

Can Observational Data Replace Randomized Trials?

Information regarding the retraction of the accompanying article (doi:10.1200/JCO.2015.62.0294) appears on page 3358

TO THE EDITOR: The increasing complexity and cost of conducting randomized trials¹ have stimulated interest in using observational data sets to evaluate cancer treatments.²⁻⁴ Determining the causal effect of treatments from observational data is, however, challenging because more aggressive treatments are selectively prescribed for patients with adverse disease characteristics or favorable comorbidity profiles. Associations may, therefore, arise between treatments and outcomes that are the result of confounding and are not causal. Relatively little is known about the extent of such confounding or the degree to which it can be removed through stratification by prognostic variables. Breast cancer is one of the commonest conditions for which radiotherapy is prescribed. We have therefore used it to examine this issue.

We analyzed data on women registered between 1990 and 2008 in the SEER public-use data set.⁵ Women were excluded if they were younger than 20 or older than 80 years when diagnosed, had previous cancer, unknown cancer laterality, bilateral cancer, or unknown radiotherapy status. Each woman entered the study on the date of her breast cancer diagnosis and left on the earliest of the following events: death, loss to follow-up, turning age 85 years, or January 1, 2009. Two analyses were conducted. In the first, deaths and person-years were stratified by five basic variables; the second stratification also included all available prognostic variables. Mortality ratios were estimated by maximum likelihood using Poisson regression. Calculations were performed using STATA version 12 (STATA, College Station, TX). Information was also collated from the Early Breast Cancer Trialists' Collaborative Group (EBCTCG) meta-analyses of randomized trials of radiotherapy versus not.⁶⁻⁸ Analyses were stratified by trial, individual follow-up year, age at randomization, and nodal status.

After breast-conserving surgery, with only basic variables in the stratification, radiotherapy was associated with reduced mortality for all the causes examined (Appendix Table A1, online only). Stratifying for all available prognostic variables changed the estimates, but all remained significantly below one. Notably, the rate ratio for mortality from all causes except breast cancer (0.69) was lower than that for breast cancer (0.74). After mastectomy in node-positive disease, with only basic variables in the stratification, radiotherapy was associated with increased breast cancer mortality and all-cause mortality, and decreased mortality from all causes except breast cancer. Stratifying for prognostic variables changed the death rate ratios for breast cancer from 1.35 to 0.89 and all causes of death from 1.21 to 0.85, but the rate ratio for mortality from all causes except breast cancer (0.74) was still lower than that for breast cancer (0.89).

Estimates of the effects of radiotherapy in the EBCTCG meta-analyses of randomized trials were then compared with

the corresponding death rate ratios in the SEER data stratified by all available variables (basic and prognostic; Fig 1). After breast-conserving surgery, the breast cancer death rate ratio in the EBCTCG data was 0.83 for all women and did not differ significantly between pN0 (node-negative disease) and pN+ (node-positive disease; $P = .76$), whereas in the SEER data the corresponding death rate ratio for all women was lower (0.74 v 0.83 ; $P = .04$), and differed significantly between pN0 and pN+ (0.84 v 0.73 ; $P = .008$). After mastectomy, the breast cancer death rate ratio in the EBCTCG data was 0.85, similar to that after breast-conserving surgery, and it did not differ significantly according to the number of positive nodes (P for heterogeneity = .53). In contrast, in the SEER data, there was substantial heterogeneity in the breast cancer death rate ratios according to the number of positive nodes (P for heterogeneity < .001).

In women with one to three positive nodes, postmastectomy radiotherapy was associated with significantly increased breast cancer mortality in the SEER data (1.10; 95% CI, 1.02 to 1.18) and significantly decreased breast cancer mortality in the EBCTCG data (0.80; 95% CI, 0.67 to 0.95). For mortality from all causes except breast cancer, there were also major qualitative differences between the EBCTCG and SEER data. In the EBCTCG data, radiotherapy was associated with significantly higher rates of mortality from all causes except breast cancer, including from heart disease and from lung cancer. In the SEER data, radiotherapy was associated with significantly lower rates of mortality from all causes except breast cancer, including from heart disease.

