Heterogeneity and the Interpretation of Treatment Effect Estimates From Risk Adjustment and Instrumental Variable Methods

John M. Brooks, PhD,* and Elizabeth A. Chrischilles, PhD†

Objectives: To contrast the interpretations of treatment effect estimates using risk adjustment and instrumental variable (IV) estimation methods using observational data when the effects of treatment are heterogeneous across patients. We demonstrate these contrasts by examining the effect of breast conserving surgery plus irradiation (BCSI) relative to mastectomy on early stage breast cancer (ESBC) survival.

Methods: We estimated discrete time survival models for 6185 ESBC patients in the 1989–1994 Iowa Cancer Registry via IV estimation using 2 distinct instruments (distance of the patient's residence from the nearest radiation center, and local area BCSI rate) and controlling for cancer stage, grade, and location; age; comorbidity; hospital access; payer; diagnosis year; and area poverty level. We then estimated comparable risk adjustment survival models using linear probability methods with robust standard errors.

Results: Risk adjustment models yielded average survival estimates similar to trial results. With favorable BCSI selection, these estimates represent an upper bound of the true effect for patients receiving BCSI. IV estimates showed a BCSI survival risk for patients whose surgery choices were affected by the instruments and these estimates varied with the instrument specification.

Conclusions: When treatment benefits are heterogeneous across patients, treatment effect estimates from observational data can still be useful to policymakers, but they must be interpreted correctly. Risk adjustment methods yield estimates that can assess whether the patients who *received* treatment benefited from the treatment, but the direction of bias must be considered. In contrast, IV estimates can assess the effect of treatment rate changes, but characteristics of patients whose choices were affected by the instruments must be considered when making such inferences.

Key Words: heterogeneity, effectiveness, instrumental variables, risk adjustment, treatment rates

(*Med Care* 2007;45: S123–S130)

From the *Program in Pharmaceutical Socioeconomics, College of Pharmacy; and †Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, Iowa.
Reprints: John M. Brooks, PhD, Program in Pharmaceutical Socioeconomics, College of Pharmacy, University of Iowa, S-515 PB, 115 S. Grand Ave., Iowa City, IA 52242. E-mail: john-brooks@uiowa.edu.

With the advent of Medicare Part D, medical care treatments will be increasingly used by patients differing from those patients in the clinical studies that demonstrated the treatment efficacy. Patients underrepresented in randomized controlled trials (RCTs) include the elderly, minorities, and patients with different comorbidities.[1–4] Given the lack of efficacy data for many patients, commentators have speculated about the extent that treatment effects vary or are *heterogeneous* across patients and the implications that heterogeneity has for treatment guideline development and evidence-based medicine.[4–9] Sources of treatment effect heterogeneity can be genetic, demographic, the severity of the underlying condition, the existence of a comorbid condition, the use of other treatments, and patient frame of mind. If treatment benefits are heterogeneous across patients, the relevant question for policymakers is often not whether a treatment should be used at all, but whether a treatment is over- or underused in practice. Wennberg correctly posed this question as "Which Rate is Right?"[10] To address this question health services research must find ways to assess the distribution of treatment effectiveness across the patient population.

If treatment effects are heterogeneous, it is impractical and probably impossible to generate sufficient RCT evidence for all patients.[11] As a result, the treatment variation in observational databases may be the only source to estimate treatment effectiveness for clinically distinct patient groups. It is well known, however, that unmeasured confounding variables can lead to incorrect casual inferences with observational data.[12–15] Risk adjustment and instrumental variable (IV) analysis methods have the potential to alleviate confounding problems.[16–19] However, if treatment effects are heterogeneous, the elimination of confounding is not the only inferential problem to be considered when using these methods. In general, estimation approaches can only identify relationships for the subset of patients generating the treatment variation,[20] and risk adjustment and IV approaches use different subsets of patients in estimation. As a result, when treatment effects are heterogeneous, these approaches yield estimates for distinct patient groups, and researchers need to understand how these estimates relate to specific policy questions. Heckman and colleagues[21,22] show that risk adjustment approaches yield average treatment effect estimates for *the subset of patients that were treated.* Angrist and colleagues[23,24] show that IV methods yield estimates of the average treatment effect *for those patients whose treatment*

choices were affected by an instrumental variable or "instrument". To address questions of treatment effectiveness and whether existing treatment rates are optimal, researchers can gain insight by using both risk adjustment and IV estimation approaches and placing the estimates from both approaches in correct context. Both confounding risks and the extent that the estimates can be generalized across the patient population must be considered.

In this article, we discuss the concepts of Heckman and Angrist using the effect of surgery choice [breast conserving surgery plus irradiation (BCSI) vs. mastectomy] on survival for patients with early stage breast cancer (ESBC) as an example. We use data from an earlier IV article that focused on stage II ESBC patients in Iowa, that suggested that higher BCSI rates in these patients would have resulted in survival loss.[25,26] In this example, we broaden our sample to include both stage I and II ESBC patients. To contrast the ideas of Heckman and Angrist, we modify a theoretical framework developed by Winship and Morgan[27] to yield models of ESBC survival and surgery choice and derive the expected value of both the risk adjustment and IV estimates. We then estimate risk adjustment and IV survival models and contrast our findings in terms of these theoretical findings.

