# Limits of Observational Data in Determining Outcomes From Cancer Therapy

**Sharon H. Giordano, MD, MPH**[1], **Yong-Fang Kuo, PhD**[2], **Zhigang Duan, BMed, MS**[1], **Gabriel N. Hortobagyi, MD**[1], **Jean Freeman, PhD**[1], and **James S. Goodwin, MD**[2]

[1]Department of Breast Medical Oncology, TheUniversity of Texas M. D. Anderson Cancer Center, Houston, Texas

[2]Sealy Center on Aging, Department of InternalMedicine, The University of Texas Medical Branchat Galveston, Galveston, Texas

## Abstract

**Background**—Observational data are used increasingly to assess the effectiveness of therapies. However, selection biases are likely to have an impact on results and threaten the validity of these studies.

**Methods**—The primary objective of the current study was to explore the effect of selection biases in observational studies of treatment effectiveness in cancer care. Patients were identified from the Surveillance, Epidemiology, and End Results-Medicare linked database. The following groups of patients were included: 5245 men treated with and without androgen deprivation for locally advanced prostate cancer, 43,847 men with active treatment versus observation for low- and intermediate-risk prostate cancer, and 4860 patients with lymph node-positive colon cancer who were treated with and without fluorouracil chemotherapy. Patients were compared by therapy for the outcomes of cancer-specific mortality, othercause mortality, and overall mortality.

**Results**—In all comparisons, the observational data produced improbable results. For example, when evaluating outcomes of men who were treated with and without androgen deprivation for locally advanced prostate cancer, men who underwent androgen deprivation had higher prostate cancer mortality (hazard ratio, 1.5; 95% confidence interval, 1.29–1.92) despite clinical trial evidence that this treatment improves cancer mortality. Controlling for comorbidity, extent of disease, and other characteristics by multivariate analyses or by propensity analyses had remarkably small impact on these improbable results.

**Conclusions**—The current results suggested that the results from observational studies of treatment outcomes should be viewed with caution.

### Keywords

androgen deprivation; goserelin; observational studies; prostate cancer mortality; selection bias; Surveillance; Epidemiology; and End Results

There has been a growing interest in using observational data to study cancer outcomes. This interest is driven in part by the availability of population-based data—in particular, data from the Surveillance, Epidemiology, and End Results (SEER) Tumor Registry that have been merged with Medicare charge data.[1] These databases have the advantages of excellent

external validity, and they allow for the study of populations that often are not included in clinical trials, such as the elderly, minorities, and patients with higher burdens of comorbidities. In addition, large administrative databases can provide information on patterns of care and treatment compliance[2–4]; can detect rare toxicities and assess treatment toxicities in representative, population-based cohorts[5–9]; and can permit the comparison of toxicities across different patient populations.[5,8–10] However, more recently, administrative datasets are being used to compare the effects of different treatments on overall survival. This approach has been used across many tumor types, including breast, lung, colon, rectal, prostate, and ovarian cancers.[11–19]

Selection biases, particularly confounding by indication, are the primary threat to the validity of using observational data to estimate benefits of therapies.[20,21] These biases can operate in several ways. For example, in a comparison between therapies where 1 therapy is considered potentially more efficacious (eg, adjuvant chemotherapy vs no chemotherapy), a bias may be expected whereby patients with poorer prognosis cancers would be more likely to receive that therapy. Alternatively, in a comparison involving potentially more toxic treatments versus less toxic treatments (eg, invasive surgery vs radiation treatment or chemotherapy vs no chemotherapy), a selection bias may be expected whereby patients with better underlying health—those considered more likely to tolerate the treatment—would be more likely to receive the more toxic therapy. Investigators clearly are aware of these potential biases and use statistical techniques to address them. Multivariate analyses, stratification, matching, restricting, and propensity analyses often are used adjusting for information available in the datasets, such as age; ethnicity; neighborhood socioeconomic level; and prior diagnoses, procedures, and hospitalizations.[11–13,22,23] Nevertheless, unmeasured confounders are likely to persist.

In this article, we explore the strong effects of selection biases in observational studies. We hypothesized that we would obtain results from observational analyses that were implausible when considered in the light of published data from clinical trials. We also hypothesized that the usual means of dealing with selection biases, such as controlling for patient and tumor characteristics using multivariate and propensity analyses, would not eliminate the improbable results. We present several examples, including reanalyses of previously published data, to illustrate the effects of common selection biases. For the first example, we selected a situation in which we believed that selection biases might produce implausible results compared with results from a randomized controlled trial. For the second and third examples, we reanalyzed previously published data. In all cases, we examined cancer mortality, noncancer mortality, and overall mortality. We reasoned that any real benefit of cancer therapy could be manifested only through differences in *cancer-specific* mortality but that selection biases might result in differences in noncancer mortality that would be as great or greater than the differences in cancer mortality calling into question the reliability of using mortality endpoints to assess treatment efficacy in nonrandomized data.

## Materials and Methods

### Data Source

Our approach was similar in all cases and was analogous to approaches that have been used by previous investigators,[11–17,24–27] in that we compared the effectiveness of different cancer therapies on outcomes. We used the merged SEER-Medicare database as our data source, which also has been used in prior outcome studies.,[11–17,24–27] The SEER Program is a national population-based tumor registry run by the National Cancer Institute that collects information on incident cancer cases. Patients in the SEER database who are eligible for Medicare have been linked to their Medicare records.[1]

