Annu. Rev. Public Health. 1998. 19:17–34

ECONOMETRICS IN OUTCOMES RESEARCH: The Use of Instrumental Variables

Joseph P. Newhouse[1] *and Mark McClellan*[2]
[1]Division of Health Policy Research and Education, Harvard University, 180 Longwood Avenue, Boston, Massachusetts 02115; [2]Department of Economics, Stanford University, Stanford, California 94305;
e-mail: newhouse@hcp.med.Harvard.edu

KEY WORDS: outcomes research, econometrics, instrumental variables, acute myocardial infarction

ABSTRACT

We describe an econometric technique, instrumental variables, that can be useful in estimating the effectiveness of clinical treatments in situations when a controlled trial has not or cannot be done. This technique relies upon the existence of one or more variables that induce substantial variation in the treatment variable but have no direct effect on the outcome variable of interest. We illustrate the use of the technique with an application to aggressive treatment of acute myocardial infarction in the elderly.

Outcomes research has come to mean many things to many people. To some it means an observational study to establish the consequences of some therapeutic intervention(s), typically using a large administrative data base (12). One of the aims behind the founding of the Agency for Health Care Policy and Research in the late 1980s was to promote such studies in the hope they would inform both coverage decisions in programs such as Medicare as well as improve clinical decision making generally. Many of the Patient Outcome Research Teams (PORTs) supported by the Agency have carried out such studies.

To others outcomes research means employing endpoints other than mortality, especially functional status, when evaluating treatment alternatives. The

0163-7525/98/0510-0017$08.00

Medical Outcome Study, for example, had additional endpoints as one of its principal aims (see Reference 14, preface).

For the purposes of this chapter we have in mind the former meaning of outcomes research and in particular the statistical tools that can be used to exploit the information in large administrative data bases. We write from our experience with the Harvard PORT on Acute Myocardial Infarction (AMI).

Some proponents of outcomes research, meaning an observational study of a clinical intervention, hope that such research can serve as a substitute for a randomized controlled trial (RCT) when a trial has not been or cannot be carried out, as is often the case. A trial, for example, may be not be feasible because a well-established technique is widely believed to be efficacious; as a result, it would be unethical to randomize an individual to no treatment. Indeed, a trial must come in that window of time where there is enough belief that a treatment is efficacious so that it is considered ethical to randomize patients to the treatment group but not sufficient belief in the efficacy of the treatment that it would be considered unethical to withhold the treatment.

Trials, of course, have been carried out in many areas of medicine, including the case we examine in detail below, that of catheterization and revascularization for AMI. Do observational studies have a role to play if results from a trial are available? Although many clinical researchers may answer this question negatively, the trial results may be less than satisfactory for a number of well-known reasons. First, trials tend to be performed in major clinical centers, and results may not be similar if the same procedure is performed elsewhere. Indeed, this is such a common problem that it has a well-recognized name; the trial may demonstrate that a procedure is *efficacious* (i.e. obtains desired results under optimal conditions), but it will not necessarily show that it is *effective* (obtains desired results under typical or standard conditions). Or, somewhat related to this point, in the time since the trial was conducted physicians may have become better at performing a procedure, such that the results of the trial are no longer relevant to current practice.

Furthermore, the population included in trials is often not representative of the population under treatment, meaning that the results of the trial cannot necessarily be generalized to the population actually being treated. For example, until relatively recently, women were underrepresented in many trials. Even today the elderly are frequently underrepresented. Those with comorbidities are often excluded. In the jargon of evaluation research this problem also has a name; the results of the trial may have internal validity (comparisons between the treatment and control groups are unbiased for the population being studied) but not external validity (results do not necessarily apply to other populations). For all of these reasons, clinical trials cannot or do not address all the relevant questions pertaining to the effects of therapeutic interventions.

As a result, it was the hope of many that observational studies or outcomes research could serve either as a substitute for a clinical trial, if a trial could not be carried out, or alternatively as a complement to a trial, if a trial had been carried out but raised questions of generalizability. This hope, however, was greeted with great skepticism on the part of many in the medical community (10). RCTs had become the gold standard of clinical research for good reason. In observational data patients would often be treated differently based on their underlying condition. In such cases, comparing outcomes across patients would confound the effect of the treatment with the effect of the underlying condition. As we shall show, this confounding, often referred to as a selection problem in the econometric literature, certainly exists in the case of catheterization of AMI patients. By contrast, in a well-executed RCT, patient condition is independent of treatment, and one can therefore reasonably attribute observed effects to the particular variation in treatment being studied.

In this review we describe a technique, long used in econometrics, that addresses this confounding problem. This technique is termed instrumental variables or IV. We shall show, however, that the IV technique may or may not be useful or even applicable in any particular outcomes research problem and that in any event, it is in general addressed to a somewhat different question than the trial is seeking to answer.

