, it is clear that patients who have catheters placed differ from patients who do not along many clinically significant dimensions. Table 1 compares patients on the basis of demographic variables and the primary diagnosis at admission. Catheterized patients are more likely to be male (by 4.5%), privately insured (by 6.3%), richer (less likely to have an income of less than $11,000 per year by 6.2%), and have more schooling (0.3 years more on average). Catheterized patients are more likely to have congestive heart failure (CHF) as the primary admitting diagnosis (by 2.6%) and much more likely to have MOSF with sepsis as an admitting diagnosis (by 17.3%), but are less likely to have coma (by 5.3%) or chronic obstructive pulmonary disease (COPD—by 8.5%) as admitting diagnoses.

Table 2 compares patients on the basis of disease history prior to admission, and functional status measures as determined by the proxy interviews. Catheterized patients are more likely to have cardiac diagnoses (by 4.4%), including a heart attack (by 1.3%) in their medical history. They are less likely to have a malignant cancer (by 4.3%), dementia (by 4.7%), and other psychiatric diagnosis (by 3.5%) in their history. Connors et al. (1996) use a standard ICU mortality prediction model to summarize each patient's severity of illness upon admission to the ICU.[15] Catheterized patients have a 4% lower two month predicted survival rate upon admission than non-catheterized patients—clearly catheterized patients are observably more

[15]Connors et al. (1996) provide details about the calculations of predicted survival.

severely ill than non-catheterized patients. This fact can also be seen by comparing mean values of the acute physiology score, another standard ICU measure of severity of illness.

Table 3 compares catheterized and non-catheterized patients' laboratory values at admission, as well as any secondary diagnoses these patients may have had at admission. Among the laboratory values, all the clinically significant and interpretable differences point toward the conclusion that catheterized patients are observably sicker. For example, the ratio of serum oxygen levels to administered oxygen levels ($PAO_2/(0.01 * FiO_2)$) are higher among catheterized patients—an indication that they are more likely to require oxygen supplementation. Catheterized patients also have lower hematocrits and are thus more likely to be anemic. They have higher levels of bilirubin, and are thus more likely to have liver disease.

Among the secondary diagnoses, catheterized patients are 9% more likely to have sepsis—a serious blood-borne bacterial infection. They are also more likely to have renal secondary diagnoses (by 2.7%) and gastrointestinal secondary diagnoses (by 4.5%). On the other hand, they are less likely to have respiratory secondary diagnoses (by 12.8%) and neurological diagnoses (by 10.8%). On balance, these secondary diagnoses also suggest that catheterized patients were observably sicker on admission to the ICU.

Thus far we have ignored statistical complications stemming from the multiplicity of the comparisons we are making between catheterized and non-catheterized patients. Because of the large number of comparisons we are making, it is likely that at conventional significance levels, we will reject several hypotheses of equal means falsely. As a result of this concern, we use the Holm (1979) multiple hypothesis testing procedure to make the comparisons formally.[16] These results are summarized in Table 4 below. This procedure is more powerful than the conventional Bonferonni procedure for multiple hypothesis testing in the sense that it too controls the probability of even one false rejection at the desired level, but always rejects at least as many hypotheses.

At the .05 level, we find that patients who are catheterized differ from those who are not catheterized along 41 of 63 possible variables. If we restrict attention to particular primary diagnoses, we reject fewer hypotheses, but many important differences between the two groups still remain. Among patients whose primary diagnosis is congestive heart failure, for example, catheterized patients differ from non-catheterized patients among nine of 51 possible variables, including demographic characteristics like age and income as well as some clinical variables like blood pressure. If we restrict attention to only massive organ system failure diagnoses, we find that catheterized patients differ from non-catheterized patients among the following five of 50 possible variables: sepsis, creatinine levels, pulmonary function test results, acute physiology

[16] Suppose one is interested in testing the s hypotheses $H_1, \ldots, H_s$ using p-values $\hat{p}_1, \ldots, \hat{p}_s$ in a way that controls for the familywise error rate at a level α (that is, the probability of making one or more false rejections is no more than α). To this end, define the constants $\alpha_1 \leq \ldots \leq \alpha_s$ by the rule: $\alpha_i = \frac{\alpha}{s-i+1}$. Denote by $\hat{p}_{(1)} \leq \ldots \leq \hat{p}_{(s)}$ the ordered values of the p-values and by $H_{(1)}, \ldots, H_{(s)}$ the corresponding hypotheses. The Holm (1979) stepdown procedure accepts all hypotheses if $\hat{p}_{(1)} \geq \alpha_1$; otherwise, it rejects hypotheses $H_{(1)}, \ldots, H_{(r)}$, where r is the largest index such that $\hat{p}_{(1)} < \alpha_1, \ldots, \hat{p}_{(r)} < \alpha_r$.

score, and predicted survival. Finally, examining only patients whose primary diagnosis is colon cancer, we reject eight of 55 hypotheses, finding that the two groups differ noticeably among pulmonary function test results, respiratory rate, blood pressure, acute physiology score, and predicted survival. Hence, even after accounting for the multiplicity of the comparisons, we maintain our earlier conclusion that catheterized patients are significantly sicker than non-catheterized patients.

6 Instrumenting with Admission Day

It should be clear from Tables 1 - 3 that a direct comparison of outcomes between catheterized and non-catheterized patients is unlikely to yield the causal effects of catheterization. Even if a full set of controls, such as all the variables in those tables, were included in the analysis, the results would be unconvincing. If catheterized and non-catheterized patients differ on so many *observed* dimensions, it seems highly unlikely that they do not differ on *unobserved* dimensions as well.[17] In this section, we develop suggestive evidence that day-of-the-week of admission is an appropriate instrument to determine the causal effect of catheterization on patient mortality.

6.1 Admission Day of Week Predicts Catheterization

We first establish that patients who are admitted to the ICU on a Saturday, Sunday, or Monday are substantially less likely to be catheterized on the day of admission than patients admitted on other days of the week. Figure 2 shows how the unconditional probability that a patient is catheterized upon admission varies by day of the week. The lowest catheterization rates are on Saturday—approximately 32% of all SUPPORT ICU patients admitted on a Saturday had a Swan-Ganz catheter placed on day of admission. The catheterization rate rises on Sunday and Monday to about 36%, and then rises again on Tuesday through Friday to about 40%. The highest rates of about 41% are on Friday.

Figure 3 shows catheterization rates by day-of-the-week for four different and important clinical subgroups, delineated by their primary diagnosis upon admission to the ICU. For patients with congestive heart failure (CHF) or with multi-organ system failure (MOSF) with sepsis as an admitting diagnosis, the likelihood of being catheterized decreases dramatically on weekends. For patients with acute respiratory failure or MOSF with malignancy, the weekend decline in catheterization rates is less pronounced, though still substantial—it is roughly on par with the decline for the typical ICU patient.

Figure 4 shows catheterization rates by day-of-the-week for four other clinically important subgroups, again delineated by primary diagnosis upon ICU admission: chronic obstructive

[17] See Altonji et al. (2005) for a formal justification of this argument.

pulmonary disease (COPD), cirrhosis, coma, and lung cancer. For patients with these diagnoses, the pattern of higher catheterization probabilities on weekdays is less clear, and in some cases (such as for COPD), reversed. Since the instrument is apparently not strongly correlated with the treatment for these diagnosis groups, we do not attempt to estimate separate treatment effect bounds for these groups.

6.2 Patient Health and Day of Week of Admission

One disconcerting possibility for our study is that patients admitted to the ICU on a weekday differ systematically from patients admitted on weekends. If that were true, then day-of-the-week would be a poor candidate for an instrument since it would be correlated with unobserved determinants of ICU patient mortality, such as health status. Our introspection leads us to think that there should be no correlation between day-of-the-week of admission and unobserved health, since the health crises that precipitate ICU admissions are unlikely to respect distinctions between weekdays and weekends. However, we recognize that introspection alone is an insufficient basis on which to make our case.

Tables 5 - 7 divide patients into two groups on the basis of the day-of-the-week instrument: those who were admitted on a weekday and those who were admitted on a weekend.[18] The tables present mean values of each variable on which we make comparisons, along with standard deviations of those variables and the p-value from a test of the hypothesis that the means of the variables are equal.

Table 5 compares patients based on demographic variables on the primary diagnosis. We reject the hypothesis that patients admitted on a weekend are equally likely as those admitted on a weekday to have congestive heart failure as a diagnosis, but we find no other statistically significant differences between the two groups (when testing at the $\alpha = 0.01$ level). Table 6, on the other hand, compares patients based on disease history and functional status. Weekend patients are less likely to have cardiac disease and congestive heart failure in their medical history, which is consistent with the finding that they are less likely to have congestive heart failure as an admitting diagnosis. On the other hand, they have no other significant medical history or functional status measure differences. Importantly, there is little difference between these groups in predicted 2-month mortality, or in acute physiology score. Finally, Table 7 compares weekend and weekday patients on the basis of laboratory tests at admission and on secondary diagnoses. There are no statistically significant differences between the two groups on these bases.

Again, we have thus far failed to account for the multiplicity of the comparisons we are making. We do so now by applying the Holm (1979) multiple hypothesis testing procedure. These results are summarized in Table 4. Comparing patients admitted to the ICU on weekends

[18] We obtain similar results if we divide up the patients based upon whether they were admitted on Saturday or Sunday against those admitted on Monday to Friday.

with patients who are admitted on weekdays at the 0.05 level, we find that they differ only along three of 63 variables: diagnosis of congestive heart failure, cardiac history, and history of congestive heart failure. Otherwise, weekend and weekday patients appear to have similar observed severity of illness.