**Example 1: Androgen-deprivation therapy versus none after primary radiation therapy for locally advanced prostate cancer—**This analysis included 5245 men with locally advanced prostate cancer (either tumor [T] classified as T2/T3 with a Gleason score of 8–10 or tumor classified as T4) in the SEER-Medicare database who were recipients of primary radiation therapy, aged 66 years, and diagnosed between 1992 and 1999. Patients were excluded if they had primary surgical therapy, if they had health maintenance organization (HMO) coverage, or if they were not enrolled in Medicare Parts A and B for the 12 months before to the 6 months after their cancer diagnosis. Prostatectomy, radiation therapy, and androgen deprivation were defined as described previously.[10] Briefly, radical prostatectomy was defined from SEER coding on site-specific surgery or any of the following codes from Medicare claims: Current Procedural Terminology (CPT) codes 55,810, 55,812, 55,815, 55,801, 55,821, 55,831, 55,842, 55,845; or International Classification of Diseases, ninth revision (ICD-09) procedure code 60.5. Radiation therapy was identified from SEER coding on site-specific surgery or any of the following codes from Medicare claims: CPT codes 77,401 through 77,499 and codes 77,750 through 77,799; and ICD-09 codes 92.21 through 92.29, V58.0, V66.1, and V67.1. Androgen deprivation was defined as either orchiectomy or at least 1 claim for a gonadotropin-releasing hormone (GNRH) agonist within 6 months of diagnosis. GNRH codes are any of the following Healthcare Common Procedure Coding System (HCPCS) codes: J9202, J1950, J9217, J9218, and J9219. Comorbidity scores were calculated by using Klabunde's adaptation of the Charlson comorbidity index.[28,29]

A series of Cox proportional-hazards models were developed that incorporated increasing numbers of covariates. Men who underwent androgen deprivation were compared with men who did not receive androgen deprivation for the outcomes of mortality from prostate cancer, other-cause mortality, and all cause mortality. We adjusted for the following variables: T classification, histologic grade (low, moderate, poorly differentiated; unknown), year of diagnosis, age (continuous), comorbidity (0, 1–2, 3), ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, other), SEER region, census tract education (percent of individuals living in a given census tract with <12 years of education, divided into quartiles), census tract poverty (percent of individuals living in a given census tract living below the poverty level, divided into quartiles), number of claims for prostate-specific antigen measurements in the 12 months before diagnosis (continuous), and number of provider visits in the 12 months before diagnosis (continuous). Missing values were coded as unknown and were included in the analyses. Follow-up was through December 31, 2000.

The propensity that a patient would receive adjuvant androgen deprivation was generated from the logistic regression model that incorporated the potential confounding factors listed in Table 1.[13,30–32] Then, we grouped the patients into 5 strata representing quintiles of the propensity score. The Cochran Mantel-Haenszel chi-square test was used to determine whether the covariates were balanced after adjusting for propensity quintiles. The covariates that retained a significant difference between the treated and untreated groups were adjusted together with propensity scores in the Cox proportionalhazards model. We also noted the association between treatment and mortality within each stratum of propensity quintiles.

**Example 2: Active treatment versus observation for men with localized prostate cancer—**For this example, we reanalyzed previously published data.[19] The original study compared survival between men who were treated actively for prostate cancer (surgery or radiation) with men who were observed. We replicated the methods in that study as described below. In brief, the study population included men between ages 65 years and 80 years with an incident prostate cancer diagnosed between 1991 and 1999 in the SEER-Medicare database. Men who had moderately to well differentiated, nonmetastatic T1 or T2 tumors were included. Men were excluded if they were diagnosed at autopsy or death, if

they had Medicare entitlement based on endstage renal disease, or if they died within 1 year of diagnosis. Patients were excluded if they had HMO coverage or if they were not enrolled in Medicare Parts A and B from the 3 months before to the 6 months after their cancer diagnosis. Patients were considered to have received active treatment if they received external-beam radiation therapy, had radiation implants, or underwent radical prostatectomy. Cox models were developed to compare outcomes of men with active treatment versus observation. The final models adjusted for year of diagnosis, age, race, urban residence, marital status, income, education, SEER region, tumor size, tumor grade, and patient comorbidity, as described in detail in the original study.

We replicated the previously published analyses and added the following additional analyses: In addition to the endpoint of overall survival, we developed Cox models for the endpoints of prostate cancer survival and other-cause mortality. We plotted survival curves from stratified Cox proportional-hazards models, adjusting for age, comorbidity, SEER region, and year of diagnosis. These survival curves are presented for type of therapy (radical prostatectomy, radiation, or observation) and for a noncancer control population. The noncancer control population was selected as follows: From the 5% sample of Medicare beneficiaries who did not have any cancer in the SEER-Medicare data, we selected men ages 65 years to 80 years who were resident in a SEER area from 1991 through 1999. Noncancer controls were assigned randomly to match the distribution of year of diagnosis for the cancer cohort. If men were not enrolled in Medicare Parts A and B, if they were enrolled in an HMO from 3 months before study entry to 6 months after study entry, or if the died within 1 year of study entry, then they were excluded. From these noncancer controls, a cohort of 43,847 men was built to match the age distribution of the patients with prostate cancer. Among these, 12,234 men died during follow-up. We also developed Cox models with the endpoints of mortality from heart disease, other cancers, cardiovascular disease, chronic obstructive pulmonary disease, pneumonia, diabetes, accident, other infections, and dementia to illustrate the effects of confounding.

**Example 3—**In our third example, we also reanalyzed previously published data.[13] The original study evaluated survival associated with 5-fluorouracil (5-FU)-based adjuvant chemotherapy among elderly patients with lymph node-positive colon cancer. We replicated the methods as described in the original report.[13] Patients were included who met the following criteria: first diagnosis of primary colon cancer between 1992 and 1996, aged 65 years, and stage III disease. Men were excluded if they were enrolled in an HMO or if they were not covered by Medicare Parts A and B from 12 months before diagnosis until 16 months after diagnosis. Adjuvant chemotherapy with 5-FU was identified by claims with an HCPCS J-code of J9190 within 120 days of diagnosis. We constructed Cox models to estimate both overall survival, as reported previously, and other-cause and colon cancer-specific survival for patients who did and did not receive adjuvant chemotherapy. Cox models were adjusted for year of diagnosis, age, sex, urban residence, SEER region, lymph nodes, tumor grade, extent of disease, comorbidity, and propensity score.