Instrumental Variables

The IV technique has been known for over half a century and is discussed in almost every econometrics textbook (e.g. 2–6). It is widely used in economics—some might say overused—because of the difficulty of doing controlled experiments in economics. Despite its popularity in economics, it was little known in the statistical and biostatistical literature until recently (1, 11). We do not give a formal exposition of IV here; for that the reader is referred to any of the textbooks just cited. Rather we present a nontechnical, intuitive explanation of the technique and then illustrate it using an application we have presented in greater detail elsewhere (7–9).

An Intuitive Explanation of Instrumental Variables

We begin by further explaining the problem that we sketched above, analyzing observational data with standard methods. Ordinary least squares (OLS) regression analysis is a widely used method for estimating constants, generally symbolized as β's, in the following kind of equation:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i, \qquad 1.$$

where the x's are variables that explain variation in another variable y, u is a random error term, and i indexes observations. For example, y may be a medical

outcome, such as functional status, x_1 may be a treatment designed to affect functional status, such as a drug, and x_2 may be an explanatory variable such as the age of the i^{th} person.

In a clinical trial the treatment (e.g. the drug) is assigned to persons randomly; with sufficient sample size, randomization ensures that with high probability the group of those receiving and the group of those not receiving the drug are similar. As a result, with data from a clinical trial one can usually estimate the effect of the treatment reasonably well simply by subtracting the means of the treatment and control groups. For example, if 90% of those in the treatment group recovered compared with 80% in the control group, one would estimate the effect of the treatment as a gain of 10 percentage points in the likelihood of recovery.

If one estimated an equation such as Equation 1 using OLS and included on the right-hand side only a variable such as x_1 that measured whether a subject was on the experimental treatment (as well as an intercept term), the estimated coefficient β_1 would equal the difference in sample means or ten percentage points in the above example.[1]

One could improve on the naïve estimator of subtracting sample means (or equivalently estimating Equation 1 with only an intercept and x_1 on the right-hand side) by estimating an equation such as Equation 1 but also including covariates such as age; this would correct for any small imbalances between the experimental and control group that remained after randomization. For example, if the variable x_1 was coded as 0 for those who did not receive the treatment and 1 for those who did and the variable x_2 was coded as age in years, the OLS estimate of β_1 would equal the difference between the treatment and control groups after controlling for age. But because randomization would insure that the distribution of age in the treatment and control groups would be approximately the same, the estimated β_1 with age in the regression should approximately equal the difference in sample means between the two groups; that is, an OLS regression of clinical trial data including covariates should yield an estimate of the therapy's effect that is similar to the naïve estimator resulting from simply comparing treatment and control group means (or from just including x_1 on the right-hand side).

In observational data the treatment is not allocated randomly. As a result, the characteristics of those obtaining the treatment will generally differ from the characteristics of those who do not (the controls). These differences may be in observable characteristics such as age, in which case a regression equation such as Equation 1 can potentially control for them.[2]

[1]The variable x_1 would be coded 1 if the subject were in the experimental group, 0 otherwise.

[2]We say potentially because there must be some overlap in characteristics between the treatment and control groups to avoid confounding; for example, if the experimental group consisted entirely

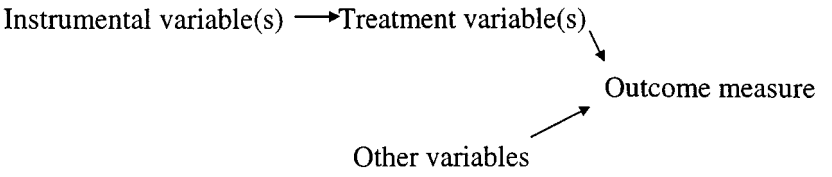

Figure 1 Schematic of Instrumental Variable Estimation.

More threateningly, the differences may be in *unobservable* (to the analyst) characteristics that affect who obtained the treatment. For example, patients who are more severely ill in ways known to their physicians but not to the analyst might not get the treatment, or vice versa. If so, the effect of the treatment on the outcome is confounded with the severity of illness. This effect is well understood by clinicians and epidemiologists and is a principal reason why RCTs are regarded as the gold standard in clinical research.

How does the IV approach address the confounding problem? The main idea is to define a variable or variables—the instruments—that have two properties. First, they affect (cause variation in) the variable whose effects we want to know something about, here the treatment variable, and second, they have no direct effect on the outcome measure (y in Equation 1). One can then estimate how much the variation in the treatment variable that is induced by the instrument—and only that induced variation—affects the outcome measure. In econometric jargon this induced variation is called the *exogenous variation* and it is said to *identify* the desired estimate. Figure 1 diagrams the required assumptions.

