Journal of Econometrics 168 (2012) 223–243

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Treatment effect bounds: An application to Swan–Ganz catheterization☆

Jay Bhattacharya [a], Azeem M. Shaikh [b], Edward Vytlacil [c,*]

[a] School of Medicine, Stanford University, United States
[b] Department of Economics, University of Chicago, United States
[c] Department of Economics, Yale University, United States

ARTICLE INFO

Article history:
Received 28 February 2008
Received in revised form
1 January 2012
Accepted 4 January 2012
Available online 18 January 2012

Keywords:
Partial identification
Average treatment effect
Swan–Ganz catheterization
Threshold crossing model
Simultaneous equation model

ABSTRACT

We reanalyze data from the observational study by Connors et al. (1996) on the impact of Swan–Ganz catheterization on mortality outcomes. The study by Connors et al. (1996) assumes that there are no unobserved differences between patients who are catheterized and patients who are not catheterized and finds that catheterization increases patient mortality. We instead allow for such differences between patients by implementing both the instrumental variable bounds of Manski (1990), which only exploits an instrumental variable, and the bounds of Shaikh and Vytlacil (2011), which exploit mild nonparametric, structural assumptions in addition to an instrumental variable. We propose and justify the use of indicators of weekday admission as an instrument for catheterization in this context. We find that in our application, the Manski (1990) bounds do not indicate whether catheterization increases or decreases mortality, where as the Shaikh and Vytlacil (2011) bounds reveal that at least for some diagnoses, Swan–Ganz catheterization reduces mortality at 7 days after catheterization. We show that the bounds of Shaikh and Vytlacil (2011) remain valid under even weaker assumptions than those described in Shaikh and Vytlacil (2011). We also extend the analysis to exploit a further nonparametric, structural assumption – that doctors catheterize individuals with systematically worse latent health – and find that this assumption further narrows these bounds and strengthens our conclusions. In our analysis, we construct confidence regions using the methodology developed in Romano and Shaikh (2008). We show in particular that the confidence regions are uniformly consistent in level over a large class of possible distributions for the observed data that include distributions where the instrument is arbitrarily "weak".

1. Introduction

We reanalyze data from a well known observational study by Connors et al. (1996) on the impact of Swan–Ganz catheterization on mortality outcomes. The Swan–Ganz catheter is a device placed in patients in the intensive care unit (ICU) to guide therapy. Connors et al. (1996) examine data on mortality outcomes among a population of patients admitted to the ICU and 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 was very surprising to ICU doctors, many of whom continue to use the Swan–Ganz catheter to guide therapy in the ICU.

The statistical strategy used by Connors et al. (1996) – the propensity score matching method – assumes away the possibility of unobserved differences between catheterized and non-catheterized patients. Our analysis, by comparison, permits the possibility of unobserved differences. We rely on an instrument for Swan–Ganz catheterization to bound the average effect of catheterization on mortality. We consider the bounds of Shaikh and Vytlacil (2011), which exploit not only an instrumental variable, but also threshold crossing properties for both the treatment and outcome variables. The assumptions underlying these bounds are therefore stronger than those underlying the instrumental variable bounds of Manski (1990). We show that the bounds of Shaikh and Vytlacil (2011) remain valid under even weaker assumptions than those described in Shaikh and Vytlacil (2011) and also extend the analysis to exploit the assumption that doctors tend to catheterize patients who have

☆ We thank Ivan Canay, Seung-Hyun Hong, Hidehiko Ichimura, Ralph Horwitz, and seminar participants at the Brigham Young University, the University of Chicago, Harvard/MIT, Michigan Ann-Arbor, Michigan State, the ZEW 2nd Conference on Evaluation Research in Mannheim, the Stanford University department of medicine, and the "Applications of Recent Methodologies in Project Evaluation" session at the Winter 2008 ASSA meeting in New Orleans for helpful comments. This research was supported by the NSF Grants SES-05-51089 and DMS-0820310. The STATA code used in our analysis is available at http://www.stanford.edu/~jay.

* Corresponding author. Tel.: +1 203 432 3244.
E-mail addresses: jay@stanford.edu (J. Bhattacharya), amshaikh@uchicago.edu (A.M. Shaikh), edward.vytlacil@yale.edu (E. Vytlacil).

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.01.001

worse latent health. In each case, we construct confidence regions using the methodology developed in Romano and Shaikh (2008). We show in particular that the confidence regions are uniformly consistent in level over a large class of possible distributions for the observed data that include distributions where the instrument is arbitrarily "weak".

We use the day of the week that the patient was admitted to the ICU as an instrument for Swan–Ganz catheterization. This same variable has been used as an instrument for treatment by Hamilton et al. (2000) in their study of the effect of queueing time on mortality in a Canadian population undergoing hip-fracture surgery. We argue that this variable meets the two crucial requirements for an instrument's validity. First, it is correlated with the application of the treatment: on weekends, patients are less likely to be catheterized. Second, it is uncorrelated with potential outcomes, i.e., mortality rates have little to do with the particular day of the week that a patient is admitted to the ICU and more to do with the arc of the patient's medical condition.

We find that the bounds of Manski (1990) do not permit us to say whether catheterization increases or decreases mortality—stronger assumptions are needed. In contrast, our application of the bounds of Shaikh and Vytlacil (2011), which imposes mild structural assumptions in addition to those of Manski (1990), shows that at least for some diagnoses, Swan–Ganz catheterization reduces mortality at 7 days after catheterization. Imposing the additional assumption that doctors catheterize individuals with the worst latent health further narrows these bounds.

Treating the decision to catheterize as based upon patient-specific factors that are, in part, unobservable to us clears up an economic mystery—why would ICU doctors catheterize their patients at all if doing so increases patient mortality? For every diagnosis we analyze, the answer we find is that catheterization either decreases mortality or has an effect of indeterminant direction on mortality in the short run (while the patient is still in the ICU). After the patient has left the ICU and is no longer under the care of an ICU physician, however, we find (like Connors et al., 1996) that catheterization, for some diagnoses, increases mortality.

2. Background on Swan–Ganz catheterization

The placement of Swan–Ganz catheters is common among ICU patients—over 2 million patients in North America are catheterized each year. A Swan–Ganz catheter is a slender tube with sensors that measure hemodynamic pressures in the right side of the heart and in the pulmonary artery. Once in place, the catheter is often left in place for days, so it can continuously provide information to ICU doctors. This information is often used to make decisions about treatment, such as whether to give the patient medications that affect the functioning of the heart.

While there are some risks associated with the placement of the catheter itself, such complications are rare. Rather, the greater risk may come from successful catheter placement. Information from Swan–Ganz catheterization may, for example, lead to false diagnoses of heart failure, which in turn may lead doctors to administer inappropriate treatments. Our goal in this paper is to estimate the treatment effect of catheterization on patient mortality for ICU patients with different primary diagnoses. Given the nature of the Swan–Ganz intervention, we interpret the treatment effect that we are measuring as an amalgam of the effect of catheterization itself plus the therapies that the information gleaned from catheterization make possible.

Before Connors et al. (1996), Gore et al. (1985) and Zion et al. (1990) also found that catheterization increases mortality. Dalen (2001) criticized both studies because they did not control for clinically important differences between the patients who had catheters placed and those who did not. The Connors et al. (1996) study was conceived in part as a response to this criticism. They included a dizzying array of clinical variables designed to control as exhaustively as possible for observed differences between catheterized and non-catheterized patients. In addition, Connors et al. (1996) expanded the set of ICU patients beyond just heart attack patients to all ICU patients. Ironically, Weil (1998) argued that because Connors et al. (1996) expanded the set of patients considered, they failed to take account of important unobserved clinical variables in their statistical work.

Despite substantial criticism, the publication of the Connors et al. (1996) study was seminal in the Swan–Ganz catheterization literature. Subsequent studies have focused on expanding the set of ICU patients considered in the analysis and on minimizing the possibility of selection bias. There has been one reanalysis of the Connors et al. (1996) study. Hirano and Imbens (2001) modify the propensity score matching method by using a model selection procedure to determine which regressors to include in the propensity score model. 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. See, for example, Fowler and Cook (2003) and Guyatt (1991). Since Connors et al. (1996), there have been at least two randomized trials on specialized ICU populations: Sandham et al. (2003) and Richard et al. (2003). Neither finds statistically significant differences in mortality between catheterized and non-catheterized patients. While it would be appealing to compare our results with these trials, substantial differences between the populations studied in the trials and this study preclude a direct comparison.

In recent work motivated in part by this paper, Li et al. (2008) and Altonji et al. (2008) also use the data of Connors et al. (1996) to evaluate the effect of Swan–Ganz catheterization. Li et al. (2008) apply the methodology of Li et al. (2009), which, like Connors et al. (1996), rules out the possibility of selection on unobservable characteristics. Altonji et al. (2008) use the methodology of Altonji et al. (2005), which allows for selection on unobservable characteristics, but involves restrictions on the relationship between the strength of selection on observable characteristics and the strength of selection on unobservable characteristics. In contrast to these papers, our methodology allows for selection on unobservable characteristics and requires no such restrictions on the strength of selection on these characteristics relative to the strength of selection on observable characteristics, but, in order to do so, it requires an instrumental variable.

3. Notation and assumptions

In this section, we define our notation and assumptions. Let Y be an indicator for patient death within the given number of days after admission into the ICU unit, and let D be an indicator for catheterization. Let Y_1 denote the potential outcome that would be observed if the individual receives treatment, and let Y_0 denote the potential outcome that would be observed if the individual does not receive treatment. Only Y_1 is observed for individuals who receive catheterization, and only Y_0 is observed for individuals who did not receive catheterization, so that $Y = (1 - D)Y_0 + DY_1$. The effect of catheterization on mortality is $Y_1 - Y_0$, and the average effect of the catheterization on mortality is $E[Y_1 - Y_0] = P\{Y_1 = 1\} - P\{Y_0 = 1\}$. Let X be observed individual characteristics determining mortality and let Z be observed individual characteristics

determining catheterization. We assume that Y_0, Y_1 and D are determined by threshold crossing models, i.e., for $d \in \{0, 1\}$,

$$\begin{aligned} Y_d^* &= r(X, d) - \epsilon_d \\ Y_d &= \mathbf{1}\{Y_d^* \geq 0\} \end{aligned} \tag{1}$$

and

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

where $\mathbf{1}\{A\}$ is the indicator function of the event A and ϵ_0, ϵ_1 and ν are unobserved random variables. The latent indices Y_1^*, Y_0^* may be interpreted as unobserved measures of health status with and without the treatment, and the latent index D^* may be interpreted as an unobserved measure of the desire by hospital staff to conduct the catheterization.

We assume further that $(X, Z) \perp\!\!\!\perp (\epsilon_0, \epsilon_1, \nu)$. We thus allow catheterization to be endogenous, reflecting the possible dependence between ϵ_0, ϵ_1, and ν, but we assume that all other regressors are exogenous. We further impose the rank similarity assumption of Chernozhukov and Hansen (2005):

$$\epsilon_d | \nu \sim \epsilon | \nu \tag{3}$$

for $d \in \{0, 1\}$. This restriction is obviously weaker than assuming that $\epsilon_0 = \epsilon_1$, in which case we obtain the triangular system of equations considered in Shaikh and Vytlacil (2011). We also assume that (ϵ, ν) has a strictly positive density with respect to Lebesgue measure on $\mathbf{R}^2$. This assumption eases the exposition but is not essential. We also require that there is at least one variable in Z that is not in X, i.e., there is some variable that affects the decision to perform catheterization, but does not directly affect mortality. Such a variable is often referred to as an instrumental variable. In our application, we will use an indicator variable for whether the patient was admitted into the ICU on a weekend (rather than a weekday) for this purpose.

Remark 3.1. Vytlacil (2002) establishes the equivalence between the threshold crossing model defined in (2) and the monotonicity assumption of Imbens and Angrist (1994). Using the potential treatment notation of Imbens and Angrist (1994), we have that $D_z = \mathbf{1}\{s(z) - \nu \geq 0\}$ for $z \in \{0, 1\}$. The special case of our model with $\epsilon_0 = \epsilon_1$ is equivalent to the monotonicity assumption of Imbens and Angrist (1994) holding on Y as well as on D. □

Remark 3.2. An important special case of our model is the bivariate probit model with structural shift of Heckman (1978), which imposes the further assumptions that $\epsilon_0 = \epsilon_1$, $r(X, D) = X\beta + D\alpha$, $s(Z) = Z\gamma$, and (ϵ, ν) is distributed as a bivariate normal vector with zero means and unit variances. Our model nests this model as a special case, but does not require any of its parametric assumptions. □

Remark 3.3. Shaikh and Vytlacil (2011) require that $\epsilon_1 = \epsilon_0$, which implies that the sign of the treatment effect does not vary among patients with the same observable characteristics. In our application, such a restriction would rule out the possibility that, among patients with the same observable characteristics, Swan–Ganz catheterization may result in negative effects for some patients while having positive effects for other patients. In contrast, the rank similarity assumption in Eq. (3) does not rule out this possibility. Note, however, that the restriction does impose that the sign of $P\{Y_1 = 1|X, \nu\} - P\{Y_0 = 1|X, \nu\}$, referred to as the "Marginal Treatment Effect" in Heckman and Vytlacil (2005), does not vary with ν. On the other hand, $P\{Y_1 = 1|X, \nu\} - P\{Y_0 = 1|X, \nu\}$ may be a non-trivial function of ν. In this sense, the rank similarity assumption allows for the possibility that doctors observe factors related to latent health status and factors related to the effect of treatment that are not observed by the econometrician. □

4. Bounds on the average treatment effect

In this section, we develop several different bounds on the average treatment effect. For ease of exposition, suppose that there are no X covariates and that Z is a binary random variable. See Remark 4.5 for a discussion of how the results below would change if these assumptions were relaxed. We assume further that Z is ordered so that $P\{D = 1|Z = 1\} > P\{D = 1|Z = 0\}$. In our application, $Z = 1$ therefore corresponds to a admission into an ICU on a weekday while $Z = 0$ corresponds to admission on a weekend.

4.1. Bounds of Manski (1990)

Manski (1990) only assumes that Y_1 and Y_0 are (mean) independent of Z, i.e., $P\{Y_0 = 1 \mid Z\} = P\{Y_0 = 1\}$ and $P\{Y_1 = 1 \mid Z\} = P\{Y_1 = 1\}$. Note that

$$\begin{aligned} P\{Y_1 = 1 \mid Z = z\} &= P\{D = 1, Y_1 = 1 \mid Z = z\} \\ &\quad + P\{D = 0, Y_1 = 1 \mid Z = z\}. \end{aligned}$$

Since $Y = Y_1$ when $D = 1$, $P\{D = 1, Y_1 = 1 \mid Z = z\} = P\{D = 1, Y = 1 \mid Z = z\}$ is immediately identified from the distribution of the observed data. $P\{D = 0, Y_1 = 1 \mid Z = z\} = P\{D = 0 \mid Z = z\}P\{Y_1 = 1 \mid D = 0, Z = z\}$, on the other hand, is not identified from the distribution of the observed data since we never observe Y_1 for individuals with $D = 0$. But $0 \leq P\{Y_1 = 1 \mid D = 0, Z = z\} \leq 1$, so

$$\begin{aligned} P\{D = 1, Y = 1|Z = z\} &\leq P\{Y_1 = 1|Z = z\} \\ &\leq P\{D = 1, Y = 1|Z = z\} \\ &\quad + P\{D = 0|Z = z\}. \end{aligned}$$

The same argument *mutatis mutandis* can be used to derive similar bounds on $P\{Y_0 = 1|Z = z\}$. Since Y_0 and Y_1 are (mean) independent of Z by assumption, we have

$$B_M^L \leq E[Y_1 - Y_0] \leq B_M^U$$

where

$$\begin{aligned} B_M^L &= \max_z\{P\{D = 1, Y = 1|Z = z\}\} \\ &\quad - \min_z\{P\{D = 0, Y = 1|Z = z\} + P\{D = 1|Z = z\}\} \\ B_M^U &= \min_z\{P\{D = 1, Y = 1|Z = z\} + P\{D = 0|Z = z\}\} \\ &\quad - \max_z\{P\{D = 0, Y = 1|Z = z\}\}. \end{aligned}$$