## Results

### Androgen Deprivation for Prostate Cancer

In our first analysis, we evaluated the outcomes of men who did or did not receive androgen deprivation after primary radiation therapy for locally advanced (stage III) prostate cancer. Randomized clinical trial data have demonstrated a survival benefit for androgen deprivation in this population.[33–35] Patients who underwent androgen deprivation had higher grade tumors, were more educated, had more frequent physician visits, and were diagnosed more recently than patients who did not receive androgen deprivation (*P* <.0001 for each) (Table

1). After adjusting for propensity score, imbalances remained in the year of diagnosis and in the number of provider visits before diagnosis. Then, we performed a series of Cox models that incorporated increasing numbers of covariates. The results of those survival analyses are shown in Table 2. In the unadjusted analysis, men who underwent androgen deprivation had a *higher* risk of death from prostate cancer (hazard ratio [HR], 1.35; 95% confidence interval [95% CI], 1.11–1.64). After adjusting for all measurable confounders, a persistent effect of higher prostate cancer mortality was observed among men who underwent androgen deprivation (HR, 1.63; 95% CI, 1.32–2.01). There was no significant difference observed in other-cause mortality between men who did and men who did not undergo androgen deprivation.

Next, we conducted a propensity analysis to adjust for unmeasured confounders.[30–32] Propensity scores are an individual patient's likelihood of receiving a treatment calculated from a logistic regression model that is based on their covariate information. We show the results first with the propensity score as a covariate in our model and then with the results stratified into quintiles based on propensity score. In each analysis, prostate cancer mortality consistently was higher among men underwent androgen deprivation. In the model that incorporated propensity score along with the imbalanced covariates, the HR for prostate cancer mortality was 1.65 (95% CI, 1.33– 2.03).

### Active Therapy Versus Observation for Localized Prostate Cancer

In the second example, we compared the outcomes of men who received active therapy versus men who were observed for localized prostate cancer. We replicated the methods in the original study and were able to reproduce the cohort of patients and point estimates of survival.[19] Then, we expanded on the previously published results by performing Cox analyses that demonstrated mortality from prostate cancer and mortality from other causes in addition to all-cause mortality. Table 3 presents the results of the survival analyses. In the unadjusted and adjusted analyses, patients who received active therapy had significantly lower all-cause mortality (adjusted HR, 0.68; 95% CI, 0.65–0.70) compared with patients on observation from prostate cancer, as reported in the original study. However, patients who received active therapy had had lower mortality from all other causes (unadjusted HR, 0.52; 95% CI, 0.50–0.54; adjusted HR, 0.68; 95% CI, 0.65–0.71). The confounding between overall health status and active therapy for prostate cancer also is demonstrated in Figure 1, which shows that patients who underwent radical prostatectomy for prostate cancer actually had better survival than a control population *without* cancer. To explore the association further between active therapy and cause of death, we performed Cox models for other individual causes of death. The HRs with 95% CIs are plotted in Figure 2. For each individual cause of death, such as diabetes or pneumonia, active treatment for prostate cancer was associated with a significant mortality benefit, similar to the benefit observed for overall mortality or mortality from prostate cancer.

### Chemotherapy for Lymph Node-positive Colon Cancer

In our last example, we also reanalyzed data that were published previously in an analysis of outcomes for patients with lymph node-positive colon cancer.[13] The original study reported that patients who received fluorouracil-based chemotherapy had a significantly lower hazard of death (HR, 0.66; 95% CI, 0.60–0.73) than patients who did not receive chemotherapy. We extended on the previous work and performed Cox regression models to estimate colon cancer-specific mortality and other-cause mortality in addition to the previously published overall mortality. We reasoned that, if the lower hazard of death was because of treatment alone, then deaths from other causes would not be related to chemotherapy use. The results are shown in Table 4. We observed a strong association between fluorouracil-based chemotherapy and other-cause mortality (HR, 0.48; 95% CI, 0.41–0.56). For colon cancer

mortality, fluorouracil chemotherapy also was associated with a survival benefit (HR, 0.80; 95% CI, 0.72–0.89), although the effect was not as strong as that for overall survival.

## Discussion

In the current study, we selected several examples in which we had a priori reasons to suspect that selection biases would influence outcomes. We believed that there may have been strong selection biases both on extent and aggressiveness of the tumor and on the underlying health of the patient. We reasoned that the bias for patients with more aggressive cancer to receive more therapy would result in an implausibly worse survival among more extensively treated patients. The selection bias favoring the treatment of healthier patients would result in improved survival among treated patients. These biases could work in isolation or could be present simultaneously. Because these biases could have opposite effects on survival and, thus, tend to cancel each other out, we segregated survival by measuring 3 types of mortality: allcause mortality, mortality from cancer, and mortality from all causes other than the cancer. This allowed us to estimate the impact of the 2 proposed selection biases. Selection biases for poorer prognosis tumors would be reflected best in cancer-specific mortality, whereas biases involving the selection of healthier patients would be reflected in mortality from all other causes.

In the comparison of men with prostate cancer who received either active therapy or observation, we hypothesized that healthier men who were diagnosed with prostate cancer would be more likely to receive active treatment, which would result in improvements in both all-cause survival and in deaths from causes other than prostate cancer. We did demonstrate a large effect of active therapy on deaths from all causes. In fact, active therapy for prostate cancer had at least as much effect on deaths from diseases like pneumonia and cardiovascular disease as it did on deaths from prostate cancer. It is important to consider how odd these results actually are. It is not plausible that prostate cancer therapy improves survival from causes other than prostate cancer. The most likely explanation is that selection biases are responsible for the effects observed. More noteworthy, these biases persist after statistical adjustment for all measured confounders. It is also possible that active therapy is a marker for overall quality of care.

Two potential explanations for why controlling for reported comorbidity does not adequately control for selection biases are the lack of information on functional status and the lack of information on selfreported health. Measures of functional status, such as the Activities of Daily Living score, the Karnofsky performance status scale, or the Barthel index, independently can predict future physical function, morbidity, and mortality.[36–38] Self-rated health, which typically is assessed by using a 4- or 5-point scale from excellent to poor, also has been demonstrated as a strong predictor of survival in several observational studies.[39–42] Most relevant to our current st0075dies, self-rated health remains a strong predictor of survival even after controlling for comorbidity and all other measurable factors that may affect survival. The best example is the Cardiovascular Health Study, which included a rich variety of clinical information from physical examination, laboratory assessments, and noninvasive testing, such as cardiac ejection fraction.[43] Self-rated health still was a strong, independent predictor of survival. This means that there is information available to individual patients about their health that is not captured even with extensive medical assessment and yet is reflected in a simple, subjective health assessment. The information reflected by patients' self-rated health also presumably is accessible to the clinicians advising them if the physicians inquire. That information could guide treatment decisions, and patients who have more robust underlying health may be more likely to choose more invasive and more extensive treatments. Given the effect of competing risks on outcomes of treatment, such decision making may be entirely appropriate.[44]