METHODS

Background and Theoretical Framework

Patients with ESBC have a choice of mastectomy or BCSI for local tumor control. ESBC patients with localized tumors less than 2 cm and no lymph node involvement are classified as stage I. If patients have either a localized tumor less than 2 cm with positive lymph node metastasis on the same side, or a tumor between 2 and 5 cm with no lymph node involvement, they are classified as stage IIa. Stage IIb patients have either a localized tumor between 2 and 5 cm with positive lymph node metastasis on the same side, or a tumor greater than 5 cm with no lymph node involvement. Several RCTs demonstrated the survival equivalence of BCSI and mastectomy for the average ESBC patient in these trials.[28–32] Based on these results, the National Institutes of Health (NIH) issued a guideline recommending BCSI over mastectomy for most ESBC patients and stated that patients should be educated and make a surgery choice based on their preferences.[33] However, the validity of generalizing the RCT results to ESBC patients across disease stages is unclear. Two of the RCTs included only stage I patients and showed no survival benefit of mastectomy over BCSI.[28,32] The remaining studies contained stage I and II patients but each estimated a single treatment effect.[29–31] No study contained only stage II patients, and in the study with the most stage II patients, tumor size and nodal involvement increased the risk of local recurrence for BCSI patients but not for patients receiving mastectomy.[34] After release of the NIH guideline, BCSI rates for ESBC patients increased, but not as much as expected, and BCSI rates varied regionally and were affected by nonclinical factors.[35–38] Several commentators attributed the slow and varied rate of BCSI diffusion in the United States to lack of provider knowledge of the evidence, and educational interventions were suggested to increase BCSI rates.[39–41] An alternative explanation for the slow diffusion by BCSI may be that many providers believed the relative benefits of BCSI and mastectomy are heterogeneous across ESBC patients and that RCT evidence cannot be generalized to many ESBC patients with severe disease. The results of the earlier IV article reinforced this notion of heterogeneity by showing higher BCSI survival risk relative to mastectomy for stage II ESBC patients.[26]

In this article, we adapted the equation-based framework used by Winship and Morgan[27] for this scenario to illustrate the parameter interpretations of Heckman and Angrist.[21,23] The survival risk of BCSI relative to mastectomy is modeled as heterogeneous across the ESBC population. Both procedures are assumed to have equal survival benefit at lower severity levels, but the survival benefit associated with BCSI relative to mastectomy decreases as severity increases. In addition, patients with more severe disease are assumed to have lower survival odds regardless of treatment. Given these circumstances, the survival equation is written:

$$Y = b_0 + (b_1 L) \cdot S + b_2 L + e, \tag{1}$$

where $Y = 1$ if patient survived a given time period after diagnosis, 0 otherwise; $S = 1$ if the patient received BCSI, 0 if mastectomy; L is a measure of disease severity that increases with severity level; e is the error term; $(b_1 L)$ represents the effect of BCSI relative to mastectomy on survival that depends on disease severity; and b_2 represents the direct effect of L on Y. To fit our heterogeneity assumption, we envision b_1 as an infinitesimally small negative number, so that when disease severity is low the survival difference between BCSI and mastectomy is negligible, and as L increases BCSI poses a survival risk relative to mastectomy. In addition, we expect $b_2 < 0$, the probability of surviving decreases with severity regardless of treatment. If patients in concert with their providers believe that the survival risk of BCSI relative to mastectomy increases with disease severity, this leads to the following surgery choice model as a function of disease severity:

$$S = c_0 + c_1 L + c_2 W + v, \tag{2}$$

where S and L are defined as above; W represents factors other than disease severity that affected surgery choice; v is the error term; and c_1 is the effect of severity on surgery choice. One would expect the signs associated with b_1 from Eq. (1) and c_1 to be the same. If the survival risk of BCSI is thought to increase with L ($b_1 < 0$), the patients with higher disease severity will be less likely to choose BCSI ($c_1 < 0$).

Given this framework, suppose a researcher has data on treatment choice and survival for a sample of breast cancer patients but no information on disease severity and estimates the following model:

$$Y = a_0 + a_1 S + z \tag{3}$$

where z contains the previous error term and the variation in surgery effectiveness associated with L. If ESBC patients choose treatments based on Eq. (2), standard estimation of

Eq. (3) yields an estimate of a_1 with the following expected value:

$$E[\tilde{\mathbf{a}}_1] = b_1 E[L|S = 1] + c_1 b_2. \qquad (4)$$

This estimate is the average effect of BCSI relative to mastectomy on survival for the patients who received BCSI, but will be biased high as $c_1 < 0$ and $b_2 < 0$. This result follows Heckman's insight that, when treatment effects are heterogeneous, the estimated treatment effect will reflect the characteristics of the patients who received the treatment. The term $E[L|S = 1]$ is the expected severity level for the patients who received BCSI ($S = 1$), and therefore $b_1 E[L|S = 1]$ equals the average treatment effect of BCSI relative to mastectomy for the patients who received BCSI. Now, because patients in our framework choose surgery based on treatment effectiveness, our estimate for the patients who chose BCSI will be biased high ($c_1 b_2 > 0$) because of favorable treatment selection. ESBC patients receiving BCSI have lower unmeasured disease severity than the ESBC patients receiving mastectomy, leading to higher survival probabilities regardless of treatment. Therefore, the estimate of $\tilde{\mathbf{a}}_1$ in Eq. (4) should be interpreted as an upper-bound estimate of the average survival effect of BCSI relative to mastectomy for the ESBC patients who received BCSI.