4.2. Bounds of Shaikh and Vytlacil (2011)

We now construct bounds under the assumptions described in Section 3. Shaikh and Vytlacil (2011) construct bounds on the average treatment effect under the assumptions described in Section 3 with the additional restriction that $\epsilon_0 = \epsilon_1$. As we show below, their bounds continue to hold under the weaker rank similarity assumption in Eq. (3). These assumptions, while remaining nonparametric in nature, are stronger than those imposed by Manski (1990). Under the assumptions of Section 3,

$$\begin{aligned} &P\{Y = 1 \mid Z = z\} \\ &\quad = P\{D = 1, Y = 1 \mid Z = z\} + P\{D = 0, Y = 1 \mid Z = z\} \\ &\quad = P\{D = 1, Y_1 = 1 \mid Z = z\} + P\{D = 0, Y_0 = 1 \mid Z = z\} \\ &\quad = P\{\nu \leq s(z), \epsilon_1 \leq r(1)\} + P\{\nu > s(z), \epsilon_0 \leq r(0)\} \\ &\quad = P\{\nu \leq s(z), \epsilon \leq r(1)\} + P\{\nu > s(z), \epsilon \leq r(0)\}, \end{aligned}$$

with the last equality following from (3). Recall that we have ordered Z so that $P\{D = 1 \mid Z = 1\} > P\{D = 1 \mid Z = 0\}$, which, under our assumptions, implies $s(1) > s(0)$. Thus, if $r(1) > r(0)$,

$$\begin{aligned} &P\{Y = 1 \mid Z = 1\} - P\{Y = 1 \mid Z = 0\} \\ &\quad = P\{s(0) < \nu \leq s(1), r(0) < \epsilon \leq r(1)\}, \end{aligned}$$

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

$$P\{Y = 1 \mid Z = 1\} - P\{Y = 1 \mid Z = 0\} = -P\{s(0) < \nu \leq s(1), r(1) < \epsilon \leq r(0)\}.$$

It follows that

$$P\{Y = 1 \mid Z = 1\} > P\{Y = 1 \mid Z = 0\} \Longleftrightarrow r(1) > r(0)$$
$$P\{Y = 1 \mid Z = 1\} < P\{Y = 1 \mid Z = 0\} \Longleftrightarrow r(1) < r(0).$$

It follows that if $P\{Y = 1 \mid Z = 1\} \geq P\{Y = 1 \mid Z = 0\}$, for example, then

$$P\{Y = 1 \mid D = 1, Z = z\} \geq P\{Y_0 = 1 \mid D = 1, Z = z\}$$
$$P\{Y = 1 \mid D = 0, Z = z\} \leq P\{Y_1 = 1 \mid D = 0, Z = z\}.$$

The resulting bounds on the average treatment effect are the same as those derived by Shaikh and Vytlacil (2011), and are given by

$$B^L_{SV} \leq E[Y_1 - Y_0] \leq B^U_{SV},$$

where

$$B^L_{SV} = P\{Y = 1 \mid Z = 1\} - P\{Y = 1 \mid Z = 0\}$$
$$B^U_{SV} = P\{D = 1, Y = 1 \mid Z = 1\} + P\{D = 0 \mid Z = 1\} - P\{D = 0, Y = 1 \mid Z = 0\}$$

when $P\{Y = 1 \mid Z = 1\} > P\{Y = 1 \mid Z = 0\}$,

$$B^L_{SV} = P\{D = 1, Y = 1 \mid Z = 1\} - P\{D = 0, Y = 1 \mid Z = 0\} - P\{D = 1 \mid Z = 0\}$$
$$B^U_{SV} = P\{Y = 1 \mid Z = 1\} - P\{Y = 1 \mid Z = 0\}$$

when $P\{Y = 1 \mid Z = 1\} < P\{Y = 1 \mid Z = 0\}$, and $B^L_{SV} = B^U_{SV} = 0$ when $P\{Y = 1 \mid Z = 1\} = P\{Y = 1 \mid Z = 0\}$.

Remark 4.1. The Shaikh and Vytlacil (2011) bounds always lie on one side of zero, unless $P\{Y = 1 \mid Z = 1\} = P\{Y = 1 \mid Z = 0\}$, in which case the average treatment effect is identified to be zero. To see this, note that if $P\{Y = 1 \mid Z = 1\} > P\{Y = 1 \mid Z = 0\}$, then the lower bound on the average treatment effect is $P\{Y = 1 \mid Z = 1\} - P\{Y = 1 \mid Z = 0\} > 0$. Conversely, if $P\{Y = 1 \mid Z = 1\} < P\{Y = 1 \mid Z = 0\}$, then the upper bound on the average treatment effect is $P\{Y = 1 \mid Z = 1\} - P\{Y = 1 \mid Z = 0\} < 0$. The bounds of Shaikh and Vytlacil (2011) therefore always identify the sign of the average treatment effect. □

Remark 4.2. Under the assumptions that D is given by (2) and that the unobservables are independent of Z, it follows from the Theorem 2 of Heckman and Vytlacil (2001) that the bounds of Manski (1990) may be written as

$$B^L_M = P\{D = 1, Y = 1|Z = 1\} - P\{D = 0, Y = 1|Z = 0\} - P\{D = 1 \mid Z = 0\}$$
$$B^U_M = P\{D = 1, Y = 1|Z = 1\} + P\{D = 0 \mid Z = 1\} - P\{D = 0, Y = 1|Z = 0\}.$$

Note that if $P\{Y = 1 \mid Z = 1\} \geq P\{Y = 1 \mid Z = 0\}$, then $B^U_{SV} = B^U_M$. The upper bounds on the average treatment effect is therefore the same. On the other hand,

$$\begin{aligned} B^L_{SV} - B^L_M &= P\{D = 0, Y = 1 \mid Z = 1\} - P\{D = 1, Y = 1 \mid Z = 0\} + P\{D = 1 \mid Z = 0\} \\ &= P\{D = 0, Y = 1 \mid Z = 1\} + P\{D = 1, Y = 0 \mid Z = 0\} \geq 0, \end{aligned}$$

so $B^L_{SV} \geq B^L_M$. Typically, the inequality will in fact be strict. Conversely, if $P\{Y = 1 \mid Z = 1\} \leq P\{Y = 1 \mid Z = 0\}$, then $B^L_{SV} = B^L_M$ and $B^U_{SV} \leq B^U_M$. The bounds of Shaikh and Vytlacil (2011) imposing threshold crossing on both Y and D are therefore smaller than those of Manski (1990). This result is in contrast to the results of Heckman and Vytlacil (2001), who show that imposing threshold crossing only on D alone does not narrow the bounds of Manski (1990), but rather implies restrictions on the observable data that simplifies the form of the bounds. □

Remark 4.3. Manski and Pepper (2000) consider a "monotone instrumental variables" (MIV) assumption and a "monotone treatment response" (MTR) assumption. The MIV assumption is a weaker form of the instrumental variable assumption found in Manski (1990). The MTR assumption requires that one knows *a priori* that $Y_1 \geq Y_0$ for all individuals or that one knows *a 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 MTR is unpalatable since it would involve directly imposing the answer to the question of interest. The relationship of the analysis of Shaikh and Vytlacil (2011) with the analysis of Manski and Pepper (2000) is studied in Bhattacharya et al. (2008). As discussed by Bhattacharya et al. (2008), it is not possible to determine the sign of the treatment effect in the same way as Shaikh and Vytlacil (2011) under the assumptions of Manski and Pepper (2000). Current work by Machado et al. (2011) develops the sharp bounds for the average treatment effect under the restriction that the outcome is monotone in the treatment, but without assuming the direction of the monotonicity *a priori* or that the treatment is monotone in the instrument. They show further that the sharp bounds under the assumptions of Manski and Pepper (2000) without assuming that the direction of the effect is known *a priori* does not correspond to the bounds of Shaikh and Vytlacil (2011). See also Blundell et al. (2007) for related analysis in the context of bounding changes in the distribution of wages when wages are only observed for workers and there is non-random selection into employment. In this paper, since we do not assume that $\epsilon_0 = \epsilon_1$, we do not require that $Y_1 \geq Y_0$ for all individuals or that $Y_0 \geq Y_1$ for all individuals. Thus, we impose neither that the direction of the effect is the same for all individuals nor that the direction of the average effect is known *a priori*. □

4.3. An extension of Shaikh and Vytlacil (2011)

In this section, we extend the analysis of Shaikh and Vytlacil (2011) to exploit the additional assumption that doctors catheterize individuals with the worst latent health. This restriction is referred to as "monotone treatment selection" by Manski and Pepper (2000), and is analogous to the stochastic dominance restriction considered by Blundell et al. (2007). Formally, we assume that ϵ and ν are positive quadrant dependent (PQD), i.e.,

$$P\{\epsilon \leq t_0 \mid \nu \leq t_1\} \geq P\{\epsilon \leq t_0\} \quad \text{for all } t_0, t_1.$$

Positive quadrant dependence is a relatively weak measure of positive dependence between two random variables. See Joe (1997) for a discussion of the relationship between positive quadrant dependence and other concepts of positive dependence. Put differently, this assumption requires that individuals with unobserved characteristics that make them more likely to be catheterized (have a low value of ν) are individuals with unobserved characteristics that make them more likely to suffer mortality (have a low values of ϵ).

The PQD assumption implies that

$$P\{\epsilon \leq t_0 \mid \nu \leq t_1\} \geq P\{\epsilon \leq t_0 \mid \nu > t_1\} \quad \text{for all } t_0, t_1.$$

It follows that

$$\begin{aligned} P\{Y = 1|D = 1, Z = z\} &= P\{\epsilon_1 \leq r(1)|\nu \leq s(z)\} \\ &= P\{\epsilon \leq r(1)|\nu \leq s(z)\} \\ &\geq P\{\epsilon \leq r(1)|\nu > s(z)\} \\ &= P\{Y_1 = 1|D = 0, Z = z\}. \end{aligned}$$

Similarly, we have that

$$P\{Y_0 = 1|D = 1, Z = z\} \geq P\{Y = 1|D = 0, Z = z\}.$$

It therefore follows from the analysis of the preceding section that if $P\{Y = 1 \mid Z = 1\} \geq P\{Y = 1 \mid Z = 0\}$, then

$$\begin{aligned} P\{Y = 1 \mid D = 1, Z = z\} &\geq P\{Y_1 = 1 \mid D = 0, Z = z\} \\ &\geq P\{Y = 1 \mid D = 0, Z = z\} \\ P\{Y = 1 \mid D = 1, Z = z\} &\geq P\{Y_0 = 1 \mid D = 1, Z = z\} \\ &\geq P\{Y = 1 \mid D = 0, Z = z\}; \end{aligned}$$

if, on the other hand, $P\{Y = 1 \mid Z = 1\} \leq P\{Y = 1 \mid Z = 0\}$, then

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

These results bound $P\{Y_0 = 1\}$ and $P\{Y_1 = 1\}$. If, for example, $P\{Y = 1 \mid Z = 1\} > P\{Y = 1 \mid Z = 0\}$, then

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

which implies that

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

Using arguments given in Shaikh and Vytlacil (2011), it is possible show that

$$\min_z\{P\{Y = 1 \mid D = 1, Z = z\}\} = P\{Y = 1 \mid D = 1, Z = 1\}.$$

The bounds resulting from this line of reasoning are given by

$$B^L_{PQD} \leq E[Y_1 - Y_0] \leq B^U_{PQD},$$

where

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

when $P\{Y = 1 \mid Z = 1\} > P\{Y = 1 \mid Z = 0\}$,

$$\begin{aligned} B^L_{PQD} &= P\{D = 1, Y = 1 \mid Z = 1\} - P\{D = 0, Y = 1 \mid Z = 0\} \\ &\quad - P\{D = 1 \mid Z = 0\} \\ B^U_{PQD} &= P\{D = 1, Y = 1 \mid Z = 1\} + P\{D = 0 \mid Z = 1\} \\ &\quad \times \min\{P\{Y = 1 \mid D = 1, Z = 1\}, P\{Y = 1 \mid D = 0, Z = 1\}\} \\ &\quad - P\{D = 0, Y = 1 \mid Z = 0\} - P\{D = 1 \mid Z = 0\} \\ &\quad \times \max\{P\{Y = 1 \mid D = 1, Z = 0\}, P\{Y = 1 \mid D = 0, Z = 0\}\}, \end{aligned}$$

when $P\{Y = 1 \mid Z = 1\} < P\{Y = 1 \mid Z = 0\}$, and $B^L_{PQD} = B^U_{PQD} = 0$ when $P\{Y = 1 \mid Z = 1\} = P\{Y = 1 \mid Z = 0\}$.

Remark 4.4. The PQD bounds are (weakly) narrower than the SV bounds. To see this, first suppose that $P\{Y = 1 \mid Z = 1\} > P\{Y = 1 \mid Z = 0\}$. In this case, $B^L_{SV} = B^L_{PQD}$, but

$$\begin{aligned} B^U_{SV} - B^U_{PQD} &= P\{D = 0|Z = 1\} \times P\{Y = 0|D = 1, Z = 1\} \\ &\quad + P\{D = 1|Z = 0\} \times P\{Y = 1|D = 0, Z = 0\} \\ &\geq 0, \end{aligned}$$

so $B^U_{SV} \geq B^U_{PQD}$. Similarly, if $P\{Y = 1 \mid Z = 1\} < P\{Y = 1 \mid Z = 0\}$, then it is possible to show that $B^U_{SV} = B^U_{PQD}$, but $B^L_{SV} \leq B^L_{PQD}$. Typically, these inequalities will in fact be strict. If $P\{Y = 1 \mid Z = 1\} = P\{Y = 1 \mid Z = 0\}$, then the average treatment effect is identified to be zero and the two sets of bounds coincide. □

Remark 4.5. Throughout Section 4, we have assumed that there are no X covariates and that Z is binary. Relaxing these assumptions is straightforward. If X is contained in Z, then all of the analysis can simply be carried out conditional on X. If, on the other hand, there exists a component of X that is not contained in Z, then it is possible to further narrow the bounds on the average treatment effect. If there is a continuous component of X that is not contained in Z, then it is possible to obtain point identification. For further details, see Shaikh and Vytlacil (2011) and Vytlacil and Yildiz (2007). If Z is not binary, then all of the analysis can be carried out with z_1 in place of 1 and z_0 in place of 0, where z_1 maximizes $P\{D = 1|Z = z\}$ and z_0 minimizes $P\{D = 1|Z = z\}$. Shaikh and Vytlacil (2011) show that the resulting bounds are sharp under the additional assumption that the support of the distribution of (X, Z) equals the product of the support of the distribution of X and the support of the distribution of Z. On the other hand, Chiburis (2010) shows that these bounds need not be sharp without this additional restriction. □

5. Inference

In this section, we discuss inference for each of the bounds described in the preceding section. For ease of exposition, we assume again that there are no X covariates. We also assume, as in the preceding section, that Z is ordered so that $P\{D = 1|Z = 1\} > P\{D = 1|Z = 0\}$. Finally, we assume further that $0 < P\{Y = y, D = d, Z = z\} < 1$ for all $(y, d, z) \in \{0, 1\}^3$.

Let (Y_i, D_i, Z_i), $i = 1, \ldots, n$ be an i.i.d. sample of random variables with common distribution given by the distribution of (Y, D, Z). For some pre-specified $\alpha \in (0, 1)$, we construct below random sets $\mathcal{C}_n$ such that for each θ between the upper and lower bounds

$$\liminf_{n\to\infty} P\{\theta \in \mathcal{C}_n\} \geq 1 - \alpha. \tag{4}$$

Following Romano and Shaikh (2008), who build upon earlier work by Chernozhukov et al. (2007), our construction will be based upon test inversion. In each case, our confidence region will therefore be of the form

$$\mathcal{C}_n = \{-1 \leq \theta \leq 1 : T_n(\theta) \leq \hat{c}_n(\theta, 1 - \alpha)\} \tag{5}$$

for an appropriate choice of test statistic $T_n(\theta)$ and critical value $\hat{c}_n(\theta, 1-\alpha)$. The critical value $\hat{c}_n(\theta, 1-\alpha)$ will be constructed using subsampling. In order to describe the construction, we require some further notation. Let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b_n/n \to 0$. Index by $i = 1, \ldots, N_n = \binom{n}{b}$ the different subsets of $\{1, \ldots, n\}$ of size b. Denote by $T_{n,b,i}(\theta)$ the test statistic $T_n(\theta)$ computed using only the ith subset of data of size b. Let $\hat{c}_n(\theta, 1-\alpha)$ denote the (smallest) $1 - \alpha$ quantile of the distribution

$$L_n(x, \theta) = \frac{1}{N_n} \sum_{1\leq i\leq N_n} \mathbf{1}\{T_{n,b,i}(\theta) \leq x\}. \tag{6}$$

Romano and Shaikh (2008) show that $\mathcal{C}_n$ defined by (5) satisfies the coverage property (4) under weak conditions on the distribution of $T_n(\theta)$. In each of the applications below, it is straightforward to show that these conditions hold under the above assumptions using arguments similar to those given in Section 3.2 of Romano and Shaikh (2008). Furthermore, Theorem A.1 in Appendix A establishes that the confidence regions defined below behave well uniformly over a large class of possible distributions for (Y, D, Z). In particular, the class of distributions we consider allows for the instrument to be "weak" in the sense that the (strict) inequality in $P\{D = 1|Z = 1\} > P\{D = 1|Z = 0\}$ may be arbitrarily close to an equality. See Imbens and Manski (2004) and Romano and Shaikh (2008) for further discussion of the importance of confidence regions that behave well in this sense.