A similar line of reasoning can be used to explain the inability to completely control for the selection biases whereby those with more aggressive tumors tend to receive more extensive treatments (ie, confounding by indication). The information reported in SEER on tumor characteristics is extensive and includes tumor stage, size, histologic type, histologic grade, and the number of positive lymph nodes. However, it would be naive to assume that experienced clinicians would not make more subtle distinctions in tumor prognosis than could be made based only on the information found in SEER. The example of androgen deprivation for the treatment of prostate cancer illustrates how difficult it can be to control for tumor aggressiveness. We observed that, whereas clinical trials have demonstrated a survival benefit of androgen deprivation, our observational data indicate that androgen deprivation is associated with worse tumor-specific survival.[17–19] Presumably, the finding that men with more aggressive tumors were more likely to receive androgen deprivation could not be captured entirely by the extent-of-disease characteristics available in the SEER data.

Just as adjusting for comorbidity and tumor characteristics did not completely remove selection biases, statistical adjustment using propensity scores did not substantially alter the findings. Some reports have suggested that propensity scores can eliminate up to 90% of the bias resulting from confounding covariates.[45–48] In our examples, adjustment for propensity score had little effect on the HRs. A recent study evaluated the effects of variable choice in propensity analyses.[49] The authors suggested that variables unrelated to the exposure but related to the outcome always should be included in propensity score analyses, because they will decrease the variance without increasing bias. It is not clear that the datasets currently used for observational studies of cancer treatment, such as the SEER-Medicare linked data, can furnish such a variable. None of the cancer treatment outcome studies that we reviewed included such a variable.[11–19] Other techniques, like instrumental variable analyses, take an analogous approach to minimize bias in observational studies. However, statistical techniques cannot eliminate all bias and confounding.

Our final example of fluorouracil-based chemotherapy for colon cancer is a less extreme and perhaps more representative situation. We note that our analyses lead to the same conclusion as the original study— that chemotherapy for lymph node-positive colon cancer is associated with improved survival. These observational analyses in older patients are consistent with data from randomized clinical trials in younger patients, which have demonstrated conclusively that 5-FU chemotherapy for colon cancer is associated with a 33% lower mortality rate.[50] Other well designed observational studies also have produced results that closely approximate data from randomized clinical trials.[51,52] However, in the colon cancer example, the strength of the association between chemotherapy and survival is strongest for noncancer deaths, which presumably are not being prevented with chemotherapy. Thus, our findings call into question the reliability of using overall survival as the primary endpoint. Our results with the other examples raise the possibility that this finding may have resulted from a chance alignment of confounders.

We have drawn several major conclusions. First, the results of observational studies that compare outcomes of different therapies should be viewed with some skepticism. Such publications may result from an interaction of selection biases with publication bias. Analyses that make sense are followed up and ultimately published, whereas analyses that produce implausible results, such as some of the examples presented here, are more likely to be discarded or rejected for publication if they are pursued.

Second, any analyses of observational data for treatment outcome, at a minimum, should attempt to segregate the outcome measurements into those that possibly may be caused by the treatments versus those that could not be caused by the treatments. Most prior

publications on cancer treatment outcomes only assessed all-cause mortality.[11–17] Examination of treatment effects on mortality from cancer versus other causes may produce clues for the presence of unmeasured selection biases. This would hold not just for cancer therapies but also for observational studies of outcomes from therapies for any condition. There are many clinical situations, particularly in the treatment of the elderly, in which data from clinical trials are nonexistent, and observational studies may be the only potential method to assess benefits of treatment. We suggest that disease-specific survival, other-cause survival, and overall survival all should be provided in any studies of treatment outcomes. Finally, the strong yet implausible treatment effects observed in our analyses should reinforce the caution and modesty of all investigators assessing outcomes from observational data.

## Acknowledgments

## References

1. Potosky AL, Riley GF, Lubitz JD, Mentnech RM, Kessler LG. Potential for cancer related health services research using a linked Medicare-tumor registry database. Med Care. 1993; 31:732–748. [PubMed: 8336512]

2. Du X, Freeman JL, Goodwin JS. The declining use of axillary dissection in patients with early stage breast cancer. Breast Cancer Res Treat. 1999; 53:137–144. [PubMed: 10326790]

3. Du X, Goodwin JS. Patterns of use of chemotherapy for *breast cancer* in older women: findings from Medicare claims data. J Clin Oncol. 2001; 19:1455–1461. [PubMed: 11230491]

4. Du X, Goodwin JS. Increase of chemotherapy use in older women with breast carcinoma from 1991 to 1996. Cancer. 2001; 92:730–737. [PubMed: 11550141]

5. Du XL, Osborne C, Goodwin JS. Population-based assessment of hospitalizations for toxicity from chemotherapy inolder women with breast cancer. J Clin Oncol. 2002; 20:4636–4642. [PubMed: 12488407]

6. Patt DA, Goodwin JS, Kuo YF, et al. Cardiac morbidity of adjuvant radiotherapy for breast cancer. J Clin Oncol. 2005; 23:7475–7482. see comment. [PubMed: 16157933]

7. Doyle JJ, Neugut AI, Jacobson JS, Grann VR, Hershman DL. Chemotherapy and cardiotoxicity in older breast cancer patients: a population-based study. J Clin Oncol. 2005; 23:8597–8605. [PubMed: 16314622]

8. Baxter NN, Tepper JE, Durham SB, Rothenberger DA, Virnig BA. Increased risk of rectal cancer after prostate radiation: a population-based study. Gastroenterology. 2005; 128:819–824. see comment. [PubMed: 15825064]

9. Baxter NN, Habermann EB, Tepper JE, Durham SB, Virnig BA. Risk of pelvic fractures in older women following pelvic irradiation. JAMA. 2005; 294:2587–2593. see comment. [PubMed: 16304072]

10. Shahinian VB, Kuo YF, Freeman JL, Goodwin JS. Risk of fracture after androgen deprivation for prostate cancer. N Engl J Med. 2005; 352:154–164. [PubMed: 15647578]