There is no reason *a priori* to believe that differences in these three variables imply that weekday patients are sicker, since congestive heart failure is no more serious a cause of admission to the ICU than the other causes (such as multi-organ system failure). Moreover, after restricting attention to particular primary diagnoses, even differences in these three variables essentially disappear. Among patients whose primary diagnosis is congestive heart failure, we find only one statistically significant difference.[19] Examining patients whose primary diagnosis is MOSF with malignancy or MOSF with sepsis, there are no statistically significant differences. Similarly, among patients whose primary diagnosis is acute respiratory failure, we find no statistically significant differences. Hence, we conclude that patients admitted on weekends are observably similar to patients admitted on weekdays, especially within particular primary diagnosis subgroups.

6.3 Day of Week, Hospital Staffing, and Outcomes

Though non-specialists sometimes find it shocking, it is well known in the health services literature that medical staffing can have a major effect on treatment decisions.[20] In a study aimed at understanding why some patients get catheterized, while others do not, Rapoport et al. (2000) emphasize both patient and hospital characteristics. In their data, patients admitted to ICUs that staff a full time ICU physician are much less likely to be catheterized (by two-thirds) than those admitted to ICUs with no full time physician.[21] Clearly, medical staffing decisions affect the decision to catheterize. However, Rapoport et al. (2000) do not report whether day-of-the-week of admission to the ICU helps determine the placement of a catheter.

Although the SUPPORT data do not permit a direct examination of the issue, it seems clear that the reason that catheterization rates vary by day-of-the-week relates closely to ICU staffing patterns. Medical personnel, just like most other workers, prefer having weekends off. The small increase in catheterization rates on Friday, relative to other weekdays, is probably due to an effort to reduce the stock of patients before the weekend. Whether this fact threatens

[19] Weekend ICU patients with congestive heart failure are more likely to earn between $11,000 and $25,000 per year; weekday ICU CHF patients are more likely to earn between $25,000 and $50,000 per year. This is not an important clinical difference.

[20] For example, Dickert-Conlin and Chandra (1999) report that babies are less likely to be born on weekends and holidays than on regular weekdays, which they attribute to staffing decisions by hospitals. In the United States, the least likely day for a baby to be born is Thanksgiving holiday. See also Bell and Redelmeier (2001).

[21] They also find that white patients, patients with private insurance coverage, and patients admitted to a surgical ICU (as opposed to a medical ICU) were more likely to be catheterized.

the validity of our instrument depends upon whether there are unobserved differences in treatment between weekday and weekend admissions, unassociated with catheterization, that help determine if a patient lives or dies. If so, then admission day would not be a valid instrument.

Evaluating the import of treatment differences (catheterization) between weekend and weekday admissions is made more complicated by the fact that Swan-Ganz catheterization itself is a gateway to a large number of other treatments. For example, ICU physicians often use the information from catheterization to titrate the dose of inotropic drugs, such as dopamine and dobutamine, that are designed to improve cardiac contractility. These drugs have a narrow therapeutic range, and are thus getting the dose right can be the difference between killing or inadequately treating a patient. Given our finding that catheterization is less likely on weekends, it would thus be unsurprising to find decreased use of inotropes on weekends as well. However, such treatment differences, which are causally related to the catheterization decision, do not pose a problem for our instrument. We can simply reinterpret our treatment effect bounds as measuring the effect of catheterization *and all the other treatments it enables or encourages* on mortality, rather than the isolated treatment effect of catheterization by itself.

However, if staffing differences lead to weekend-weekday treatment differences for those treatments that are not downstream from catheterization and which have substantial effects on the mortality of ICU patients, then our treatment effect bounds will be biased. Unfortunately, we cannot check this directly with the SUPPORT data that is available to us; while there is considerable laboratory and diagnostic information, the only treatment information available is on Swan-Ganz catheterization.

There is, however, an extensive medical literature on the effects of weekend-weekday staffing differences on ICU mortality risk that suggests that (non-Swan-Ganz related) weekend-weekday treatment differences are not important for patient outcomes. For example, Ensminger et al. (2004) analyze 29,084 ICU patients from the Mayo Clinic. They find that, while patients with weekend ICU admissions had a 4.3 percentage point higher in-hospital mortality when compared with ICU patients admitted on a weekday, this difference can be explained entirely by weekend-weekday differences in patient health status. Similarly, Wunsch et al. (2004) find that raw weekday-weekend in-hospital mortality differences disappear in models that control flexibly for patient health status.

Looking at mortality in the emergency room, rather than in the ICU, Dobkin (2003) finds that controlling for observed patient characteristics alone does not erase weekend-weekday mortality differences one day after hospitalization. He argues that unobserved weekday-weekend differences in patient health status, caused by the triaging of more severely ill patients on a weekday, explain this mortality difference entirely. He concludes that there is "no evidence of excess mortality on the weekend in the California hospital system for people admitted through the Emergency Department. This is despite significant delay in diagnostic and treatment procedures for patients admitted on the weekend."

This literature suggests, though it does not prove, that the effect of weekday-weekend ICU treatment differences have small effects on in-hospital or one-day post admission mortality, at least in the samples examined. The best explanation for this finding is that medical personnel do a reasonable job triaging patients, and thus it is patients most likely to benefit from catheterization who will be catheterized on Saturday, Sunday, or Monday.[22] If doctors successfully identify the most severely ill weekend patients for more intensive intervention, then the effect of decreased staffing on the weekend on patient mortality in the short run (until a weekday arrives) will be small. This does not mean that catheterization, or other treatments, have a small mortality effect at other horizons after hospitalization. It does suggest that the existence of weekend-weekday staffing differences do not, by themselves invalidate our instrument.

7 Results

In this section, we show the results from two different approaches to estimating treatment effects. In the first, more traditional approach, we estimate linear parametric models by ordinary least squares and report confidence intervals around the treatment effect estimates implied by those models. In the second, we report estimates from the three different bounding strategies that we have considered in this paper: the Manski IV bounds, the Shaikh-Vytlacil bounds, and our extension of the Shaikh-Vytlacil analysis that additionally imposes the PQD assumption.

7.1 Linear Parametric Models

In Section 5, we have seen that catheterized patients differ in many ways from non-catheterized patients. Figure 5 shows 95% confidence intervals around treatment effect estimates from an estimator that ignores these differences altogether. The main outcome that we analyze is mortality at varying periods after admission to the ICU—7 days, 30 days, 60 days, 90 days, and 180 days. Let $Y(t)$ represent the event of a patient dying t or more days after admission to the ICU. Let D represent the event of a patient being catheterized on the day of admission to the ICU. Figure 5 shows the 95% confidence interval around the coefficient obtained from a ordinary least squares (OLS) linear regression of $Y(t)$ on D for different values of t. This regression coefficient represents the estimated treatment effect of catheterization: $\Pr[Y(t) = 1|D = 1] - \Pr[Y(t) = 1|D = 0]$. The panel on the right shows the treatment effect estimates for a regression with no covariates other than D, while the panel on the left shows these estimates for a regression with the full set of covariates listed in Tables 1 - 3. In all

[22] It is not clear to us why Monday catheterization rates are more similar to Sunday rates than they are to other weekday rates, though that is clearly what the data suggest. Accordingly, we will refer to admissions on Saturday, Sunday, and Monday as "weekend" admissions.

cases, the point estimate of the treatment effect indicates that catheterized patients are more likely to die. Including health covariates reduces the size of the increase in mortality from catheterization but does not alter its sign. Based upon these results, one might conclude that excess mortality due to catheterization 7 days after ICU admission is about two percentage points, and that it increases to about six percentage points by 30 days after admission.

Figure 6 shows 95% confidence intervals around treatment effect estimates from linear regressions for subsets of patients with different diagnoses at admission. For ICU patients with MOSF with sepsis as the primary admitting diagnosis, the regression results suggest catheterization increases mortality. The inclusion of covariates in the regression does not alter this finding. For patients with congestive heart failure, the results without covariates suggest the same conclusion, but when covariates are included, the confidence intervals span zero. For patients with multi-organ system failure with malignancy, the point estimate at 7 days suggests that catheterization may be clinically beneficial, though the result is not statistically significant. Between 30 and 90 days, one cannot reject a hypothesis of no effect (in the model with covariates included).

The problem with the OLS estimates is that they ignore the possibility of unobserved selection. That the inclusion of covariates makes such a large difference in the treatment effect estimates suggests substantial bias due to important controls omitted from the regression. We attempt to address this deficiency with instrumental variables methods, using admission day-of-week as our instrument. Because the error term of the model is not additively separable from the regressors, standard linear instrumental variables is not a consistent estimator of the average effect of catheterization in this context and in general access to an instrument does not allow one to point identify or point estimate the average effect of catheterization on mortality. We thus turn instead to bounding approaches that exploit access to an instrument.

7.2 Bounding Approaches

We consider three alternative sets of IV-bounds. Figure 7 shows the Manski and the Shaikh-Vytlacil bounds for the whole sample, calculated using an indicator for whether the patient was admitted to the ICU on a weekday as the instrumental variable. The panel on the left represents the Manski IV bounds, while the panel on the right represents the Shaikh-Vytlacil IV bounds. The Manski IV bounds are uniformly wide, and always straddle zero; in fact they are little better than the naive Manski bounds, which have a width of one (since $Y(t)$ is bounded above by one and below by zero). The Shaikh-Vytlacil IV bounds, by construction, do not straddle zero. The Shaikh-Vytlacil IV bounds suggest that catheterization reduces mortality at 7 days, while it increases mortality at 30 days and later. Recall that Connors et al. (1996) found that catheterization increases mortality at 7 days using this same data set that we use here, but a different statistical method that assumes that there are no unobserved differences between catheterized and non-catheterized patients. The width of the Shaikh-Vytlacil bounds are about half the width of the Manski IV bounds.

