HEALTH ECONOMICS
Health Econ. **12**: 171–186 (2003)
Published online 31 May 2002 in Wiley InterScience (www.interscience.wiley.com). **DOI:**10.1002/hec.710

OUTCOME MEASUREMENT

An exploratory instrumental variable analysis of the outcomes of localized breast cancer treatments in a medicare population

Jack Hadley[a,b,*], Daniel Polsky[c,d], Jeanne S. Mandelblatt[e], Jean M. Mitchell[f], Jane C. Weeks[g], Qin Wang[h], Yi-Ting Hwang[i] and the OPTIONS Research Team
[a] *The Urban Institute, Washington, DC, USA*
[b] *Center for Studying health System Change, Washington, DC, USA*
[c] *University of Pennsylvania, USA*
[d] *Leonard Davis Institute of Health Economics, Philadelphia, PA, USA*
[e] *Department of Human Oncology and Medicine, Georgetown University School of Medicine, Lombardi Cancer Center, Washington, DC, USA*
[f] *Public Policy Institute, Georgetown University, Washington, DC, USA*
[g] *Department of Adult Oncology, Dana-Farber Cancer Institute, Harvard University School of Medicine, Boston, MA, USA*
[h] *Institute for Health Care Research and Policy, Georgetown University, Washington, DC, USA*
[i] *Department of Human Oncology, Georgetown University School of Medicine, Lombardi Cancer Center, Washington, DC, USA*

Summary

This study is motivated by the potential problem of using observational data to draw inferences about treatment outcomes when experimental data are not available. We compare two statistical approaches, ordinary least-squares (OLS) and instrumental variables (IV) regression analysis, to estimate the outcomes (three-year post-treatment survival) of three treatments for early stage breast cancer in elderly women: mastectomy (MST), breast conserving surgery with radiation therapy (BCSRT), and breast conserving surgery only (BCSO). The primary data source was Medicare claims for a national random sample of 2907 women (age 67 or older) with localized breast cancer who were treated between 1992 and 1994.

Contrary to randomized clinical trial (RCT) results, analysis with the observational data found highly significant differences in survival among the three treatment alternatives: 79.2% survival for BCSO, 85.3% for MST, and 93.0% for BCSRT. Using OLS to control for the effects of observable characteristics narrowed the estimated survival rate differences, which remained statistically significant. In contrast, the IV analysis estimated survival rate differences that were not significantly different from 0. However, the IV-point estimates of the treatment effects were quantitatively larger than the OLS estimates, unstable, and not significantly different from the OLS results. In addition, both sets of estimates were in the same quantitative range as the RCT results.

We conclude that unadjusted observational data on health outcomes of alternative treatments for localized breast cancer should not be used for cost-effectiveness studies. Our comparisons suggest that whether one places greater confidence in the OLS or the IV results depends on at least three factors: (1) the extent of observable health information that can be used as controls in OLS estimation, (2) the outcomes of statistical tests of the validity of the instrumental variable method, and (3) the similarity of the OLS and IV estimates. In this particular analysis, the OLS estimates appear to be preferable because of the instability of the IV estimates.

Keywords instrumental variable analysis; observational data bias; outcomes of localized breast cancer; health outcomes research

---

*Correspondence to: The Urban Institute, 2100 M St., NW, Washington, DC 20037, USA. E-mail: Jhadley@ui.urban.org

Received 23 October 2000
Accepted 3 January 2002

Introduction

Cost-effectiveness analysis and evidence-based medicine are increasingly being used as tools to help decide 'what works best' in medicine [1]. Randomized clinical trials (RCTs) generally provide the best information on clinical outcomes for use in cost-effectiveness and evidence-based studies [2]. However, for ethical, practical, and cost reasons, it is often not feasible to conduct RCTs of alternative treatments. Rather, observational data are frequently used as a substitute [2]. Some even envision a national health outcomes data base that would record medications used and results for people treated in non-experimental settings [3].

As is well known, however, analyses of observational data face a fundamental difficulty. The treatment received may be correlated with unobserved health or other characteristics that also influence subsequent health outcomes, costs, and health utilities. Where such correlations exist, it is difficult to disentangle the effects of the treatment from the effects of unobserved characteristics that influence both the treatment received and subsequent outcomes. This problem is typically referred to as observational data bias or selection bias.

Instrumental variable (IV) methods have been used to overcome selection bias and provide consistent estimates of the causal relationship between treatment and outcome for several conditions and treatments [4–7]. Analyzing observational data on the outcome of localized breast cancer treatments provides a good opportunity to explore further the potential value of IV methods because extensive RCTs of alternative breast cancer treatments were conducted during the 1980s.

Several RCTs compared survival rates for breast conserving surgery with radiation therapy (BCSRT), breast conserving surgery only (BCSO), and mastectomy (MST) in the treatment of women with localized (stage 1 or 2) breast cancer. The largest trial (1843 women of all ages evenly divided among the three treatment arms) found that there were no statistically significant differences in five-year survival rates, which were 75.9% for MST, 79.8% for BCSRT, and 84.2% for BCSO [8]. Other RCTs, which had smaller enrollments and compared only MST and BCSRT survival over (6–15) years, also found statistically insignificant survival differences ranging from −3% to 8% [9].

The RCT results create a strong prior hypothesis of no difference in outcomes between the treatments of BCSRT, BCSO, and MST. If analysis of observational data finds survival differences between these treatments, there would be reason to believe that these differences are driven by selection bias rather that a true causal relationship between treatment and survival. Further, if the estimates of the observational data using IV methods were similar to those of the RCTs, this example would reinforce the use of IV methods as a substitute for RCTs in situations where they are too costly, too impractical, or unethical. Alternatively, if the estimates using the IV methods are substantially different from those suggested by the RCTs, the analysis would demonstrate the limits of IV analysis.

Specifically, we hypothesize that IV estimates of the differences in outcomes of alternative breast cancer treatments in elderly women should be small and statistically insignificant. We also hypothesize that if treatment selection is correlated with underlying health in the observational data, then the treatment received by the healthiest women should be associated with the best survival, and the treatment received by the sickest women should be associated with the worst survival.

We analyze data for a sample of Medicare beneficiaries (age 67 and older) who were treated for localized breast cancer between 1992 and 1994. Our goal is to assess the validity of an IV approach to specifying treatment received by comparing the IV and observational data results of an analysis of the effect of treatment (BCSRT, BCSO, or MST) on survival. We conduct the standard statistical tests for the quality of the IV estimates [10,11]. We also test the sensitivity of our results to different specifications of the functional form and the exogenous identifying instruments used to construct the IV treatment estimates.