In contrast, the IV approach estimates a_1 by exploiting the surgery variation from measured factors within W in Eq. (2) that are assumed to be uncorrelated with unmeasured confounders such as disease severity and affect survival only through their effects on surgery choice. If Z (an instrumental variable or "instrument") represents a measured factor within W that is assumed to have this characteristic and X the factors within W that do not, Eqs. (1) and (2) can be rewritten in terms of measured variables:

$$S = c_0 + c_2 X + c_2 Z + t, \qquad (5)$$

$$Y = a_0 + a_1 S + a_2 X + r, \qquad (6)$$

where t and r contain the original error terms plus terms related to severity. A 2-stage approach is used to estimate IV models. In the first stage, Eq. (5) is estimated and a Chow F-test[42] can be used to assess whether the instrument (Z) describes a significant portion of the variation in choice of surgery. In the second stage, Eq. (6) is estimated using the predicted BCSI propensity for each patient from Eq. (5)—$\hat{\mathbf{S}}$. Using this process, only the variation in S that stems from changes in Z is used to estimate a_1. As Z is assumed to be unrelated to L, it essentially provides a natural experiment in S,[43] and the IV estimate of a_1—$\hat{\mathbf{a}}_{IIV}$—is a consistent estimate of the survival effects of BCSI relative to mastectomy. However, if the survival effect of BCSI relative to mastectomy is heterogeneous across patients, following Angrist and colleagues, the resulting estimate is a local average treatment effect that can be strictly generalized only to the patients whose surgery choices were affected by the instrument[23,24]:

$$E[\hat{\mathbf{a}}_{IIV}] = b_1 E[L|S(Z)], \qquad (7)$$

where $E[L|S(Z)]$ is the expected severity level of the subset of patients whose treatment choices were affected Z. Note that

the expected severity level in Eq. (7) differs from the expected severity level in Eq. (4), and so even without the confounding bias in Eq. (4), the estimates yielded by both approaches would differ because information from a different set of patients was used in their respective estimation.

The result in Eq. (7) also means that IV estimates of a_1 may vary with the instrument or set of instruments specified in the model as individual instruments may affect the treatment choices of different patient subsets. In our analysis, we demonstrate the effect of instrument choice on IV estimates using 2 distinct instruments. Our first instrument was developed using the notions of regional treatment "signatures" or "philosophies."[44–46] We theorized that regional differences in BCSI rates may stem from region-specific provider extrapolations of the RCT evidence to the ESBC patients unlike those in the trials. Figure 1 illustrates this idea. Suppose that the population of ESBC patients is distributed across the x-axis based on disease severity and disease severity increases (eg, larger vs. smaller tumor size) as we move to the right on the axis, and that that the BCSI treatment rate in this population is U. The y-axis is the expected survival risk of BCSI relative to mastectomy associated with a unit increase in the BCSI rate. Further assume that patients live in either of 2 geographic areas and that the distribution of disease severity across patients is the same in both geographic areas. Providers in both areas are assumed to have consistent beliefs on the relative effectiveness of BCSI and mastectomy for patients like those in the RCTs, but they differ across areas in how they extrapolate the RCT evidence to patients with more severe disease. The solid curve represents provider beliefs in an area with a pessimistic extrapolation of the survival effects of BCSI relative to mastectomy. These providers believe that BCSI and mastectomy have equal survival benefit for the first U_P percent of patients, but for patients beyond U_P, they believe BCSI has survival risk relative to mastectomy. The dotted curve represents the average provider beliefs in an area with an optimistic extrapolation of the survival effects of BCSI relative to mastectomy. In this area, providers believe
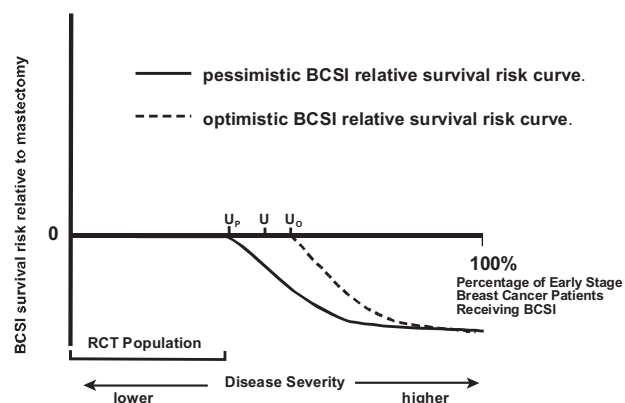

FIGURE 1. Hypothetical relationship between breast conserving surgery plus irradiation (BCSI) survival beliefs relative to mastectomy and BCSI treatment rates. U_P = BSCI rate in an area pessimistic about BCSI; U_o = BCSI rate in an area optimistic about BCSI; U = overall BCSI rate.

that BCSI and mastectomy have equal survival effects for the first U_O percent of patients, with survival risk for patients beyond U_O who receive BCSI.

