Effectiveness of Radiation Therapy for Elderly Patients with Unresected Early Stage Lung Cancer

Juan P. Wisnivesky, MD, DrPH, Ethan Halm, MD, MPH, Marcelo Bonomi, MD, Charles Powell, MD, Emilia Bagiella, PhD

Online Supplement

Methods

Patients were identified from the SEER registry linked to Medicare files. The SEER program collects data on new cases of cancer from 17 population-based registries covering approximately 26% of the United States population (E1). These registries collect uniform information on all cancers diagnosed within their geographic regions, capturing about 97% of all incident cases in those areas. The distribution of age, sex, education level, and socioeconomic status for individuals in the SEER areas is comparable to that of the US population (E2). Our study sample consisted of 6,065 unresected patients >65 years of age with histologically confirmed stage I-II NSCLC diagnosed between 1992 and 2002. We excluded patients who received chemotherapy and those who were diagnosed at autopsy or from death certificate data. In order to assess comorbidities and chemotherapy use, we also excluded patients who were enrolled in a health maintenance organization or who were not covered by Medicare Parts A and B.

Sociodemographic information (age, sex, race/ethnicity, marital status and estimated income) was obtained from SEER and Medicare databases. Socioeconomic status was estimated based on the median income for the ZIP code of the patient's residence available in the Medicare file (E3). To evaluate the burden of comorbidities, we used the Deyo adaptation of the Charlson comorbidity index as described by Klabunde et al (E4, E5). To calculate the comorbity index for each patient, we multiplied condition-specific weights, developed for lung cancer patients, by their corresponding dichotomous condition indicators (such as chronic obstructive pulmonary disease, diabetes mellitus, etc) and summed them to create the index. Stage was classified according to the American Joint Committee on Cancer criteria (E6). Data regarding tumor location, size, extension, and histology was obtained from SEER. Histology subtypes were classified into categories of adenocarcinoma, bronchioloalveolar carcinoma, squamous cell carcinoma, large cell and other histologic types. Cancer histology was determined according to the ICD-O-2 morphology codes available in SEER (E7).

RT use was ascertained from SEER and Medicare claims. Virnig et al showed that a combination of these sources provides the most complete ascertainment of RT use for cancer (E8). Patients were considered as RT treated if they were coded by SEER as having received external beam radiation or if Medicare inpatient, outpatient, or physician claims contained any code indicating RT use (International Classification of Diseases 9th edition procedure codes: 92.21-92.29, Current Procedural Terminology and Healthcare Common Procedure Coding System codes codes: 77401-77499 and 77750-77799, and revenue center codes 0330 and 0333).

Survival was determined as the interval from the date of cancer diagnosis to the Medicare date of death. Those surviving past December 31, 2004 were classified as censored (alive at the end of follow-up). The cause of death was coded according to SEER, which uses state death certificates as a primary source.

Statistical Analysis

Differences in distribution of baseline characteristics between patients who received or did not receive RT were evaluated using the χ^2 test. The Kaplan-Meier method was used to estimate unadjusted survival rates among patients in the two treatment groups (E9).

We used two methods to control for potential selection bias; propensity score and instrumental variable (IV) analysis (E10, E11). Propensity scores can be viewed as a measure of the likelihood that a patient will receive certain treatment solely on the basis of their covariate information; thus, patients with similar propensity scores are expected to have equivalent outcomes. To perform the propensity score analyses, first we estimated the probability that each patient would receive RT using logistic regression. The model included variables for the patients' sociodemographic characteristics, comorbidities, and cancer-related factors (stage, tumor size, location, and histology). Once the model was fitted, we used regression analyses to evaluate whether the baseline covariates were balanced across study groups after adjusting for the estimated propensity scores.

Cox regression analysis was used to compare overall and lung cancer-specific survival of patients treated with and without RT, adjusting for propensity scores in two ways. First, we included the propensity score as a continuous covariate in a Cox model comparing survival of patients who receive RT and those who did not, using data from the entire cohort. In a second approach, Cox models were estimated within strata defined by propensity score quintiles. For the models estimating lung cancer-specific survival, deaths attributed to causes other than lung cancer were censored at the date of death.