Methods

Sample

The sample was constructed in two steps because it was drawn from Medicare claims, which neither confirm a cancer diagnosis nor identify disease stage. First, potentially eligible cases were drawn based on reported diagnosis and surgical procedures. Second, we surveyed the physicians who performed the breast surgery to verify diagnosis and determine disease stage.

Initial patient sample selection

Following selection strategies used in earlier studies [12–14], we obtained all inpatient, outpatient, and physician Part-B claims which had either a breast cancer diagnosis or surgery procedure code (Table A1 of Appendix A) for calendar years 1992–1994 from the Health Care Financing Administration's (HCFA) national claims data base for a 5% random sample of all Medicare beneficiaries.

In order to limit the sample to women for whom clinical conditions do not strongly favor either BCS or MST, we excluded cases for the following reasons: history of cancer diagnosis (4.5%), CIS diagnosis (3.0%), metastasis diagnosis (0.6%), or bilateral procedure (0.1%). We also excluded cases without a surgical procedure code (4.0%) or missing a physician identifier (1.7%), and if age was less than 67 (14.2%). (The lower age boundary was selected in order to have up to two years of prior Medicare claims for another analysis.) We also followed earlier studies and deleted cases for whom breast surgery was not the primary procedure code (16.2%) or breast cancer was not the primary diagnosis (13.9%), because they are less likely to be bona fide cases of new breast cancer. Lastly, we excluded 5.5% of cases because the physician provider number could not be matched to an identifiable individual physician in HCFA's provider data base. Applying these exclusion criteria resulted in a preliminary sample of 10 695 women from HCFA's 5% random sample for 1992–1994.

Physician survey to determine patient eligibility

We surveyed the physicians identified on the surgical claims to verify study eligibility based on the presence of localized invasive disease (tumor size T_1 or T_2; nodal status N_0, N_1, or N_2; and metastasis status M_0) and the absence of the exclusion criteria listed above. The patient sample generated 6698 unique physicians who had performed surgery on at least one sample patient. Our goal was to identify approximately 3000 eligible women who would serve as the sampling frame for a follow-up telephone survey in 1997. Physicians received two mailings, with subsequent follow-up contacts made by telephone. 80.7% provided information for at least one patient, 8.6% were unable to provide the information, 3.6% could not be reached, and 7.1% refused. After deletions for post-facto exclusions, 2907 women with complete data for all variables were classified as eligible.

Other data sources

We used Medicare's 100% National Claims History file for 1994 to extract all breast surgery claims to construct average Medicare payments to physicians for mastectomy and breast conserving surgery across geographic areas defined by 3-digit zip codes. The Medicare program was also the source of information for physicians' year of graduation from medical school, which was obtained from Medicare's Provider of Services file for physicians, and the geographic input price index. Information on hospitals with radiation therapy facilities was obtained from the American Hospital Association's Annual Survey of Hospitals. We used 1990 Census data by zip code to construct area-level data on per capita income and the percentage of the population with a college education, which were linked to sample cases by zip code.

Model specification

Equation (1) summarizes the outcome model:

$$O_i = \beta_0 + \beta_1 T_i + \beta_2 Y_i + U_{0i} \quad (1)$$

where outcome, O, is a linear function of the treatment received, T, observable health and sociodemographic characteristics, Y, and a random error term U_0. Outcome is measured by survival status three years after the date of surgery. The error term U_0 captures unobserved health characteristics. If selection bias affects the treatment received, T and U_0 will be correlated, resulting in a biased estimate of β_1.

To correct for this bias we employ, two-stage least squares (2SLS), an instrumental variables approach. Equations (2a) and (2b) summarize the 2SLS treatment and outcome models. In Equation (2a) we assume that the actual treatment received, T (MST – mastectomy, BCSRT – breast conserving surgery with radiation therapy, or

BCSO – breast conserving surgery only) can be characterized by a latent-index framework [4], where the index, T^*, depends linearly on a vector, **Y**, of observable health and sociodemographic factors, a vector, **X**, of exogenous factors (i.e. the instrumental variables) that are unrelated to the woman's health condition (i.e. uncorrelated with U_2) but influence the treatment received, and a random error term, U_1, that captures unobservable health and other factors that influence treatment received.

In Equation (2b), Outcome, O, is assumed to depend linearly on the estimate of treatment from the first stage, the same vector of observable health and sociodemographic characteristics, Y, and a random error term U_2. The estimate of $\beta_{1\text{iv}}$ is consistent because the error term U_2 is uncorrelated with T^* as a result of the fact that the instrumental variables used in the first stage prediction of T are uncorrelated with $U_{2.}$.

For patient i,

$$T_{\text{i}}^* = \alpha_0 + \alpha_1\mathbf{X}_{\text{i}} + \alpha_2\mathbf{Y}_{\text{i}} + U_{1i},$$

and

$$T_i = \begin{cases} 1 & \text{if } T_{\text{i}}^* > 0, \quad \text{(patient gets treatment)} \\ 0 & \text{if } T_{\text{i}}^* \leqslant 0 \end{cases} \tag{2a}$$

$$O_i = \beta_0 + \beta_{1\text{iv}}T_i^* + \beta_2 Y_i + U_{2\text{i}} \tag{2b}$$

The exogenous identifying variables in the vector **X** include:

- Medicare's average physician fees (payments) in 1994 for breast conserving surgery and mastectomy, calculated for treating physicians' 3-digit zip code areas.
- The distance between the population centroids of the woman's 5-digit residential zip code and the 5-digit zip code of the nearest hospital with a radiation therapy facility (as a proxy for travel and inconvenience costs to receive radiation therapy).
- A regional dummy variable to represent regional variations in practice patterns.
- A measure of input prices faced by physicians, the Geographic Adjustment Factor used by Medicare to calculate fees paid under the Medicare Fee Schedule.
- Two dummy variables indicating when the treating surgeon graduated from medical school (post-1979, or between 1965–1979, relative to pre-1965).
- Two dummy variables for the year of surgery (1992 or 1993) to capture secular changes in treatment patterns.

Medicare fees are exogenous and independent of unobservable health because they were determined by a combination of the resource-based fee specified by the Medicare Fee Schedule, which is independent of any particular physician's or patient's characteristics, and the average historical Medicare payment in the geographic area. Implementation of the Medicare Fee Schedule began in 1992 and was carried out over five years, with the weight for local component of the fee going to zero over the transition period [15].