If providers recommend BCSI only to those patients they believe have no survival risk relative to mastectomy, the area with optimistic beliefs will have a higher treatment rate (U_O) than the area with pessimistic beliefs (U_P). In this case, ESBC patients with the lowest disease severity and those with the highest disease severity would receive consistent surgery recommendations across areas, whereas patients represented by the severity levels between U_P and U_O in Figure 1 would have received different recommendations from providers leading to different surgery choices. The use of area treatment rates as an instrument will yield the average survival effect of BCSI relative to mastectomy for these patients.

For the second instrument, we used the distance from a patient's residence at the time of diagnosis to the nearest radiation treatment center. A longer distance from the patient to the nearest radiation treatment center increases the cost of BCSI treatment to ESBC patients. We envisioned that higher treatment access costs affects the surgery choices of all ESBC patients regardless of severity including many patients like those in RCTs. As a result, the subset of ESBC patients whose treatment choices were affected by the distance to radiation facilities would have lower average disease severity than the subset of patients whose surgery choices were affected by area treatment rates. If this theory is correct, from Eq. (7), we expect that IV estimates using area treatment rates as an instrument will be larger in absolute value (larger $E[L|S(Z)]$) than IV estimates using distance to the nearest radiation treatment center.

Sample and Variable Definitions

The data used in this study are more fully described elsewhere.[25,26] Our sample includes all patients with a diagnosis of first-primary ESBC listed in the Iowa Cancer Registry from 1989 to 1994. Registry data were merged with Iowa Hospital Association inpatient discharge abstract files, providing a sample of 6185 patients in either stage I (N = 3280) or II (N = 2905). We created binary variables defining surgery choice, survival (alive 1, 2, 3, and 4 years after diagnosis), cancer stage, grade and tumor location, age, payer, comorbidities, patients distance to nearest hospital, and the poverty percentage in the patient's zip code. We calculated the BCSI percentage of ESBC surgeries for all other ESBC patients living in a 50-mile radius around each patient's residence in their diagnosis year. We calculated the distance from each patient to the nearest radiation treatment center in the diagnosis year based on the zip code centroids.

Analytic Approach

For IV estimation, we used a nonparametric 2-stage least squares (2SLS) approach that has been used previously in healthcare research.[26,45–50] In the first estimation stage of 2SLS, the probability of BCSI was estimated using ordinary least squares as a function of measured confounding variables (cancer stage, grade and location, age, comorbidity, hospital access, payer, diagnosis year, area poverty level) and a series of binary variables that grouped patients based on their

instrument values. Binary variables for the instruments were constructed based on percentiles across the sample. In separate analyses, we varied the number of patient groups (2, 4, 8, and 12 groups) constructed for each instrument to assess the robustness of our findings. In the second stage of 2SLS, we estimated survival models using 4 different survival measures (1, 2, 3, and 4 years). Each survival model specified the set of measured confounders and the predicted BCSI probability from the first stage regression. To provide a direct comparison to the IV estimates, we then estimated comparable risk adjustment survival models using linear probability models.[49] Estimation was performed using STATA software (IVREG and REG) with robust standard errors.

RESULTS

Table 1 compares the ESBC patients in our sample by surgical choice, the BCSI percentage in the area around their residence, and the distance from the patient's residence to the nearest radiation treatment center. The patients who received BCSI were younger with lower staged disease, lower tumor grades, and had fewer comorbidities. Both instruments were related to whether a patient received BCSI and provided a more balanced distribution of measured confounders between groups than grouping patients by surgery. Differences remained in the distributions of age and tumor grade across patients grouped by the instruments. Iowa Cancer Registry officials suggested that grade distribution differences reflected different reporting practices across Iowa and were not related to disease severity, and the differences in the age distributions reflected pockets of rural elderly in Iowa. We controlled directly for these variables in our IV analysis.

Table 2 contains the Chow F-statistics for the instruments across several specifications of the first-stage BCSI choice model. The Chow F-statistics enable us to test whether the specified instrument described a statistically significant portion of the variation in BCSI choice after controlling for the other measured confounders. We report the F-statistics for specifications differentiated by the instruments specified in the model, the number of patient groups delineated by the instruments, and cancer stage. We present the F-statistics by cancer stage to help assess the average disease severity of the patients whose treatment choices were affected by each instrument. Both instruments described a statistically significant portion of the variation in BCSI choice across the entire ESBC sample. When we focus on the estimates by cancer stage, however, we find that distance from the radiation center affected the surgery choices for both stage I and stage II patients, whereas the local area BCSI rate affected only the surgery choices for stage II patients and did not affect the surgery choices for stage I patients.