One can think of the instrumental variable as a device that achieves a pseudo-randomization. Indeed, the actual randomization in an RCT is a special case of IV. Imagine, for example, that one tosses an unbiased coin to assign people to treatment or control groups at random. The outcome of the coin toss, heads or tails, is the IV, a variable that induces variation in the treatment variable.

A key assumption of the IV technique is that in Figure 1 there is no arrow running from the instrumental variable to the outcome except through the treatment variable. That is, the instrument has no independent effect on the outcome, something that is obviously satisfied in the coin toss example because whether the coin comes up heads or tails does not by itself affect outcomes. Rather outcome is affected, if at all, only by the treatment being evaluated. A second key assumption is that variation in the IV causes substantial variation in the treatment variable. This assumption is also satisfied in the coin toss example,

of one race and the control group of another, one could not disentangle the effect of race from that of the treatment. Also, for simplicity, we assume the other variables are independent of the random error term; in econometric jargon, that they are exogenous. This is the case for demographic variables such as age, sex, and race.

because a large number of coin tosses will result in something approximating half the sample in the experimental group and half in the control group (we assumed a fair coin!). If there exists an instrumental variable or variables that satisfy these two assumptions, and one has sufficient sample size, one can obtain reasonably good estimates of the effect of the treatment on the outcome variable.

If the first assumption does not hold, that is, if a variable that is treated as an instrument actually affects the outcome directly or if there are no variables observed by the analyst that do not affect the outcome directly, then the results from IV estimation will be biased, and the effect of the treatment is said to be unidentified or underidentified. It is as if the randomization in an RCT failed and assignment to the treatment group was related to a factor affecting outcomes.

If the second assumption does not hold, that is, if the variation in the IV does not induce much variation in the treatment variable, then the random error term will tend to mask the effect of the treatment variable, and the IV technique will tend to produce results similar to OLS (13). In short, unless these two assumptions are satisfied, the IV technique is not helpful.

An Example of Instrumental Variable Estimation: The Decision to Catheterize in the Treatment of Acute Myocardial Infarction

The treatment of acute myocardial infarction (AMI) consists of many decisions, but an important set revolves around catheterization and revascularization. Catheterization is a diagnostic technique that provides images of how blood flow to the heart may be compromised. Revascularization comprises one of two techniques, coronary artery bypass grafting (CABG) or percutaneous transluminal coronary angioplasty (PTCA). CABG circumvents an occlusion of the coronary arteries by splicing around it (bypassing it) using a piece of an artery or vein taken from elsewhere in the body. PTCA consists of threading into the occluded artery a balloon-like material that is then expanded in order to improve blood flow through the artery. Revascularization is never done without prior catheterization, because one must determine the sites of occlusions in order to revascularize.

We now illustrate the use of IV in outcomes research using catheterization and associated revascularization in the treatment of AMI as an example. More details of this example are References 7–9.

Our sample consists of almost all elderly (those over the age of 65) who suffered an AMI in 1987. The few exclusions are described in the cited references; they relate to data availability and should not compromise generalizability. The data on whether a patient received catheterization and revascularization come

from Medicare claims data; because the coding of these procedures affects payment for both the hospital and the physician, their coding in claims data is thought to be accurate. There are criminal penalties for fraud in overreporting, and substantial revenues would be foregone if the procedures are underreported. Data on mortality come from the Social Security Administration, which obtains the data in order to terminate pension payments and make any death benefit payments. Again, because substantial sums of money turn on these determinations, coding is thought to be accurate.

OLS METHODS The problem we address is to evaluate the effect of catheterization and any associated revascularization on mortality. Table 1 shows simple descriptive results for two groups of elderly with an AMI, those who did and did not receive catheterization. Looking first at the bottom row, we see that just under 23% of the elderly with an AMI received catheterization in 1987; of

Table 1 Characteristics of sample, by catheterization status*

	No catheterization	Catheterization in 90 days
Female	53.5	39.7
White	90.4	91.8
Age in years	77.4	71.6
Urban	69.6	73.8
Cancer	2.2	0.85
Pulmonary disease, uncomplicated	11.1	9.3
Diabetes	18.3	17.1
Cerebrovascular disease	5.4	2.8
Admit to catheterization hospital[a]	40.9	62.9
Admit to revascularization hospital[a]	21.6	41.6
Admit to high-volume hospital[a]	50.0	58.0
One-day mortality	10.3	0.9
7-day mortality	22.0	3.3
30-day mortality	26.6	7.4
1-year mortality	47.1	16.6
2-year mortality	55.3	21.3
4-year mortality	66.7	29.9
Number of observations	158,261	46,760

*%, except for age and number of observations.