5.1. Bounds of Manski (1990)

Let

$$n_z = |\{1 \le i \le n : Z_i = z\}| \tag{7}$$

and define

$$\begin{aligned}\delta_{1,n}(z_1, z_2) &= \frac{1}{n_{z_1}} \sum_{1\le i\le n: Z_i = z_1} D_i Y_i \\ &\quad - \frac{1}{n_{z_2}} \sum_{1\le i\le n: Z_i = z_2} ((1 - D_i)Y_i + D_i)\end{aligned} \tag{8}$$

$$\begin{aligned}\delta_{2,n}(z_1, z_2) &= \frac{1}{n_{z_1}} \sum_{1\le i\le n: Z_i = z_1} (D_i Y_i + (1 - D_i)) \\ &\quad - \frac{1}{n_{z_2}} \sum_{1\le i\le n: Z_i = z_2} (1 - D_i)Y_i.\end{aligned} \tag{9}$$

If $z_1 \neq z_2$, then define

$$s_{1,n}(z_1, z_2) = \sqrt{\frac{\hat\sigma^2_{n,DY|Z=z_1}}{n_{z_1}} + \frac{\hat\sigma^2_{n,(1-D)Y+D|Z=z_2}}{n_{z_2}}} \tag{10}$$

$$s_{2,n}(z_1, z_2) = \sqrt{\frac{\hat\sigma^2_{n,DY+(1-D)|Z=z_1}}{n_{z_1}} + \frac{\hat\sigma^2_{n,(1-D)Y|Z=z_2}}{n_{z_2}}}; \tag{11}$$

if $z_1 = z_2$, then define

$$s_{1,n}(z_1, z_2) = \sqrt{\frac{\hat\sigma^2_{n,DY-(1-D)Y-D|Z=z_1}}{n_{z_1}}}$$

$$s_{2,n}(z_1, z_2) = \sqrt{\frac{\hat\sigma^2_{n,DY+(1-D)-(1-D)Y|Z=z_1}}{n_{z_1}}}.$$

Finally, for $-1 \le \theta \le 1$, define

$$\begin{aligned}T_n(\theta) &= \sum_{(z_1,z_2)\in\{0,1\}^2} \left(\frac{\delta_{1,n}(z_1, z_2) - \theta}{s_{1,n(z_1,z_2)}}\right)^2_+ \\ &\quad + \sum_{(z_1,z_2)\in\{0,1\}^2} \left(\frac{\theta - \delta_{2,n}(z_1, z_2)}{s_{2,n(z_1,z_2)}}\right)^2_+.\end{aligned}$$

Remark 5.1. Imbens and Manski (2004) discuss the construction of confidence regions with the coverage property (4) for partially identified models where the identified set is an interval whose upper and lower endpoints are means or at least behave like means asymptotically. Although the identified set here is also an interval, the upper and lower endpoints do not have this property, so their analysis is not applicable here. Chernozhukov et al. (2009), on the other hand, develop methods that would be applicable in our context. □

5.2. Bounds of Shaikh and Vytlacil (2011)

Let

$$\Delta_n = \frac{1}{n_1} \sum_{1\le i\le n: Z_i = 1} Y_i - \frac{1}{n_0} \sum_{1\le i\le n: Z_i = 0} Y_i, \tag{12}$$

where n_z is given by (7), and define

$$s_n = \sqrt{\frac{\hat\sigma^2_{n,Y|Z=1}}{n_1} + \frac{\hat\sigma^2_{n,Y|Z=0}}{n_0}}. \tag{13}$$

For $0 < \theta \le 1$, define

$$T_n(\theta) = \left(\frac{-\Delta_n}{s_n}\right)^2_+ + \left(\frac{\Delta_n - \theta}{s_n}\right)^2_+ + \left(\frac{\theta - \delta_{2,n}(1, 0)}{s_{2,n}(1, 0)}\right)^2_+, \tag{14}$$

where $\delta_{2,n}(1, 0)$ is given by (9) and $s_{2,n}(1, 0)$ is given by (11); for $-1 \le \theta < 0$, define

$$T_n(\theta) = \left(\frac{\Delta_n}{s_n}\right)^2_+ + \left(\frac{\theta - \Delta_n}{s_n}\right)^2_+ + \left(\frac{\delta_{1,n}(1, 0) - \theta}{s_{1,n}(1, 0)}\right)^2_+,$$

where $\delta_{1,n}(1, 0)$ is given by Eq. (8) and $s_{1,n}(1, 0)$ is given by Eq. (10); and for $\theta = 0$, define

$$T_n(\theta) = \left(\frac{\Delta_n}{s_n}\right)^2.$$

5.3. PQD bounds

Let

$$n_{z,d} = |\{1 \le i \le n : Z_i = z, D_i = d\}|,$$

let Δ_n be given by (12), and let s_n be given by (13). Define

$$\delta_{3,n} = \frac{1}{n_{1,1}} \sum_{1\le i\le n: Z_i = 1, D_i = 1} Y_i - \frac{1}{n_0} \sum_{1\le i\le n: Z_i = 0} Y_i$$

$$\delta_{4,n} = \frac{1}{n_{1,1}} \sum_{1\le i\le n: Z_i = 1, D_i = 1} Y_i - \frac{1}{n_{0,0}} \sum_{1\le i\le n: Z_i = 0, D_i = 0} Y_i$$

$$\delta_{5,n} = \frac{1}{n_1} \sum_{1\le i\le n: Z_i = 1} Y_i - \frac{1}{n_{0,0}} \sum_{1\le i\le n: Z_i = 0, D_i = 0} Y_i,$$

and

$$s_{3,n} = \sqrt{\frac{\hat\sigma^2_{Y|Z=1,D=1}}{n_{1,1}} + \frac{\hat\sigma^2_{Y|Z=0}}{n_0}}$$

$$s_{4,n} = \sqrt{\frac{\hat\sigma_{Y|Z=1,D=1}}{n_{1,1}} + \frac{\hat\sigma_{Y|Z=0,D=0}}{n_{0,0}}}$$

$$s_{5,n} = \sqrt{\frac{\hat\sigma_{Y|Z=1}}{n_1} + \frac{\hat\sigma_{Y|Z=0,D=0}}{n_{0,0}}}.$$

For $0 < \theta \le 1$, define

$$T_n(\theta) = \left(\frac{-\Delta_n}{s_n}\right)^2_+ + \left(\frac{\Delta_n - \theta}{s_n}\right)^2_+ + \left(\frac{\theta - \delta_{4,n}}{s_{4,n}}\right)^2_+;$$

for $-1 \le \theta < 0$, define

$$\begin{aligned}T_n(\theta) &= \left(\frac{\Delta_n}{s_n}\right)^2_+ + \left(\frac{\delta_{1,n}(1, 0) - \theta}{s_{1,n}(1, 0)}\right)^2_+ + \left(\frac{\theta - \Delta_n}{s_n}\right)^2_+ \\ &\quad + \left(\frac{\theta - \delta_{3,n}}{s_{3,n}}\right)^2_+ + \left(\frac{\theta - \delta_{4,n}}{s_{4,n}}\right)^2_+ + \left(\frac{\theta - \delta_{5,n}}{s_{5,n}}\right)^2_+,\end{aligned}$$

where $\delta_{1,n}(1, 0)$ is given by (8) and $s_{1,n}(1, 0)$ is given by (10); and for $\theta = 0$, define

$$T_n(\theta) = \left(\frac{\Delta_n}{s_n}\right)^2.$$

6. A test of threshold crossing

In this section, we briefly discuss a means of testing for the threshold crossing structure on the treatment equation. As discussed in Remark 4.2, Heckman and Vytlacil (2001) show that when D is given by (2) and that the unobservables are independent of Z the bounds of Manski (1990) may be written as

$$\begin{aligned}B^L_M &= P\{D = 1, Y = 1|Z = 1\} - P\{D = 0, Y = 1|Z = 0\} \\ &\quad - P\{D = 1 \mid Z = 0\}\end{aligned}$$

$$\begin{aligned}B^U_M &= P\{D = 1, Y = 1|Z = 1\} + P\{D = 0 \mid Z = 1\} \\ &\quad - P\{D = 0, Y = 1|Z = 0\}.\end{aligned}$$

Table 1
Catheterized vs. not catheterized; demographic and diagnostic comparisons.

Variable	Not catheterized [s.d.]	Catheterized [s.d.]	*p*-value
Age	61.7 [0.323]	60.7 [0.292]	0.020
Male	53.8% [0.009]	58.5% [0.011]	0.001
Black	16.5% [0.007]	15.3% [0.008]	0.184
Other race	6.0% [0.004]	6.5% [0.006]	0.564
Years education	11.6 [0.052]	11.9 [0.080]	0.003
No insurance	5.3% [0.004]	6.2% [0.005]	0.160
Private insurance	27.1% [0.008]	33.5% [0.011]	<0.001
Medicare	26.7% [0.007]	23.4% [0.010]	0.003
Medicaid	12.8% [0.005]	8.8% [0.006]	<0.001
Private insurance and medicare	21.0% [0.007]	22.5% [0.008]	0.150
Family Income: <$11 k	58.8% [0.009]	52.4% [0.010]	<0.001
Family Income: $11–25 k	20.1% [0.006]	20.7% [0.009]	0.558
Family Income: $25–50 k	13.9% [0.006]	18.0% [0.010]	0.001
Weight (kg)	65.2 [0.475]	72.4 [0.626]	<0.001
Dx: Acute respiratory failure	45.0% [0.007]	41.7% [0.010]	0.007
Dx: COPD	11.4% [0.006]	2.7% [0.004]	<0.001
Dx: Congestive heart failure	7.0% [0.005]	9.6% [0.006]	<0.001
Dx: Cirrhosis	5.0% [0.003]	2.2% [0.003]	<0.001
Dx: Coma	9.7% [0.004]	4.4% [0.004]	<0.001
Dx: MOSF with malignancy	6.9% [0.005]	7.3% [0.005]	0.574
Dx: MOSF with sepsis	15.0% [0.006]	32.1% [0.010]	<0.001
N	3511 (61.7%)	2178 (38.3%)	–

Note: Each entry shows the mean and standard deviation (in brackets) for each variable. The *p*-value (not adjusted for multiple comparisons—see Remark 7.1) is for a *t*-test of the hypothesis that the means for catheterized and non-catheterized patients are equal.

It is therefore possible to test whether D is given by (2) and that the unobservables are independent of Z by comparing these expressions for the bounds of Manski (1990) with those stated in Section 4.1. These two expressions will be the same if and only if

$$\begin{aligned}
&P\{D=0, Y=1|Z=1\} - P\{D=1 \mid Z=1\} \\
&\quad \geq P\{D=0, Y=1|Z=0\} - P\{D=1 \mid Z=0\} \\
&P\{D=1, Y=1|Z=0\} + P\{D=0 \mid Z=0\} \\
&\quad \geq P\{D=1, Y=1|Z=1\} + P\{D=0 \mid Z=1\} \\
&P\{D=1, Y=1|Z=1\} \geq P\{D=1, Y=1|Z=0\} \\
&P\{D=0, Y=1|Z=0\} \geq P\{D=0, Y=1|Z=1\}.
\end{aligned}$$

We now describe one of several possible ways of testing whether these inequalities hold jointly. Let n_z be given by (7) and define

$$\begin{aligned}
\psi_{1,n} &= \frac{1}{n_0} \sum_{1\leq i\leq n: Z_i=0} ((1-D_i)Y_i - D_i) \\
&\quad - \frac{1}{n_1} \sum_{1\leq i\leq n: Z_i=1} ((1-D_i)Y_i - D_i) \\
\psi_{2,n} &= \frac{1}{n_1} \sum_{1\leq i\leq n: Z_i=1} (D_iY_i + (1-D_i)) \\
&\quad - \frac{1}{n_0} \sum_{1\leq i\leq n: Z_i=0} (D_iY_i + (1-D_i)) \\
\psi_{3,n} &= \frac{1}{n_0} \sum_{1\leq i\leq n: Z_i=0} D_iY_i - \frac{1}{n_1} \sum_{1\leq i\leq n: Z_i=1} D_iY_i \\
\psi_{4,n} &= \frac{1}{n_1} \sum_{1\leq i\leq n: Z_i=1} (1-D_i)Y_i - \frac{1}{n_0} \sum_{1\leq i\leq n: Z_i=0} (1-D_i)Y_i.
\end{aligned}$$

Consider the test statistic

$$T_n = \max_{1\leq i\leq 4} \psi_{i,n}.$$

Large values of this test statistic provide evidence against the null hypothesis that all four of the above inequalities hold. We may again construct a critical value for this test statistic using subsampling as described in the beginning of Section 5. As before, the validity of such an approach can be verified using arguments similar to those given in Section 3.2 of Romano and Shaikh (2008). It may, of course, be desirable to divide each of the $\psi_{i,n}$ by an estimate of its standard error, as was done in the previous section.

7. Data

The Connors et al. (1996) data come from intensive care units (ICUs) at five prominent hospitals—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. The study admitted only severely ill patients with one of nine disease conditions: acute respiratory failure, chronic obstructive pulmonary disease (COPD), congestive heart failure, cirrhosis, nontraumatic coma, metastatic colon cancer, late-stage non-small cell lung cancer, and multiorgan system failure (MOSF) with malignancy or sepsis. 59.2% of the sample is over the age of 60. Murphy and Cluff (1990) provide a detailed description of patient recruitment procedures, including a list of exclusion criteria. Connors et al. (1996) count a patient as catheterized if the procedure was performed within 24 hours of entering the ICU.

There are 5735 patients, all of whom were admitted to or transferred to the ICU within 24 hours of entering the hospital. Because we analyze each diagnosis separately, we drop patients with a primary diagnosis of lung cancer or colon cancer because there are few patients with these diagnoses in the data (seven colon cancer and 39 lung cancer patients). After these patients are dropped, there remain 5689 patients in the data.

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. Tables 1–3 compare patients who were catheterized during their first day of admission to the ICU with those who were not catheterized. These tables present the *p*-value from a test of the hypothesis that the means of the variables are equal for catheterized and non-catheterized patients.

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 4.7%), privately insured (by 6.4%), richer (less likely to have an income of less than $11,000 per year by 6.4%), and less likely to be on Medicaid (by 4.0%). Catheterized patients

Table 2
Catheterized vs. not catheterized; disease history and functional status.

Variable	Not catheterized [s.d.]		Catheterized [s.d.]		p-value
Hx: Cardiac disease	16.1%	[0.006]	20.5%	[0.008]	<0.001
Hx: Congestive heart failure	16.9%	[0.006]	19.5%	[0.008]	0.002
Hx: Dementia	11.7%	[0.005]	6.9%	[0.005]	<0.001
Hx: Psychiatric condition	8.1%	[0.005]	4.6%	[0.005]	<0.001
Hx: Chronic pulmonary disease	21.8%	[0.007]	14.4%	[0.007]	<0.001
Hx: Renal disease	4.2%	[0.003]	4.8%	[0.004]	0.201
Hx: Liver disease	7.5%	[0.005]	6.2%	[0.006]	0.070
Hx: GI Bleed	3.7%	[0.003]	2.5%	[0.003]	0.002
Hx: Malignant cancer	23.7%	[0.008]	20.1%	[0.008]	0.004
Hx: Immunological disease	25.7%	[0.008]	29.2%	[0.009]	0.009
Hx: Acute myocardial infarction	3.0%	[0.003]	4.3%	[0.005]	0.017
Admitted via transfer	9.5%	[0.005]	15.0%	[0.007]	<0.001
2 month predicted survival	60.8%	[0.003]	56.9%	[0.004]	<0.001
Duke activity scale index	20.4	[0.090]	20.7	[0.106]	0.022
Acute physiology score	51.1	[0.319]	60.8	[0.345]	<0.001
Glasgow coma score	22.4	[0.518]	19.0	[0.619]	<0.001
Diastolic blood pressure	84.9	[0.612]	68.1	[0.731]	<0.001
Do not resuscitate order	13.9%	[0.005]	7.1%	[0.006]	<0.001

Note: Each entry shows the mean and standard deviation (in brackets) for each variable. The p-value (not adjusted for multiple comparisons—see Remark 7.1) is for a t-test of the hypothesis that the means for catheterized and non-catheterized patients are equal.

Table 3
Catheterized vs. not catheterized; lab values and secondary diagnoses.