Table 3 contains the risk adjusted linear probability model and the IV survival models. For both methods, the parameter estimates are interpreted as the change in the X-year survival rate for a 1 percentage point increase in the BCSI rate for the respective population subsets discussed earlier. After the discussion above, the linear probability model specifications yield estimates of the average BCSI survival effects for the patients treated with BCSI. These estimates will be biased to the

TABLE 1. Comparison of Patient Characteristics of ESBC Patients in Iowa Grouped by Treatment and Instruments, 1989–1994

| | Full Sample | Treatment | | Area BCSI Percentage* | | Distance From Radiation Center[†] | |
		Mastectomy	BCSI	Lower	Greater	Far	Near
BCSI (%)	9.7	0	100	11.1	16.8[‡]	10.9	17.2[‡]
Age (%)							
<65	47.1	44.3	64.1[‡]	44.5	49.5[‡]	42.6	51.5[‡]
65–74	27.3	27.5	26.0	28.1	26.5	28.0	26.6
75+	25.7	28.2	9.9[‡]	27.4	24.1[‡]	29.5	22.0[‡]
Tumor size (%)							
T1 (<2 cm)	66.4	64.1	80.3[‡]	65.9	67.0	66.2	66.6
T2 (2–5 cm)	32.3	34.4	19.4[‡]	32.8	31.8	32.6	31.9
T3 (>5 cm)	1.3	1.5	0.2[‡]	1.4	1.3	1.2	1.4
Positive nodes (%)	27.8	29.2	19.5[‡]	28.7	27.1	28.0	27.7
Stage (%)							
I	53.0	50.7	67.5[‡]	51.7	54.3[§]	52.6	53.4
IIa	31.2	32.1	25.6[‡]	32.5	30.0[§]	31.7	30.7
IIb	15.8	17.2	6.9[‡]	15.9	15.7	15.6	15.9
Grade (%)							
1	6.6	6.0	10.2[‡]	4.3	8.8[‡]	5.5	7.7[‡]
2	22.9	26.7	22.2[‡]	19.3	26.2[‡]	24.7	21.1[‡]
3	26.6	27.0	24.3	27.0	26.2	26.8	26.4
4	5.1	5.4	3.1[‡]	7.1	3.3[‡]	7.1	3.2[‡]
9 (unknown)	38.9	39.4	35.8[‡]	42.4	35.6[‡]	35.9	41.7[‡]
Charlson comorbidity index >0 (%)[‖]	19.6	21.1	10.9[‡]	19.6	19.7	20.3	19.0
High poverty zip code	35.0	35.2	34.3	35.2	34.9	32.9	37.1
No. patients in group	6185	5315	870	2972	3213	3051	3134

*Patient in "lower" group if less than 20% of all ESBC surgeries (stage I and II) in the 50-mile radius around the patient's residence in the year of diagnosis were BCSI.

[†]Patient in Anear@group if distance to radiation treatment center in year of diagnosis is less than 19 miles.

[‡][§]Statistically different rate across groups at 0.99 and 0.95 confidence levels, respectively.

[‖]Modified Charlson comorbidity indices developed by mapping Clinical Classifications Software (CCS) diagnosis and procedure groups available on each HCUP discharge abstract into Charlson index groups.

extent that unmeasured confounders affect both surgery choice and survival. The first row contains the unadjusted linear probability model estimates which clearly reflect the favorable selection of patients into BCSI, because these estimates suggest that, for patients receiving BCSI, BCSI has a survival advantage over mastectomy. The second row contains the linear probability model estimates adjusted for the measured confounders. These estimates reveal no survival difference between BCSI and mastectomy for the patients receiving BCSI. Relative to Figure 1, this estimate of $\hat{a}_1$ can be interpreted as an upper-bound estimate of the average survival effect of BCSI relative to mastectomy for the ESBC patients receiving BCSI from the origin to *U*. If unmeasured confounders remain that are favorable toward BCSI, however, these estimates represent an upper bound on the survival effects of BCSI relative to mastectomy.

The IV estimates show a negative effect of BCSI on survival relative to mastectomy that is consistent across specifications. The magnitude of these estimates and the level of statistical significance, however, varied with the instrument specification. The use of local BCSI rate as an instrument produced the largest estimated survival impacts of BCSI and the specifications with distance from the radiation center

the smallest. Including both instruments in the specification yielded estimates between the estimates found with the instruments specified individually. Relative to the earlier published survival estimates for stage II patients alone, both the unadjusted linear probability model and the IV estimates using combined stage I and stage II patients are consistently smaller.[26]

DISCUSSION

Our objective was to demonstrate how treatment effect heterogeneity affects the interpretation of the treatment effect estimates using risk adjustment and IV methods on observational data. Risk adjustment estimation approaches yield average estimates for the set of patients who received a given treatment, whereas IV methods yield average estimates for patients whose treatment choices were affected by instrumental variables.[18,24,43,51] If treatment effects are heterogeneous across a population, these estimates will vary with the patients whose treatment choices were used to estimate the treatment effects. In our example, we theorize that patients receiving BCSI will be favorably selected with respect to