[a]Catheterization hospital is a hospital with five or more catheterization procedures on patients in our sample that is not a revascularization hospital; revascularization hospital is a hospital with ten or more revascularizations on patients in our sample; high-volume hospital is a hospital treating 75 or more AMI admissions in our sample.

this group, roughly half went on to receive a revascularization procedure (30% received CABG and 21% received PTCA within 90 days of their AMI, data not shown).[3]

Looking next at the mortality data near the bottom of the table, it appears as if catheterization and the associated revascularization is a "slam-bang" treatment. By four years following the AMI two thirds of the group that did not receive catheterization are dead, compared with only 30% of the catheterized group. If these two groups had been formed by a randomized trial, one would estimate that catheterization and any associated revascularization had a 37-percentage point effect on four-year mortality ($37 = 66.7 - 29.9$).

But the groups do not come from a randomized trial, and there are many indications in Table 1 that the groups are dissimilar in ways that may affect mortality independent of the treatment. First, the group not receiving catheterization is markedly older, and it has a higher prevalence of each of the four comorbidities shown in Table 1. Second, there is a large difference in mortality at day one, 10% versus 1%, and almost no revascularization was done on day one in 1987. Because catheterization is a diagnostic procedure that by itself does not reduce mortality (indeed, there is a very small mortality risk associated with the procedure), this difference in mortality at day one must stem principally from underlying differences in health status between the two groups. Further, this mortality difference at day one tends to increase as time passes.

Finally, Table 1 shows that the group being catheterized was admitted to a different set of hospitals. Specifically, they were more likely to be admitted to catheterization, revascularization, and high-volume hospitals. As we use the terms, catheterization hospitals were hospitals that did five or more catheterizations in 1987 on patients in our sample, but did not do ten or more revascularizations; hospitals that did ten or more revascularizations we term revascularization hospitals. These two groups of hospitals are thus mutually exclusive. Irrespective of whether the hospital is a catheterization or revascularization hospital, we classify high-volume hospitals as those that treated 75 or more AMI patients in our sample; the value of 75 was chosen as one that approximately divided our sample into two equal groups.

Not surprisingly, both catheterization and revascularization hospitals are more likely than the remaining hospitals to have other sophisticated treatments available to their patients, and their physicians and nurses may be more highly trained (data not shown). The other capabilities available at these two types of hospitals may affect outcome independent of the catheterization and revascularization procedures.

[3]Of the persons in the non-catheterization group, 1.5% had a revascularization within 90 days. These individuals most likely had an outpatient catheterization, which our data did not record. Reclassifying these individuals in the catheterization group does not appreciably change our results.

Moreover, if practice makes perfect, as is generally the case in medicine, high-volume hospitals will obtain better results than other hospitals. Because catheterization and revascularization hospitals are larger, they are also more likely to be high-volume hospitals. In short, some of the differences in observed mortality between the two groups could result from differences in the capabilities of the hospitals and physicians treating the two groups of patients, independently of the procedures given them.

A possible way to disentangle these effects is to attempt to control for the observed differences between the groups by employing OLS and including the observable characteristics of both the patients (e.g. age) and the hospitals to which they were admitted (e.g. high-volume) as explanatory variables, as well as including a variable indicating catheterization. That is, one would estimate an equation such as Equation 1, where x_1 would take the value one if the patient were catheterized and zero otherwise, and x_2, x_3, and so forth, would be variables such as age of patient and characteristics of the hospital.

We have estimated such an equation (not shown). Controlling for several patient characteristics leaves catheterization and associated revascularization looking like a highly successful treatment. In particular, after controlling for age, sex, race, rural residence, comorbidities, and state of residence, there is still a 28-percentage point difference in mortality at four years. In other words, controlling for these observable characteristics reduces the uncontrolled 37-percentage point difference between the two groups to 28 percentage points, but that large difference leaves aggressive treatment of AMIs among the elderly looking very promising.

The problem with this approach to the confounding problem is there may still be differences between the two groups that are not controlled. For example, after controlling for these observed covariates there is still a 7-percentage point difference in mortality at one day that increases to 13 percentage points at 7 days, and 18 percentage points at 30 days. This increasing spread is highly suggestive of remaining unobserved differences between the two groups, so that the 28-percentage point difference at four years should not be treated as the size of the true catheterization effect. OLS does not appear to be sufficient to resolve the confounding problems.

More detailed clinical data on the two groups would further reduce the estimated effect, but even such data are not likely to suffice for the purpose of estimating the effect of aggressive treatment. In explaining who obtains such treatment, clinical data from a detailed review of medical records typically can only explain about a quarter of the variance, suggesting that even the availability of more detailed clinical data would leave a substantial confounding or selection problem. Such a conclusion, of course, would not surprise those who believe RCTs are the only valid approach to evaluating medical treatments;

they would never have attempted to use observational data with covariates in the first place for precisely this reason.