11. Schrag D, Rifas-Shiman S, Saltz L, Bach PB, Begg CB. Adjuvant chemotherapy use for Medicare beneficiaries with stage II colon cancer. J Clin Oncol. 2002; 20:3999–4005. [PubMed: 12351597]

12. Earle CC, Tsai JS, Gelber RD, Weinstein MC, Neumann PJ, Weeks JC. Effectiveness of chemotherapy for advanced lung cancer in the elderly: instrumental variable and propensity analysis. J Clin Oncol. 2001; 19:1064–1070. [PubMed: 11181670]

13. Sundararajan V, Mitra N, Jacobson JS, Grann VR, Heitjan DF, Neugut AI. Survival associated with 5-fluorouracil-based adjuvant chemotherapy among elderly patients with nodepositive colon cancer. Ann Intern Med. 2002; 136:349–357. summary for patients in *Ann Intern Med*. 2002;136:I19]. [PubMed: 11874307]

14. Neugut AI, Fleischauer AT, Sundararajan V, et al. Use of adjuvant chemotherapy and radiation therapy for rectal cancer among the elderly: a population-based study. J Clin Oncol. 2002; 20:2643–2650. [PubMed: 12039925]

15. Du XL, Jones DV, Zhang D. Effectiveness of adjuvant chemotherapy for node-positive operable breast cancer in older women. J Gerontol A Biol Sci Med Sci. 2005; 60:1137–1144. [PubMed: 16183952]

16. Hershman D, Fleischauer AT, Jacobson JS, Grann VR, Sundararajan V, Neugut AI. Patterns and outcomes of chemotherapy for elderly patients with stage II ovarian cancer: a population-based study. Gynecol Oncol. 2004; 92:293–299. [PubMed: 14751173]

17. Ramsey SD, Howlader N, Etzioni RD, Donato B. Chemotherapy use, outcomes, and costs for older persons with advanced nonsmall-cell lung cancer: evidence from Surveillance, Epidemiology and End Results-Medicare. J Clin Oncol. 2004; 22:4971–4978. [PubMed: 15611512]

18. Giordano SH, Duan Z, Kuo YF, Hortobagyi GN, Goodwin JS. Use and outcomes of adjuvant chemotherapy in older women with breast cancer. J Clin Oncol. 2006; 24:2750–2756. see comment. [PubMed: 16782915]

19. Wong YN, Mitra N, Hudes G, et al. Survival associated with treatment vs observation of localized prostate cancer in elderly men. JAMA. 2006; 296:2683–2693. [PubMed: 17164454]

20. Grimes DA, Schulz KF. Bias and causal associations in observational research. Lancet. 2002; 359:248–252. see comment. [PubMed: 11812579]

21. Wen SW, Hernandez R, Naylor CD. Pitfalls in nonrandomized outcomes studies. The case of incidental appendectomy with open cholecystectomy. JAMA. 1995; 274:1687–1691. [PubMed: 7474273]

22. Schrag D, Cramer LD, Bach PB, Begg CB. Age and adjuvant chemotherapy use after surgery for stage III colon cancer. J Natl Cancer Inst. 2001; 93:850–857. [PubMed: 11390534]

23. Sundararajan V, Hershman D, Grann VR, Jacobson JS, Neugut AI. Variations in the use of chemotherapy for elderly patients with advanced ovarian cancer: a population-based study. J Clin Oncol. 2002; 20:173–178. [PubMed: 11773167]

24. El-Serag HB, Siegel AB, Davila JA, et al. Treatment and outcomes of treating of hepatocellular carcinoma among Medicare recipients in the United States: a populationbased study. J Hepatology. 2006; 44:158–166. see comment.

25. Iwashyna TJ, Lamont EB. Effectiveness of adjuvant fluorouracil in clinical practice: a population-based cohort study of elderly patients with stage III colon cancer. J Clin Oncol. 2002; 20:3992–3998. see comment. [PubMed: 12351596]

26. Krzyzanowska MK, Weeks JC, Earle CC. Treatment of locally advanced pancreatic cancer in the real world: population-based practices and effectiveness. J Clin Oncol. 2003; 21:3409–3414. [PubMed: 12972517]

27. Urbach DR, Swanstrom LL, Hansen PD. The effect of laparoscopy on survival in pancreatic cancer. Arch Surg. 2002; 137:191–199. [PubMed: 11822959]

28. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987; 40:373–383. [PubMed: 3558716]

29. Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. J Clin Epidemiol. 1993; 46:1075–1079. discussion1081–1090. [PubMed: 8410092]

30. Rubin DB, Thomas N. Matching using estimated propensity scores: relating theory to practice. Biometrics. 1996; 52:249–264. [PubMed: 8934595]

31. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med. 1997; 127(8 pt 2):757–763. [PubMed: 9382394]

32. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a nonrandomized control group. Stat Med. 1998; 17:2265–2281. [PubMed: 9802183]

33. Bolla M, Gonzalez D, Warde P, et al. Improved survival in patients with locally advanced prostate cancer treated with radiotherapy and goserelin. N Engl J Med. 1997; 337:295–300. see comment. [PubMed: 9233866]

34. Pilepich MV, Caplan R, Byhardt RW, et al. Phase III trial of androgen suppression using goserelin in unfavorableprognosis carcinoma of the prostate treated with definitive radiotherapy: report of Radiation Therapy Oncology Group Protocol 85–31. J Clin Oncol. 1997; 15:1013–1021. [PubMed: 9060541]

35. Pilepich MV, Winter K, John MJ, et al. Phase III Radiation Therapy Oncology Group (RTOG) trial 86–10 of androgen deprivation adjuvant to definitive radiotherapy in locally advanced carcinoma of the prostate. Int J Radiat Oncol Biol Phys. 2001; 50:1243–1252. [PubMed: 11483335]

36. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Md State Med J. 1965; 14:61–65. [PubMed: 14258950]

37. Walter LC, Brand RJ, Counsell SR, et al. Development and validation of a prognostic index for 1-year mortality in older adults after hospitalization. JAMA. 2001; 285:2987–2994. see comment. [PubMed: 11410097]

38. Karnofsky, DA.; Burchenal, JH. The clinical evaluation of chemotherapeutic agents in cancer. In: Macleod, CM., editor. Evaluation of Chemotherapeutic Agents. New York, NY: Columbia University Press; 1949. p. 199-205.