The woman's place of residence is also plausibly independent of underlying health. It is highly unlikely that a woman will change her residence, as recorded in Medicare's administrative data base, in the time between diagnosis and treatment of breast cancer in order to be closer to a radiation therapy hospital or to seek a particular treatment in another region. The distance to the nearest hospital with radiation therapy represents the time and inconvenience costs of follow-up radiation therapy, which usually consists of multiple treatments over several weeks. Greater distance, therefore, should be associated with a greater likelihood of receiving mastectomy or breast conserving surgery without radiation as the local treatment. A second geographic factor that is plausibly independent of underlying health but appears to influence treatment, perhaps through a historical practice style effect, is geographic area. In particular, we observe that women residing in the Northern half of the US (Northeast and Midwest census regions) and the Pacific census division are more likely to receive BCSRT and less likely to be treated by MST.

Another exogenous factor that might influence the surgeon's treatment decision is the price of inputs. Although the Medicare payment is designed to incorporate variations in input prices, this adjustment may be imperfect for specific procedures. Since mastectomy and breast conserving surgery differ in their use of both physician time inputs and hospital inputs, variations in input prices may affect the treatment received independently of the woman's health. Input prices are measured by Medicare's Geographic Adjustment Factor (GAF), which is a weighted average of hospital salaries, office rents, and the value of physician time, that Medicare uses to determine payments under the Medicare Fee Schedule.

Finally, we assume that both the surgeon's year of graduation from medical school and the year the woman was treated may influence the treatment received independently of the woman's health. Surgeons who graduated from medical school more recently may have adopted the NIH consensus guidelines more quickly than physicians who entered practice when mastectomy may have been the preferred treatment alternative. Similarly, the diffusion of the NIH guidelines may have caused secular changes in treatment patterns across all patients and surgeons.

The observable health and sociodemographic measures in the vector **Y** include age, race, disease stage, dummy variables for different levels of Medicare payments in the year prior to treatment, and area-level measures of income and education. The variables measuring prior Medicare payments reflect women's prior use of medical services, which we hypothesize is correlated with their comorbidities and prior health status.

Statistical analysis

Although the treatment and outcome variables are dichotomous, we used linear probability specifications, rather than nonlinear functional forms such as the logistic transformation. We chose the linear specification because conventional two-stage least squares using a linear probability model in the first stage is consistent whether or not the first-stage conditional expectation function is linear, while a nonlinear first stage equation is inconsistent unless the nonlinear function is correct [16]. In a sensitivity analysis, however, we did estimate the first stage using a multinomial logit specification to see if the results are sensitive to this particular nonlinear functional form. Finally, both models were estimated using White's efficient estimator to correct for heteroskedasticity of the error terms.

It has been shown that if the exogenous factors used to construct the instrumental variable are only weakly related to the endogenous variable replaced by the instrument, then the resulting parameter estimates will be biased toward the ordinary least-squares (OLS) (observational data) estimates, even if the instrumental variable is not correlated with the error term of the health outcome equation [10,11]. To evaluate whether there may be a bias from weak instruments, we performed an F-test of the null hypothesis that the exogenous identifying variables have no power to predict the treatment choice.

We then tested the assumption that the instrument is uncorrelated with the error term of the health outcome model using a test of the over-identifying restrictions [17]. We regressed the residual from the second-stage model estimated with the instrumental variables against all of the exogenous variables from the first- and second-stage equations. The test statistic for the null hypothesis of no correlation between the instruments and the error term is constructed by multiplying the R^2 from this regression times the number of observations. It has a chi-squared distribution with degrees of freedom equal to the difference between the number of exogenous identifying variables and number of instrumental variables. If the value of the test statistic exceeds the critical value for the chi-squared, then one rejects the null hypothesis that the instrument is uncorrelated with the error term.

The OLS model, however, is more efficient than the IV model. If treatment choice was not found to be endogenous, then the OLS estimates might be preferred. To evaluate this possibility, we performed the augmented Hausman specification test [18]. Rejecting the null hypothesis of no endogeneity would suggest there is sufficient difference between the coefficients of the IV and OLS models to reject the OLS model in favor of the alternative.

Results

Descriptive comparisons

Table 1 defines all of the study variables and reports their means for the full sample of all eligible women and by treatment received. Approximately 65% of women received MST, 25% BCSRT, and 10% BCSO. Overall, 86.6% of the women were still alive three years after surgery. However, the survival rates differed significantly by treatment received. Women who received BCSO had the lowest survival rate, 79.2%. Survival was 6.1% higher for MST cases, and 13.8% higher for BCSRT cases. In comparison, the randomized trials reported very small and

Table 1. Variable definitions and mean values, by treatment received

Variable[a]	All cases	Treatment received		
		Mastectomy (MST)	Breast conservation & radiation therapy (BCSRT)	Breast conservation only (BCSO)
Number of cases	2907	1875	715	317
Endogenous				
Treatment received				
Mastectomy	0.645	1.00	0.00	0.00
Breast conserving and radiation therapy	0.246	0.00	1.00	0.00
Breast conservation only	0.109	0.00	0.00	1.00
Survival three years post-treatment	0.866	0.853	0.930	0.792
Exogenous identifying variables				
Medicare fee for breast conserving surgery ($)[b]	369.44	363.92	379.50	379.45
Medicare fee for mastectomy ($)	843.18	838.80	847.40	859.64
Distance to nearest hospital with a radiation therapy Capacity (mi.)[c]	17.2	18.6	13.8	16.7
Treated in 1992	0.357	0.363	0.307	0.435
Treated in 1993	0.336	0.330	0.364	0.306
Treated in 1994[d]	0.308	0.308	0.329	0.259
North/Pacific areas	0.615	0.588	0.675	0.634
Other areas	0.385	0.412	0.325	0.366
Input price index	0.990	0.985	0.999	1.004
Surgeon's medical school graduation				
Pre-1965[d]	0.296	0.289	0.284	0.360
1965–1979	0.444	0.433	0.482	0.426
Post-1979	0.260	0.277	0.234	0.215
Other exogenous variables				
Percent of population who are college graduates (%)[e]	0.191	0.184	0.198	0.208
Per capita income ($)[e]	13 848	13 489	14 492	14 514
Age 67–74[d]	0.561	0.543	0.668	0.429
Age 75–79	0.232	0.239	0.221	0.218
Age 80+	0.207	0.219	0.111	0.353
Disease stage 1[d]	0.573	0.510	0.714	0.628
Disease stage 2a	0.330	0.367	0.241	0.312
Disease stage 2b	0.097	0.123	0.045	0.059
Nonwhite race	0.066	0.065	0.057	0.088
Medicare payments in prior year				
Less than $3000[d]	0.479	0.388	0.665	0.596
$3000–6,000	0.224	0.261	0.161	0.142
$6000–13,000	0.192	0.236	0.112	0.114
$13000–22,000	0.059	0.066	0.031	0.082
$22000 or more	0.047	0.050	0.031	0.062

[a] All variables are dichotomous unless otherwise noted.
[b] Calculated for 3-digit zip code of treating physicians.
[c] Calculated from the population centroids of the woman's 5-digit residential zip code and the hospital's 5-digit zip code.
[d] Omitted reference group.
[e] From 1990 census data for the woman's 5-digit zip code of residence.