INSTRUMENTAL VARIABLES In order to proceed with IV estimation, we must find a variable or variables that have no direct effect on AMI mortality but that do affect the likelihood of catheterization. The instrumental variables that we will use in this application are the differential distances to catheterization, revascularization, and high-volume hospitals. We define differential distance as the additional distance, if any, beyond the distance to the nearest hospital to reach a hospital of the given characteristic (e.g. catheterization hospital). To calculate differential distance we begin by calculating the distance in miles between the person's place of residence and the nearest hospital (5-digit zip code centroid to 5-digit zip code centroid). If the nearest hospital is in the zip code of residence, this distance is of course zero. The differential distance to a catheterization hospital is then the additional distance, if any, from the nearest hospital to a catheterization hospital. Thus, if the nearest hospital is a catheterization hospital, the differential distance to a catheterization hospital is zero, irrespective of whether the nearest hospital is in the same zip code. Note that differential distance is defined simply with respect to the location of facilities relative to the location of residence; it does not take into account the hospital to which the patient was actually admitted.

Because patients with an AMI tend to go to the nearest hospital, differential distance is highly predictive of whether the patient was admitted to a catheterization hospital. The greater the differential distance, the less likely it is that a patient will be admitted to a catheterization hospital. Admission for these purposes is defined as the hospital of initial admission; if the patient was discharged from a non-catheterization hospital and subsequently readmitted to a catheterization hospital or if the patient was transferred to a catheterization hospital, we treat that patient as having been admitted to a non-catheterization hospital.

Table 2 shows a different division of the sample than Table 1, namely according to the differential distance to a catheterization hospital. The division at 2.5 miles is chosen to divide the sample into two approximately equal-size groups. Thus, the group facing a differential distance of 2.5 miles or less either had a catheterization hospital as their nearest hospital or had to travel less than 2.5 miles further than the distance to their nearest hospital to reach a catheterization hospital.

The data show that the second assumption required for IV estimation, that variation in the IV causes variation in the treatment variable, is satisfied. The half of the sample that is closer to a catheterization hospital has a 6.7-percentage point greater chance ($6.7 = 26.2 - 19.5$) of receiving a catheterization within

Table 2 Characteristics of sample, by differential distance*

	Differential distance $\leq$ 2.5 miles	Differential distance $>$ 2.5 miles
Female	51.3	49.5
White	89.0	92.3
Age in years	76.1	76.1
Cancer	1.9	1.9
Pulmonary disease, uncomplicated	10.4	11.0
Diabetes	18.1	18.0
Cerebrovascular disease	4.8	4.8
Admit to catheterization hospital	45.4	5.0
Admit to revascularization hospital	41.7	10.7
Admit to high-volume hospital	67.1	36.5
90-day catheterization	26.2	19.5
1-day mortality	7.50	8.88
7-day mortality	16.80	18.59
30-day mortality	24.86	26.35
1-year mortality	39.79	40.54
2-year mortality	47.20	47.89
4-year mortality	58.06	58.52
Number of observations	102,516	102,505

*%, except for age and number of observations.

90 days following the AMI (and about a 3.5-percentage point greater chance of receiving a subsequent revascularization) than the half of the sample who live closer to hospitals without catheterization capabilities.

The first assumption, that differential distance affects outcomes only through its effect on the likelihood of receiving the treatment, must be partly taken on faith. One can, however, partly test this assumption and can also ask how reasonable it is a priori. The assumption would be satisfied if a person's place of residence, or more precisely the person's differential distance, was not associated with the clinical severity of the heart attack, the primary unobserved variable that will determine treatment. This assumption appears reasonable to us. It can be buttressed in several ways with data, however.

If differential distance is independent of the severity of the heart attack, something that is unobserved by us the analysts, then it should also be independent of observed variables such as age and comorbidities that are associated with health status and hence the likelihood of receiving a catheterization. The data in Table 2 suggest that this independence is approximately the case; the mean age is the same between the two groups, and the prevalence of comorbidities is

very nearly the same. Certainly the prevalence of comorbidities is much more similar than between the two groups in Table 1.

Assuming no direct effect of differential distance on mortality but only an indirect effect through how it affects treatment, a simple IV estimator of the effect of catheteriztation in this case is simply the difference in four-year mortality between the two groups shown in Table 2 divided by the difference in the catheterization rates:

$$\begin{aligned}\text{Effect of catheterization} &= \Delta\text{mortality}/\Delta\text{catheterization rate} \\ &= (58.06 - 58.52)/(26.2 - 19.5) \\ &= -0.46/6.7 = -0.069. \end{aligned} \qquad 2.$$

From this calculation one can infer that the additional 6.7% of patients who were catheterized in the close-by group had, on average, a 6.9-percentage point additional chance of surviving to four years (the negative sign means that mortality was lower in the close-by group).