SEER is one of the largest, most detailed data sets. If we had used these SEER analyses to draw conclusions about the causal effects of radiotherapy, we would have concluded that radiotherapy after breast-conserving surgery is more effective in node-positive than in node-negative disease, and that radiotherapy after mastectomy in women with one to three positive nodes causes death from breast cancer. We would have also concluded that radiotherapy prevents mortality from all causes except breast cancer, including from heart disease and from accidents and violence (Appendix Table A1). These results contradict those of the randomized trials. We conclude, as have others,⁹ that nonrandomized comparisons are liable to provide misleading estimates of treatment effects. Therefore, they need careful justification every time they are used.

Paul McGale, David Cutter, Sarah C. Darby, and Katherine E. Henson

University of Oxford, Oxford, United Kingdom

Reshma Jagsi

University of Michigan, Ann Arbor, MI

Carolyn W. Taylor

University of Oxford, Oxford, United Kingdom

ACKNOWLEDGMENT

Supported by Cancer Research UK (grant no. C8225/A21133) as well as core funding from Cancer Research UK, the UK Medical Research Council, and the British Heart Foundation to the Oxford University Clinical Trial Service Unit.

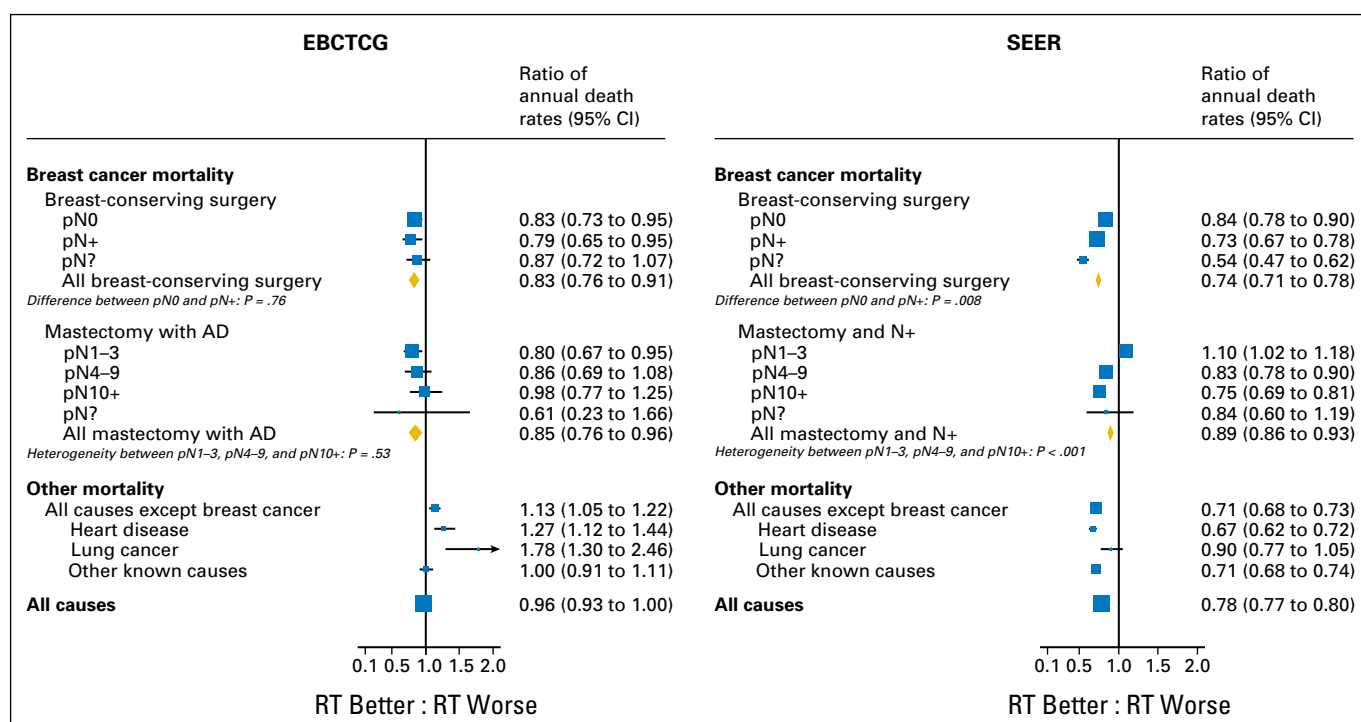


Fig 1. Ratios of annual death rates in women randomly assigned to receive radiotherapy (RT) or not in the Early Breast Cancer Trialists' Collaborative Group (EBCTCG) meta-analyses (left panel) and in women recorded as receiving RT or not in the SEER cancer registries (right panel). For the EBCTCG data, numbers of women randomized are as follows: breast-conserving surgery, $n = 10,801$ in 17 trials; mastectomy and axillary dissection (AD), $n = 3,131$ in 14 trials; other mortality, $n = 42,080$ in 78 trials. Stratification is by trial, individual year of follow-up, age at randomization (< 40 , 40-49, 50-59, 60-69, 70+ years), and nodal status (breast conserving surgery [BCS]: negative [pN0], positive [pN+], unknown [pN?]; mastectomy: one to three positive nodes [pN1-3], four to nine positive nodes [pN4-9], 10 or more positive nodes [pN10+], positive but number of positive nodes unknown [pN?]). For the SEER data, numbers of women are as follows: breast-conserving surgery ($n = 186,571$ [RT]; $n = 54,146$ [no RT]); mastectomy and node-positive ($n = 32,664$ [RT]; $55,854$ [no RT]); stratification is by both basic and prognostic variables (definitions are in Appendix Table A1).