Figure 8 shows the Manski IV and Shaikh-Vytlacil IV bounds for the same diagnosis subgroups as in the figures above: acute respiratory failure, congestive heart failure (CHF), MOSF with malignancy, and MOSF with sepsis. The Manski IV bounds are uniformally uninformative with a width of nearly one for all these diagnosis groups. The Shaikh-Vytlacil results, by contrast, in many cases provide some clear guidance for ICU doctors. They suggest that:

- Catheterization reduces mortality at 7 days for patients with an admitting diagnosis of MOSF with malignancy or CHF, but increases it at 30 days and after. The same results hold for ICU patients as a whole (recall Figure 7).
- Catheterization increases mortality at all horizons for patients with MOSF with sepsis or acute respiratory failure, though the lower end of the bound is close to zero.[23]

Finally, Figures 9 and 10 show the PQD bounds,

$$B^L_{PQD} \leq E(Y_1 - Y_0) \leq B^U_{PQD},$$

which impose the restriction that doctors are effective at triaging patients so that it is those patients with the worst health who are actually catheterized. These figures show that imposing this seemingly plausible restriction decreases the width of the treatment effect bounds, often dramatically. Of course, by construction, the PQD bounds are always on the same side of zero as the Shaikh-Vytlacil bounds. The reduction in bound width is greatest when the treatment effect estimate is positive, that is when the finding is that catheterization increases mortality. This is to be expected, as the PQD restriction rules out the possibility that doctors cause great harm to large numbers of their patients. The reduction in bound width is smaller when the treatment effect estimate suggests catheterization saves lives, but clinically important. If catheterization saves lives, yet has a near zero effect, it might not be worthwhile given the monetary costs. Since the PQD bounds (relative to the Shaikh-Vytlacil bounds) push the upper bound further away from zero when catheterization saves lives, it may permit researchers to render judgement that catheterization is cost-effective even as the Shaikh-Vytlacil bounds permits the possibility that it may not be.

8 Conclusion

While direct comparisons of the mortality of catheterized and non-catheterized patients lead to the conclusion that catheterization increases mortality, we show evidence that this result is

[23] Recall that by construction, the Shaikh-Vytlacil bounds can never cross zero. If the estimate of $\Pr[Y = 1|Z = 1] - \Pr[Y = 1|Z = 0]$ is sufficiently close to zero, the estimate of the bounds will be the point zero, i.e., point estimation at zero. For any finite N, how one defines close to zero is arbitrary, and we have decided that the estimated $\Pr[Y = 1|Z = 1] - \Pr[Y = 1|Z = 0]$ is sufficiently far from zero in all cases not to set the bounds to point estimation at zero.

due to profound differences between the catheterized and non-catheterized patients; the former are much more severely ill than the latter.

We provide suggestive evidence that weekday admission can serve as an instrumental variable for catheterization. Patients admitted on a weekday are about four to eight percentage points more likely to be catheterized than patients admitted on a weekend. Yet, weekday and weekend patients appear similar in health status along a large number of dimensions. Exploiting an instrumental variable permits us to address the unobserved differences between catheterized and non-catheterized ICU patients.

We turn to bounding approaches that exploit access to our instrument, including a new bounds estimator introduced by Shaikh and Vytlacil (2004), which we compare with the Manski (1990) IV bounds. We also extend the analysis of Shaikh and Vytlacil (2004) to consider bounds that impose the additional assumption that doctors catheterize individuals with systematically worse latent health We find that, while the Manski bounds are almost entirely uninformative, the Shaikh-Vytlacil bounds typically produces a clearer answer—catheterization reduces mortality at 7 days, and increases it at 30 days and after, though the lower bound of the treatment effect is always close to zero. Applying a further nonparametric structural assumption that doctors catheterize individuals with worse health further narrows these bounds and strengthens these conclusions.

Our primary finding that catheterization improves mortality outcomes only in the short run, and may increase it in the long run is intuitively appealing because it suggests an explanation for the fact that many ICU doctors are deeply committed to the use of the Swan-Ganz catheter. Since most ICU patients leave the ICU well before 30 days after admission have elapsed, ICU doctors never observe the increase in mortality. They do, however, observe the decline in mortality at 7 days.

Catheterization has long been a critical tool for ICU doctors, thought to help stabilize severely ill patients. The information on heart and lung function that catheterization yields can often lead doctors to the right treatment that keeps a patient alive in the ICU. Our results suggest two possibilities, which are not mutually exclusive. The treatments and medications that doctors feel emboldened to administer by the Swan-Ganz catheter may stabilize patients in the short run, but may have malign longer run effects. Connors et al. (1996) suggest this mechanism to explain their finding of longer run mortality increases from catheterization, but their finding of increases in short run mortality leaves open the mystery of why doctors would ever apply such stabilizing treatments. A second possibility is a simple selection story. Catheterization saves the lives, in the short run, of the most severely ill patients, but the deaths of these patient cannot be staved off for long. Disentangling these possibilities will require even more detailed data and further research. In addition, formal inference for the Shaikh-Vytlacil bounds and our extension of them has not yet been developed and is an additional topic for further research.

References

Altonji, J., Elder, T., and Taber, C. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*. forthcoming.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91:444–455.

Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92:1171–1176.

Bell, C. and Redelmeier, D. (2001). Mortality among patients admitted to hospitals on weekends as compared with weekdays. *New England Journal of Medicine*, 345(9):663–668.

Connors, A., Speroff, T., Dawson, N., Thomas, C., Harrell, F., Wagner, D., Desbiens, N., Goldman, L., Wu, A., Califf, R., Fulkerson, W. J., Vidaillet, H., Broste, S. Bellamy, P., Lynn, J., and Knaus, W. (1996). The effectiveness of right heart catheterization in the initial care of critically ill patients. support investigators. *Journal of the American Medical Association*, 276(11):889–897.

Dalen, J. (2001). The pulmonary arter catheter-friend, foe, or accomplice? *Journal of the American Medical Association*, 286(3):348–350.

Dickert-Conlin, S. and Chandra, A. (1999). Taxes and the timing of births. *Journal of Political Economy*, 107(1):161–177.

Dobkin, C. (2003). Hospital staffing and inpatient mortality. mimeo, UC Santa Cruz.

Ensminger, S., Morales, I., Peters, S., Keegan, M., Finkielman, J., Lymp, J., and Afessa, B. (2004). The hospital mortality of patients admitted to the icu on weekends. *Chest*, 126(4):1292–1298.

Fowler, R. and Cook, D. (2003). The arc of the pulmonary artery catheter. *Journal of the American Medical Association*, 290(20):2732–2734.

Gore, J., Goldberg, R., Spodick, D., Aplert, J., and Dalen, J. (1985). A community-wide assessment of the use of pulmonary artery catheters in patients with acute myocardial infarction: A prospective autopsy study. *Chest*, 88:567–572.

Guyatt, G. (1991). A randomized trial of right heart catheterization in critically ill patients. *Journal of Intensive Care Medicine*, 6:91–95. for the Ontario Intensive Care Study Group.

Hamilton, B., Ho, V., and Goldman, D. (2000). Queing for surgery: Is the u.s. or canada worse off? *Review of Economics and Statistics*, 82(2):297–308.

Heckman, J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 46:931–959.

Heckman, J. and Vytlacil, E. (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences*, 96:4730–4734.

Heckman, J. and Vytlacil, E. (2000). The relationship between treatment parameters within a latent variable framework. *Economics Letters*, 66(1):33–39.

Heckman, J. and Vytlacil, E. (2001). Instrumental variables, selection models, and tight bounds on the average treatment effect. In Lechner, M. and Pfeiffer, F., editors, *Econometric Evaluations of Active Labor Market Policies in Europe*. Physica-Verlag, Heidelberg; New York:.

Hirano, K. and Imbens, G. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes Research Methodology*, 2:259–278.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(65-70).

Hotz, J., Mullins, C., and Sanders, S. (1997). Bounding causal effects using data from a contaminated natural experiment: Analyzing the effects of teenage childbearing. *Review of Economic Studies*, 64:575–603.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, New York.

Manski, C. (1989). Anatomy of the selection problem. *Journal of Human Resources*, 24:343–360.

Manski, C. (1990). Nonparametric bounds on treatment effects. *American Economic Review, Papers and Proceedings*, 80:319–323.

Manski, C. and Pepper, J. (2000). Monotone instrumental variables: With an application to the returns to schooling. *Econometrica*, 68:997–1010.

McClellan, M., McNeil, B., and Newhouse, J. (1994). Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? analysis using instrumental variables. *Journal of the American Medical Association*, 272(11):859–866.

Murphy, D. and Cluff, L. (1990). Support: Study to understand prognoses and preferences for outcomes and risks of treatments-study design. *Journal of Clinical Epidemiology*, 43(suppl):1S–123S.

Polanczyk, C., Rohde, L., Goldman, L., Cook, E., Thomas, E., Marcantonio, E., Mangione, C., and Lee, T. (2001). Right heart catheterization and cardiac complications in patients undergoing noncardiac surgery: an observational study. *Journal of the American Medical Association*, 286(3):309–314.