We used IV analysis to obtain a consistent estimate of the effectiveness of RT even in the presence of unmeasured confounders. We used Health Care Service Areas (HCSA) as our IV as we observed considerable geographic variation in the use of RT. HCSAs represents regions with certain characteristics of health care availability and have been used to study geographic variation in health care utilization (E12, E13). The areas range in size from parts of a city to a substantial portion of less populous states. We classified patients as residing in a high-utilization or low-utilization area based on the proportion of patients in the HCSA that received RT for the treatment of unresected early stage NSCLC. Areas where the proportion of patients treated with RT was above the median were classified as high-utilization areas. To avoid bias in the instrument, we excluded HCSAs with ≤5 patients. We compared the baseline characteristics of patients in high and low RT utilization areas, to evaluate if the two groups were reasonably matched for prognostic features.

As proposed by Earle et al, the IV was calculated as the difference between the adjusted survival in the high- and low-utilization areas, divided by the probability of undergoing RT among NSCLC patients in those regions (E14). Thus, the IV estimate represents the absolute difference in survival at these time points among patients treated with and without RT. Adjusted survival was estimated using Cox proportional hazards models, controlling for sociodemographic characteristics, comorbidities, and stage. The adjusted survival was then calculated for a 70-year-old white female with no comorbidities. The confidence interval of the IV estimate was obtained using bootstrap. One thousand subsamples were drawn with replacement from the original population and the IV estimate was calculated on each subsample. The 2.5% and 97.5% percentiles of this analysis were used as the confidence limits of the estimate. All analyses were performed using SAS (SAS, Cary, NC) statistical package.

References for Online Supplement

E1. Surveillance, epidemiology, and end results (seer) program populations (1969-2006) (www.Seer.Cancer.Gov/popdata), national cancer institute, dccps, surveillance research program, cancer statistics branch, released february 2009.

E2. Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the seer-medicare data: Content, research applications, and generalizability to the united states elderly population. Med Care 2002;40:IV-3-18.

E3. Bach PB, Guadagnoli E, Schrag D, Schussler N, Warren JL. Patient demographic and socioeconomic characteristics in the seer-medicare database applications and limitations. Med Care 2002;40:IV-19-25.

E4. Klabunde CN, Potosky AL, Legler JM, Warren JL. Development of a comorbidity index using physician claims data. *J Clin Epidemiol* 2000;53:1258-1267.

E5. Klabunde CN, Legler JM, Warren JL, Baldwin LM, Schrag D. A refined comorbidity measurement algorithm for claims-based studies of breast, prostate, colorectal, and lung cancer patients. *Ann Epidemiol* 2007;17:584-590.

E6. Mountain CF. Revisions in the international system for staging lung cancer. *Chest* 1997;111:1710-1717.

E7. Percy C, Van Holten V, Muir C, editors. International classification of diseases for oncology. Geneva, Switzerland, World Health Organization; 1990.

E8. Virnig BA, Warren JL, Cooper GS, Klabunde CN, Schussler N, Freeman J. Studying radiation therapy using seer-medicare-linked data. Med Care 2002;40:IV-49-54.

E9. Kaplan EL, Meier P. Nonparametric estimation for incomplete obserbations. J Am Stat Assoc 1958;53:457-481.

E10. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997;127:757-763.

E11. Zohoori N, Savitz DA. Econometric approaches to epidemiologic data: Relating endogeneity and unobserved heterogeneity to confounding. *Ann Epidemiol* 1997;7:251-257.

E12. Hammond JR. Substate district, HSA, and PSRO area designations. Am J Public Health 1976;66:788-790.

E13. Ballard-Barbash R, Potosky AL, Harlan LC, Nayfield SG, Kessler LG. Factors associated with surgical and radiation therapy for early stage breast cancer in older women. J Natl Cancer Inst 1996;88:716-726.

E14. 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.