TABLE 2. Chow[42] *F*-Statistics Testing Whether Instruments Affected the BCSI Choice for ESBC Patients in Iowa 1989–1994*

Instruments Specified	No. Patient Groups Per Instrument	ESBC Patients		
		All (N = 6185)	Stage I (N = 3280)	Stage II (N = 2905)
Local area BCSI rate	2	11.87†	3.74	8.57†
	4	4.95†	1.26	5.19†
	8	2.98†	0.69	3.43†
	12	2.41†	1.31	3.00†
Distance from the radiation center	2	27.73†	9.79†	21.79†
	4	9.39†	3.58‡	7.52†
	8	5.51†	3.36†	3.30†
	12	5.03†	3.30†	2.94†
Local area BCSI rate and distance from the radiation center	2	16.62†	5.61†	13.08†
	4	5.54†	1.90	4.99†
	8	3.44‡	1.77‡	2.76†
	12	3.37†	2.20†	2.74†

*Models also specified binary variables for age groups (<50, 50–64, 65–69, 70–74, 75–79, 80–84, 85+), tumor sizes (<2, 2–5, and 5+ cm), positive lymph node involvement, tumor grade groups (1, 2, 3, 4, 9-unknown), tumor location groups (nipple, central portion, upper-inner quad, lower-inner quad, upper-outer quad, lower-outer quad, axillary tail, overlapping lesion, not-stated), Charlson comorbidity index (0, 1, 2, 3+), residence zip code poverty percentage (#7, 7–10, 10–13, 13–20, >20), distance from residence to nearest hospital (#2.83, 2.83–9, 9–15, >15), payer (Medicaid, Medicare, Blue Cross/Blue Shield, other private, other government, self-pay), and year of diagnosis (1989, 1990, 1991, 1992, 1993, 1994).
†‡Statistically significant at 0.99 and 0.95 confidence level, respectively.

TABLE 3. Instrumental Variable and Risk-Adjusted Linear Probability Model (LPM) Estimates of the Effectiveness of BCSI on Survival Relative to Mastectomy for ESBC Patients in Iowa, 1989-1994

Row	Analysis Method	Instruments Specified	No. Groups Per Instrument	Instrument *F*-Statistic	After Diagnosis, Effect of BCSI on Patient Survival			
					1 yr	2 yr	3 yr	4 yr
1	Unadjusted LPM	None	NA	NA	0.012*	0.025†	0.053†	0.070†
2	Adjusted LPM‡	None	NA	NA	−0.001	−0.003	0.003	0.002
3	Instrumental variable estimates‡	BCSI rate	2	11.87†	−0.20	−0.34	−0.35	−0.20
4			4	4.95†	−0.24	−0.38*	−0.30	−0.28
5			8	2.98†	−0.23*	−0.36*	−0.29	−0.26
6			12	2.41†	−0.18*	−0.30*	−0.25	−0.10
7		Radiation distance	2	27.73†	−0.12	−0.03	−0.07	−0.13
8			4	9.39†	−0.12	−0.10	−0.18	−0.31
9			8	5.51†	−0.12	−0.07	−0.09	−0.12
10			12	5.03†	−0.06	−0.06	−0.15	−0.25
11		BCSI rate and radiation distance	2	16.62†	−0.14*	−0.11	−0.14	−0.14
12			4	5.54†	−0.16*	−0.16	−0.20	−0.29
13			8	3.44*	−0.16†	−0.15	−0.09	−0.09
14			12	3.37†	−0.09*	−0.12	−0.11	−0.15

*†Statistically significant at 0.95 and 0.99 confidence level, respectively.
‡Models also specified binary variables for age groups (<50, 50–64, 65–69, 70–74, 75–79, 80–84, 85+), tumor sizes (<2, 2–5, and 5+ cm), positive lymph node involvement, tumor grade groups (1, 2, 3, 4, 9-unknown), tumor location groups (nipple, central portion, upper-inner quad, lower-inner quad, upper-outer quad, lower-outer quad, axillary tail, overlapping lesion, not-stated), Charlson comorbidity index (0, 1, 2, 3+), residence zip code poverty percentage (#7, 7–10, 10–13, 13–20, >20), distance from residence to nearest hospital (#2.83, 2.83–9, 9–15, >15), payer (Medicaid, Medicare, Blue Cross/Blue Shield, other private, other government, self-pay), and year of diagnosis (1989, 1990, 1991, 1992, 1993, 1994).

survival and that providers believe that the survival benefit of BCSI relative to mastectomy is similar for ESBC patients with less severe disease, but that the survival benefit of BCSI relative to mastectomy diminishes as severity increases.