39. Lee Y. The predictive value of self-assessed general, physical, and mental health on functional decline and mortality in older adults. J Epidemiol Commun Health. 2000; 54:123–129.

40. Mossey JM, Shapiro E. Self-rated health: a predictor of mortality among the elderly. Am J Public Health. 1982; 72:800–808. [PubMed: 7091475]

41. Vuorisalmi M, Lintonen T, Jylha M. Global self-rated health data from a longitudinal study predicted mortality better than comparative self-rated health in old age. J Clin Epidemiol. 2005; 58:680–687. [PubMed: 15939219]

42. Benyamini Y, Blumstein T, Lusky A, Modan B. Gender differences in the self-rated health-mortality association: is it poor self-rated health that predicts mortality or excellent self-rated health that predicts survival? Gerontologist. 2003; 43:396–405. discussion 372–375. [PubMed: 12810904]

43. Fried LP, Kronmal RA, Newman AB, et al. Risk factors for 5-year mortality in older adults: the Cardiovascular Health Study. JAMA. 1998; 279:585–592. see comment. [PubMed: 9486752]

44. Welch HG, Albertsen PC, Nease RF, Bubolz TA, Wasson JH. Estimating treatment benefits for the elderly: the effect of competing risks. Ann Intern Med. 1996; 124:577–584. [PubMed: 8597322]

45. Cochran WG. The effectiveness of adjustment by subclassi-fication in removing bias in observational studies. Biometrics. 1968; 24:295–313. [PubMed: 5683871]

46. Austin PC, Mamdani MM, Stukel TA, Anderson GM, Tu JV. The use of the propensity score for estimating treatment effects: administrative versus clinical data. Stat Med. 2005; 24:1563–1578. [PubMed: 15706581]

47. Sturmer T, Schneeweiss S, Brookhart MA, Rothman KJ, Avorn J, Glynn RJ. Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal anti-inflammatory drugs and short-term mortality in the elderly. Am J Epidemiol. 2005; 161:891–898. [PubMed: 15840622]

48. Reeves BC, van Binsbergen J, van Weel C. Systematic reviews incorporating evidence from nonrandomized study designs: reasons for caution when estimating health effects. Eur J Clin Nutr. 2005; 59(suppl 1):S155–S161. [PubMed: 16052184]

49. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. Am J Epidemiol. 2006; 163:1149–1156. [PubMed: 16624967]

50. Moertel CG, Fleming TR, Macdonald JS, et al. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. N Engl J Med. 1990; 322:352–358. see comment. [PubMed: 2300087]

51. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000; 342:1887–1892. see comment. [PubMed: 10861325]

52. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med. 2000; 342:1878–1886. [PubMed: 10861324]

**Figure 1.**
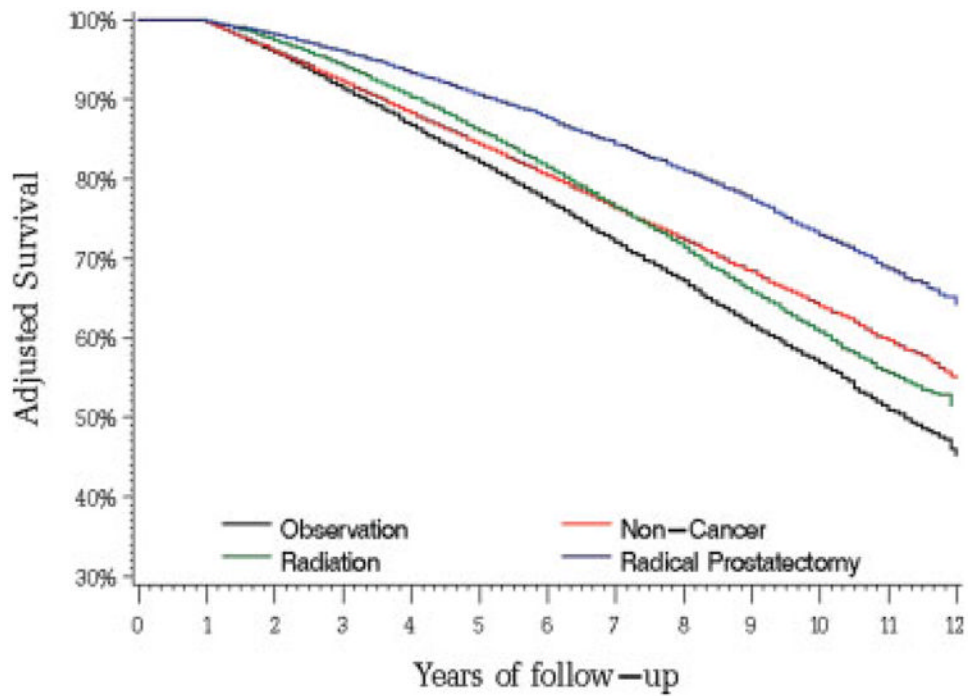Adjusted overall survival curves from Cox models for men with localized prostate cancer stratified by therapy and for a matched noncancer control population.
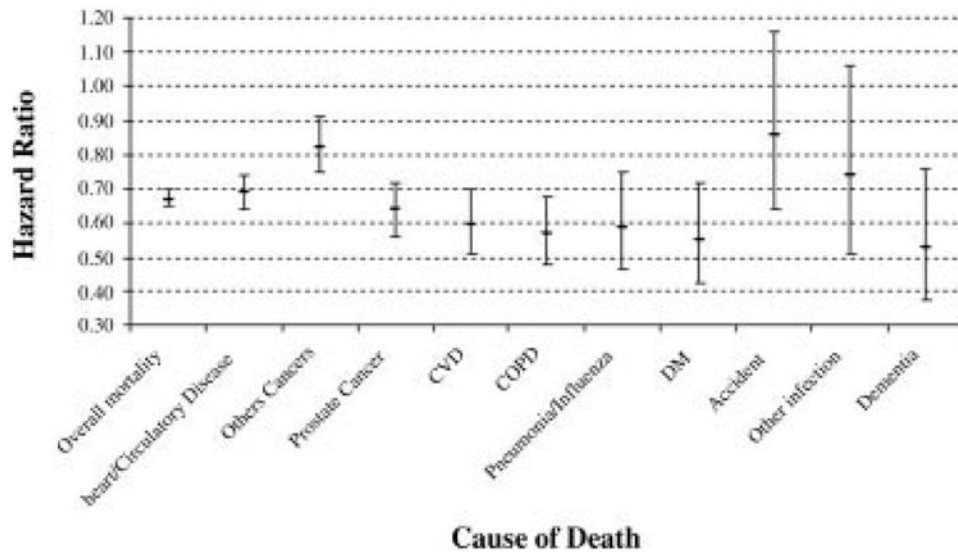
**Figure 2.**
Plot of hazard ratio for death comparing active therapy versus observation in men with localized prostate cancer. CVD indicates cardiovascular disease; COPD, chronic obstructive pulmonary disease; DM, diabetes mellitus.