Table 2. Mean value of selected observable health measures, treatment received, and survival, by distance to nearest radiation therapy hospital and by geographic area

	Distance to radiation therapy hospital		Geographic area	
	Near[a]	Far	North or Pacific	Other
Exogenous identifying variable				
Distance to radiation therapy hospital (mi.)	3.8	35.1	14.9	19.5
North/Pacific areas	0.672	0.539	1.000	0.000
Observable health measure[b]				
Age 67–74	0.551	0.575	0.555	0.571
Age 80+	0.211	0.201	0.214	0.201
Disease stage 1	0.570	0.577	0.571	0.576
Disease stage 2b	0.098	0.096	0.100	0.093
Prior medicare payments				
Less than $3000	0.484	0.471	0.475	0.485
More than $13 000	0.103	0.110	0.105	0.108
Treatment received[c]				
Mastectomy	0.608	0.696	0.618	
Breast conservation and radiation	0.279	0.201	0.270	0.207
Breast conservation only	0.114	0.103	0.113	0.104
Three-year survival[b]	0.860	0.873	0.862	0.871

[a] Less than 10 miles.
[b] $p > 0.29$ for all tests of significant differences.
[c] Differences in treatment patterns are significantly different from 0, $p < 0.01$.

statistically insignificant differences in survival rates (for women of all ages) after three years [8].

The data in Table 1 also suggest that there were substantial differences among treatment groups in the observable health indicators. The BCSRT group had the highest proportions of cases in the youngest age group (67–74), with stage 1 disease, and with the lowest Medicare payments in the prior year. In contrast, women treated by BCSO were much more likely to be in the very oldest age group and in the two highest prior year's Medicare spending categories, 14.4% compared to 6.2% for BCSRT. Women who received MST had the highest proportions of stages 2a and 2b disease.

In Table 2, we divided the sample into groups based on their values of two of the exogenous variables used to construct the instrumental variable treatment measures. These comparisons illustrate the logic underlying the IV estimation approach. In columns 1 and 2, the sample was grouped by whether they lived 'near' (less than 10 miles) or 'far' (more than 10 miles) from a hospital with a radiation therapy facility. In columns 3 and 4, cases were grouped by whether or not they lived in the Northern or Pacific areas of the US. If these factors are satisfactory exogenous identifying variables, then the treatment pattern should vary between the groups, but observable (and presumably unobservable) health characteristics should be similar, thereby approximating the effect of randomizing people to treatment groups in an RCT [8, Table 3]. While unobservable health characteristics are, by definition, unobserved, if observable health characteristics are similar we would expect unobservable characteristics to be similar as well.

Looking first at the observable health characteristics, we see that the distributions of women by age, disease stage, and prior Medicare spending are very similar across each of the groupings and are not significantly different from each other. For example, the proportion of women 80 or older ranged between 20.1 and 21.4%, compared to a more than threefold difference (11.1–35.3%) by actual treatment received (Table 1). Similarly, the proportion of women with stage 1 disease was between 57.0% and 57.7%, compared to a difference of more than 20 percentage points by

Table 3. Treatment received, first stage linear probability models

Variable	Parameter estimates (standard errors) by treatment received ($N=2907$)		
	Mastectomy (MST)	Breast conservation and radiation (BCSRT)	Breast conservation only (BCSO)
Identifying variables			
Distance to radiation therapy hospital	0.00083[a]	−0.00103[b]	0.00020
	(0.00040)	(0.00035)	(0.00027)
Medicare fee for breast conserving surgery	−0.00075[b]	0.00078[b]	−0.00003
	(0.00022)	(0.00020)	(0.00015)
Medicare fee for mastectomy	0.00029[c]	−0.00051[b]	0.00022[a]
	(0.00017)	(0.00015)	(0.00011)
Input price index	−0.17449	0.08819	0.08630
	(0.16749)	(0.15080)	(0.11366)
Physician graduated post-1979	−0.04234[c]	−0.00591[c]	−0.03643[a]
	(0.02256)	(0.02031)	(0.01531)
Physician graduated 1965–1979	−0.01210	0.03435	−0.02225[c]
	(0.01993)	(0.01794)	(0.01352)
North or pacific areas	−0.03910[a]	0.04553[b]	−0.00643
	(0.01853)	(0.01669)	(0.01258)
Treated in 1992	0.01577	−0.06046[a]	0.04469[b]
	(0.02066)	(0.01860)	(0.01402)
Treated in 1993	−0.00557	−0.00411	0.00968
	(0.02093)	(0.01885)	(0.01421)
Other exogenous variables			
Age 75–79	0.04089[b]	−0.06220[b]	0.02131
	(0.02073)	(0.01866)	(0.01407)
Age 80+	0.05410[b]	−0.15942[b]	0.10532[b]
	(0.02165)	(0.01949)	(0.01469)
Nonwhite race	0.00123	−0.04452	0.04329[c]
	(0.03410)	(0.03070)	(0.02314)
Percent with college degrees	−0.03151	−0.08986	0.12137[c]
	(0.012461)	(0.09225)	(0.06953)
Per capita income	−1.20E-6	2.55E-6	−1.36E-6
	(1.89E-6)	(1.70E-6)	(1.28E-6)
Prior year medicare payments			
$3000–6000	0.23972[b]	−0.17463[b]	−0.06509[b]
	(0.02145)	(0.01931)	(0.01455)
$6000–13 000	0.26743[b]	−0.19753[b]	−0.06991[b]
	(0.02264)	(0.02038)	(0.01536)
$13 000–22 000	0.18713[b]	−0.195178[b]	0.00805
	(0.03653)	(0.03289)	(0.02479)
$22 000+	0.15029[b]	−0.17241[b]	0.02212
	(0.04060)	(0.03655)	(0.02755)
Disease stage 2a	0.14626[b]	−0.12271[b]	−0.02355[c]
	(0.01835)	(0.01652)	(0.01835)
Disease stage 2b	0.24563[b]	−0.18603[b]	−0.05960[b]
	(0.02913)	(0.02623)	(0.01977)
Intercept	0.65710[b]	0.48797[b]	−0.14506[c]
	(0.12785)	(0.11511)	(0.08676)
Adjusted R^2	0.115	0.113	0.039
F-test for overidentifying restrictions (9, 2886)	6.13	7.76	3.53

[a] Significantly different from 0, $p < 0.05$.
[b] Significantly different from 0, $p < 0.01$.
[c] Significantly different from 0, $p < 0.10$.

actual treatment. In the RCTs, the proportion of women with tumors less than 2 cm in size ranged from 49.4–52.9% across the three treatment groups [8].