The 6.9-percentage point effect of catheterization on survival is clearly much less than the 28-percentage point effect estimated using OLS with observed covariates as controls, but it is still a substantial clinical effect. If the estimate were valid, catheterization would probably be generally indicated for these patients. In fact, however, there are two reasons to believe the 6.9-percentage point figure is overstated.

First, the figure is analogous to one that would come from a simple regression of mortality on the catheterization rate, one that did not consider either patient or hospital characteristics. Even though patient characteristics are not controlled for in this calculation, they appear approximately balanced between the two groups. As a result, they are unlikely to be causing the 6.9-percentage point difference. But the data in Table 2 show that the close-by group is almost twice as likely to be admitted to a high-volume hospital. In other words, even though patient characteristics are approximately balanced between the two groups, hospital characteristics are not, and the differences will tend to inflate the estimate.

By using other instrumental variables, however, one can simultaneously control for admission to a high-volume hospital while estimating the effect of the procedure.[4] Doing so, as well as controlling for demographic variables pertaining to the patient including age, sex, race, urban or rural residence, yields the estimates in Table 3.[5] In other words, the estimates of the catheterization effect

[4]Technically, admission to a high-volume hospital is endogenous but it is identified through the use of additional instruments, namely differential distance to other types of hospitals.

[5]These estimates improve on the simple estimate of −6.9% given above by also breaking up the differential distance measures into several discrete groups instead of the two-way classification (over or under 2.5 miles) used in Equation 2. (For details, see Reference 9.)

Table 3 Instrumental variable estimates of effect of catheterization, admission to a high-volume hospital, and rural residence on mortality*

	Received catheterization	Admit high-volume	Rural residence
1-day mortality	−5.0	−0.88	0.57
	(1.1)	(0.24)	(0.19)
7-day mortality	−8.0	−1.23	0.49
	(1.8)	(0.33)	(0.26)
30-day mortality	−6.8	−1.45	0.50
	(2.6)	(0.38)	(0.30)
1-year mortality	−4.8	−1.07	−0.15
	(3.2)	(0.42)	(0.33)
2-year mortality	−5.4	−0.88	−0.02
	(3.3)	(0.43)	(0.33)
4-year mortality	−5.1	−0.75	0.14
	(3.2)	(0.42)	(0.32)

*In percentage points, standard errors in parentheses.

in Table 3 come from a multiple regression with all of the foregoing variables entered as covariates.

The results in Table 3 show that controlling for admission to a high-volume hospital, as well as controlling for rural residence, reduces the estimated effect of catheterization at four years from around the 7 percentage points shown in Equation 2 to around 5 percentage points. They also show that admission to a high-volume hospital results in about a 1-percentage point reduction in mortality, so the volume-outcome relationship that appears elsewhere in medicine is present here as well.

Most interestingly, however, the results in Table 3 show that the 5-percentage point reduction from catheterization appears at day one, before the procedure itself could be having any appreciable life-saving effect.[6] Our interpretation of the results in Table 3, therefore, is that something that is associated with the procedure, but not the procedure itself, is having a beneficial effect on mortality.

Other work suggests to us that the beneficial effect is associated with the catheterization unit and the treatment of AMI at the hospital and is not attributable to generally better hospitals having catheterization and revascularization units (8). Our guess is that hospitals with catheterization units are more likely to have cardiologists and cardiac nurses treating AMI patients, and that use of medications by these personnel may be better than is the case with less specialized personnel.

[6]Only 15% of all the revascularization procedures were performed on the first day; even if these all saved lives (an extreme assumption), this would not be sufficient to explain the first-day difference in mortality rates.

We show elsewhere that the gain in survival comes at a cost of around $45,000 to $100,000 per person surviving at least one year (8). Because we do not know how much longer those who survived four years lived, we cannot compute a more usual cost-per-year-of-life-saved figure, but it would be lower than the $45,000 to $100,000 range just given.

Finally, the results in Table 3 show that rural residence increases mortality through 30 days, but that this effect has vanished at one year. Our guess is that this short-term gain is attributable to better emergency medical systems in urban areas.

LIMITATIONS OF THE IV ESTIMATES We began this chapter by asking whether IV estimates could be used with observational data to substitute for estimates of a clinical trial. Even in cases in which the two key IV assumptions are satisfied, as is probably the case in our example, the answer to this question will generally be no, though that does not mean the IV estimates are not useful.

First, in the classic clinical trial everyone in a defined group is randomized to either receive or not receive the treatment. As noted above, if the trial is well designed and executed, one can derive a good estimate of the effect of the treatment by subtracting the means of the outcome variable(s) in the experimental group and the control group. By definition, however, the mean is the *average* effect; thus, one derives an estimate from the clinical trial of the average effect of the treatment in the entire population eligible for the trial or in a subgroup if subgroup analysis is conducted.