**Table 1**

**Patient Characteristics**

| Characteristic | Adjuvant androgen deprivation versus none for locally advanced prostate cancer, % | | P | |
| | Yes, N = 1863 | No, N = 3382 | Before adjustment[*] | After adjusting for propensity score[†] |
| --- | --- | --- | --- | --- |
| Age, y | | | | |
| 66–69 | 20 | 20.3 | .8850 | .5313 |
| 70–74 | 37.3 | 37.7 | | |
| 75–79 | 31.3 | 31.3 | | |
| 80 | 11.4 | 10.7 | | |
| Clinical stage | | | | |
| Incidental, clinically/radiographically apparent tumor | 35.1 | 33.1 | .1521 | .7055 |
| Localized | 29 | 28.4 | | |
| Extension beyond prostate | 35.9 | 38.5 | | |
| Grade | | | | |
| Well differentiated | 1.1 | 2.8 | <.0001 | .6304 |
| Moderately differentiated | 17 | 22.3 | | |
| Poorly differentiated | 80.7 | 73.7 | | |
| Unknown | 1.2 | 1.2 | | |
| Race | | | | |
| White | 81.2 | 80.5 | <.0001 | .9256 |
| Black | 7.1 | 10.6 | | |
| Hispanic | 4.2 | 3.1 | | |
| Other/unknown | 7.5 | 5.8 | | |
| SEER region | | | | |
| San Francisco | 10.2 | 8.9 | <.0001 | .3755 |
| Connecticut | 13.7 | 15.3 | | |
| Michigan | 16.9 | 21.1 | | |
| Hawaii | 4.1 | 4.8 | | |
| Iowa | 10.7 | 12.2 | | |
| New Mexico | 2.6 | 4.1 | | |
| Seattle | 12.8 | 10.8 | | |
| Utah | 5.3 | 4.3 | | |
| Georgia | 3 | 4.7 | | |
| San Jose | 4.2 | 2.8 | | |
| Los Angeles | 16.5 | 11 | | |
| Year of diagnosis | | | | |
| 1992 | 8.2 | 27.9 | <.0001 | .0418 |
| 1993 | 7.4 | 21.7 | | |
| 1994 | 7.8 | 15.7 | | |

| Characteristic | Adjuvant androgen deprivation versus none for locally advanced prostate cancer, % | | P | |
| | Yes, N = 1863 | No, N = 3382 | Before adjustment[*] | After adjusting for propensity score[†] |
| --- | --- | --- | --- | --- |
| 1995 | 8.7 | 10.5 | | |
| 1996 | 13.5 | 7.5 | | |
| 1997 | 14.8 | 6.4 | | |
| 1998 | 18.4 | 5 | | |
| 1999 | 21.2 | 5.3 | | |
| Census tract education, % of adults with <12 y of education | | | | |
| Lowest quartile | 27.3 | 23.4 | <.0001 | .7326 |
| 2nd Quartile | 27.2 | 23.8 | | |
| 3rd Quartile | 23.7 | 27.5 | | |
| 4th Quartile | 20.2 | 23.8 | | |
| Unknown | 1.6 | 1.5 | | |
| Census tract poverty, % of adults living below poverty line | | | | |
| Lowest quartile | 24.8 | 24.5 | .0740 | .9345 |
| 2nd Quartile | 28 | 27.2 | | |
| 3rd Quartile | 25 | 22.9 | | |
| 4th Quartile | 20.6 | 23.9 | | |
| Unknown | 1.6 | 1.5 | | |
| Comorbidity index | | | | |
| 0 | 76.2 | 76.9 | .7830 | .5360 |
| 1 | 14.3 | 14.5 | | |
| 2 | 3.9 | 3.5 | | |
| 3 | 5.6 | 5.1 | | |
| No. of provider visits in the 12 mo before diagnosis | | | | |
| 3 | 25.8 | 28 | .0004 | .0758 |
| 4–7 | 24.3 | 27.6 | | |
| 8–12 | 22.6 | 21.7 | | |
| 12 | 27.3 | 22.7 | | |
| Mean no. of PSA tests in first 6 mo after diagnosis | | | | |
| 0 | 34.8 | 56.4 | <.0001 | .3138 |
| 1 | 32 | 26.6 | | |
| 2 | 33.2 | 17 | | |

SEER indicates Surveillance, Epidemiology, and End Results Program; PSA, prostate-specific antigen.

[*] Mantel-Haenszel Chi-square.

[†] Cochran Mantel-Haenszel Chi-square, adjusting for propensity quintiles.