Based on Heckman's insights, our linear probability models yield estimates of the average survival estimate of BCSI relative to mastectomy for the patients who received BCSI. The unadjusted linear probability model estimates in Table 3 clearly reflect the effects of favorable selection, as they suggest that BCSI has a protective survival effect relative to mastectomy. Risk adjusting for measured confounders seems to eliminate the favorable selection bias because the

adjusted linear probability model results are comparable to the RCT findings of no survival difference between BCSI and mastectomy for patients with less serious disease. In contrast, the IV estimates show a consistently negative survival risk of BCSI relative to mastectomy, which, following Angrist's ideas, can only be generalized to the patients whose surgery choices were affected by the instrument specified. The magnitude of the effect was greatest when the area BCSI rate was the instrument and the smallest when distance from the radiation center was specified. This result is consistent with our theoretical framework and the findings in Table 2, which showed that distance from the radiation center affected the BCSI choice of patients in both stages I and II, whereas the area BCSI rate only affected BCSI choice for stage II patients.

In this study, we showed how inferences of treatment effectiveness can be made by applying risk adjustment and IV models to retrospective data when the treatment effect is thought to be heterogeneous across patients. Theoretical models of treatment choice and outcome coupled with assumptions of the relationship between unmeasured confounders, treatment choice, and outcomes can be used to bound estimates of the treatment effect for treated patients who come from risk adjustment models. IV estimates provide information on the effectiveness of treatment for those patients whose treatment choices would be likely affected by a change in treatment rates. However, in the application of IV estimates, policymakers must consider whether the patients whose treatment choices were affected by individual instruments are similar to those patients whose treatment choices are apt to change as the result of a policy under consideration.

REFERENCES

1. Gross CP, Filardo G, Mayne ST, et al. The impact of socioeconomic status and race on trial participation for older woman with breast cancer. *Cancer*. 2004;103:483–491.
2. Gross CP, Herrin J, Wong N, et al. Enrolling older persons in cancer trials: the effect of socioeconomic, protocol, and recruitment center characteristics. *J Clin Oncol*. 2005;23:4755–4763.
3. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *JAMA*. 2004;291:2720–2726.
4. Upshur REG. Looking for rules in a world of exceptions. *Perspect Biol Med*. 2005;48:477–489.
5. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q*. 2004;82:661–687.
6. Starfield B. Threads and yarns: weaving the tapestry of comorbidity. *Ann Fam Med*. 2006;4:101–103.
7. Rothwell PM. Subgroup analysis in randomized controlled trials: importance, indications, and interpretation. *Lancet*. 2005;365:176–186.
8. Lohr KN, Eleazer K, Mauskopf J. Health policy issues and applications for evidence-medicine and clinical practice guidelines. *Health Policy*. 1998;46:1–19.
9. Steinberg EP, Luce BR. Evidence based? Caveat emptor! *Health Aff*. 2005;24:80–92.
10. Wennberg JE. Which rate is right? *N Engl J Med*. 1986;315:810–815.
11. Saver JL, Kalafut M. Combination therapies and the theoretical limits of evidence-based medicine. *Neuroepidemiology*. 2001;20:57–64.
12. Byar DP. Problems with using observational databases to compare treatments. *Stat Med*. 1991;10:663–666.
13. Jollis J, Ancukiewicz M, DeLong ER, et al. Discordance of databases designed for claims payment versus clinical information systems. *Ann Int Med*. 1993;119:844–850.
14. Doll R. Summation of a conference, doing more good than harm: the evaluation of health care interventions. *Ann N Y Acad Sci*. 1994;705:310–313.
15. Hornberger J, Wrone E. When to base clinical policies on observational versus randomized trial data. *Ann Int Med*. 1997;127:697–703.
16. Harris KM, Remler DK. Who is the marginal patient? Understanding instrumental variables estimates of treatment effects. *Health Serv Res*. 1998;33:1337–1360.
17. Newhouse J, McClellan M. Econometrics in outcomes research: the use of instrumental variables. *Annu Rev Public Health*. 1998;19:17–34.
18. Heckman JJ, Robb R. Alternative methods for evaluating the impact of interventions. In: Heckman JJ, Singer B, eds. *Longitudinal Analysis of Labor Market Data*. New York, NY: Cambridge University Press; 1985:156–245.
19. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for casual effects. *Biometrika*. 1983;70:41–55.
20. Manski CF. *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard University Press; 1995.
21. Heckman J. Instrumental variables: a study of implicit behavioral assumptions used in making program evaluations. *J Hum Resour*. 1997;32:441–462.
22. Heckman JJ, Urzua S, Vytlacil EJ. Understanding instrumental variables in models with essential heterogeneity. NBER Working Paper #12574. National Bureau of Economic Research, Inc.; 2006.
23. Angrist JD. Treatment effect heterogeneity in theory and practice. *Econ J*. 2004;114:C52–C83.
24. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica*. 1994;62:467–475.
25. Brooks JM, Chrischilles E, Scott S, et al. Information gained from linking SEER cancer registry data to state-level hospital discharge abstracts. *Med Care*. 2000;38:1131–1140.
26. Brooks JM, Chrischilles E, Scott SD, et al. Was breast conserving surgery underutilized for early stage breast cancer? Instrumental variables evidence for stage II patients from Iowa. *Health Serv Res*. 2003;38(6 Part I):1385–1402. [Erratum appears in *Health Serv Res*. 2004;39:693].
27. Winship C, Morgan SL. The estimation of causal effects from observational data. *Annu Rev Sociol*. 1999;25:659–706.
28. Arriagada R, Le MG, Rochard F, et al. Concervation treatment versus mastectomy in early breast cancer: patterns of failure with 15 years of follow-up data. *J Clin Oncol*. 1996;14:1558–1564.
29. Fisher B, Anderson S, Redmon CK, et al. Reanalysis and results after 12 years of follow-up in a randomized clinical trial comparing total mastectomy with lumpectomy with or without irradiation in the treatment of breast cancer. *N Engl J Med*. 1995;333:1456–1461.
30. Jacobson JA, Danforth DN, Cowan KH, et al. Ten-year results of a comparison of conservation with mastectomy in the treatment of stage I and II breast cancer. *N Engl J Med*. 1995;332:907–911.
31. van Dongen JA, Voogd AC, Fentiman IS, et al. Long-term results of a randomized trial comparing breast-conserving therapy with mastectomy: European Organization for Research and Treatment of Cancer 10801 Trial. *J Natl Cancer Inst*. 2000;92:1143–1150.
32. Veronesi U, Banfi A, Salvadori B, et al. Breast conservation is the treatment of choice in small breast cancer: long-term results of a randomized trial. *Eur J Cancer*. 1990;26:668–670.
33. NIH Consensus Conference. Treatment of early-stage breast cancer. *JAMA*. 1991;265:391–397.
34. van Dongen JA, Bartelink H, Fentiman IS, et al. Factors influencing local relapse and survival and results of salvage treatment after breast-conserving therapy in operable breast cancer: EORTC trial 10801, breast concervation compared with mastectomy in TNM stage I and II breast cancer. *Eur J Cancer*. 1992;28A:801–805.
35. Gilligan MA, Kneusel RT, Hoffman RG, et al. Persistent differences in socioeconomic determinants of breast conserving surgery adoption despite overall increased adoption. *Med Care*. 2002;40:181–189.
36. Du X, Freeman DH, Syblik DA. What drove changes in the use of breast conserving surgery since the early 1980s? The role of the clinical trial, celebrity action and an NIH consensus statement. *Breast Cancer Res Treat*. 2000;62:71–79.
37. Riley GF, Potosky AL, Klabunde CN, et al. Stage at diagnosis and treatment patterns among older women with breast cancer: an HMO and fee-for-service comparison. *JAMA*. 1999;281:720–726.