Rapoport, J., Teres, D., Steingrub, J., Higgins, T., McGee, W., and Lemeshow, S. (2000). Patient characteristics and ICU organizational factors that influence frequency of pulmonary artery catheterization. *Journal of the American Medical Association*.

Richard, C., Warszawski, J., Anguel, N., Deye, N., Combes, A., Barnoud, D., Boulain, T., Lefort, Y., Fartoukh, M., Baud, F., Boyer, A., Brochard, L., and Teboul, J. (2003). Early use of the pulmonary artery catheter and outcomes in patients with shock and acute respiratory distress syndrome: A randomized controlled trial. *Journal of the American Medical Association*. for the French Pulmonary Artery Catheter Study Group.

Sandham, J., Hull, R., Brant, R., Knox, L., Pineo, G., Doig, C., Laporta, D., Viner, S., Passerini, L., Devitt, H., Kirby, A., and Jacka, M. (2003). A randomized, controlled trial of the use of pulmonary-artery catheters in high-risk surgical patients. *New England Journal of Medicine*. for the Canadian Critical Care Clinical Trials Group.

Shaikh, A. and Vytlacil, E. (2004). Limited dependent variable models and bounds on treatment effects: A nonparametric analysis. mimeo, Stanford University.

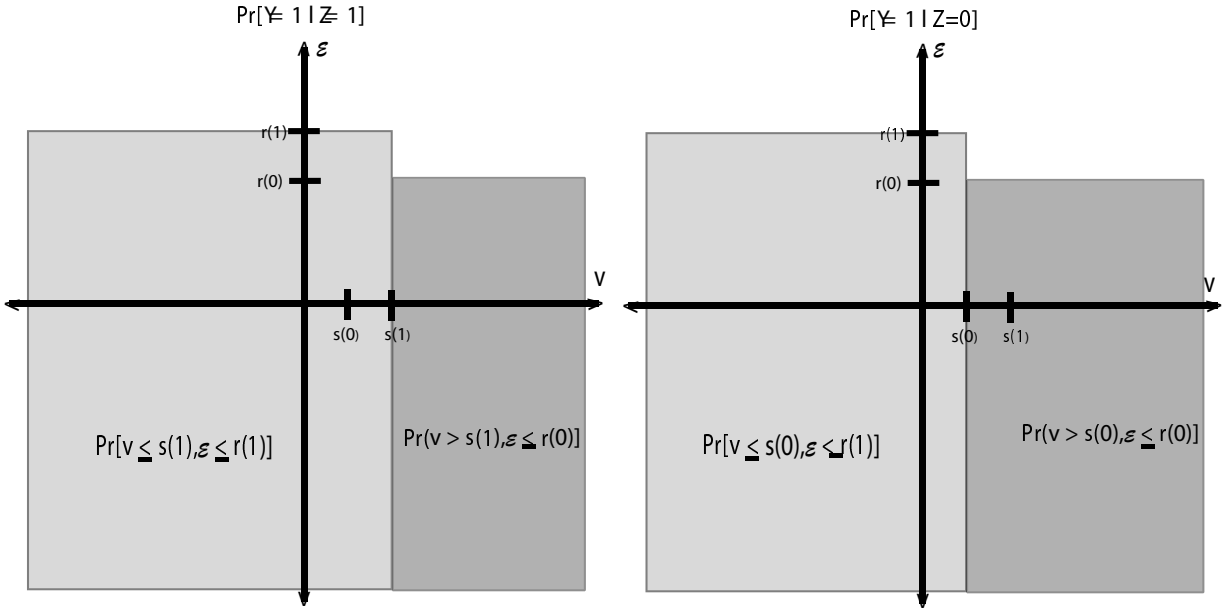
Vytlacil, E. and Yildiz, N. (2004). Dummy endogenous variables in weakly separable models. mimeo, Stanford University.

Weil, M. (1998). The assault on the swan-ganz catheter: A case history of constrained technology, constrained bedside clinicians, and constrained monetary expenditures. *Chest*, 113(5):1379–1386.

Wunsch, H., Mapstone, J., Brady, T., Hanks, R., and Rowan, K. (2004). Hospital mortality associated with day and time of admission to intensive care units. *Intensive Care Medicine*, 30(5):895–901.

Zion, M., Balkin, J., Rosenmann, D., Goldbourt, U., Reicher-Reiss, H., Kaplinsky, E., and Behar, S. (1990). Use of pulmonary artery catheters in patients with acute myocardial infarction. analysis of experience in 5,841 patients in the sprint registry. *Chest*, 98:1331–1335.

Figure 1:


Pr[Y= 1 | Z= 1]
ε
r(1)
r(0)
v
s(0)
s(1)
Pr[v ≤ s(1),ε ≤ r(1)]
Pr(v > s(1),ε ≤ r(0)]
Pr[Y= 1 | Z=0]
ε
r(1)
r(0)
v
s(0)
s(1)
Pr[v ≤ s(0),ε ≤ r(1)]
Pr(v > s(0),ε ≤ r(0)]

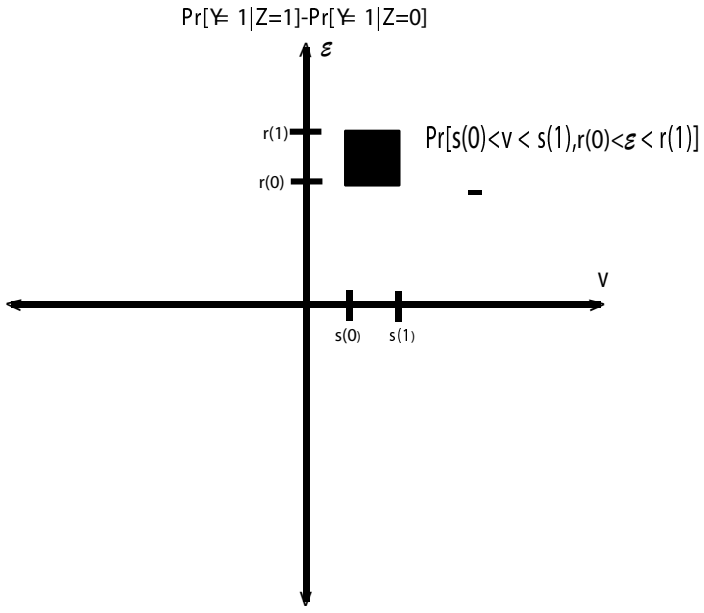

Pr[Y= 1|Z=1]-Pr[Y= 1|Z=0]
ε
r(1)
r(0)
Pr[s(0)<v < s(1),r(0)<ε < r(1)]
v
s(0)
s(1)

Figure 2: % Catheterized by Day-of-Week of Admission to ICU

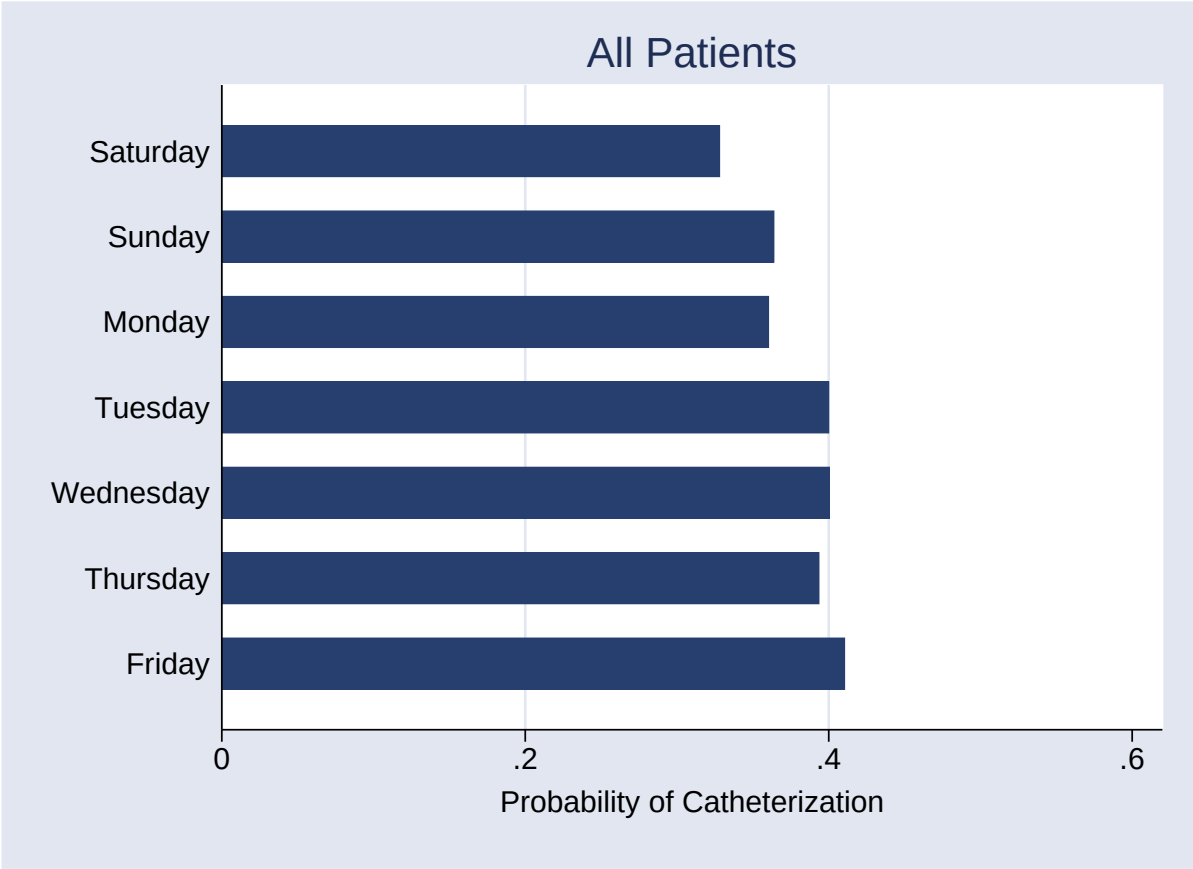


Table 1: Catheterized vs. Not Catheterized; Demographic and Diagnostic Comparisons