Comparing the distributions of treatments shows that there were significant differences in the proportions receiving either BCSRT or MST, but smaller differences in the proportion receiving BCSO. About 28% of women living either 'near' a hospital with a radiation therapy facility or living in the North or Pacific areas received BCSRT, compared to about 20% in the 'far' distance or 'other' areas groups, who were treated by BCSRT. The proportion receiving MST shows the opposite pattern. The fact that the BCSO proportion does not vary as much is consistent with the possibility that the decision to pursue the least aggressive therapy depends primarily on the woman's clinical condition and preferences, rather than on external factors that influence the convenience or payment associated with the alternative surgeries [19]. Finally, the differences in the three-year survival rates are much smaller across the groups in Table 2 than they are in Table 1. The survival difference was 1.3 percentage points for the groups stratified by distance to a hospital with radiation therapy, and 0.9 percentage points when stratified by region.

First-stage treatment choice models

Table 3 reports the coefficients of the first-stage linear probability regression models used to construct the instrumental variables for treatment received. Although the models did not explain a high proportion of the variation in treatments received (11% for the MST and BCSRT models, and 4% for the BCSO model), several of the exogenous identifying variables were statistically significant. Increasing the distance to the nearest hospital with a radiation therapy facility increased the probability of receiving MST. The Medicare fee variables have opposite and statistically significant effects in the BCSRT and MST equations. Living in the North or Pacific areas and having been treated in 1992 are also statistically significant in the BCSRT and MST equations, while the year of surgery and the physician's graduation era are significant in the BCSO equation. The input price index was not statistically significant in any of the models.

The observable health variables were highly significant and had the expected effects. Increasing age, advanced disease stage, and greater Medicare payments in the year before treatment all had substantial negative effects on the probability of receiving BCSRT, which clearly suggest that the healthiest women were most likely to be treated by BCSRT. Age, disease stage, and prior Medicare payments had different effects on the other two treatments, with age having the strongest impact on the probability of receiving BCSO, while disease stage and prior Medicare spending had greater positive effects on the probability of MST.

The *F*-statistics are reported at the bottom of Table 3. The null hypothesis that the parameters of the exogenous identifying variables are jointly equal to 0 is rejected in all three cases. However, as a conservative rule of thumb, an *F*-statistic less than 10 is an indicator of a weak instrument [11]. All three *F*-statistics are less than 10. The test statistics for MST and BCSRT were 6.13 and 7.76, while the *F*-statistic for the BCSO model was only 3.53. The bias from weak instruments may be an issue, particularly for BCSO.

Second-stage survival rate models

Table 4 shows the parameter estimates for the linear probability models of three-year survival. Column 1 reports the OLS results using the observational data on treatment received and column 2 reports the IV estimates of the effects on survival of receiving BCSRT or BCSO (relative to MST, as the omitted reference treatment). Adding the observable health characteristics to the observational data reduced, but did not eliminate the differences in survival rates between the three treatments. The survival rate was 4.5% lower for women treated by BCSO, and 3.4% higher for women who received BCSRT relative to treatment with MST. The unadjusted differences were −6.1% and 7.7%, respectively. From these results, one would still conclude that there are significant differences in survival among the three treatment alternatives.

The effects of the observable health characteristics were as expected. Survival declined significantly with increasing age, greater Medicare spending in the prior year, and more advanced disease stage. The sociodemographic variables and year of treatment did not appear to be associated with survival, although education and income are

Table 4. Three year survival, linear probability models

Variable		Parameter estimates (standard errors) by estimation method	
		Observational	Instrumental variable
Treatment received			
Breast conservation only	(−0.061[a])	−0.04500[b]	0.08537
		(0.02030)	(0.19630)
Breast conservation and radiation	(0.077[a])	0.03433[b]	−0.05252
		(0.01520)	(0.09979)
Age 75–79		−0.03291[b]	−0.04027[c]
		(0.01502)	(0.01540)
Age 80 +		−0.19115[c]	−0.21814[c]
		(0.01593)	(0.02884)
Nonwhite race		−0.00591	−0.01570
		(0.02455)	(0.02612)
Percent with college degree		0.05897	0.03777
		(0.07423)	(0.07527)
Per capita income		−1.60E-6	1.17E-6
		(1.33E-6)	(1.46E-6)
Prior year Medicare payments			
$3000–6000		0.00366	−0.00312
		(0.01588)	(0.02941)
$6000–13 000		−0.00901	−0.01773
		(0.01678)	(0.03101)
$13 000–22 000		−0.05147[b]	−0.07050[b]
		(0.02662)	(0.03603)
$22 000 +		−0.13621[c]	−0.15297[c]
		(0.02950)	(0.04215)
Disease stage 2a		−0.04723[c]	−0.05453[c]
		(0.01344)	(0.01911)
Disease stage 2b		−0.13685[c]	−0.14456[c]
		(0.02137)	(0.03655)
Intercept		0.95960[c]	0.98104[c]
		(.01733)	(0.05157)
Adjusted R^2		0.084	0.058

[a] Unadjusted difference.
[b] Significantly different from 0, $p < 0.05$.
[c] Significantly different from 0, $p < 0.01$.

measured with error, since they are area rather than the individual level variables.

The IV point estimates of the treatment effects (column 2) reverse the order of the survival rates and suggest that BCSRT has the lowest survival and BCSO the highest. Moreover, the standard errors of the estimates increased substantially. Although the ordering of the treatment effects changes, they do not approach statistical significance and one cannot infer that the treatments have significantly different survival rates. Lastly, the IV estimates reinforce the inference that the observational results are subject to selection bias, and that the significantly better survival associated with BCSRT may largely reflect differences in unobservable health. The other parameter estimates in the IV model change only slightly.