Variable	Not Catheterized [s.d.]		Catheterized [s.d.]		p-value
WBC count	15.3	[0.172]	16.3	[0.280]	0.002
Heart rate	113.0	[0.684]	118.9	[0.870]	<0.001
Respiratory rate	29.0	[0.272]	26.7	[0.253]	<0.001
Temperature (°C)	37.6	[0.027]	37.6	[0.040]	0.320
$PAO_2/(0.01^*FiO_2)$	240.0	[2.055]	192.5	[2.409]	<0.001
Albumin	3.2	[0.012]	3.0	[0.024]	<0.001
Hematocrit	32.7	[0.151]	30.5	[0.133]	<0.001
Bilirubin	2.0	[0.078]	2.7	[0.105]	<0.001
Creatinine	1.9	[0.036]	2.5	[0.046]	<0.001
Sodium	137.0	[0.136]	136.3	[0.160]	0.002
Potassium	4.1	[0.020]	4.1	[0.020]	0.315
PACO2	40.0	[0.216]	36.8	[0.193]	<0.001
Serum Ph	7.4	[0.002]	7.4	[0.002]	<0.001
2nd Dx: Respiratory	41.9%	[0.009]	28.9%	[0.010]	<0.001
2nd Dx: Neurological	16.2%	[0.005]	5.4%	[0.005]	<0.001
2nd Dx: Gastrointestinal	14.8%	[0.007]	19.2%	[0.008]	<0.001
2nd Dx: Renal	4.2%	[0.003]	6.7%	[0.005]	<0.001
2nd Dx: Metabolic	4.8%	[0.004]	4.3%	[0.004]	0.260
2nd Dx: Hematological	6.8%	[0.004]	5.2%	[0.004]	0.004
2nd Dx: Sepsis	14.6%	[0.007]	23.7%	[0.008]	<0.001

Note: Each entry shows the mean and standard deviation (in brackets) for each variable. The p-value (not adjusted for multiple comparisons—see Remark 7.1) is for a t-test of the hypothesis that the means for catheterized and non-catheterized patients are equal.

also weigh more than non-catheterized patients (by 7.2 kg). The primary diagnosis on ICU admission plays an important role in the probability of catheterization. Catheterized patients are less likely to have COPD, cirrhosis, or coma as an admitting diagnosis, but more likely to have congestive heart failure or MOSF with sepsis.

Table 2 compares patients on the basis of disease history prior to admission and functional status. Catheterized patients are more likely to have had a history (Hx) of cardiac disease (by 4.4%), but less likely to have a history of dementia, psychiatric disease, or COPD. Catheterized patients are more likely than non-catheterized patients to have arrived at the hospital by transfer from another hospital. Catheterized patients have a 3.9% lower two month predicted survival rate upon admission than non-catheterized patients and a worse acute physiology score. Catheterized patients have a lower Glasgow coma score and are less likely to have requested a "do not resuscitate" (DNR) order. Patients with DNR orders typically want to avoid aggressive therapies, including many of the sort supported by catheterization. Clearly, catheterized patients are observably more ill than non-catheterized patients.

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.

Remark 7.1. Because of the large number of comparisons we are making (there are 59 variables in Tables 1–3), it is likely that we will reject several hypotheses falsely. We use the multiple testing procedure of Holm (1979) to make the comparisons while controlling the familywise error rate – the probability of even one false rejection – at level α. Let $\hat{p}_{(1)} \leq \cdots \leq \hat{p}_{(s)}$ denote the ordered values of the p-values and let $H_{(1)}, \ldots, H_{(s)}$ denote the corresponding null hypotheses. If $\hat{p}_{(1)} \geq \alpha/s$, then the procedure rejects no null hypotheses; otherwise, it rejects null hypotheses $H_{(1)}, \ldots, H_{(r)}$, where r is the largest index such that $\hat{p}_{(i)} \leq \alpha/(s - i + 1)$ for all $i \leq r$. This procedure always rejects at least as many null hypotheses as the Bonferroni procedure, which simply rejects any null hypothesis H_i for which the corresponding $\hat{p}_i \leq \alpha/s$.

At the $\alpha = 0.05$ significance level, we find that patients who are catheterized differ from those who are not catheterized along 32 of 59 possible variables. The results are qualitatively similar at the $\alpha = 0.01$ significance level. Hence, even after accounting for the multiplicity of comparisons, we maintain our earlier conclusion that catheterized patients are significantly different from non-catheterized patients. □

8. Instrumenting with admission day

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 were included in the analysis, the results would be unconvincing. If catheterized and non-catheterized patients differ on so many *observed* dimensions, it is unlikely that they do not differ on *unobserved* dimensions as well. See Altonji et al. (2005) for a formal justification of this argument. 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.

8.1. 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. The results remain similar if we exclude Monday from the definition of the weekend. Fig. 1 shows catheterization rates by day-of-the-week for patients, divided upon the basis of primary diagnosis upon ICU admission. For patients with acute respiratory failure, congestive heart failure, MOSF with malignancy, or MOSF with sepsis the probability of being catheterized decreases on weekends. A t-test of the difference in probability of catheterization between weekend and weekday rejects equality at the $\alpha = 0.05$ significance level for all four groups. However, the same is not true for patients with COPD, cirrhosis, or coma: there is no statistically significant difference, though the point estimates suggest that weekend admissions are less likely to result in catheterization, even in these groups. As described in Section 5, the inference method we use is robust to the presence of "weak" instruments, so we analyze all of these patient groups.

Remark 8.1. Since the bounding analysis includes no covariates (except for patient diagnosis), the analogous first stage regression is a linear regression of the treatment on the instrument and indicators for each diagnosis. The F-statistic from a test of the null hypothesis that the coefficient on the instrument in such a regression is zero is 14.53, which suggests that the instrument is not "weak" by conventional standards. On the other hand, if separate regressions are run for each diagnosis, then the F-statistic does not exceed 10 for any of the regressions, which suggests that the instrument is "weak" by conventional standards for a diagnosis-by-diagnosis analysis. This fact is problematic for standard two-stage least squares estimator in the diagnosis-by-diagnosis analysis, but not for our approach. □

8.2. Patient health and day of week of admission

If patients admitted to the ICU on a weekday differed systematically from patients admitted on weekends, then day-of-the-week would be a poor instrument since it would be correlated with unobserved determinants of ICU patient mortality such as health status. We believe that there should be no such correlation, since the health crises that precipitate ICU admissions are unlikely to respect distinctions between weekdays and weekends. We now present suggestive evidence in favor of this view.

We divide patients up on the basis of whether they were admitted on a weekend or a weekday. Unlike Tables 1–3, where there were many statistically significant differences between catheterized and non-catheterized patients, we find no statistically significant differences between weekend and weekday patients on the basis of the 59 variables listed in Tables 1–3 at the $\alpha = 0.05$ significance level (after adjusting for the fact that we are testing for multiple hypotheses using the Holm (1979) procedure that we outline in Remark 7.1). Importantly, there is no statistically significant difference between these groups on the basis of laboratory tests at admission or other objectively measured variables. We further examine the possibility of differences between the two groups by regressing the instrument on all of the 59 variables, and we fail to reject the null hypothesis that the coefficients on all of these 59 variables are zero (the corresponding F-statistic is 1.11 and the p-value is 0.266).

Remark 8.2. Our belief that the day of week of admission is uncorrelated with the unobserved determinants of mortality relies a great deal on the empirical fact that it is difficult for a doctor to control the course of a very sick patient's condition to coincide with a weekday. While medical providers may desire patients arrive at the ICU at a convenient time and day, in most cases this is beyond the ability of doctors to determine. Card et al. (2007) similarly suggest that medical conditions for which admission rates do not depend on the day of week of admission are exogenous to the control of doctors. In their study of the effect of Medicare coverage on patient mortality, they include patients in their analysis on the basis of a test of admission rates on weekend and weekdays. We conduct a similar test for each admitting condition in our data. For every admitting diagnosis, we fail to reject the null hypothesis that admission rates for the condition are equal on weekends and weekdays at the $\alpha = 0.05$ significance level. □

Remark 8.3. Even though health status at admission does not appear to vary by the day of week of admission to the ICU, death rates will vary if catheterization rates depend on day of week of admission and mortality is affected by catheterization. Figs. 2 and 3 show estimates mean mortality rates at 7, 30, 60, 90, and 180 days after ICU admission for patients admitted on weekends and weekdays as well as 95% confidence intervals for the mean. Fig. 2 shows that for patients admitted for acute respiratory failure, congestive heart failure, MOSF with malignancy, or MOSF with sepsis, observed mortality is the same at 7 days after ICU admission for weekend and weekday patients, but higher for weekday patients at 30, 60, 90, and 180 days post-ICU admission. Taken individually, none of these mortality differences between weekend and weekday patients are statistically significant at the $\alpha = 0.05$ significance level. However, when grouped together, the differences at 30, 60, 90, and 180 days are statistically significant at the $\alpha = 0.05$ significance level. By contrast, Fig. 3 shows that for patients admitted with COPD, coma, or cirrhosis, patients admitted on a weekday have a lower mortality than patients admitted on a weekend, at nearly every time interval. This suggests that for these patients, catheterization is protective, even after discharge from the ICU. As before, none of the differences are statistically significant when taken diagnosis by diagnosis, but are statistically significant when these three groups of patients are grouped together. □

Remark 8.4. Empirically, it is interesting to know for which groups of patients day of the week of admission is likely to be influential in the decision to catheterize. The threshold crossing assumption implies that there are three different groups of

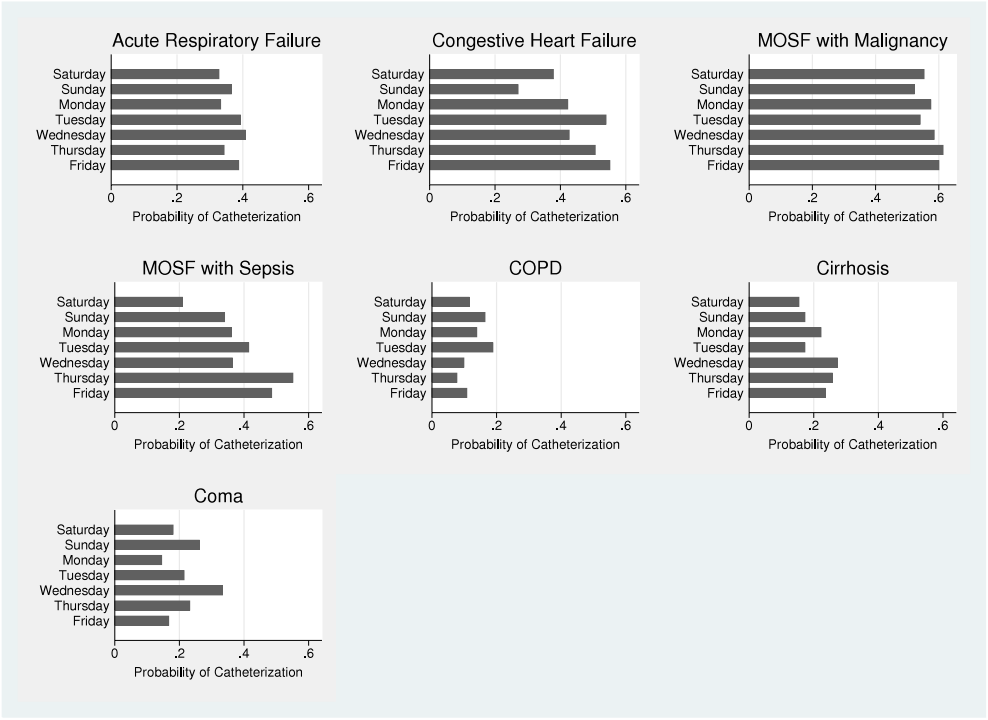

Fig. 1. % Catheterized by day-of-week of admission by diagnosis.

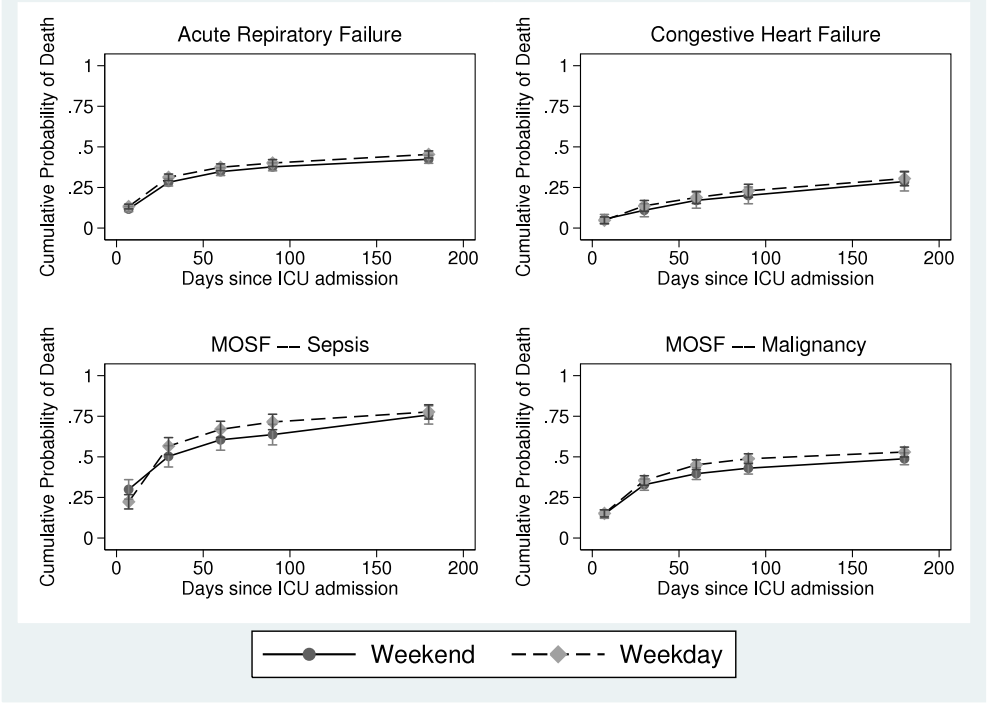

Fig. 2. Mortality rates for weekend vs. weekday admissions (part I).

patients that differ on this basis: one group would be catheterized on any day of the week, a second group would not be catheterized on any day of the week, and a third group would be catheterized on weekdays but not on weekends. This third group, in language of Angrist et al. (1996), are called "compliers" and are interesting because the local average treatment effect equals the average treatment effect for them.

Almond and Doyle (2008) develop a method to measure the expected value of covariates among compliers. See Eq. (2) and the discussion that precedes it in their paper for details. In our context, this equation says that the expected value of covariates among compliers is simply

$$\frac{\pi_C + \pi_A}{\pi_C}\left(E[X|D=1,Z=1] - \frac{\pi_A}{\pi_C+\pi_A}E[X|D=1,Z=0]\right),$$

where $\pi_A = P\{D=1|Z=0\}$, $\pi_N = P\{D=1|Z=1\}$, and $\pi_C = 1-\pi_A-\pi_N$. Using this equation, for each covariate in

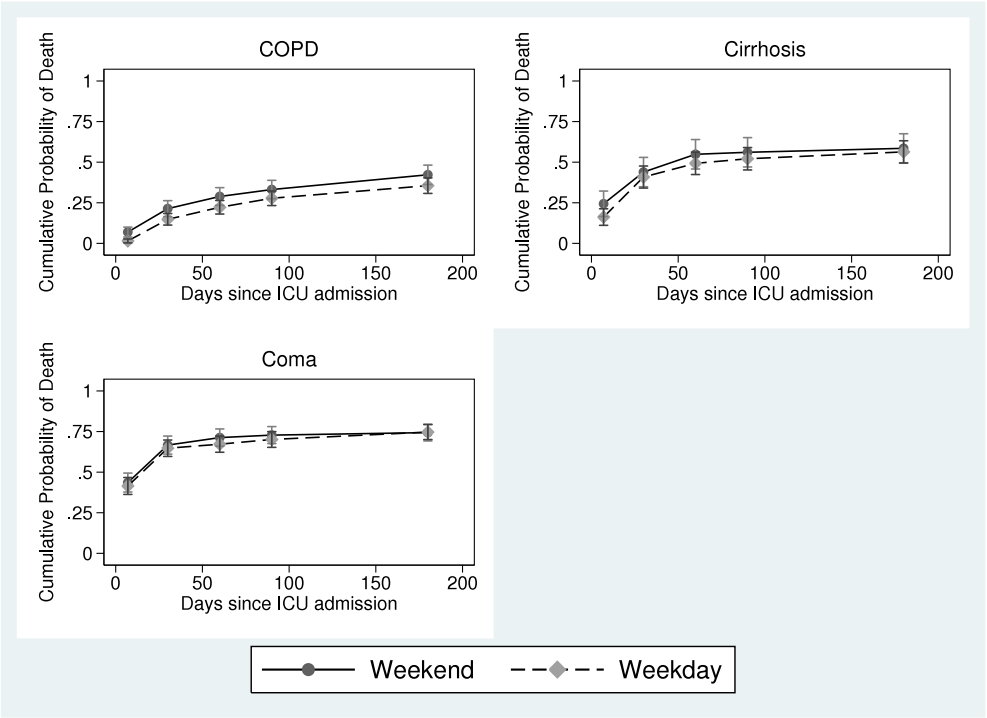

Fig. 3. Mortality rates for weekend vs. weekday admissions (part II).