38. Morrow M, White J, Moughan J, et al. Factors predicting the use of breast-conserving therapy in stage I and II breast cancer. *J Clin Oncol*. 2001;19:2254–2262.

39. Benedict S, Cole DJ, Baron L, et al. Factors influencing choice between mastectomy and lumpectomy for women in the Carolinas. *J Surg Oncol*. 2001;76:6–12.

40. Kelemen JJ, Poulton T, Swartx MT, et al. Surgical treatment of early-stage breast cancer in the Department of Defense Healthcare System. *J Am Coll Surg*. 2001;192:293–297.

41. Nold RJ, Beamer RL, Helmer SD, et al. Factors influencing a woman's choice to undergo breast-conserving surgery versus modified radical mastectomy. *J Surg*. 2001;180:413–418.

42. Chow G. Tests of equality between sets of coeffiecient in two linear models. *Econometrica*. 1960;28:591–605.

43. Angrist JD, Krueger AB. Instrumental variables and the search for identification: from supply and demand to natural experiments. *J Econ Persp*. 2001;15:69–85.

44. Phelps CE. Diffusion of information in medical care. *J Econ Perspect*. 1992;6(3 Summer):23–42.

45. Baicker K, Chandra A, Skinner JS. Geographic variation in health care and the problem of measuring racial disparities. *Perspect Biol Med*. 2005;48(1 Suppl):S42–S53.

46. Mandelblatt JS, Berg CD, Meropol NJ, et al. Measuring and predicting surgeon's practice styles for breast cancer treatment in older women. *Med Care*. 2001;39:228–242.

47. Beck CA, Penrod J, Gyorkos TW, et al. Does aggressive AMI care reduce mortality? *Health Serv Res*. 2003;38:1423–1440.

48. Brooks JM, McClellan M, Wong HS. The marginal benefits of invasive treatments for acute myocardial infarction: does insurance coverage matter? *Inquiry*. 2000;37:75–90.

49. McClellan M, McNeil B, Newhouse J. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA*. 1994;272:859–866.

50. McClellan M, Newhouse JP. The marginal cost-effectiveness of medical technology: a panel instrumental-variables approach. *J Econom*. 1997;77:39–64.

51. Heckman JJ, Vytlacil E. The relationship between treatment parameters within a latent variable framework. *Econ Lett*. 2000;66:33–39.