The second major test for the quality of an IV is that it be uncorrelated with the error term in the second-stage model. The chi-squared test statistic has a value of 5.2, which is less than the critical value of 14.1 with 7 degrees of freedom at the 0.05 level of confidence. Thus, we can be reasonably confident that the instruments satisfy the second test criterion. We also conducted the augmented Hausman test for the difference between the observational and IV parameter estimates. Given the large standard errors of the IV estimates, we

Table 5. Comparison of alternative IV parameter estimates (and standard errors), by first-stage functional form and specification (with and without input price measure

	LPM[a] with input price	LPM[a] without input price	MNL[b] with input price	MNL[b] without input price
Estimated IV treatment effects[c]				
BCSO	0.085	0.141	−0.214	−0.164
	(0.196)	(0.198)	(0.165)	(0.166)
BCSRT	−0.052	−0.032	−0.103	−0.090
	(0.100)	(0.101)	(0.084)	(0.085)
Weak instrument test				
F-statistic/chi-square				
BCSO	3.53 (*F*-test)	3.90 (*F*-test)	30.27 (chi-sq.)	29.68 (chi-sq.)
BCSRT	7.76	8.69	69.74	68.61
MST	6.13	6.76	—	—
Exogeneity test				
Chi-square	5.23	2.32	7.84	4.65
Augmented Hausman test				
Chi-square	1.70	1.90	1.45	—

[a] LPM – linear probability model in first stage.
[b] MNL – multinomial logistic model in first stage.
[c] Estimated difference in survival compared to MST, from second-stage linear probability model.

cannot reject the hypothesis that the two sets of parameters are significantly different from each other.

Alternative IV estimates

For the first stage of the two-stage least-squares model, if the logistic distribution is indeed the correct conditional expectation function then the multinomial logistic would also produce consistent estimates and it would be preferred over the linear probability specification. We explored this alternative specification of the first-stage model. The first-stage estimates (available on request) were qualitatively similar, with the same pattern of statistically significant exogenous identifying variables as in the linear probability models. In Table 5 we report the resulting IV treatment effects and statistical tests for the validity of the IV estimates. Using a multinomial logistic specification reverses the sign of the BCSO treatment effect and increases the magnitudes of the point estimates for both treatments relative to MST. BCSO now has the poorest estimated survival. The fact that the point estimates of the multinomial logistic differ from the linear model suggests that the conditional expectation function of the multinomial logistic may not be the appropriate functional form. However, even with the greater magnitudes of the point estimates, their standard errors remain large and the estimates are not significantly different from 0.

We also estimated a specification without the input price index, which was never statistically significant. Dropping the input price index from the first-stage models improves the outcomes of the weak-instrument and over-identification tests. However, the estimated treatment effects remain statistically insignificant with relatively large increases in the magnitudes of the BCSO estimates. None of the IV estimates are significantly different from the more efficient OLS estimates. In fact, the OLS point estimates of differences in survival, 3.4% and −4.5%, are closest in magnitude to the statistically insignificant differences found by the RCTs. Thus, while the alternative set of IV estimates appears to do better with respect to the statistical tests for over-identification and independence, ambiguity regarding the values of the point estimates remains.

Discussion

Our study was motivated by the potential problem of having to use observational data to draw inferences about treatment outcomes when experimental data are not available. We compared two statistical approaches, ordinary least-squares (OLS) and instrumental variables (IV) regression analysis, to estimate the outcomes of three treatments (mastectomy, MST, breast conserving surgery with radiation therapy, BCSRT, and breast conserving surgery only, BCSO) for early stage breast cancer in elderly women. The OLS approach assumes that the treatments received are independent of any unobserved differences among patients, while the IV approach 'simulates' randomization by creating variations in treatment patterns based on factors that are plausibly unrelated to differences in patients' characteristics. The advantage of comparing these approaches with data on breast cancer treatment outcomes is that experimental data from several RCTs provide a benchmark for assessing the OLS and IV statistical approaches.

The RCTs found that differences in survival rates among the three treatment groups are quantitatively small ($\pm$3–8 percentage point differences in the percentage of patients surviving from 5 to 18 years) and not statistically significant [9]. The largest trial, which included just over 1800 women of all ages divided approximately evenly among the three treatments, found statistically insignificant differences in survival rates of 8.4% for (BCSO–MST) and 3.9% for (BCSRT–MST).

We found evidence of selection bias based on underlying health into the treatment options. Tabular comparisons of women grouped by actual treatment received clearly indicated that women who received BCSRT had the best observable health characteristics, suggesting that their unobservable health characteristics were probably also better than those of women who received the other treatments. Controlling for observable health characteristics (age, disease stage, and prior Medicare spending) reduced the apparent observational data bias, from -6.1% to -4.5% for (BCSO–MST), and from 7.7 to 3.4% for (BCSRT–MST). However, with some important health characteristics remaining unobserved, the differences remained statistically significant and BCSRT, the treatment associated with better underlying health, still had the best survival rate.

Using a standard parametric instrumental variable method (two-stage least squares) to adjust for observational data bias reversed the order of the treatment outcomes, but increased the magnitudes of the differences in survival rates, to 8.5% for (BCSO–MST) and to -5.2% for (BCSRT–MST). However, because IV is generally less efficient than OLS, the estimated differences were no longer statistically significant. While the lack of statistical significance is consistent with the RCT results in a narrow statistical sense, we were also unable to reject the hypothesis that the IV results were different from the OLS results, even though the instrumental variables used appeared reasonable in that they were not correlated with observable health characteristics and they were predictive of the treatment received.

What conclusions can we draw from these comparisons? First, the similar magnitudes of the OLS and IV-point estimates of the differences in survival rates, and their similarity to the RCT estimates of survival differences, suggest that neither approach clearly dominates the other in this case. While there was clear evidence of selection bias in the unadjusted observational data, controlling for observable differences in health and other characteristics reduced the extent of bias. Other research suggests that detailed clinical information from medical records can in fact provide adequate controls [20]. The quantitative similarity of the OLS and IV-point estimates suggests that investigators could, and perhaps should, use both sets of estimates to establish a range of effectiveness outcomes.