Table 4
Compliers vs. overall sample: significant differences.

Variable	Overall [s.d.]	Compliers [s.d.]	*p*-value
Dx: COPD	0.080 [0.004]	0.011 [0.017]	<0.001
$PACO_2$	38.757 [0.146]	35.191 [0.988]	<0.001
2nd Dx: Respiratory	0.369 [0.007]	0.212 [0.043]	<0.001

Note: Please see Remark 8.4. Each entry shows the mean and standard deviation (in brackets) for each variable. For each of the 59 variables in Tables 1–3, we conduct a test of the null hypothesis that the mean of the indicated variable among "compliers" equals the mean of the indicated variable among all patients. Here, we report the variables that show a statistically significant difference at the $\alpha = 0.05$ significance level after adjustment for multiple comparisons. Please see Remark 7.1. After such adjustment, only three of these variables show a statistically significantly difference – $PACO_2$, 2nd Dx: Respiratory, and Dx: COPD. At the $\alpha = 0.01$ significance level, there is only one statistically significant difference – Dx: COPD.

Tables 1–3, we test whether the expected value of the covariate among compliers equals the expected value of the covariate among compliers and non-compliers. After applying the multiple testing procedure described in Remark 7.1, we find significant differences (at the $\alpha = 0.05$ level) for only three covariates: proportion of patients with a primary diagnosis of COPD, proportion of patients with adverse values of $PACO_2$ (a measure of lung function), and proportion of patients with a secondary diagnosis of respiratory disease. See Table 4. At the $\alpha = 0.01$ significance level, only the proportion of patients with a diagnosis of COPD remains statistically different between compliers and the overall population.

For brevity, we do not show the analogous table within each diagnosis category. These results show that for nearly every diagnosis category, once we adjust for multiple testing, there are no differences among compliers and non-compliers. There are only two exceptions: Within the COPD diagnostic category, complies differ from non-compliers in terms of the likelihood of having a secondary diagnosis of a neurological condition, and within the coma category, compliers differ from non-compliers in terms of two-month predicted survival rate at admission to the ICU. □

8.3. Day of week, hospital staffing, and outcomes

Although non-specialists sometimes find it surprising, it is well known in the health services literature that medical staffing can have a major effect on treatment decisions, including the decision to catheterize a patient. Rapoport et al. (2000), for example, find that patients admitted to ICUs that staff a full time ICU physician are two-thirds less likely to be catheterized than those admitted to ICUs with no full time physician. Whether this fact threatens 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 patient mortality. If so, then admission day would not be a valid instrument.

Evaluating the importance of differences in treatment between weekend and weekday admissions is 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, which are designed to improve cardiac contractility. These drugs have a narrow therapeutic range, and thus small differences in the dosing can be the difference between killing and inadequately treating a patient. Since catheterization is less likely on weekends, it would be unsurprising to find decreased use of inotropes on weekends as well. We can accommodate such differences in treatment between weekend and weekday admissions by simply interpreting the treatment effect as catheterization *and all the other treatments it enables or encourages* on mortality, rather than catheterization by itself.

It is possible that weekend–weekday staffing differences, for reasons having nothing to do with catheterization or its downstream consequences, may lead to higher patient mortality. If so, then our instrument would be invalid. Since staffing tends to be sparser on weekends, one would expect that mortality rates would be higher then. In fact, in our data mortality rates are higher on weekdays for some diagnoses and lower for others,

which is inconsistent with a direct mortality effect of staffing. Furthermore, several studies have found no evidence that staffing differences explain weekend–weekday mortality differences in ICUs. See Ensminger et al. (2004), Wunsch et al. (2004) and Dobkin (2003).

9. Results

In this section, we report results from the three different bounds described in Section 4. We analyze outcomes t days after admission to the ICU separately for five different values of t—7, 30, 60, 90, and 180 days. Each of the bounds except one that we report rely upon our instrumental variable—an indicator for whether the patient was admitted to the ICU on a Tuesday through Friday. For each of the bounds, we display 95% confidence intervals for the average treatment effect computed as described in Section 5. Computational details are described below in Remark 9.4.

Figs. 4 and 5 show the bounds of Manski (1990) (labeled "Manski bounds") for the treatment effect of catheterization on mortality for patient groups divided up based on admitting diagnosis. These figures also show the 95% confidence band around these bounds. In every case, and at every time horizon, the Manski (1990) bounds have a width of nearly one and thus always fail to exclude zero. Apparently, our instrument plus the fact that probabilities lie between zero and one is insufficient to determine whether catheterization increases or reduces mortality. See Remark 9.1 below for further discussion.

Figs. 6 and 7 show the bounds of Shaikh and Vytlacil (2011) (labeled "SV bounds") for each admitting diagnosis (along with the 95% confidence bands around these bounds). The Shaikh and Vytlacil (2011) bounds are considerably more informative than the Manski bounds in several cases. For instance, for patients with acute respiratory failure, the SV bounds suggest that catheterization increases mortality at 30 and 180 days post ICU admission, but that one cannot rule out at the $\alpha = 0.05$ significance level that it has no effect on mortality (or even decreases mortality) at 7 days. A similar story can be told for patients with MOSF with sepsis. For patients admitted with MOSF with malignancy, the SV bounds show a decrease in mortality at 7 days, but an increase at 90 days (both statistically significant at the $\alpha = 0.05$ significance level). For these groups of patients, catheterization causes a short term improvement in survival, but a longer term loss.

For patients with an admitting diagnosis of COPD or cirrhosis, the SV bounds in Fig. 7 suggest that catheterization reduces mortality at 7 days (and even at 30 and 60 days for COPD patients), but that one cannot rule out at the $\alpha = 0.05$ significance level an increase in mortality at longer intervals. Of course, one cannot rule out a decrease in mortality or no effect on mortality at those longer intervals.

Recall that Connors et al. (1996) found that catheterization increases mortality even 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. Their result raises the question of why ICU doctors do not observe the increased mortality from catheterization and react accordingly. The Shaikh and Vytlacil (2011) bounds provides a possible answer—ICU doctors do not see rise in mortality which happens only after many patients have been released from the ICU.

Figs. 8 and 9 show the bounds from the extension of Shaikh and Vytlacil (2011) described in Section 4.3 (labeled "PQD bounds"). These bounds 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 plausible restriction decreases the width of the treatment effect bounds, often dramatically.

By construction, these bounds are always on the same side of zero as the Shaikh and Vytlacil (2011) bounds. The reduction in the width of the bounds is greatest when the average treatment effect is positive, i.e., when 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. On the other hand, the PQD bounds have a lower upper bound than the Shaikh and Vytlacil (2011) bounds when the average treatment effect is negative, i.e., when catheterization decreases mortality, so it may permit researchers to conclude, for example, that the intervention is cost-effective even as the Shaikh and Vytlacil (2011) bounds permit the possibility that it may not be.

Remark 9.1. Despite the evidence presented in Section 8, it is interesting to consider how our inferences would change if we did not rely upon our instrumental variable. One possible answer is to rely on the bounds of Manski (1990), which may be constructed without an instrument. In that case, the width of the bounds is always exactly one and thus always fail to exclude zero. A second possible answer is given by the analysis of Section 4.3, which may also be constructed without an instrument. In that case, the PQD assumption reduces to $P\{Y_1 = 1 \mid D = 1\} \geq P\{Y_1 = 1 \mid D = 0\}$ and $P\{Y_0 = 1 \mid D = 1\} \geq P\{Y_0 = 1 \mid D = 0\}$, which implies the following bounds on the average treatment effect:

$$\begin{aligned} &P\{Y_1 = 1 \mid D = 1\}P\{D = 1\} \\ &\quad - P\{Y_0 = 1 \mid D = 0\}P\{D = 0\} - P\{D = 1\} \\ &\leq E[Y_1 - Y_0] \leq P\{Y_1 = 1 \mid D = 1\} - P\{Y_0 = 1 \mid D = 0\}. \end{aligned}$$

Figs. 10 and 11 shows these bounds and associated 95% confidence intervals. In every case, the bounds cross zero, though their width is substantially less than one. The PQD assumption by itself is therefore not enough to identify the direction of the treatment effect. □

Remark 9.2. Heckman and Vytlacil (2001) show that the threshold crossing structure on D implies that $B_M^U - B_M^L = 1 - P\{D = 1|Z = 1\} + P\{D = 1|Z = 0\}$, where Z is ordered such that $P\{D = 1|Z = 1\} > P\{D = 1|Z = 0\}$. If $P\{D = 1|Z = 1\}$ is close to one and $P\{D = 1|Z = 0\}$ is close to zero, then the bounds will have width close to zero. In contrast, if $P\{D = 1|Z = 1\}$ is close to $P[D = 1|Z = 0\}$, then the width of the bounds will be nearly one, i.e., almost as wide as the naive bounds that do not impose or exploit an instrument described in Remark 9.1. Our empirical result that the width of the bounds is close to one is a direct result of the instrument being "weak" in the sense that $P\{D = 1|Z = 1\}$ is close to $P\{D = 1|Z = 0\}$. □

Remark 9.3. We also implement the test of the threshold crossing assumption on treatment selection that is described in Section 6. For each value of t, we fail to reject the inequalities shown in that section at the $\alpha = 0.10$ significance level. This provides at least weak evidence in favor of the assumptions underlying the bounds described in Sections 4.2 and 4.3. □

Remark 9.4. For the results we reported above, we used a subsample size of $b = 50$. In results not reported here, we also tried different subsample sizes ranging from 25 to 75 and found that our results are remained similar for these values of b. Finally, because N_n is large, we used an approximation to (6) in which we randomly chose with replacement $B_n = 1000$ of the N_n possible subsamples. It follows from Corollary 2.4.1 of Politis et al. (1999) that critical values constructed in this way remain valid provided that B_n tends to infinity. □

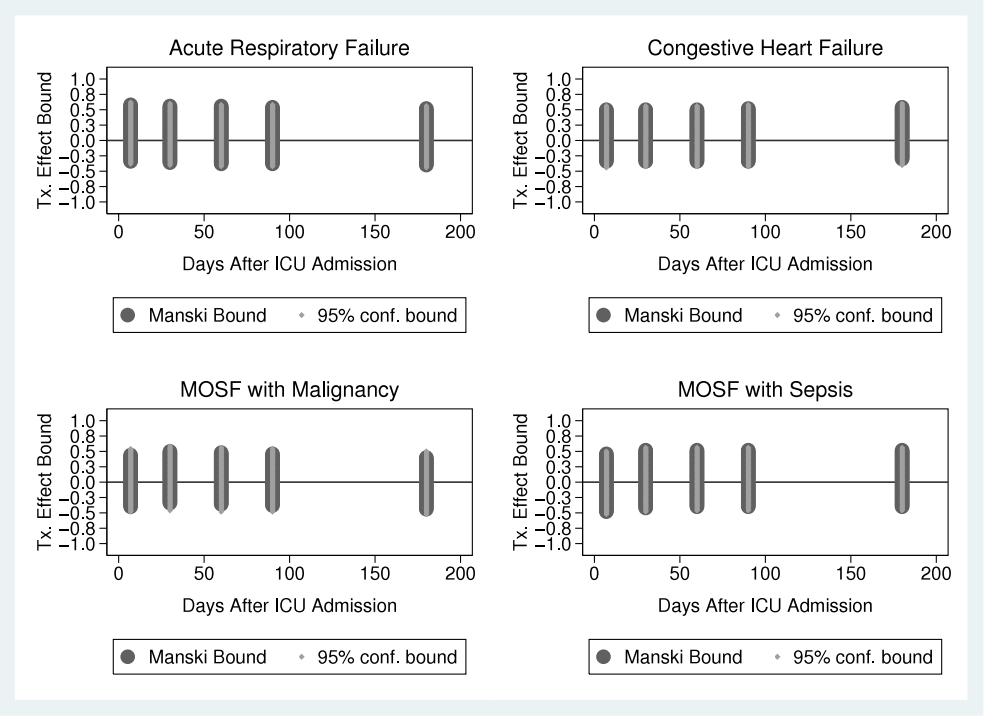

Fig. 4. Manski (1990) bounds (part I).

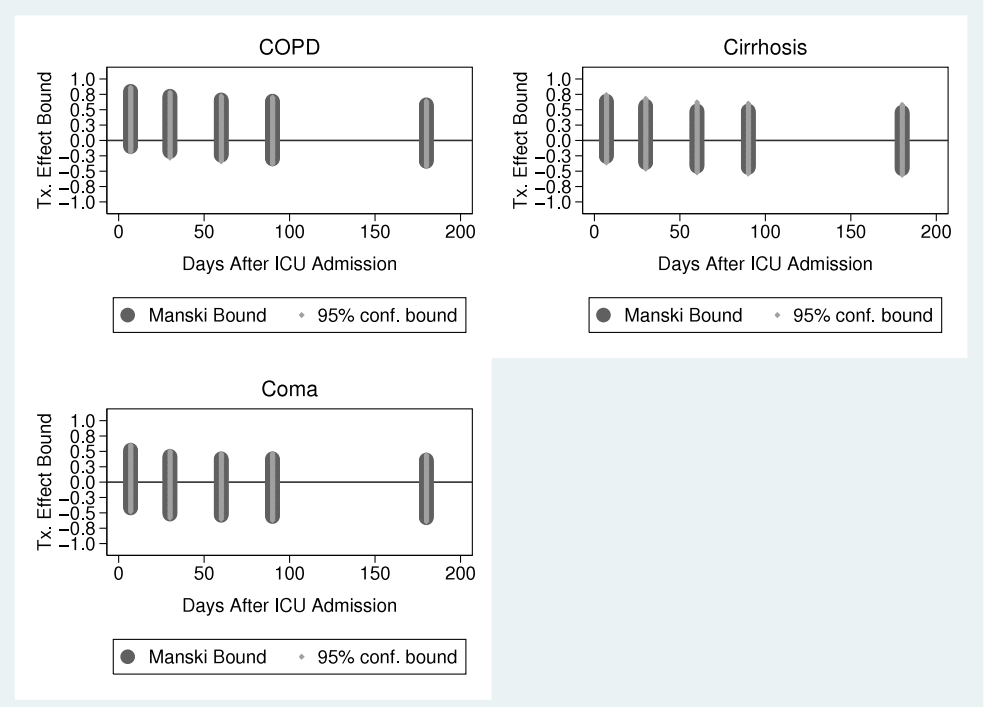

Fig. 5. Manski (1990) bounds (part II).

10. 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 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.

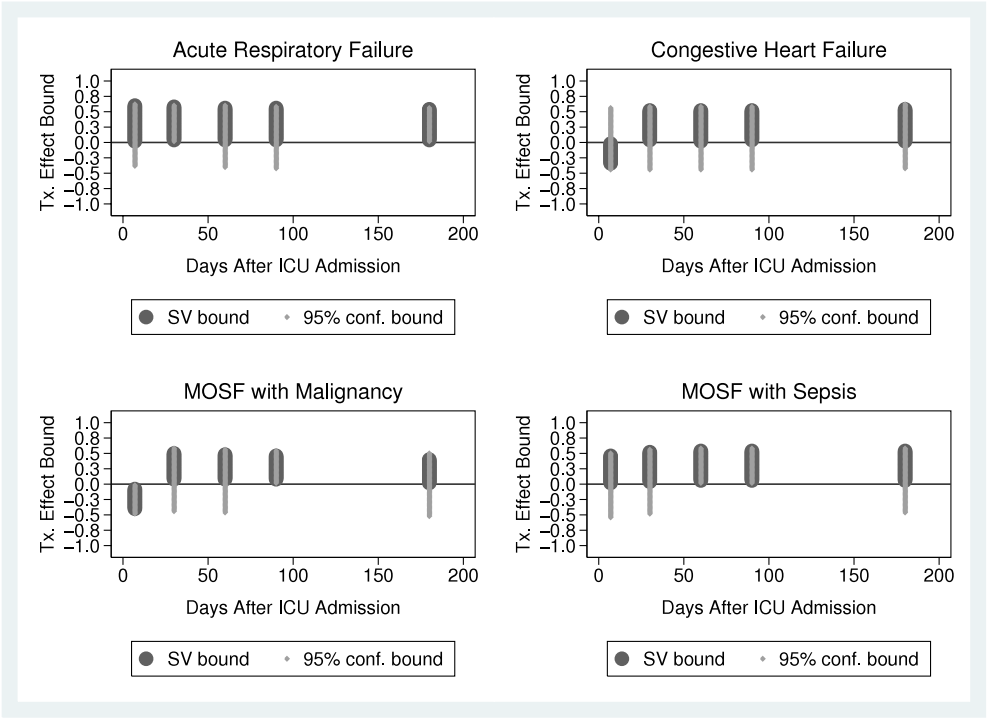

Fig. 6. Shaikh and Vytlacil (2011) bounds (part I).