**Table 2**

**Outcomes Among Men Treated With and Without Androgen Deprivation for Locally Advanced Prostate Cancer After Primary Radiation Therapy***

| Adjuvant androgen deprivation versus none (Referent category) | Mortality from prostate cancer | | Mortality from other causes | | All cause mortality | |
|---|---|---|---|---|---|---|
| | HR | 95% CI | HR | 95% CI | HR | 95% CI |
| Unadjusted | 1.35 | 1.11–1.64 | 0.95 | 0.83–1.10 | 1.07 | 0.95–1.20 |
| Adjusted for stage, grade, and y of diagnosis | 1.49 | 1.21–1.84 | 0.97 | 0.83–1.13 | 1.12 | 0.99–1.27 |
| Adjusted as above plus age and comorbidity score | 1.49 | 1.21–1.83 | 0.98 | 0.84–1.14 | 1.13 | 1.00–1.28 |
| Adjusted as above plus ethnicity, region, census tract education, and poverty | 1.57 | 1.27–1.94 | 0.99 | 0.85–1.16 | 1.16 | 1.02–1.32 |
| Adjusted as above plus no. of PSA tests and no. of provider visits | 1.63 | 1.32–2.01 | 1.00 | 0.86–1.18 | 1.18 | 1.04–1.34 |
| Adjusted for y of diagnosis, no. of provider visit, and propensity score | 1.65 | 1.33–2.03 | 1.00 | 0.85–1.16 | 1.18 | 1.04–1.33 |
| Cox regression adjusted for age, y of diagnosis, and no. of provider visits stratified by propensity score | | | | | | |
| 1 (Lowest) | 1.16 | 0.69–1.95 | 1.30 | 0.93–1.82 | 1.25 | 0.94–1.66 |
| 2 | 2.57 | 1.80–3.66 | 0.86 | 0.62–1.21 | 1.33 | 1.05–1.68 |
| 3 | 1.19 | 0.80–1.76 | 0.95 | 0.71–1.27 | 1.03 | 0.82–1.30 |
| 4 | 1.59 | 0.91–2.81 | 0.98 | 0.68–1.41 | 1.13 | 0.84–1.53 |
| 5 (Highest) | 7.17 | 0.96–53.55 | 0.92 | 0.55–1.53 | 1.20 | 0.74–1.93 |

HR indicates hazard ratio; 95% CI, 95% confidence interval.

*The total number of deaths from all causes was 393 with androgen deprivation and 1169 with none; the number of prostate cancer deaths was 146 with androgen deprivation and 350 with none; and the number of deaths from other causes was 247 with androgen deprivation and 819 with none.

**Table 3**

**Outcomes of Men With Active Treatment Versus Observation for Clinically Localized Prostate Cancer**

| Active treatment versus observation (Referent category) | Mortality from prostate cancer | | Mortality from other causes | | All cause mortality | |
|---|---|---|---|---|---|---|
| | HR | 95% CI | HR | 95% CI | HR | 95% CI |
| Unadjusted | 0.62 | 0.56–0.69 | 0.52 | 0.50–0.54 | 0.53 | 0.51–0.55 |
| Adjusted[†] | 0.64 | 0.56–0.72 | 0.68 | 0.65–0.71 | 0.67 | 0.65–0.70 |
| Cox regression adjusted for age, tumor size, grade, comorbidity income, and propensity score | 0.64 | 0.57–0.73 | 0.68 | 0.65–0.71 | 0.68 | 0.65–0.70 |
| Cox regression stratified by quintile of propensity score | | | | | | |
| 1 (Lowest) | 0.79 | 0.63–0.98 | 0.70 | 0.65–0.75 | 0.71 | 0.66–0.76 |
| 2 | 0.84 | 0.67–1.04 | 0.65 | 0.60–0.71 | 0.67 | 0.62–0.72 |
| 3 | 0.54 | 0.42–0.70 | 0.70 | 0.63–0.78 | 0.68 | 0.62–0.74 |
| 4 | 0.40 | 0.29–0.54 | 0.69 | 0.60–0.80 | 0.64 | 0.56–0.72 |
| 5 (Highest) | 0.42 | 0.28–0.61 | 0.61 | 0.50–0.73 | 0.57 | 0.48–0.67 |

HR indicates hazard ratio; 95% CI, 95% confidence interval; PSA, prostate-specific antigen.

*
Overall, 33,108 men received active treatment, and 10,739 men received observation. The total number of deaths from all causes was 7969 with active treatment and 4388 with observation; the number of prostate cancer deaths was 1009 with active treatment and 475 with observation; and the number of deaths from other causes was 6960 with active treatment and 3913 with observation.

†
Adjusted for year of diagnosis; age; race; urban residence; marital status; income; education; Surveillance, Epidemiology, and End Results region; tumor size; tumor grade; and comorbidity score.

*Cancer*. Author manuscript; available in PMC 2013 December 09.

**Table 4**

**Outcomes Among Patients Treated With and Without Fluorouracil-based Chemotherapy for Lymph Node-positive Colon Cancer**[*]

| Fluorouracil versus none (Referent category) | Mortality from colon cancer | | Mortality from other causes | | All cause mortality | |
|---|---|---|---|---|---|---|
| | HR | 95% CI | HR | 95% CI | HR | 95% CI |
| Unadjusted | 0.74 | 0.67–0.82 | 0.35 | 0.31–0.41 | 0.57 | 0.53–0.62 |
| Adjusted[†] | 0.78 | 0.70–0.87 | 0.48 | 0.41–0.56 | 0.66 | 0.61–0.72 |
| Cox regression adjusted for age and propensity score | 0.80 | 0.72–0.89 | 0.48 | 0.41–0.56 | 0.67 | 0.62–0.74 |
| Cox regression adjusted for age, stratified by propensity score | | | | | | |
| 1 (Lowest) | 0.94 | 0.68–1.31 | 0.54 | 0.35–0.83 | 0.75 | 0.58–0.97 |
| 2 | 0.72 | 0.58–0.89 | 0.38 | 0.28–0.52 | 0.57 | 0.48–0.68 |
| 3 | 0.77 | 0.61–0.96 | 0.56 | 0.42–0.75 | 0.68 | 0.57–0.82 |
| 4 | 0.87 | 0.68–1.10 | 0.51 | 0.37–0.71 | 0.73 | 0.60–0.88 |
| 5 (Highest) | 0.81 | 0.63–1.04 | 0.48 | 0.31–0.75 | 0.72 | 0.58–0.90 |

HR indicates hazard ratio; 95% CI, 95% confidence interval.

[*] Overall, 2464 men received chemotherapy, and 2396 did not. The total number of deaths from all causes was 1103 with chemotherapy and 1539 without; the number of colon cancer deaths was 805 with chemotherapy and 876 without; and the number of deaths from other causes was 298 with chemotherapy and 663 without.

[†] Adjusted for year of diagnosis; age; sex; urban residence; Surveillance, Epidemiology, and End Results region; no. of lymph nodes; tumor grade; extent of disease; and comorbidity score.