Second, several aspects of the IV results are troubling. For one, the lack of statistical significance was due primarily to the substantial inflation in the values of the standard errors, which is a result typically associated with the use of IVs. Thus, the lack of statistical significance by itself should not be used to infer that the IV results are truly closer to the RCTs than are the OLS results. Moreover, the fact that the point estimates of the treatment effects generally increased in magnitude, rather than moving closer to 0 in value, and were sensitive to the specification of the first-stage model should make investigators very cautious. Thus, this particular (IV) application could not produce reliable estimates of the true, presumably small and insignificant differences in survival rates among the alternative treatments.

Although the ultimate results or our IV analysis may be ambiguous with regard to the objective of obtaining a precise point estimate of the treatment effect, we believe that several useful implications can be drawn. First, it is important to remember that the consistency of IV estimates is a large sample property. Our sample of 2907 cases, of whom only 317 received one of the treatments (breast conserving surgery without radiation therapy), may simply have been too small to produce unbiased and stable point estimates of the true treatment effects. In contrast, the IV analysis of treatments for acute myocardial infarction was based on over 200 000 observations [4]; and the IV analysis of longer postpartum stays on outcomes for newborns employed over 100 000 observations [6].

Second, instrumental variables that are predictive of treatment choice, but are considered to be weak instruments based on established criteria are unlikely to produce stable results [11]. In the absence of convincing and statistically sound exogenous identifying factors, the IV approach is unlikely to be successful. One should have greater confidence in the IV results if the collateral statistical tests are strong, and if the IV results are significantly different from the potentially biased OLS results. If the OLS and IV results are similar, then either the remaining unobservable bias is small, or the IV instruments are too weak to be reliable.

Third, the IV analysis of multiple treatment alternatives may be considerably more complex than a bivariate analysis of whether a single particular treatment or set of related treatments is received. In particular, our instrument for BCSO appeared to be much weaker than the instruments for BCSRT and MST. We have speculated elsewhere that this may reflect the fact that the initial treatment decision is between a less aggressive therapy (BCSO) and more aggressive therapy (MST or BCSRT), and that this decision is largely driven by the woman's health condition [19].

Fourth, this analysis highlights the importance of having access to information on fees, distances to available services, and other factors, such as insurance coverage and provider characteristics, that are plausibly unrelated to people's unobserved health states and can be used to construct instrumental variables for treatments received. Traditionally, little attention has been given to collecting this type of information as part of cost-effectiveness or evidence-based studies of medical treatments. Although there are significant issues of data confidentiality, cost, and complexity, it is clear that simply using observational data may result in highly erroneous conclusions.

Finally, results from observational studies should not be dismissed merely because they do not agree with results from randomized clinical trials. Trial results may not be generalizable to the time period or the subgroup of the observational study. During the time period of this study, 1992–1994, breast conserving surgery was widely diffused [13], which may have resulted in improved outcomes relative to the clinical trials when breast conserving surgery was still relatively new [12]. Also, this study looks exclusively at the surgery outcomes of elderly women which could differ from the surgery outcomes of the general population of women with breast cancer sampled in the clinical trials. The trials did not have the power to analyze whether the treatment effect for the subgroup of elderly women was consistent with the trial-wide effect [21]. Finally, it may be control over the choice itself that has an influence on the treatment effect [22, 23]; randomization eliminates this effect.

While potential selection bias needs to be addressed in any study using observational data, results should not be dismissed only because they do not agree with earlier trial evidence. Nor should observational data be dismissed automatically in the absence of any RCT evidence. Clearly, researchers using observational data need to assess the relationship between observable characteristics and treatment received. Our analysis suggests that if one exists, controlling for observable characteristics in a regression model may reduce the extent of bias, and other research suggests that the more detailed the controls, the greater the probable reduction in bias [20].

At the same time, in the absence of sufficient sample size and strong exogenous predictors of variations in treatment received, IV methods may not be a surefire cure. If the two approaches produce similar, albeit not identical estimates, and if the IV results are based on 'strong' instruments, then the combination of OLS and IV approaches may provide useful boundaries on the magnitude of the true effect. Conversely, if there is strong evidence of selection bias, but limited direct controls for underlying health differences, and if the OLS and IV results are very different, then the IV estimates, if based on strong instruments, may be preferred [24].

Acknowledgments

The authors would like to thank Lynne Eggert, FNP, MPH, Project Coordinator for field implementation of the protocol, the OPTIONS support staff who made data collection possible, Mathematica Policy Research, Inc for data management and collection, Mary Cummings, Ph.D. of the Agency for Healthcare Research and Quality for administrative and programmatic support and oversight, the OPTIONS National Advisory Committee for helpful suggestions on the conduct of the project, Ms. Sandy Fournier for administrative support, and Ms. Theresa Jordan and Ms Kathy Summers for manuscript preparation.

The OPTIONS (Outcomes and Preferences for Treatment in Older Women Nationwide Study) team includes:

Deborah Axelrod, MD	Beth Israel Medical Center
Frederick Barr, MD	Sibley Memorial Hospital
Christine Berg, MD	National Cancer Institute[†]
Caroline Burnett, RN, ScD	Georgetown University Medical Center
Anthony Cahan, MD	Beth Israel Medical Center
Stephen Edge, MD	Roswell Park Cancer Institute
Lynne Eggert, CRNP, MPH	Georgetown University Medical Center
Karen Gold, Ph.D.	Georgetown University Medical Center
Luther Gray, Jr., MD	Sibley Memorial Hospital
Jackie Dunmore-Griffith, MD	Howard University Hospital
Ed Guadagnoli, Ph.D.	Harvard Medical School
Jack Hadley, Ph.D.	Georgetown University Medical Center
Marilyn Halper, MPH	Beth Israel Medical Center
Nuhad Ibrahim, MD	MD Anderson Cancer Center
Claudine Isaacs, MD	Georgetown University Medical Center
Jon Kerner, Ph.D.	Georgetown University Medical Center
Jack Lynch, MD	Washington Hospital Center Cancer Institute
Jeanne Mandelblatt, MD, MPH	Georgetown University Medical Center
Neal Meropol, MD	Fox Chase Cancer Center
Jean Mitchell, Ph.D.	Georgetown University Medical Center
Bert Petersen, Jr., MD	Beth Israel Medical Center
Julia Rowland, Ph.D.	Georgetown University Medical Center
Kevin Schulman, MD, MBA	Georgetown University Medical Center
Ruby Senie, Ph.D.	Columbia School of Public Health
Brenda Shank, MD, Ph.D.	Doctors Medical Center
Robert Siegel, MD	George Washington University Medical Center
Rebecca Silliman, MD, MPH, Ph.D.	Boston University Medical Center
Juliana Simmons, MD	Washington Hospital Center Cancer Institute
Theodore Tsangaris, MD	Georgetown University Medical Center
Qin Wang	Georgetown University Medical Center
Jane Weeks, MD	Dana-Farber Cancer Institute
Rodger Winn, MD	MD Anderson Cancer Center