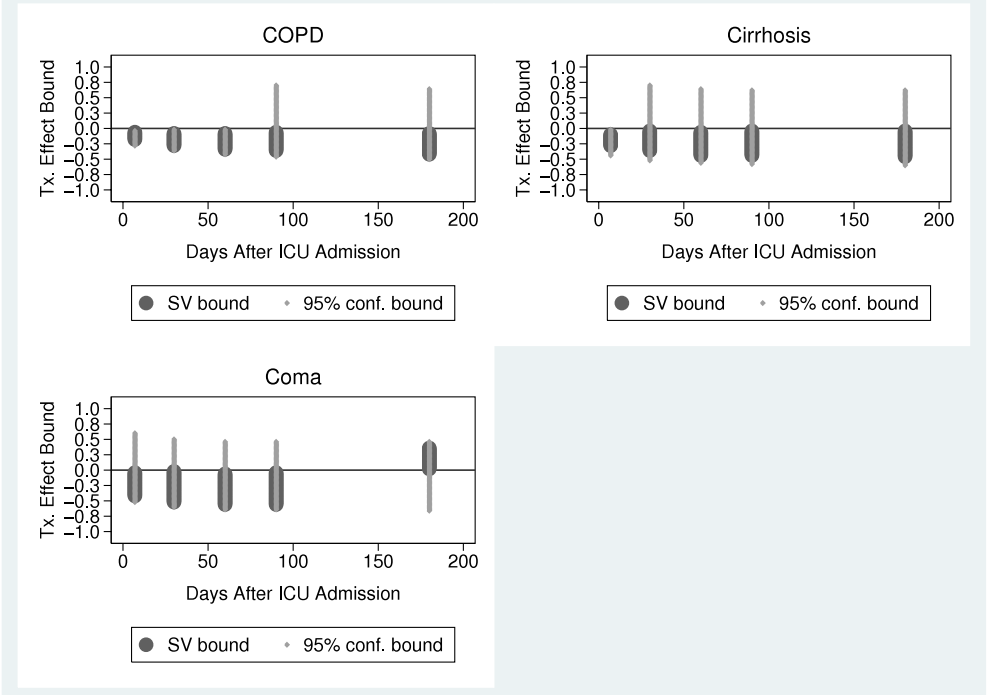

Fig. 7. Shaikh and Vytlacil (2011) bounds (part II).

We apply different bounding approaches that exploit access to our instrument, in particular, the recent approach introduced by Shaikh and Vytlacil (2011), which we compare with the approach of Manski (1990). We find that, while the bounds of Manski (1990) always straddle zero, the bounds of Shaikh and Vytlacil (2011) produce a clearer answer at least for some diagnoses—catheterization reduces mortality at 7 days after admission to the ICU. We extend the analysis of Shaikh and Vytlacil (2011) to exploit a further nonparametric structural assumption – that doctors catheterize individuals with systematically worse latent health – and find that this assumption further narrows these bounds and strengthens these conclusions.

The main theme of the paper is the trade-off induced by the acceptance of potentially unverifiable structural assumptions. If one is willing to accept very strong structural assumptions, such as those underlying the bivariate probit model, then one can obtain

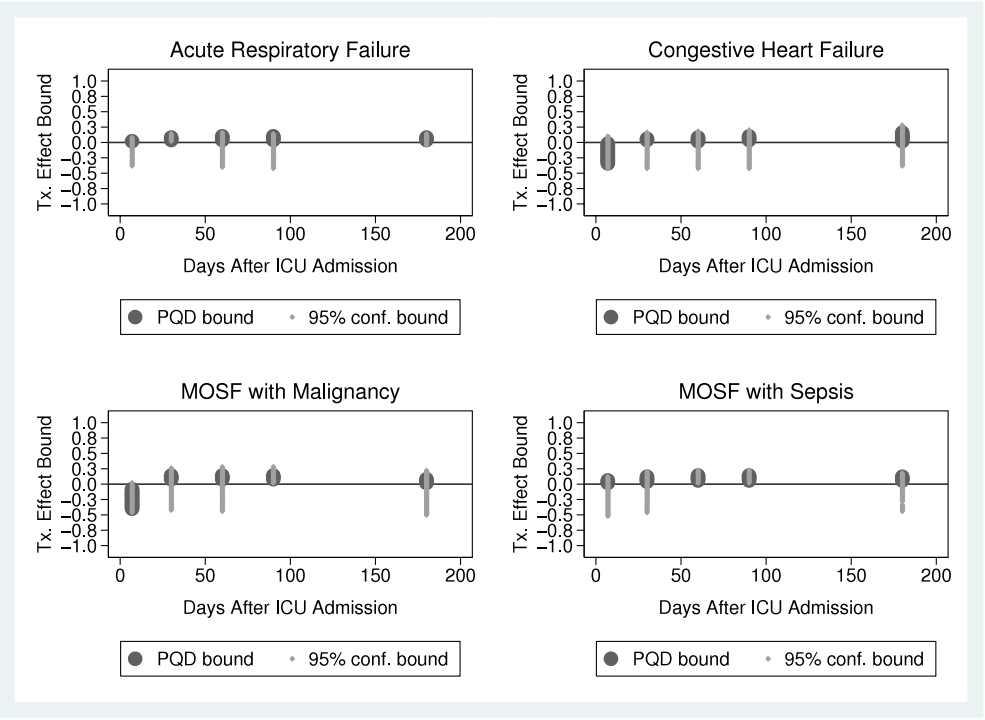

Fig. 8. PQD bounds (part I).

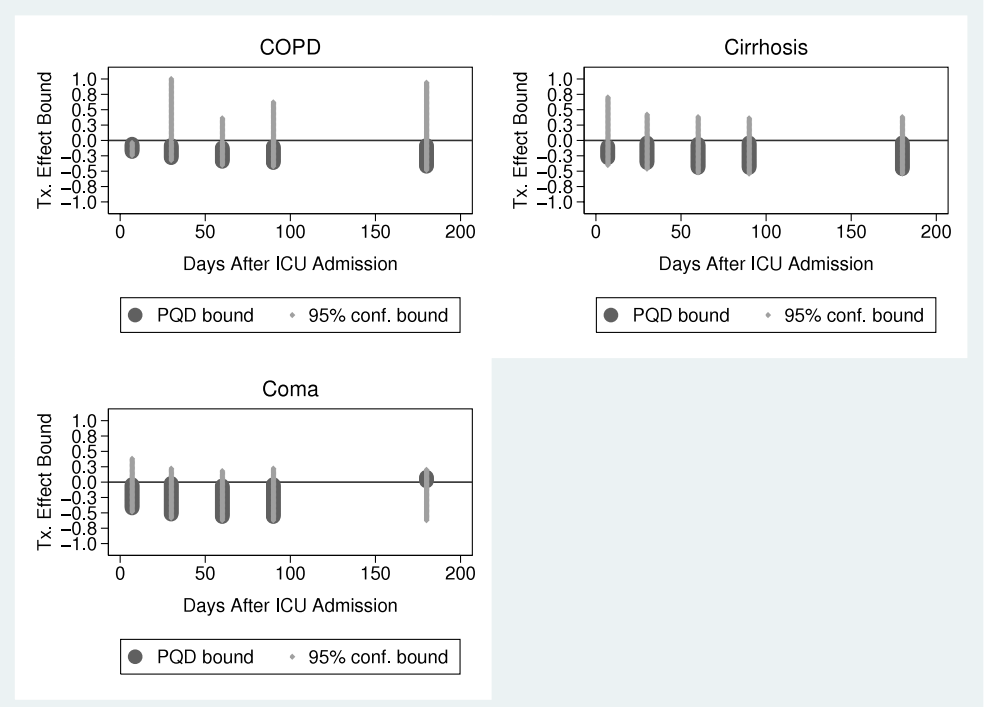

Fig. 9. PQD bounds (part II).

point identification. At the other extreme, if the only structural assumption one accepts is that probabilities lie between zero and one (such as in the Manski (1990) bounds without an instrument), then the width of the bounds on the average treatment effect is exactly one, so it is not possible to determine the sign of the average treatment effect. In between these two extremes, one may accept different nonparametric, structural assumptions, such as the validity of an instrument or threshold crossing models on the outcome or treatment variables, which may not lead to point identification, but may reduce the width of the bounds considerably, as in our empirical example, and are more palatable than the very strong parametric assumptions required for the bivariate probit model.

Our primary substantive finding is that catheterization improves mortality outcomes only in the short run, if at all, and in most cases we cannot rule out that it increases mortality in the long

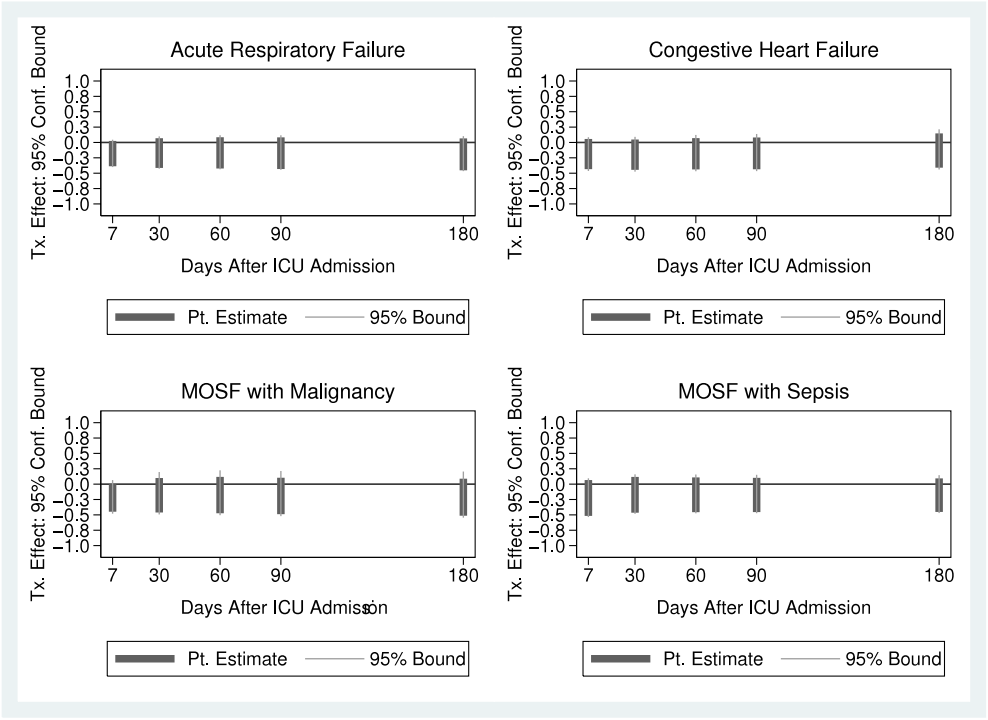

Fig. 10. PQD bounds without instrument (part I).

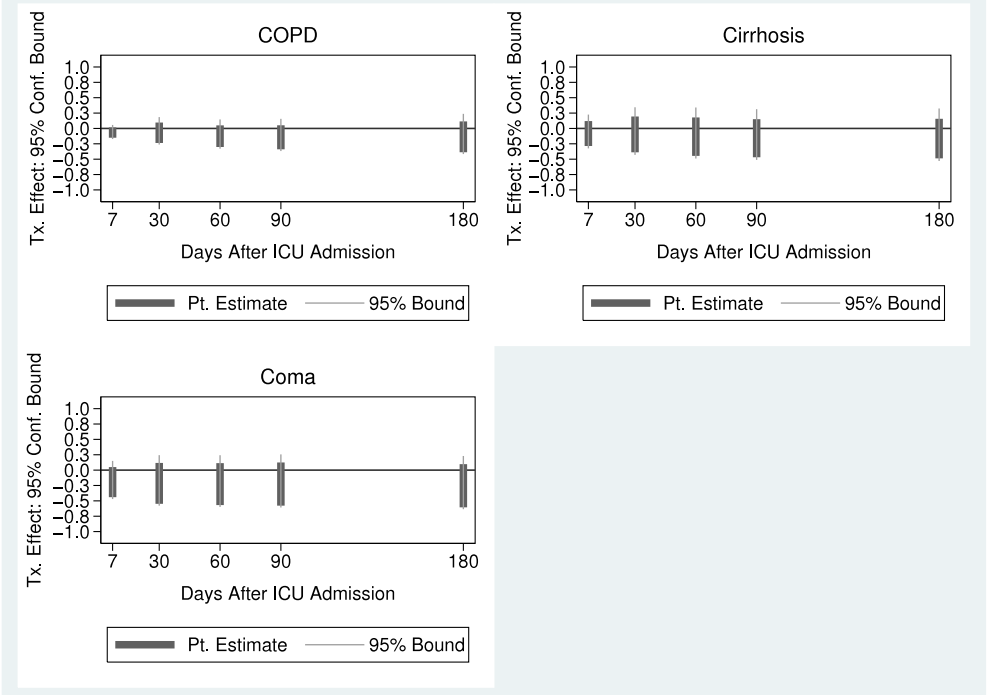

Fig. 11. PQD bounds without instrument (part II).

run. This finding is intuitively appealing because it suggests a possible explanation for the fact that many ICU doctors are 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 may never observe the increase in mortality. Our results also suggest a second (not mutually exclusive) possibility: 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.

Appendix A

In this appendix, we show that the confidence regions described in Section 5 behave well uniformly over a large class of possible distributions $\mathbf{P}$ for (Y, D, Z). Our main result is Theorem A.1 below.

Note that in the statement of the theorem we index the upper and lower bounds by $P \in \mathbf{P}$ to reflect their obvious dependence on the distribution of the observed data. The proof of the theorem utilizes Lemma B.6, which is established in the following Appendix.

Theorem A.1. *Let (Y_i, D_i, Z_i), $i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$. Suppose $\mathbf{P}$ is such that*

$$P\{Y = y, D = d, Z = z\} > \epsilon$$

for some $\epsilon > 0$ and

$$P\{D = 1|Z = 1\} > P\{D = 1|Z = 0\} \tag{15}$$

for all $(y, d, z) \in \{0, 1\}^3$ and $P \in \mathbf{P}$. Let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b_n/n \to 0$. Then, the following statements are true:

(i) *If $T_n(\theta)$ is defined as in Section 5.1, then $\mathcal{C}_n$ defined by (5) satisfies*

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} \inf_{B^L_M(P)\le\theta\le B^U_M(P)} P\{\theta \in \mathcal{C}_n\} \ge 1 - \alpha.$$

(ii) *If $T_n(\theta)$ is defined as in Section 5.2, then $\mathcal{C}_n$ defined by (5) satisfies*

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} \inf_{B^L_{SV}(P)\le\theta\le B^U_{SV}(P)} P\{\theta \in \mathcal{C}_n\} \ge 1 - \alpha.$$

(iii) *If $T_n(\theta)$ is defined as in Section 5.3, then $\mathcal{C}_n$ defined by (5) satisfies*

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} \inf_{B^L_{PQD}(P)\le\theta\le B^U_{PQD}(P)} P\{\theta \in \mathcal{C}_n\} \ge 1 - \alpha.$$

Proof. We prove only part (ii) of the theorem. The arguments for parts (i) and (iii) are very similar. To this end, suppose by way of contradiction that (ii) fails to hold. It follows that there exists $\eta < 1 - \alpha$ and a sequence $\{(P_n, \theta_n) \in \mathbf{P} \times [-1, 1] : n \ge 1\}$ with $B^L_{SV}(P_n) \le \theta_n \le B^U_{SV}(P_n)$ for all $n \ge 1$ such that

$$P_n\{\theta_n \in \mathcal{C}_n\} = P_n\{T_n(\theta_n) \le \hat{c}_n(\theta, 1 - \alpha)\} \to \eta. \tag{16}$$

By considering a subsequence if necessary, we may assume that $\theta_n > 0$, $\theta_n < 0$, or $\theta_n = 0$ for all $n \ge 1$. Suppose $\theta_n > 0$ for all $n \ge 1$. It follows that $T_n(\theta_n)$ is given by (14) and that

$$-(\mu_{Y|Z=1}(P_n) - \mu_{Y|Z=0}(P_n)) \le 0$$
$$(\mu_{Y|Z=1}(P_n) - \mu_{Y|Z=0}(P_n)) - \theta_n \le 0$$
$$\theta_n - (\mu_{DY+(1-D)|Z=1}(P_n) - \mu_{(1-D)Y|Z=0}(P_n)) \le 0.$$

Next, apply Lemma B.6 to $((-Y, Y-\theta, -DY-(1-D)), (-Y, Y, \theta-(1-D)Y), Z)$ by identifying P in Lemma B.6 with (P, θ) in the present context. It is straightforward to see that the conditions of Lemma B.6 are satisfied. It follows that (16) cannot hold. In a similar way, we reach a contradiction when $\theta_n < 0$ or $\theta_n = 0$ for all $n \ge 1$. The desired result thus follows. □

Remark A.1. Note that Theorem A.1 allows the equality in (15) to be arbitrarily close to an equality. In fact, part (i) of Theorem A.1 continues to hold even if we do not require (15) to hold for all $P \in \mathbf{P}$. □

Appendix B

In this appendix, we derive a series of lemmas, building up to Lemma B.6 which is used in the derivation of Theorem A.1.