Variable	Not Catheterized	Catheterized	p-value
N	3,551 (62%)	2,184 (38%)	
Age	61.8	60.8	
	[17.3]	[15.6]	0.0258
Male	53.9%	58.5%	
	[0.499]	[0.493]	0.0006
Black	16.5%	15.3%	
	[0.371]	[0.360]	0.2553
Other Race	6.0%	6.5%	
	[0.237]	[0.247]	0.4423
Years of Education	11.569	11.856	
	[3.134]	[3.157]	0.0008
No Insurance	5.2%	6.2%	
	[0.223]	[0.242]	0.1141
Private Insurance	27.2%	33.5%	
	[0.445]	[0.472]	< 0.0001
Medicare	26.7%	23.4%	
	[0.442]	[0.423]	0.0057
Medicaid	12.8%	8.8%	
	[0.334]	[0.284]	< 0.0001
Private Insurance & Medicare	21.0%	22.4%	
	[0.407]	[0.417]	0.2017
Family Income < \$11K per year	58.6%	52.4%	
	[0.493]	[0.500]	< 0.0001
Family Income \$11K-\$25K	20.1%	20.7%	
	[0.401]	[0.405]	0.5728
Family Income \$25K-\$50K	14.1%	18.0%	
	[0.348]	[0.384]	0.0001
Dx: Acute Respiratory Failure	44.5%	41.6%	
	[0.497]	[0.493]	0.0313
Dx: COPD	11.2%	2.7%	
	[0.316]	[0.161]	< 0.0001
Dx: Congestive Heart Failure	7.0%	9.6%	
	[0.254]	[0.294]	0.0004
Dx: Cirrhosis	4.9%	2.2%	
	[0.216]	[0.148]	< 0.0001
Dx: Coma	9.6%	4.3%	
	[0.295]	[0.204]	< 0.0001
Dx: Lung Cancer	1.0%	0.2%	
	[0.097]	[0.048]	0.0011
Dx: MOSF with malignancy	6.8%	7.2%	
	[0.252]	[0.259]	0.5177
Dx: MOSF with sepsis	14.8%	32.1%	
	[0.006]	[0.010]	< 0.0001

Table 2: Catheterized vs. Not Catheterized; Disease History and Functional Status

Variable	Not Catheterized	Catheterized	p-value
Hx: Cardiac Disease	16.0% [0.366]	20.4% [0.403]	< 0.0001
Hx: Congestive Heart Failure	16.8% [0.374]	19.5% [0.396]	0.0101
Hx: Dementia	11.6% [0.321]	6.9% [0.254]	< 0.0001
Hx: Psychiatric Condition	8.1% [0.272]	4.6% [0.209]	< 0.0001
Hx: Chronic Pulmonary Disease	21.8% [0.413]	14.4% [0.351]	< 0.0001
Hx: Renal Disease	4.2% [0.201]	4.9% [0.215]	0.2409
Hx: Liver Disease	7.5% [0.263]	6.2% [0.242]	0.0748
Hx: GI Bleed	3.7% [0.189]	2.5% [0.155]	0.0113
Hx: Immunological Disease	25.5% [0.436]	29.1% [0.454]	0.003
Transferred from another hospital	9.4% [0.292]	15.0% [0.357]	< 0.0001
Hx: Acute Myocardial Infarction	3.0% [0.169]	4.3% [0.204]	0.0052
Hx: Non-metatstatic Cancer	18.0% [0.384]	15.3% [0.360]	0.0088
Hx: Malignant Cancer	24.6% [0.430]	20.3% [0.403]	0.0002
2 Month Predicted Survival	60.7% [0.192]	56.8% [0.198]	< 0.0001
Duke Activity Scale Index	20.371 [5.482]	20.701 [5.033]	0.0227
Acute Physiology Score	50.934 [18.814]	60.739 [20.271]	< 0.0001
Glasgow Coma Score	22.253 [31.374]	18.973 [28.265]	0.0001
Diastolic Blood Pressure	84.9 [38.9]	68.2 [34.2]	< 0.0001
Do Not Resuscitate Order	14.1% [0.348]	7.1% [0.257]	< 0.0001

Table 3: Catheterized vs. Not Catheterized; Laboratory Values and Secondary Diagnoses

Variable	Not Catheterized	Catheterized	p-value
WBC Count	15.263	16.266	
	[11.414]	[12.549]	0.0019
Heart Rate	112.873	118.928	
	[40.941]	[41.471]	< 0.0001
Respiratory Rate	28.978	26.652	
	[13.946]	[14.175]	< 0.0001
Temperature (^{o}C)	37.633	37.595	
	[1.741]	[1.828]	0.4293
$PAO_2/(0.01*FiO_2)$	240.627	192.433	
	[116.661]	[105.538]	< 0.0001
Albumin	3.164	2.978	
	[0.672]	[0.925]	< 0.0001
Hematocrit	32.7	30.509	
	[8.795]	[7.415]	< 0.0001
Bilirubin	1.997	2.706	
	[4.425]	[5.329]	< 0.0001
Creatinine	1.924	2.473	
	[2.025]	[2.053]	< 0.0001
Sodium	137.037	136.333	
	[7.678]	[7.600]	0.0007
Potassium	4.077	4.05	
	[1.038]	[1.012]	0.3213
$PACO_2$	39.953	36.792	
	[14.244]	[10.973]	< 0.0001
Serum Ph	7.393	7.38	
	[0.107]	[0.114]	< 0.0001
Weight (kg)	65.04	72.36	
	[29.502]	[27.730]	< 0.0001
2nd Dx: Respiratory Diagnosis	41.7%	28.9%	
	[0.493]	[0.454]	< 0.0001
2nd Dx: Orthopedic Dx	0.1%	0.2%	
	[0.029]	[0.043]	0.2988
2nd Dx: Neurological Dx	16.2%	5.4%	
	[0.368]	[0.226]	< 0.0001
2nd Dx: GI Dx	14.7%	19.2%	
	[0.354]	[0.394]	< 0.0001
2nd Dx: Renal Dx	4.1%	6.8%	
	[0.199]	[0.251]	< 0.0001
2nd Dx: Metabolic Dx	4.8%	4.3%	
	[0.215]	[0.202]	0.3052
2nd Dx: Hematological Dx	6.7%	5.3%	
	[0.251]	[0.223]	0.0252
2nd Dx: Sepsis	14.5%	23.6%	
	[0.352]	[0.425]	< 0.0001
2nd Dx: Trauma	0.5%	1.6%	
	[0.071]	[0.124]	< 0.0001

Table 4: Number of Hypotheses Rejected by Holm Procedure

Variable	Swan-Ganz	Weekend	Hypotheses
All	41	3	63
Congestive Heart Failure	9	1	51
Multi-Organ System Failure	5	0	50
Colon Cancer	8	0	55