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at www.jco.org.

REFERENCES

1. Institute of Medicine. A national cancer clinical trials system for the 21st century: Reinvigorating the NCI Cooperative Group Program. The National Academies Press, Washington, DC, 2010
2. Meyer A-M, Wheeler SB. Weinberger system for the 21st century: Reinvigorating the NCI Cooperative Group Program. The National Academies Press, Washington, DC, 2010
3. Korn EL, Freidlin B: Methodology for comparative effectiveness research: Potential and limitations. *J Clin Oncol* 30:4185-4187, 2012
4. Lyman GH, Levine M: Comparative effectiveness research in oncology: An overview. *J Clin Oncol* 30:4181-4184, 2012
5. National Cancer Institute DCCPS Surveillance Research Program Cancer Statistics Branch: (released January 2011). Surveillance, Epidemiology

and End Results (SEER) Program [www.seer.cancer.gov Research Data (1973-2008)]

6. Early Breast Cancer Trialists' Collaborative Group: Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: An overview of the randomised trials. *Lancet* 366: 2087-2106, 2005
7. Early Breast Cancer Trialists' Collaborative Group: Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: Meta-analysis of individual patient data for 10,801 women in 17 randomised trials. *Lancet* 378:1707-1716, 2011
8. Early Breast Cancer Trialists' Collaborative Group (EBCTCG): Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: Meta-analysis of individual patient data for 8135 women in 22 randomized trials. *Lancet* 383:2127-2135, 2014
9. Giordano SH, Kuo YF, Duan Z, et al: Limits of observational data in determining outcomes from cancer therapy. *Cancer* 112:2456-2466, 2008

DOI: 10.1200/JCO.2016.68.8879; published online ahead of print at www.jco.org on July 25, 2016.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Can Observational Data Replace Randomized Trials?

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or jco.ascopubs.org/site/ifc.