Lemma B.1. *Let (X_i, Z_i), $i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\mathbf{R} \times \{0, 1\}$. Suppose*

$$\limsup_{\lambda\to\infty} \sup_{P\in\mathbf{P}} E_P\left[\left(\frac{X - \mu_{X|Z=1}(P)}{\sigma_{X|Z=1}(P)}\right)^2 \times I\left\{\left|\frac{X - \mu_{X|Z=1}(P)}{\sigma_{X|Z=1}(P)}\right| > \lambda\right\}\middle| Z = 1\right] = 0 \tag{17}$$

and for some $\epsilon > 0$ that

$$\inf_{P\in\mathbf{P}} P\{Z = 1\} > \epsilon. \tag{18}$$

Then,

$$\limsup_{\lambda\to\infty} \sup_{P\in\mathbf{P}} E_P\left[\left(\frac{(X - \mu_{X|Z=1}(P))Z}{\sigma_{(X-\mu_{X|Z=1}(P))Z}(P)}\right)^2 \times I\left\{\left|\frac{(X - \mu_{X|Z=1}(P))Z}{\sigma_{(X-\mu_{X|Z=1}(P))Z}(P)}\right| > \lambda\right\}\right] = 0. \tag{19}$$

Proof. Note that the lefthand-side of (19) equals

$$\limsup_{\lambda\to\infty} \sup_{P\in\mathbf{P}} E_P\left[\left(\frac{X - \mu_{X|Z=1}(P)}{\sigma_{X|Z=1}(P)\sqrt{P\{Z = 1\}}}\right)^2 \times I\left\{\left|\frac{X - \mu_{X|Z=1}(P)}{\sigma_{X|Z=1}(P)\sqrt{P\{Z = 1\}}}\right| > \lambda\right\}\middle| Z = 1\right] P\{Z = 1\}.$$

The desired result (19) now follows from (17) and (18). □

Lemma B.2. *Let Z_i, $i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\{0, 1\}$. Then,*

$$\limsup_{\lambda\to\infty} \sup_{P\in\mathbf{P}} E_P[|Z - \mu_Z(P)|I\{|Z - \mu_Z(P)| > \lambda\}] = 0. \tag{20}$$

Proof. Follows simply by noting that $|Z - \mu_Z(P)| \le 1$. □

Lemma B.3. *Let (X_i, Z_i), $i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\mathbf{R} \times \{0, 1\}$. Suppose for some $\epsilon > 0$ that (17) and (18) hold. Then, under any sequence $\{P_n \in \mathbf{P} : n \ge 1\}$,*

$$\hat{\sigma}^2_{n,X|Z=1}/\sigma^2_{X|Z=1}(P_n) \xrightarrow{P_n} 1.$$

Proof. Assume without loss of generality that $\mu_{X|Z=1}(P) = 0$ and $\sigma^2_{X|Z=1}(P) = 1$. Hence, $\mu_{XZ}(P) = 0$ and $\mu_{X^2Z}(P) = P\{Z = 1\}$. Note that

$$\frac{1}{n_1}\sum_{1\le i\le n: Z_i=1} X_i = \left(\frac{1}{n}\sum_{1\le i\le n} X_iZ_i\right)\Big/\left(\frac{1}{n}\sum_{1\le i\le n} Z_i\right) = \left(\frac{\frac{1}{n}\sum_{1\le i\le n} X_iZ_i}{P\{Z_i = 1\}}\right)\Big/\left(\frac{P\{Z_i = 1\}}{\frac{1}{n}\sum_{1\le i\le n} Z_i}\right).$$

From Lemma B.1, we see that (19) holds. Note that $\sigma_{XZ}(P) = \sqrt{P\{Z = 1\}}$. It therefore follows from Lemma 11.4.2 of Lehmann and Romano (2005) that

$$\frac{\frac{1}{n}\sum_{1\le i\le n} X_iZ_i}{\sqrt{P_n\{Z_i = 1\}}} \xrightarrow{P_n} 0.$$

Hence,

$$\frac{\frac{1}{n}\sum_{1\le i\le n} X_i Z_i}{P_n\{Z_i=1\}} \xrightarrow{P_n} 0.$$

From Lemma B.2, we see that (20) holds. It therefore follows from Lemma 11.4.2 of Lehmann and Romano (2005) that

$$\frac{1}{n}\sum_{1\le i\le n} Z_i - P_n\{Z_i=1\} \xrightarrow{P_n} 0.$$

Hence,

$$\frac{\frac{1}{n}\sum_{1\le i\le n} Z_i}{P_n\{Z_i=1\}} = \frac{\frac{1}{n}\sum_{1\le i\le n} Z_i - P_n\{Z_i=1\}}{P_n\{Z_i=1\}} + 1 \xrightarrow{P_n} 1.$$

Thus,

$$\frac{1}{n_1}\sum_{1\le i\le n: Z_i=1} X_i \xrightarrow{P_n} 0.$$

Next, note that

$$\begin{aligned}\frac{1}{n_1}\sum_{1\le i\le n: Z_i=1} X_i^2 &= \left(\frac{1}{n}\sum_{1\le i\le n} X_i^2 Z_i\right) \Big/ \left(\frac{1}{n}\sum_{1\le i\le n} Z_i\right) \\ &= \left(\frac{\frac{1}{n}\sum_{1\le i\le n} X_i^2 Z_i}{P\{Z_i=1\}}\right) \Big/ \left(\frac{P\{Z_i=1\}}{\frac{1}{n}\sum_{1\le i\le n} Z_i}\right).\end{aligned}$$

It follows from Lemma 11.4.3 of Lehmann and Romano (2005) that

$$\frac{\frac{1}{n}\sum_{1\le i\le n} X_i^2 Z_i}{P_n\{Z_i=1\}} \xrightarrow{P_n} 1.$$

Thus,

$$\frac{1}{n_1}\sum_{1\le i\le n: Z_i=1} X_i^2 \xrightarrow{P_n} 1.$$

The desired result now follows. □

Lemma B.4. *Let (X_i, Y_i, Z_i), $i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\mathbf{R}\times\mathbf{R}\times\{0,1\}$. Suppose that for some $\epsilon > 0$*

$$\epsilon < \inf_{P\in\mathbf{P}} P\{Z=1\} \le \sup_{P\in\mathbf{P}} P\{Z=1\} < 1-\epsilon \tag{21}$$

and that

$$\limsup_{\lambda\to\infty}\sup_{P\in\mathbf{P}} E_P\left[\left(\frac{X-\mu_{X|Z=1}(P)}{\sigma_{X|Z=1}(P)}\right)^2 \times I\left\{\left|\frac{X-\mu_{X|Z=1}(P)}{\sigma_{X|Z=1}(P)}\right| > \lambda\right\}\Big| Z=1\right] = 0$$

and

$$\limsup_{\lambda\to\infty}\sup_{P\in\mathbf{P}} E_P\left[\left(\frac{Y-\mu_{Y|Z=0}(P)}{\sigma_{Y|Z=0}(P)}\right)^2 \times I\left\{\left|\frac{Y-\mu_{Y|Z=0}(P)}{\sigma_{Y|Z=0}(P)}\right| > \lambda\right\}\Big| Z=0\right] = 0.$$

Then, the following are true:

(i) *For any sequence $\{P_n \in \mathbf{P} : n \ge 1\}$,*

$$\left(\frac{\sigma^2_{X|Z=1}(P_n)}{P_n\{Z_i=1\}} + \frac{\sigma^2_{Y|Z=0}(P_n)}{P_n\{Z_i=0\}}\right) \Big/ \left(\frac{\sigma^2_{X|Z=1}}{\frac{1}{n}\sum_{1\le i\le n} Z_i} + \frac{\sigma^2_{Y|Z=0}}{\frac{1}{n}\sum_{1\le i\le n}(1-Z_i)}\right) \xrightarrow{P_n} 1.$$

(ii) *For any sequence $\{P_n \in \mathbf{P} : n \ge 1\}$,*

$$\left(\frac{\sigma^2_{X|Z=1}(P_n)}{P_n\{Z_i=1\}} + \frac{\sigma^2_{Y|Z=0}(P_n)}{P_n\{Z_i=0\}}\right) \Big/ \left(\frac{\hat\sigma^2_{n,X|Z=1}}{\frac{1}{n}\sum_{1\le i\le n} Z_i} + \frac{\hat\sigma^2_{n,Y|Z=0}}{\frac{1}{n}\sum_{1\le i\le n}(1-Z_i)}\right) \xrightarrow{P_n} 1.$$

(iii) *For any sequence $\{P_n \in \mathbf{P} : n \ge 1\}$,*

$$\left(\frac{\sigma^2_{X|Z=1}(P_n)}{n_1} + \frac{\sigma^2_{Y|Z=0}(P_n)}{n_0}\right) \Big/ \left(\frac{\hat\sigma^2_{n,X|Z=1}}{n_1} + \frac{\hat\sigma^2_{n,Y|Z=0}}{n_0}\right) \xrightarrow{P_n} 1.$$

Proof. To establish (i), note from Lemma B.2 and Lemma 11.4.2 of Lehmann and Romano (2005) that

$$\frac{\sigma^2_{X|Z=1}(P_n)}{P_n\{Z_i=1\}} \Big/ \frac{\sigma^2_{X|Z=1}(P_n)}{\frac{1}{n}\sum_{1\le i\le n} Z_i} \xrightarrow{P_n} 1$$

$$\frac{\sigma^2_{Y|Z=0}(P_n)}{P_n\{Z_i=0\}} \Big/ \frac{\sigma^2_{Y|Z=0}(P_n)}{\frac{1}{n}\sum_{1\le i\le n}(1-Z_i)} \xrightarrow{P_n} 1.$$

Note further that for any positive real numbers a, b, c and d, that

$$\left|\frac{a+b}{c+d} - 1\right| \le \left|\frac{a}{c} - 1\right| + \left|\frac{b}{d} - 1\right|.$$

The desired result thus follows. A similar argument establishes (ii) and (iii). □

Lemma B.5. *Let (X_i, Y_i, Z_i), $i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\mathbf{R}^k \times \mathbf{R}^k \times \{0,1\}$. Suppose (21) holds for some $\epsilon > 0$ and for each $1 \le j \le k$ that*

$$\limsup_{\lambda\to\infty}\sup_{P\in\mathbf{P}} E_P\left[\left(\frac{X_j-\mu_{X_j|Z=1}(P)}{\sigma_{X_j|Z=1}(P)}\right)^2 \times I\left\{\left|\frac{X_j-\mu_{X_j|Z=1}(P)}{\sigma_{X_j|Z=1}(P)}\right| > \lambda\right\}\Big| Z=1\right] = 0 \tag{22}$$

and

$$\limsup_{\lambda\to\infty}\sup_{P\in\mathbf{P}} E_P\left[\left(\frac{Y_j-\mu_{Y_j|Z=0}(P)}{\sigma_{Y_j|Z=0}(P)}\right)^2 \times I\left\{\left|\frac{Y_j-\mu_{Y_j|Z=0}(P)}{\sigma_{Y_j|Z=0}(P)}\right| > \lambda\right\}\Big| Z=0\right] = 0. \tag{23}$$

Define $W_n(P)$ to be the vector whose jth element for $1 \le j \le k$ is given by

$$\frac{\frac{1}{n_1}\sum_{1\le i\le n: Z_i=1} X_{j,i} - \mu_{X_j|Z=1}(P) - \frac{1}{n_0}\sum_{1\le i\le n: Z_i=0} Y_{j,i} - \mu_{Y_j|Z=0}(P)}{\sqrt{\frac{\sigma^2_{X_j|Z=1}(P)}{n_1} + \frac{\sigma^2_{Y_j|Z=0}(P)}{n_0}}}$$

and

$$V(P) = D(P)\Omega_{X|Z=1}(P) + (I - D(P))\Omega_{Y|Z=0}(P), \tag{24}$$

where $\Omega_{X|Z=1}(P)$ is the correlation matrix of X conditional on $Z = 1$ under P, $\Omega_{Y|Z=0}(P)$ is the correlation matrix of Y conditional on $Z = 0$ under P,

$$D(P) = diag\left(\frac{\frac{\sigma^2_{X_1|Z=1}(P)}{P\{Z_i=1\}}}{\frac{\sigma^2_{X_1|Z=1}(P)}{P\{Z_i=1\}} + \frac{\sigma^2_{Y_1|Z=0}(P)}{P\{Z_i=0\}}}, \ldots, \frac{\frac{\sigma^2_{X_k|Z=1}(P)}{P\{Z_i=1\}}}{\frac{\sigma^2_{X_k|Z=1}(P)}{P\{Z_i=1\}} + \frac{\sigma^2_{Y_k|Z=0}(P)}{P\{Z_i=0\}}}\right),$$

and I is the k-dimensional identity matrix. Let $\{P_n \in \mathbf{P} : n \geq 1\}$ be such that $V(P_n) \to V^*$ for some matrix V^*. Then,

$$W_n(P_n) \xrightarrow{d} \Phi_{V^*}(x) \tag{25}$$

under P_n.

Proof. Assume without loss of generality that $\mu_{X|Z=1}(P) = \mu_{Y|Z=0}(P) = 0$. It follows from Lemma B.2, Lemma 11.4.2 of Lehmann and Romano (2005), and part (i) of Lemma B.4 that for any $1 \leq j \leq k$

$$\frac{\frac{1}{n_1}\sum_{1\leq i\leq n: Z_i=1} X_{j,i}}{\sqrt{\frac{\sigma^2_{X_j|Z=1}(P)}{n_1} + \frac{\sigma^2_{Y_j|Z=0}(P)}{n_0}}} = (1 + \delta_{1,n}(P))\frac{\sqrt{\frac{\sigma^2_{X_j|Z=1}(P)}{P\{Z_i=1\}}}}{\sqrt{\frac{\sigma^2_{X_j|Z=1}(P)}{P\{Z_i=1\}} + \frac{\sigma^2_{Y_j|Z=0}(P)}{P\{Z_i=0\}}}} \frac{\frac{1}{\sqrt{n}}\sum_{1\leq i\leq n}\frac{X_{j,i}Z_i}{P\{Z_i=1\}}}{\sqrt{\frac{\sigma^2_{X_j|Z=1}(P)}{P\{Z_i=1\}}}} \tag{26}$$

$$\frac{\frac{1}{n_0}\sum_{1\leq i\leq n: Z_i=0} Y_{j,i}}{\sqrt{\frac{\sigma^2_{X_j|Z=1}(P)}{n_1} + \frac{\sigma^2_{Y_j|Z=0}(P)}{n_0}}} = (1 + \delta_{0,n}(P))\frac{\sqrt{\frac{\sigma^2_{Y_j|Z=0}(P)}{P\{Z_i=0\}}}}{\sqrt{\frac{\sigma^2_{X_j|Z=1}(P)}{P\{Z_i=1\}} + \frac{\sigma^2_{Y_j|Z=0}(P)}{P\{Z_i=0\}}}} \frac{\frac{1}{\sqrt{n}}\sum_{1\leq i\leq n}\frac{Y_{j,i}(1-Z_i)}{P\{Z_i=0\}}}{\sqrt{\frac{\sigma^2_{Y_j|Z=0}(P)}{P\{Z_i=0\}}}}, \tag{27}$$

where

$$\delta_{1,n}(P_n) \xrightarrow{P_n} 0$$
$$\delta_{0,n}(P_n) \xrightarrow{P_n} 0.$$

Define $W^*_n(P)$ to be the vector whose first k elements are given by (26) for $1 \leq j \leq k$ and whose second k elements are given by (27) for $1 \leq j \leq k$.