The OPTIONS National Advisory Committee includes:

Judith Baigis, RN, Ph.D.	Georgetown University
Harold Freeman, MD	Harlem Hospital Center
Mary Jo Gibson, MA	American Association of Retired Persons
Bruce Hillner, MD	Medical College of Virginia
Joanne Lamphere	American Association of Retired Persons
Amy Langer, MBA	National Alliance of Breast Cancer Organizations
Marc Lippman, MD	Georgetown University Medical Center
Monica Morrow, MD	Prentice Women's Hospital
Jeanne Petrek, MD	Memorial Sloan-Kettering Cancer Center
Rashida Muhammad	National Council of Negro Women
Natalie Davis Spingarn	——
David Winchester, MD	American College of Surgeons

Research Support was provided by Grant #HS08395 from the Agency for Health Care Policy and Research (Agency for Health Care Research and Quality), USDHHS; Grant #17-94-J-4212 from the Department of the Army; supplemental funding to AHRQ from the Federal Coordinating Committee on Breast Cancer; and The Urban Institute.

[†] Was at Georgetown when study was completed

Appendix A

ICD-9 and HCPCS codes used to select Medicare claims for women who may have had localized breast cancer (see Table A1).

Table A1

ICD 9 Inpatient Hospital Diagnosis Codes	
174–174.9	Malignant neoplasm of female breast
233	Carcinoma
196.1–196.9	Secondary and unspecified malignant neoplasm of lymph nodes
198.8	Secondary malignant neoplasm of female breast
ICD-9 Inpatient Hospital Procedure Codes	
40.11	Biopsy of lymphatic structure
40.23	Excision of axillary lymph node
85.20	Excision or destruction of breast tissue, not otherwise specified
85.21	Local excision of lesion of breast
85.22	Resection of quadrant of breast
85.23	Subtotal mastectomy
85.41–85.47	Unilateral or bilateral, simple or radical mastectomy
HCPCS (CPT) Physician Procedure Codes	
19 120	Excision of cyst, fibroadenoma, or other benign or malignant tumor aberrant breast tissue
19 160	Mastectomy, partial
19 162	Mastectomy with axillary lymphadenectomy
19 180	Mastectomy, simple, complete
19 182	Mastectomy, subcutaneous
19 240	Mastectomy, modified radical

References

1. Russell L, for members of the Panel on Cost-Effectiveness in Health and Medicine. The role of cost-effectiveness analysis for decision making in health. *JAMA* 1996; **276**: 1172–1177.
2. Mandelblatt JS, Fryback DF, Weinstein MC *et al*. Assessing effectiveness for cost-effectiveness analyses. In *Cost Effectiveness in Health and Medicine,* Gold MR, Siegel JE, Russell LB, Weinstein MC (eds). Oxford University Press: New York, 1996; 135–175.
3. Newmyer J. Access to your medication's track record. *Los Angeles Times* 2000; August 1.
4. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA* 1994; **272**: 859–866.
5. Hadley J, Rabin DR, Epstein A *et al*. Posthospitalization home health care use and changes in functional status in a Medicare population. *Med Care* 2000; **38**: 494–507.
6. Malkin JD, Broder MS, Keeler E. Do longer postpartum stays reduce newborn readmission? Analysis using instrumental variables. *Health Serv Res* 2000; **35**: 1071–1092.
7. Howard D. The impact of waiting time on liver transplant outcomes. *Health Serv Res* 2000; **35**: 1117–1134.
8. Fisher B, Bauer M, Margolese R *et al*. Five-year results of a randomized clinical trial comparing total mastectomy and segmental mastectomy with or without radiation in the treatment of breast cancer. *N Eng J Med* 1985; **312**: 665–673.
9. Steering Committee on Clinical Practice Guidelines for the Care and Treatment of Breast Cancer. Mastectomy or lumpectomy? The choice of operation for clinical stages I and II breast cancer. *Can Med Assoc J* 1998; **158** (Suppl. 3): S15–S21.
10. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J Am Stat Assoc* 1995; **90**: 443–450.
11. Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica* 1997; **65**: 557–586.
12. Farrow DC, Hunt WC, Samet JM. Geographic variation in the treatment of localized breast cancer. *NEJM* 1992; **326**: 1097–1101.
13. Nattinger AB, Gottlieb MS, Hoffman RG *et al*. Minimal increase in the use of breast conserving surgery from 1986–1990. *Med Care* 1996; **34**: 479–489.
14. Nattinger AB, Gottlieb MS, Veum J *et al*. Geographic variation in the use of breast-conserving treatment for breast cancer. *N Eng J Med* 1992; **326**: 1102–1107.
15. Committee on Ways and Means, U.S. House of Representatives. 1993. Medicare reimbursement to physicians. *1993 Green Book*. U.S. Government Printing Office: Washington, 407ff.
16. Kelejian H. Two-stage least squares and econometric systems linear in parameters but nonlinear in the endogenous variables. *J Am Stat Assoc* 1971; **66**: 373–374.
17. Greene WH. *Econometric Analysis*. Macmillan Publishing Co.: New York, 1997; 638–639.
18. Davidson R, MacKinnon JG. *Estimation and Inference in Econometrics*. Oxford University Press: New York, 1993; 236–242.
19. Hadley J, Mandelblatt JS, Mitchell JM *et al*. The effects of Medicare breast surgery fees on the treatment received by older women with localized breast cancer. Working Paper IWP 99-115. Institute for Health Care Research and Policy, Georgetown University, 1999.
20. McClellan M, Noguchi H. The validity of Instrumental variable estimates. Unpublished manuscript.

National Bureau of Economic Research: Cambridge, MA, 2000.

21. Fisher B, Anderson S, Redmond C *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 Eng J Med* 1995; **333**: 1456–1461.
22. Fallowfield LJ, Hall A, Maguire P *et al.* Psychological effects of being offered a choice of surgery for breast cancer. *Br Med J* 1994; **309**: 448.
23. Polsky D, Keating NL, Weeks JC *et al.* The utility of choice: options for breast cancer treatment. *Mimeograph*. University of Pennsylvania, 2001.
24. McClellan M, Newhouse J. Overview of the special supplement issue. *Health Serv Res* 2000; **35**: 1061–1070.