The simple estimate of 6.9 percentage points that we derived in Equation 2 of the catheterization effect was not an average effect in the entire population because it only applied to a subpopulation, specifically it was the (average) effect in the additional 6.7% of the elderly population with an AMI who received a catheterization because they lived relatively close by a catheterization hospital but who would not have received it had they lived further away.[7] The observational data we used can tell us nothing about the effect of catheterization in the other 93% of the population. This 93% falls in the category of either "would always have received a catheterization no matter how distant" or "would never have received a catheterization no matter how close" so there is no useful variation in the observational data with which to estimate an effect. In short, the −6.9% estimate in Equation 2 is an estimate of a marginal effect, namely the effect (on average) of increasing the catheterization rate in a population from 19.5% to 26.2%. It is not an estimate of the average effect on a random patient from the entire population, only the average effect in these additional 6.7% of the population.

The procedure could have a very substantial beneficial effect on others in the population, presumably those who would have received it irrespective of

[7]Technically, it is the average marginal effect.

their location. That beneficial effect would be reflected in estimates from a clinical trial, but would not be reflected in these estimates. Moreover, unlike the population in a clinical trial, this marginal population whose treatment changes because of location is not straightforwardly identifiable by the clinician; that is, whether the patient is part of that population is not immediately obvious to the physician. As a result, these estimates are only indirectly applicable to clinical practice; they suggest that in close calls with elderly patients the clinician not use catheterization. In other words, the results in Table 3 cannot be used to say that for the average elderly person catheterization and the associated revascularization result in about a 5-percentage point reduction in mortality (even ignoring the interpretation of the one-day results above). Rather, taking the results at face value there is a 5-percentage point effect just in the population whose treatment changes by dint of their location. Moreover, taking account of the one-day results suggests that the true beneficial effect in this subgroup is due to some other factor.

But there is another difference between a clinical trial and observational data. In a well-executed clinical trial, the focus is on one treatment or a very small number of treatments. All other factors will be held constant insofar as is possible. The results thus pinpoint insofar as possible the causal factor.

In observational data it is generally impossible to hold other factors constant. In the AMI case, for example, the only hospital characteristic that has been held constant is the volume of AMI patients. But as noted above, it is likely that hospitals performing catheterization have better-trained physicians and nurses for treating AMI, something that is not controlled for in the analysis. Certainly, some factor is causing better results among the patients undergoing catheterization, as is evidenced by the 5-percentage point effect at day one that maintains itself at least for four years. If this factor is not catheterization and the associated revascularization, it must be something not in the regression that is associated with these factors. Whatever this something is would be missed by a clinical trial that focused solely on the effects of the procedure. Thus, the narrow lens of a clinical trial of aggressive treatment would not have found this effect; the wide-angle lens of the observational study picks it up but cannot precisely identify the causal factor.

The observational study also has the advantage of considering a more general population than is typically included in clinical trials. The difference in population can be substantial. We observed a 40.1% mortality rate at one year across the entire elderly population that suffered an AMI (Table 1).[8] This value is well above the mortality rates observed in control groups of clinical trials (15). For example, one meta-analysis of clinical trials yielded mortality rates in the 10% to 15% range (16). Part of this large difference is undoubtedly

[8] 40.1% is a weighted average of the 47.1% and 16.6% rates shown in Table 1.

attributable to some poorer risks being excluded from the population eligible for clinical trials, for example, those with serious comorbidities, but being included in our sample. As a result, the effects found in a clinical trial may be too optimistic for a more general population. In short, the estimates from the trial may have greater internal validity, but those from the observational study may have greater external validity.

Conclusion

What can be done to evaluate medical treatments and procedures when a clinical trial is impossible to conduct but observational data are available? The best case is where one has a variable or variables, such as differential distance in the AMI example, that predict treatment but are largely independent of any direct effect on outcomes. In that case, one can use the IV method to estimate an effect in the population whose treatment changes as a result of variation in the IV. In our example, these were the people who were catheterized because they lived relatively close to a catheterization hospital but who would not have been catheterized if they did not live relatively close.

To a clinician this information may not be immediately applicable, because the physician cannot straightforwardly identify who is a member of this group. It may, however, be helpful in suggesting that close calls be resolved in favor of less aggressive treatment. More precisely, the clinician wants to know the expected outcome of the procedure if the procedure is performed on the patient standing or lying in front of the physician. This is simply the average effect, or perhaps the average effect conditional on certain observable variables, such as age. In other words, if the patient is a random draw from a given population, the expected outcome in that patient is the result given by the clinical trial—provided the patient comes from the population eligible for the trial and is otherwise being treated like patients in the trial.

But if a clinical trial is not available or cannot be done, as is often the case, are the results from a study using IV with observational data useful? To show that they are, we illustrate how the results from the AMI study might be used. Indeed, we would make an even stronger statement: The results of a well-designed observational study are useful even if the results of a clinical trial are available.