Paul McGale

No relationship to disclose

David Cutter

No relationship to disclose

Sarah C. Darby

No relationship to disclose

Katherine E. Henson

Employment: British United Provident Association

Reshma Jagsi

Employment: University of Michigan

Honoraria: International Journal of Radiation Oncology Biology Physics, Institute for Medical Education

Consulting or Advisory Role: Eviti, Baptist Health

Research Funding: AbbVie (Inst)

Carolyn W. Taylor

Expert Testimony: Crown Prosecution Service

Appendix**Table A1.** Death Rate Ratios in Women Recorded as Receiving Radiotherapy Versus Other Women (no radiotherapy) in the SEER Cancer Registries, 1990 to 2008

Cause of Death	Basic Variables Only*	Basic Variables Plus Prognostic Variables†
Breast-conserving surgery (n = 186,571 RT; n = 54,146 no RT)		
Breast cancer	0.61 (0.59-0.64)	0.74 (0.71 to 0.78)
All causes except breast cancer	0.62 (0.60 to 0.64)	0.69 (0.66 to 0.71)
Heart disease	0.54 (0.51 to 0.58)	0.62 (0.57 to 0.67)
Lung cancer‡	0.82 (0.71-0.96)	0.81 (0.68 to 0.97)
Other cancers	0.77 (0.72-0.83)	0.83 (0.76 to 0.91)
Other diseases	0.58 (0.55 to 0.61)	0.65 (0.61 to 0.69)
Accidents and violence	0.71 (0.58 to 0.88)	0.67 (0.52 to 0.86)
All causes of death	0.61 (0.60-0.63)	0.70 (0.68 to 0.73)
Mastectomy and node positive (n = 32,664 RT; n = 55,854 no RT)		
Breast cancer	1.35 (1.31 to 1.39)	0.89 (0.86 to 0.93)
All causes except breast cancer	0.89 (0.85 to 0.94)	0.74 (0.68 to 0.80)
Heart disease	0.87 (0.78 to 0.97)	0.77 (0.66 to 0.89)
Lung cancer‡	0.83 (0.62 to 1.11)	1.07 (0.74 to 1.55)
Other cancers	1.16 (1.04 to 1.29)	0.93 (0.80 to 1.08)
Other diseases	0.79 (0.72 to 0.85)	0.61 (0.55 to 0.69)
Accidents and violence	0.92 (0.68 to 1.24)	0.69 (0.46 to 1.04)
All causes of death	1.21 (1.17 to 1.24)	0.85 (0.81 to 0.88)

NOTE. Data are ratios of annual death rates (95% CIs) in women recorded as receiving RT and in women for whom RT was not recorded (no RT).

Abbreviation: RT, radiotherapy.

*Basic variables: age at breast cancer diagnosis (< 40, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79 years), calendar year of breast cancer diagnosis (1990-1992, 1993-1997, 1998-2002, 2003-2008), years since breast cancer diagnosis (0-4, 5-9, 10-14, 15-19, 20+), stage (localized, regional), and ethnicity (black, white, other/unknown).

†Prognostic variables: tumor size (T1 [< 2 cm], T2 [2-5 cm], T3 [> 5 cm], unknown), number of involved nodes (up to 2003: exact number of nodes examined by the pathologist and found to be positive [0, 1-3, 4-9, 10+, unknown or not specified, not applicable, not stated]; from 2004 onward: American Joint Committee on Cancer nodal stage [N0, zero positive nodes; N1, one to three positive nodes; N2, four to nine positive nodes; N3, 10+ positive nodes; not applicable; unknown]), grade (low, intermediate, high, unknown), estrogen receptor status (positive, negative, unknown), quadrant (inner, outer, other), axillary clearance (yes, no, other or unknown). These variables were recorded only from 1990.

‡Includes only deaths in women with microscopically confirmed lung cancer, identified by cross-matching of recorded deaths to SEER data file of lung cancer registrations.