Figure 3: % Catheterized by Day-of-Week of Admission by Diagnosis

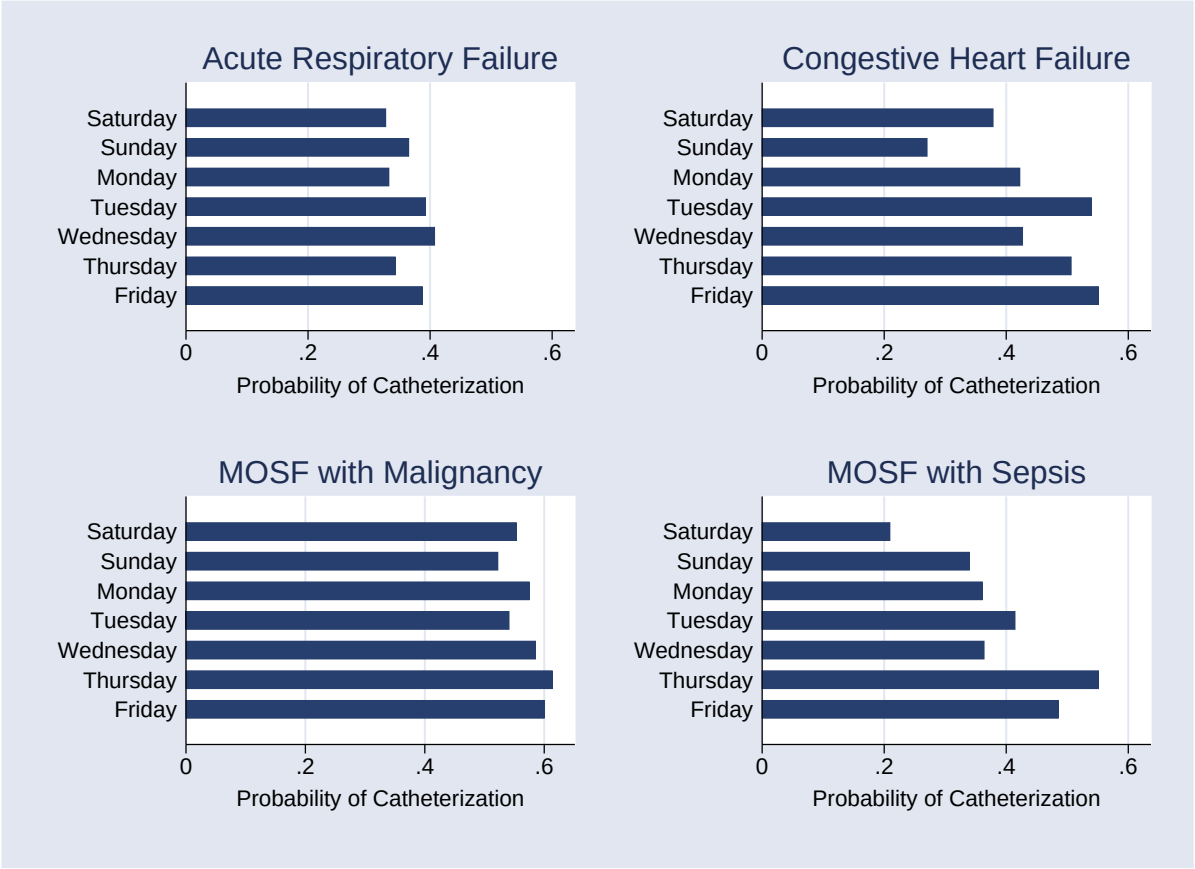


Figure 4: % Catheterized by Day-of-Week of Admission by Diagnoses (cont.)

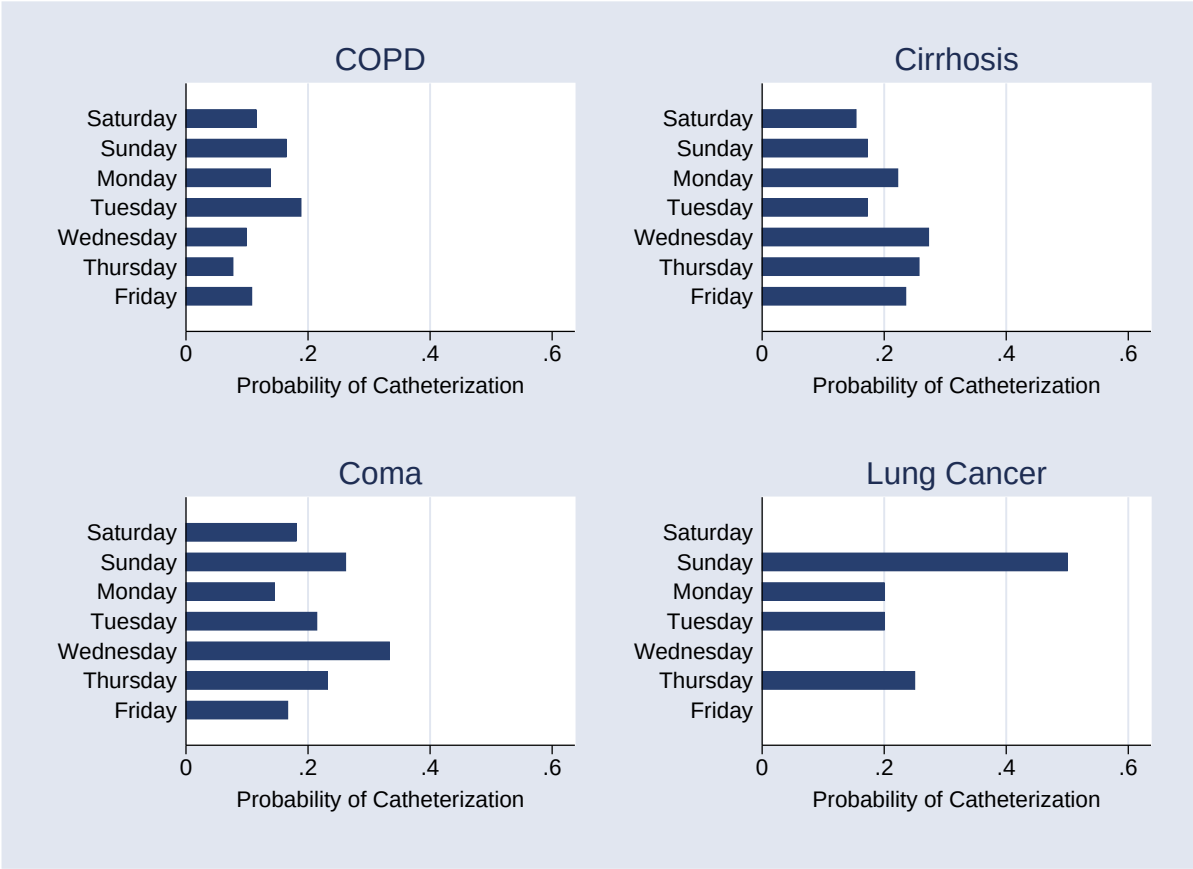


Table 5: Weekend vs. Weekday Admissions; Demographics and Diagnoses

Variable	Weekday	Weekend	p-value
Age	61.12	62.106	
	[16.709]	[16.588]	0.0501
Male	55.6%	55.9%	
	[0.497]	[0.497]	0.8385
Black	16.1%	16.0%	
	[0.367]	[0.367]	0.9645
Other Race	6.5%	5.3%	
	[0.247]	[0.224]	0.1029
Years of Education	11.663	11.722	
	[3.137]	[3.172]	0.5311
No Insurance	5.8%	5.0%	
	[0.234]	[0.219]	0.2665
Private Insurance	29.9%	28.7%	
	[0.458]	[0.453]	0.3813
Medicare	24.8%	27.2%	
	[0.432]	[0.445]	0.0621
Medicaid	11.4%	10.9%	
	[0.318]	[0.312]	0.5835
Private Insurance & Medicare	21.4%	22.1%	
	[0.410]	[0.415]	0.5323
Family Income < \$11K per year	56.3%	56.2%	
	[0.496]	[0.496]	0.9297
Family Income \$11K-\$25K	20.3%	20.4%	
	[0.402]	[0.403]	0.9443
Family Income \$25K-\$50K	15.6%	15.5%	
	[0.363]	[0.362]	0.9641
Dx: Acute Respiratory Failure	42.5%	46.1%	
	[0.494]	[0.499]	0.0141
Dx: COPD	7.9%	8.2%	
	[0.270]	[0.275]	0.6965
Dx: Congestive Heart Failure	9.0%	5.0%	
	[0.286]	[0.218]	< 0.0001
Dx: Cirrhosis	4.0%	3.7%	
	[0.195]	[0.189]	0.632
Dx: Coma	7.3%	8.5%	
	[0.260]	[0.279]	0.1409
Dx: Lung Cancer	0.8%	0.3%	
	[0.089]	[0.058]	0.0609
Dx: MOSF with malignancy	7.1%	6.5%	
	[0.257]	[0.246]	0.3774
Dx: MOSF with sepsis	21.3%	21.6%	
	[0.006]	[0.011]	0.8338

Table 6: Weekend vs. Weekday Admissions; Disease History and Functional Status

Variable	Weekday	Weekend	p-value
Hx: Cardiac Disease	19.0%	13.8%	
	[0.393]	[0.345]	< 0.0001
Hx: Congestive Heart Failure	18.9%	14.8%	
	[0.391]	[0.355]	0.0004
Hx: Dementia	10.0%	9.3%	
	[0.300]	[0.291]	0.4641
Hx: Psychiatric Condition	7.0%	5.9%	
	[0.256]	[0.235]	0.1156
Hx: Chronic Pulmonary Disease	19.0%	19.1%	
	[0.392]	[0.393]	0.8998
Hx: Renal Disease	4.7%	3.6%	
	[0.212]	[0.187]	0.0765
Hx: Liver Disease	7.2%	6.3%	
	[0.259]	[0.243]	0.2387
Hx: GI Bleed	3.2%	3.3%	
	[0.176]	[0.179]	0.8603
Hx: Malignant Cancer	22.3%	24.7%	
	[0.417]	[0.431]	0.0647
Hx: Immunological Disease	27.5%	25.2%	
	[0.447]	[0.434]	0.0884
Transferred from another hospital	11.6%	11.5%	
	[0.320]	[0.319]	0.9514
Hx: Acute Myocardial Infarction	3.4%	3.6%	
	[0.182]	[0.187]	0.7249
Hx: Non-metatstatic Cancer	16.5%	18.3%	
	[0.371]	[0.387]	0.1087
2 Month Predicted Survival	59.7%	58.1%	
	[0.196]	[0.192]	0.0074
Duke Activity Scale Index	20.378	20.836	
	[5.297]	[5.361]	0.043
Acute Physiology Score	54.415	55.389	
	[20.027]	[19.743]	0.1055
Glasgow Coma Score	20.802	21.582	
	[30.116]	[30.696]	0.3921
Diastolic Blood Pressure	78.667	78.1	
	[37.723]	[38.967]	0.621
Do Not Resuscitate Order	0.113	0.116	
	[0.317]	[0.321]	0.7453

Table 7: Weekend vs. Weekday Admissions; Laboratory Values and Secondary Diagnoses