Suppose by way of contradiction that (25) fails. Then, there exists a subsequence $\{P_{n_m} \in \mathbf{P} : m \geq 1\}$ and $x \in \mathbf{R}^k$ such that

$$P_{n_m}\{W_{n_m}(P_{n_m}) \leq x\} \not\to \Phi_{V^*}(x).$$

By considering a further subsequence if necessary, we may assume that

$$D(P_{n_m}) \to D^*$$
$$\Omega_{Y|Z=0}(P_{n_m}) \to \Omega^*_0$$
$$\Omega_{X|Z=1}(P_{n_m}) \to \Omega^*_1$$

for matrices D^*, Ω^*_0 and Ω^*_1 such that

$$V^* = D^*\Omega^*_1 + (I - D^*)\Omega^*_0.$$

It suffices to show that

$$W^*_{n_m}(P_{n_m}) \xrightarrow{d} N(0, \bar{V}) \tag{28}$$

under P_{n_m}, where

$$\bar{V} = \text{diag}(D^*\Omega^*_1, (I - D^*)\Omega^*_0).$$

From Lemma B.1, we see that

$$\limsup_{\lambda\to\infty}\sup_{P\in\mathbf{P}} E_P\left[\left(\frac{X_jZ - \mu_{X_jZ}(P)}{\sigma_{X_jZ}(P)}\right)^2 \times I\left\{\left|\frac{X_jZ - \mu_{X_jZ}(P)}{\sigma_{X_jZ}(P)}\right| > \lambda\right\}\right] = 0 \tag{29}$$

and

$$\limsup_{\lambda\to\infty}\sup_{P\in\mathbf{P}} E_P\left[\left(\frac{Y_j(1-Z) - \mu_{Y_j(1-Z)}(P)}{\sigma_{Y_j(1-Z)}(P)}\right)^2 \times I\left\{\left|\frac{Y_j(1-Z) - \mu_{Y_j(1-Z)}(P)}{\sigma_{Y_j(1-Z)}(P)}\right| > \lambda\right\}\right] = 0. \tag{30}$$

Furthermore, (29) and (30) continue to hold if X_jZ and $Y_j(1-Z)$ are replaced with $\frac{X_jZ}{P\{Z_i=1\}}$ and $\frac{Y_j(1-Z)}{P\{Z_i=0\}}$, respectively. Finally, note that

$$\sigma^2_{\frac{X_jZ}{P\{Z=1\}}}(P) = \frac{\sigma^2_{X_j|Z=1}(P)}{P\{Z=1\}}$$
$$\sigma^2_{\frac{Y_j(1-Z)}{P\{Z=0\}}}(P) = \frac{\sigma^2_{Y_j|Z=0}(P)}{P\{Z=0\}}$$

and

$$N(0, \bar{V}_{n_m}) \xrightarrow{d} N(0, \bar{V}),$$

where

$$\bar{V}_n = \text{diag}(D(P_n)\Omega_{X|Z=1}(P_n), (I - D(P_n))\Omega_{Y|Z=0}(P_n)).$$

The desired conclusion (28) now follows from Lemma 3.1 of Romano and Shaikh (2008) and Slutsky's Theorem. □

Lemma B.6. *Let (X_i, Y_i, Z_i), $i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\mathbf{R}^k \times \mathbf{R}^k \times \{0, 1\}$. Suppose (21) holds for some $\epsilon > 0$ and for each $1 \leq j \leq k$ that*

$$\mu_{X_j|Z=1}(P) - \mu_{Y_j|Z=0}(P) \leq 0$$

for all $P \in \mathbf{P}$ and that (22) and (23) hold. Define

$$T_n = \sum_{1\leq j\leq k}\left(\frac{\frac{1}{n_1}\sum_{1\leq i\leq n: Z_i=1} X_{j,i} - \frac{1}{n_0}\sum_{1\leq i\leq n: Z_i=0} Y_{j,i}}{\sqrt{\frac{\hat{\sigma}^2_{n,X_j|Z=1}}{n_1} + \frac{\hat{\sigma}^2_{n,Y_j|Z=0}}{n_0}}}\right)^2_+$$

and $J_n(x, P) = P\{T_n \leq x\}$. Let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \to 0$. Index by $i = 1, \ldots, N_n = \binom{n}{b}$ the different subsets of $\{1, \ldots, n\}$ of size b. Denote by $T_{n,b,i}$ the test statistic T_n computed using only the ith subset of data of size b. Let

$$\hat{c}_n(1-\alpha) = \inf\left\{x \in \mathbf{R} : \frac{1}{N_n}\sum_{1\leq i\leq N_n} I\{T_{n,b,i} \leq x\} \leq 1 - \alpha\right\}.$$

Then,

$$\liminf_{n\to\infty}\inf_{P\in\mathbf{P}} P\{T_n \leq \hat{c}_n(1-\alpha)\} \geq 1 - \alpha.$$

Proof. From Theorem 2.1 of Romano and Shaikh (2010b), it suffices to show that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{x\in\mathbf{R}} \{J_b(x,P) - J_n(x,P)\} \leq 0. \tag{31}$$

In order to establish (31), first note that because $T_n \geq 0$, it is enough to consider the supremum over $x \geq 0$. For $1 \leq j \leq k$, define

$$T_{n,j} = \frac{\frac{1}{n_1}\sum_{1\leq i\leq n: Z_i=1} X_{j,i} - \frac{1}{n_0}\sum_{1\leq i\leq n: Z_i=0} Y_{j,i}}{\sqrt{\frac{\hat\sigma^2_{n,X_j|Z=1}}{n_1} + \frac{\hat\sigma^2_{n,Y_j|Z=0}}{n_0}}}$$

$$T^*_{n,j}(P) = \frac{\frac{1}{n_1}\sum_{1\leq i\leq n: Z_i=1} X_{j,i} - \mu_{X_j|Z=1}(P) - \frac{1}{n_0}\sum_{1\leq i\leq n: Z_i=0} Y_{j,i} - \mu_{Y_j|Z=0}(P)}{\sqrt{\frac{\hat\sigma^2_{n,X_j|Z=1}}{n_1} + \frac{\hat\sigma^2_{n,Y_j|Z=0}}{n_0}}}.$$

Note that

$$T_{n,j} = T^*_{n,j}(P) + \hat\Delta_{n,j}(P),$$

where

$$\hat\Delta_{n,j}(P) = \frac{\mu_{X_j|Z=1}(P) - \mu_{Y_j|Z=0}(P)}{\sqrt{\frac{\hat\sigma^2_{n,X_j|Z=1}}{n_1} + \frac{\hat\sigma^2_{n,Y_j|Z=0}}{n_0}}}.$$

Further note that

$$\hat\Delta_{n,j}(P) = \delta_{n,j}(P)\Delta_{n,j}(P),$$

where

$$\delta_{n,j}(P) = \left(\sqrt{\frac{\sigma^2_{X_j|Z=1}(P)}{P\{Z_i=1\}} + \frac{\sigma^2_{Y_j|Z=0}(P)}{P\{Z_i=0\}}}\right) \Bigg/ \left(\sqrt{\frac{\hat\sigma^2_{n,X_j|Z=1}}{\frac{1}{n}\sum_{1\leq i\leq n} Z_i} + \frac{\hat\sigma^2_{n,Y_j|Z=0}}{\frac{1}{n}\sum_{1\leq i\leq n}(1-Z_i)}}\right).$$

$$\Delta_{n,j}(P) = \sqrt{n}\frac{\mu_{X_j|Z=1}(P) - \mu_{Y_j|Z=0}(P)}{\sqrt{\frac{\sigma^2_{X_j|Z=1}(P)}{P\{Z_i=1\}} + \frac{\sigma^2_{Y_j|Z=0}(P)}{P\{Z_i=0\}}}}.$$

We may therefore write

$$\begin{aligned} J_n(x,P) &= P\{T_n \leq x\} \\ &= P\left\{\sum_{1\leq j\leq k} (T_{n,j})^2_+ \leq x\right\} \\ &= P\left\{\sum_{1\leq j\leq k} (T^*_{n,j}(P) + \hat\Delta_{n,j}(P))^2_+ \leq x\right\} \\ &= P\left\{\sum_{1\leq j\leq k} (T^*_{n,j}(P) + \delta_{n,j}(P)\Delta_{n,j}(P))^2_+ \leq x\right\}. \end{aligned}$$

Since $b \leq n$ and $\mu_{X_j|Z=1}(P) - \mu_{Y_j|Z=0}(P) \leq 0$, we see that $\Delta_{n,j}(P) \leq 0$ and

$$\Delta_{b,j}(P) \geq \Delta_{n,j}(P).$$

Since $\delta_{n,j}(P) \geq 0$, it follows that

$$J_b(x,P) \leq J^*_b(x,P),$$

where

$$J^*_b(x,P) = P\left\{\sum_{1\leq j\leq k} (T^*_{b,j}(P) + \delta_{b,j}(P)\Delta_{n,j}(P))^2_+ \leq x\right\}.$$

It therefore suffices to show that

$$\sup_{P\in\mathbf{P}} \sup_{x\geq 0} |J^*_b(x,P) - J_n(x,P)| \to 0. \tag{32}$$

Suppose by way of contradiction that (32) fails to hold. It follows that there exists $\eta > 0$ and a sequence $\{P_n \in \mathbf{P} : n \geq 1\}$ such that

$$\sup_{x\geq 0} |J^*_b(x,P_n) - J_n(x,P_n)| \to \eta. \tag{33}$$

By extracting a further subsequence if necessary, we may assume that

$$V(P_n) \to V^*,$$

where $V(P)$ is given by (24), for some matrix V^*. Define $\hat W_n(P)$ to be the vector whose jth element for $1 \leq j \leq k$ is given by $T^*_{n,j}(P)$. It follows from part (iii) of Lemma B.4, Lemma B.5 and Slutsky's Theorem that

$$\hat W_n(P_n) \xrightarrow{d} N(0, V^*)$$

under P_n. Similarly, we see that

$$\hat W_b(P_n) \xrightarrow{d} N(0, V^*)$$

under P_n. There are two cases to consider. First consider the case where there is a subsequence $\{P_{n_m} \in \mathbf{P} : m \geq 1\}$ such that for all $1 \leq j \leq k$

$$\Delta_{n_m,j}(P_{n_m}) \to -\infty.$$

By Lemma B.4, for all $1 \leq j \leq k$,

$$\delta_{n_m,j}(P_{n_m}) \xrightarrow{P_{n_m}} 1$$

$$\delta_{b_{n_m},j}(P_{n_m}) \xrightarrow{P_{n_m}} 1.$$

Hence,

$$\sup_{x\geq 0} |J^*_{b_{n_m}}(x,P_{n_m}) - 1| \to 0$$

$$\sup_{x\geq 0} |J_{n_m}(x,P_{n_m}) - 1| \to 0.$$

It therefore follows from the triangle inequality that

$$\sup_{x\geq 0} |J^*_{b_{n_m}}(x,P_{n_m}) - J_{n_m}(x,P_{n_m})| \to 0. \tag{34}$$

If this is not the case, then there is a subsequence $\{P_{n_m} \in \mathbf{P} : m \geq 1\}$ and $\emptyset \neq J \subseteq \{1,\ldots,k\}$ such that for all $j \notin J$

$$\Delta_{n_m,j}(P_{n_m}) \to -\infty$$

and for all $j \in J$

$$\Delta_{n_m,j}(P_{n_m}) \to -c_j$$

for some $c_j > 0$. It follows that

$$\sum_{1\leq j\leq k} (T^*_{b_{n_m},j}(P_{n_m}) + \delta_{b_{n_m},j}(P_{n_m})\Delta_{n_m,j}(P_{n_m}))^2_+ \xrightarrow{d} \sum_{j\in J}(Z_j - c_j)^2_+ \tag{35}$$

under P_{n_m}, where $Z \sim N(0, V^*)$. Note that the distribution of the righthand-side of (35) is continuous everywhere except possibly at zero. It is straightforward to check that

$$P_{n_m}\left\{\sum_{1\leq j\leq k} (T^*_{b_{n_m},j}(P_{n_m}) + \delta_{b_{n_m},j}(P_{n_m})\Delta_{n_m,j}(P_{n_m}))^2_+ \leq 0\right\} \to P\left\{\sum_{j\in J}(Z_j - c_j)^2_+ \leq 0\right\}.$$

Hence,

$$\sup_{x\geq 0}\left|J^*_{b_{n_m}}(x,P_{n_m})-P\left\{\sum_{j\in J}(Z_j-c_j)^2_+\leq x\right\}\right|$$
$$=\sup_{x\geq 0}\left|P_{n_m}\left\{\sum_{1\leq j\leq k}(T^*_{b_{n_m},j}(P_{n_m})+\delta_{b_{n_m},j}(P_{n_m})\Delta_{n_m,j}(P_{n_m}))^2_+\leq x\right\}\right.$$
$$\left.-\ P\left\{\sum_{j\in J}(Z_j-c_j)^2_+\leq x\right\}\right|\to 0,$$

where the convergence to zero follows from Lemma A.2 of Romano and Shaikh (2010a). Similarly, we see that

$$\sup_{x\geq 0}\left|J_{n_m}(x,P_{n_m})-P\left\{\sum_{j\in J}(Z_j-c_j)^2_+\leq x\right\}\right|\to 0.$$

From the triangle inequality, we see again that (34) holds. We thus reach a contradiction to (33), from which the desired result follows. □

References

Almond, D., Doyle, J., 2008. After midnight: a regression discontinuity design in length of postpartum hospital stays. NBER Working Paper No. 13877.

Altonji, A., Elder, T., Taber, C., 2008. Using selection on observed variables to assess bias from unobservables when evaluating Swan–Ganz catheterization. American Economic Review, Papers and Proceedings 98 (2), 345–350.

Altonji, J., Elder, T., Taber, C., 2005. Selection on observed and unobserved variables: assessing the effectiveness of catholic schools. Journal of Political Economy 113, 151–184.

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

Bhattacharya, J., Shaikh, A.M., Vytlacil, E., 2008. Treatment effect bounds under monotonicity assumptions: an application to Swan Ganz catheterization. American Economic Review, Papers and Proceedings 98 (2), 351–356.

Blundell, R., Gosling, A., Ichimura, H., Meghir, C., 2007. Changes in the distribution of male and female wages accounting for employment composition using bounds. Econometrica 75 (2), 323–363.

Card, D., Dobkin, C., Maestas, N., 2007. Does medicare save lives? NBER Working Paper No. 13668.

Chernozhukov, V., Hansen, C., 2005. An iv model of quantile treatment effects. Econometrica 73 (1), 245–261.

Chernozhukov, V., Hong, H., Tamer, E., 2007. Estimation and confidence regions for parameter sets in econometric models. Econometrica 75 (5), 1243–1284.

Chernozhukov, V., Lee, S., Rosen, A., 2009. Intersection Bounds: Estimation and Inference. Mimeo, University College London.

Chiburis, R., 2010. Semiparametric bounds on treatment effects. Journal of Econometrics 159 (2), 267–275.

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., 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.

Dobkin, C., 2003. Hospital Staffing and Inpatient Mortality. Mimeo, UC Santa Cruz.

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

Fowler, R., 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., 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., Goldman, D., 2000. Queing for surgery: is the US 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., Vytlacil, E., 2005. Structural equations, treatment effects, and econometric policy evaluation. Econometrica 73 (3), 669–738.

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

Hirano, K., 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.

Imbens, G., Angrist, J., 1994. Identification and estimation of local average treatment effects. Econometrica 62, 467–476.

Imbens, G., Manski, C., 2004. Confidence intervals for partially identified parameters. Econometrica 72, 1845–1857.

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

Lehmann, E.L., Romano, J.P., 2005. Testing Statistical Hypotheses, third edition. Springer, New York.

Li, Q., Racine, J., Wooldrdige, J., 2008. Estimating average treatment effects with continuous and discrete covariates: the case of Swan–Ganz catheterization. American Economic Review, Papers and Proceedings 98 (2), 357–362.

Li, Q., Racine, J., Wooldrdige, J., 2009. Efficient estimation of average treatment effects with mixed categorical and continuous data. Journal of Business and Economic Statistics 27 (2), 206–223.

Machado, C., Shaikh, A.M., Vytlacil, E., 2011. Instrumental variables, alternative monotonicity restrictions, and identifying the sign of the average treatment effect. Unpublished working paper.

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

Manski, C., Pepper, J., 2000. Monotone instrumental variables: with an application to the returns to schooling. Econometrica 68, 997–1010.

Murphy, D., 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.

Politis, D.N., Romano, J.P., Wolf, M., 1999. Subsampling. Springer, New York.

Rapoport, J., Teres, D., Steingrub, J., Higgins, T., McGee, W., 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., 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.

Romano, J.P., Shaikh, A.M., 2008. Inference for identifiable parameters in partially identified econometric models. Journal of Statistical Planning and Inference—Special Issue in Honor of Ted Anderson 138 (9), 2786–2807.

Romano, J.P., Shaikh, A.M., 2010b. On the uniform asymptotic validity of subsampling and the bootstrap. Mimeo, Stanford University and University of Chicago.

Romano, J.P., Shaikh, A.M., 2010a. Inference for the identified set in partially identified econometric models. Econometrica 78 (1), 169–211.

Sandham, J., Hull, R., Brant, R., Knox, L., Pineo, G., Doig, C., Laporta, D., Viner, S., Passerini, L., Devitt, H., Kirby, A., 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.M., Vytlacil, E., 2011. Partial identification in triangular systems of equations with binary dependent variables. Econometrica 79 (3), 949–955.

Vytlacil, E., 2002. Independence, monotonicity, and latent index models: an equivalence result. Econometrica 70 (1), 331–341.

Vytlacil, E., Yildiz, N., 2007. Dummy endogenous variables in weakly separable models. Econometrica 75 (3), 775–779.

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., 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., Behar, S., 1990. Use of pulmonary artery catheters in patients with acute myocardial infarction. analysis of experience in 5841 patients in the sprint registry. Chest 98, 1331–1335.