First, our results using the particular IV of differential distance suggest that regionalization of catheterization facilities might save money with little adverse mortality effect because regionalization would causally increase distance. It would also have the beneficial effect of raising the fraction of procedures done in high-volume facilities. Both the internal and external validity of our study for this conclusion is probably very good. A clinical trial would have little or nothing to say about this issue.

More generally, policy issues often involve incremental decisions as opposed to all-or-nothing decisions. Thus, the results of our study may be very relevant to decisions about measures that would have the effect of somewhat increasing or decreasing the numbers of catheterization laboratories, for example, by changing reimbursement or regulation, although the internal validity of our results for that issue is more questionable than for the regionalization issue.[9] Just as with regionalization, a clinical trial is likely to have little to say about the effect of changing reimbursement or tightening regulation.

Second, at the level of the patient, as already noted, our results suggest that the marginal gain from performing the procedures is small and thus in close cases the physician might be less aggressive in treating AMI among the elderly. Moreover, it may be possible to do better with observational data. One could, in principle, collect clinical data that might predict who was in the marginal group, that is, the group whose treatment changed because of differential distance. If so, the data from studies such as this one would become even more useful to the clinician.

Furthermore, even if the results from a clinical trial are available, the patient actually in front of the physician may differ in many ways from the population included in the clinical trial; she may, for example, have a serious comorbidity. In that case the results from the trial may not be very useful to the clinician; external validity is compromised, whereas the results from the observational study may have much better external validity.

Finally, in a different example it may happen that the range of variation spanned by the IV may be greater and hence better approximate the average effect in the population. For example, if we divide the AMI sample into two equal-size groups by differential distance to a high-volume facility rather than differential distance to a hospital with a catheterization unit, the percent admitted to a high-volume facility in the close-in group is 82% and in the further away group is 21%. This 21% to 82% range is obviously much closer to the 0% to 100% range that a clinical trial (with no attrition) would span than is the 19.5 to 26.2 range of catheterization rates induced by differential distance to a catheterization hospital (Table 3). In short, as the range spanned in the observational data increases toward 100%, the results from the observational study will better approximate the average effect for the entire population that would be estimated from a clinical trial. In that case the observational study may be a good substitute for a well-executed trial. Indeed, because of its likely greater external validity, the results from the observational data might even be preferred, provided of course that a good set of instruments is available.

[9]The key issue is whether the group whose treatment would change if, say, reimbursement were to change, is similar to the group whose treatment changed in our study because differential distance changed.

Literature Cited

1. Angrist J, Imbens G, Rubin D. 1996. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* 91:444–55
2. Bowden RJ, Turkington DA. 1984. *Instrumental Variables*. Cambridge: Cambridge Univ. Press
3. Davidson R, MacKinnon JG. 1993. *Estimation and Inference in Econometrics*. New York: Oxford Univ. Press
4. Greene WH. 1990. *Econometric Analysis*. New York: Macmillan
5. Judge GG, Griffiths WE, Hill WC, Lee TC. 1980. *The Theory and Practice of Econometrics*. New York: Wiley
6. Kmenta J. 1986. *Elements of Econometrics*. New York: Macmillan. 2nd ed.
7. McClellan M, McNeil BJ, Newhouse JP. 1994. Does more intensive treatment of acute myocardial infarction reduce mortality? *JAMA* 272:859–866
8. McClellan M, Newhouse JP. 1997. The marginal costs and benefits of medical technology: a panel instrumental-variables approach. *J. Econom.* 77:39–64
9. McClellan M, Newhouse JP. 1997. The marginal benefits of medical technology. Mimeo
10. Office of Technology Assessment. 1994. *Identifying Health Technologies that Work: Searching for Evidence*. Washington, DC: GPO
11. Permutt T, Hebel JR. 1989. Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight. *Biometrics* 45:619–22
12. Roper WL, Winkenwerder W, Hackbarth GM, Krakauer H. 1988. Effectiveness in health care—an initiative to evaluate and improve medical practice. *N. Engl. J. Med.* 319(18):1197–202
13. Staiger D, Stock J. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65:557–86
14. Stewart A, Ware JE Jr, eds. 1992. *Measuring Functioning and Well-Being*. Durham: Duke Univ. Press
15. Udvarhelyi IS, Gatsonis C, Epstein AM, Pashos CL, Newhouse JP, McNeil BJ. 1992. Acute myocardial infarction in the Medicare population: process of care and clinical outcomes. *JAMA* 268:2530–36
16. Yusuf S, Sleight P, Held P, McMahon S. 1990. Routine medical management of acute myocardial infarction: lessons from overviews of recent randomized clinical trials. *Circulation* 82(Suppl. II):117–34