Variable	Weekday	Weekend	p-value
WBC Count	15.555	15.903	
	[11.264]	[13.443]	0.3303
Heart Rate	114.894	115.993	
	[41.115]	[41.618]	0.3768
Respiratory Rate	28.049	28.215	
	[14.025]	[14.233]	0.6948
Temperature (^{o}C)	37.583	37.718	
	[1.777]	[1.765]	0.0119
$PAO_2/(0.01 * FiO_2)$	223.835	217.814	
	[115.790]	[112.448]	0.0822
Albumin	3.098	3.077	
	[0.810]	[0.702]	0.3652
Hematocrit	31.855	31.896	
	[8.372]	[8.342]	0.8687
Bilirubin	2.313	2.135	
	[4.980]	[4.248]	0.2174
Creatinine	2.147	2.092	
	[2.086]	[1.955]	0.3724
Sodium	136.691	136.992	
	[7.669]	[7.614]	0.1919
Potassium	4.076	4.04	
	[1.044]	[0.983]	0.244
$PACO_2$	38.607	39.155	
	[13.090]	[13.443]	0.1676
Serum Ph	7.39	7.384	
	[0.109]	[0.113]	0.0757
Weight (kg)	67.842	67.788	
	[29.346]	[28.220]	0.9505
2nd Dx: Respiratory Dx	36.2%	38.6%	
	[0.481]	[0.487]	0.1027
2nd Dx: Orthopedic Dx	0.2%	0.0%	
	[0.041]	[0.000]	0.1173
2nd Dx: Neurologic Dx	11.7%	13.3%	
	[0.321]	[0.340]	0.0905
2nd Dx: Gastrointestinal Dx	16.5%	16.3%	
	[0.371]	[0.369]	0.8551
2nd Dx: Renal Dx	5.2%	5.0%	
	[0.222]	[0.219]	0.8391
2nd Dx: Metabolic Dx	4.6%	4.8%	
	[0.209]	[0.213]	0.7425
2nd Dx: Hematologic Dx	6.2%	6.0%	
	[0.242]	[0.237]	0.7272
2nd Dx: Sepsis	18.2%	17.2%	
	[0.386]	[0.378]	0.3744
2nd Dx: Trauma	0.9%	1.0%	
	[0.093]	[0.100]	0.6297

Figure 5: OLS Treatment Effect

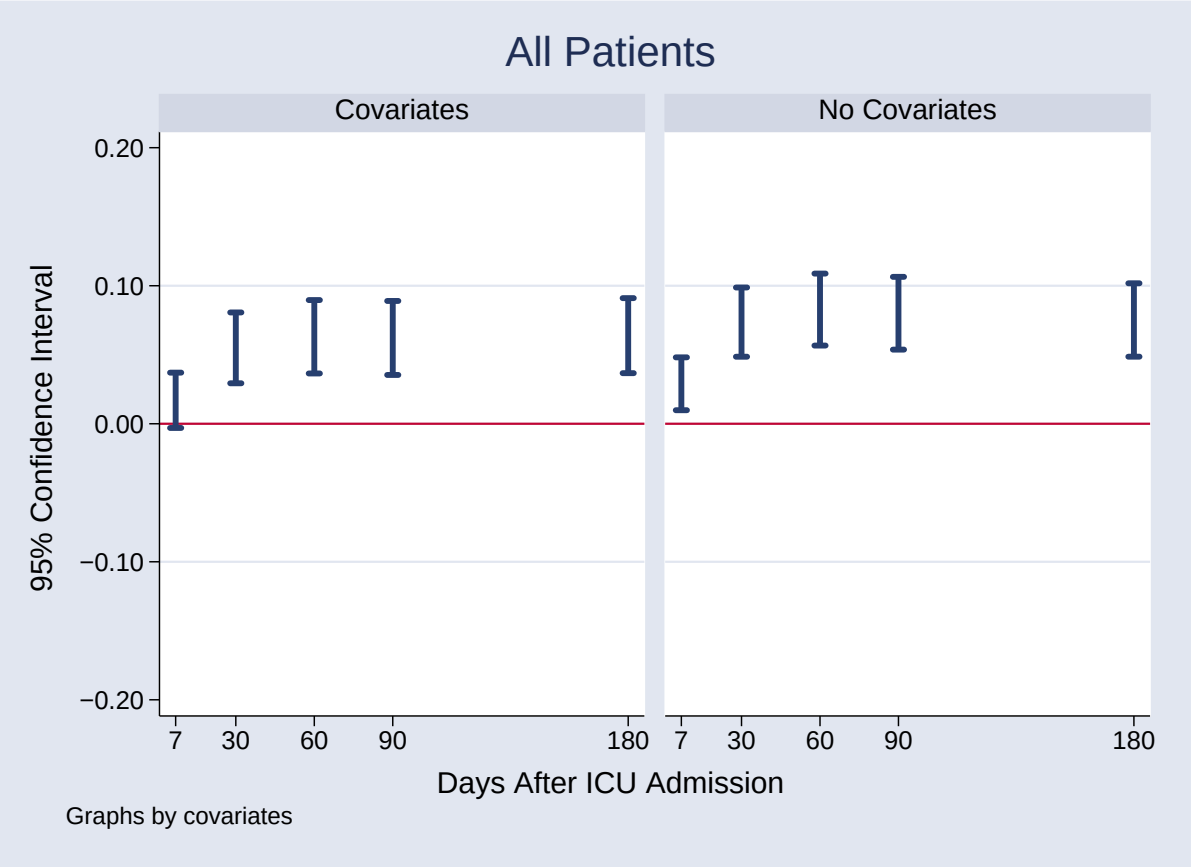


Figure 6: OLS Treatment Effect by Diagnosis

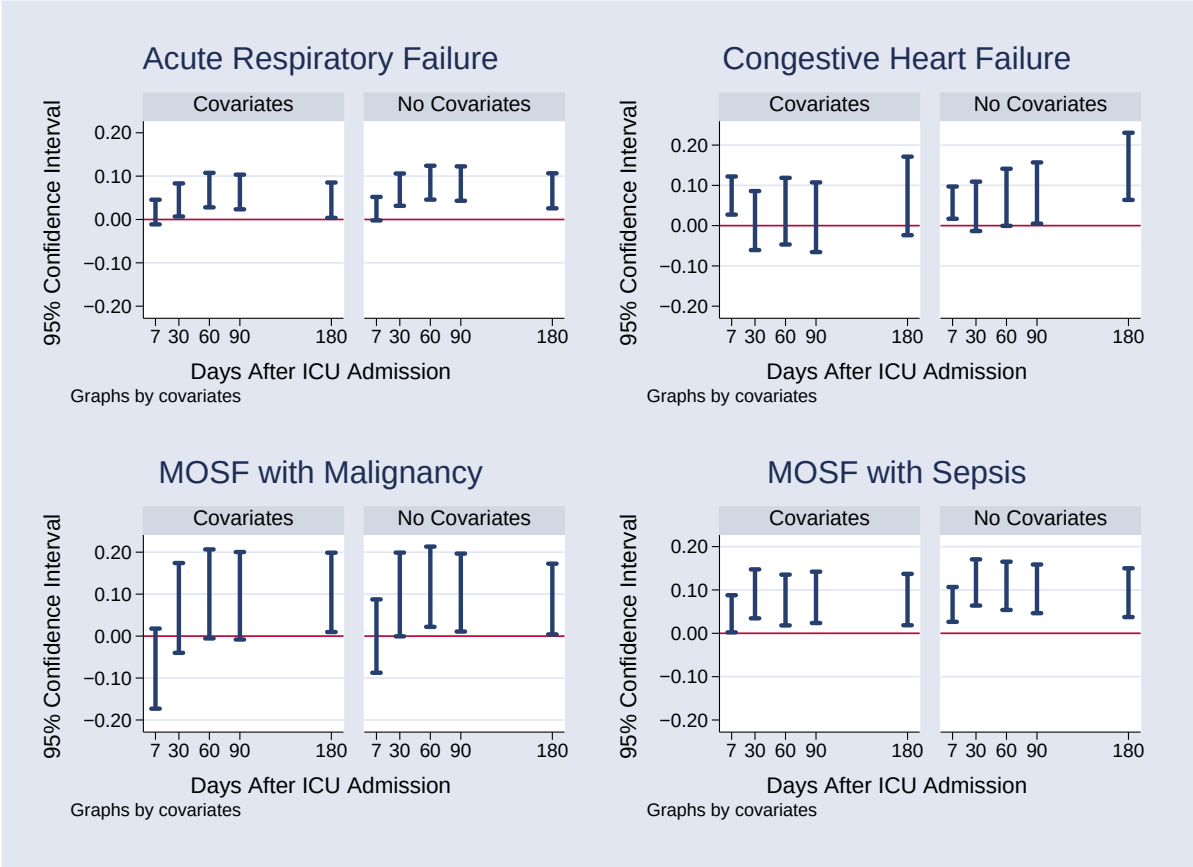


Figure 7: Manski IV and Shaikh-Vytlacil Bounds

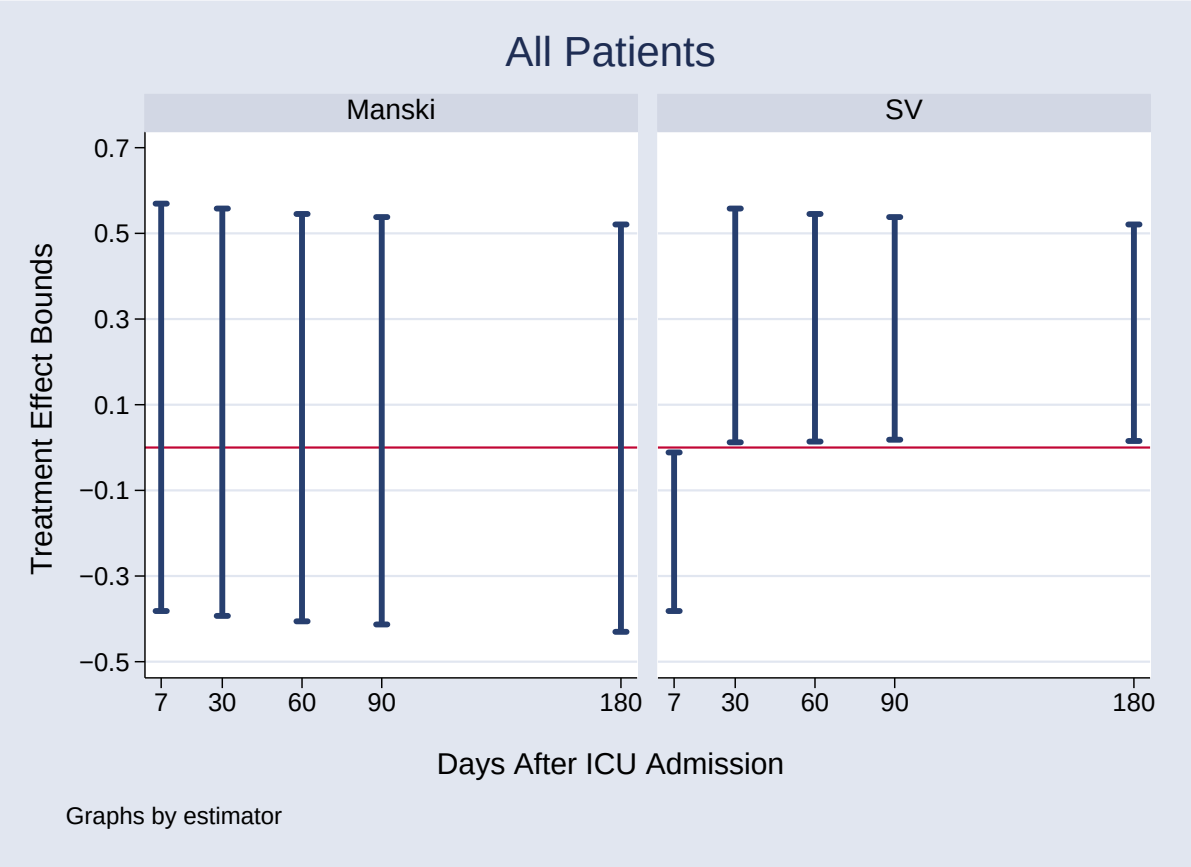


Figure 8: Manski and Shaikh-Vytlacil Bounds by Diagnosis

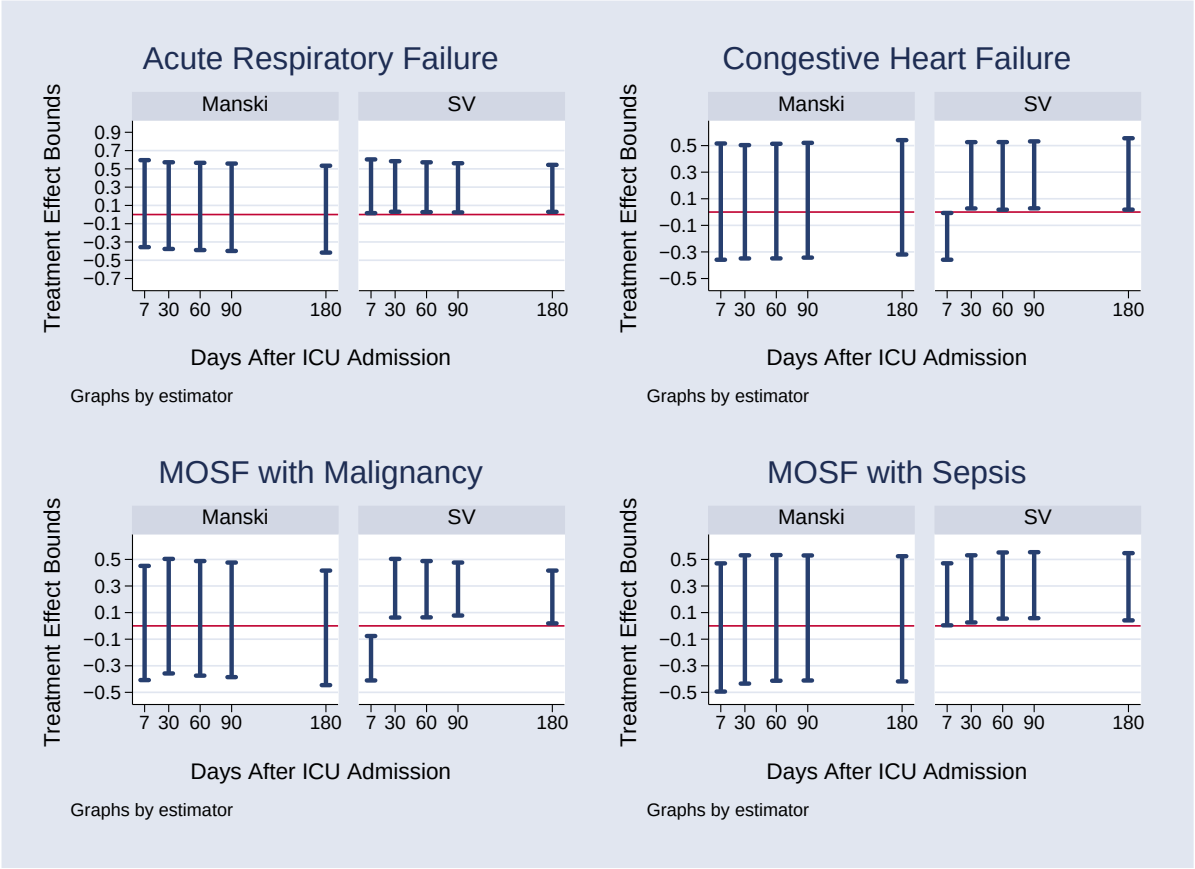


Figure 9: PQD Bounds

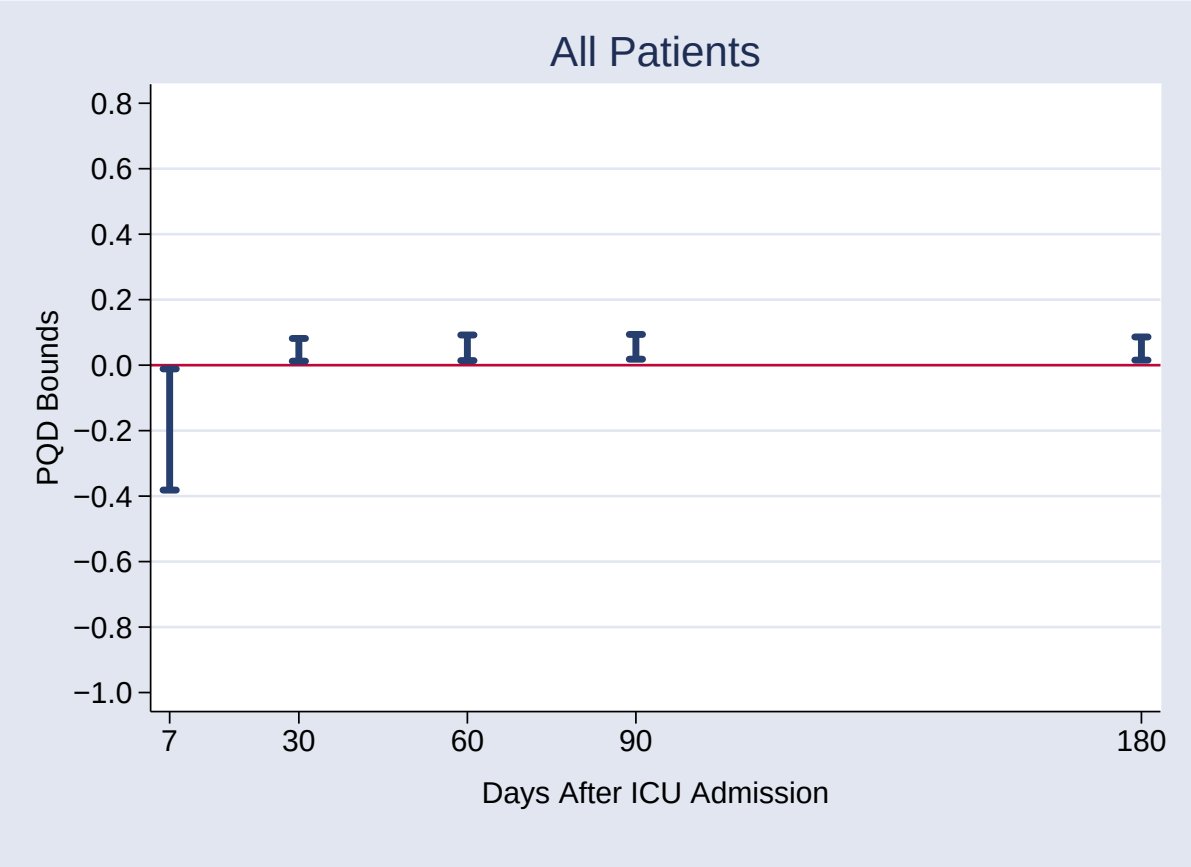


Figure 10: PQD Bounds by Diagnosis