ELSEVIER

Seminars in
RADIATION
ONCOLOGY

Population-Based Observational Studies in Oncology: Proceed With Caution

Payal D. Soni, MD,* and Daniel E. Spratt, MD†

A principle goal of research in Oncology is to determine the optimal treatment for our patients. This often takes the form of comparing 2 existing therapies to one another to determine which is superior, or to introduce a new therapy and determine if it is superior or noninferior to the existing standard of care. This type of research is termed comparative effectiveness research (CER), and the gold-standard is through the conduct of randomized trials. This is the preferred approach, and the only true methodologic approach that can assign a cause-and-effect relationship between a treatment effect and the observed outcome. An alternative approach that is gaining popularity is the use of population-based registry analysis given that it is quicker, cheaper, and easier to perform. However, there are unavoidable forms of bias and confounding that exist when using observational research to perform CER, and recent evidence suggests that population-based CER often results in erroneous results, and that statistical methods to minimize bias are ineffective to overcome the numerous limitations of these databases. In this article, the strengths and weaknesses of both randomized and observational research will be discussed.
Semin Radiat Oncol 29:302–305 Published by Elsevier Inc.

Comparative effectiveness research (CER) is an integral part of medicine that allows us to identify best practices, integrate them into healthcare policy, and ultimately improve clinical outcomes. Randomized controlled trials (RCT) have long been the gold standard for CER. Given their rigorous methodology and ability to minimize bias, RCTs can be used to infer causal relationships between treatments and survival outcomes (Table 1). However, the financial burden of RCTs, concerns regarding their external validity, the desire for faster results, and the growth of electronic medical records facilitating access to "big data" in healthcare, have fostered an interest in complementing, or even replacing, RCTs with population-based observational research. Population data provide a remarkable resource for evaluating prevalence and distribution of disease, patterns and quality of healthcare, and clinical outcomes in the real world. Population research plays a valuable role in identifying gaps in care where interventions can make a drastic impact on the overall health of a population. However, their ability to isolate treatment effects and infer causal relationships is limited, and thus the increasing use of population registries in CER is concerning.

The goal of CER in oncology is to demonstrate whether a treatment regimen independently impacts an oncologic endpoint. The challenges of assessing intermediate endpoints are unique from those affecting the assessment of survival endpoints. Herein, we focus on studies investigating survival endpoints. In a perfect scenario, all other factors affecting survival must be controlled. This is easier to do in the laboratory, and much more challenging with patients. However, randomization, especially double-blind randomization, is the most effective method to minimize patient factor (i.e., age, comorbidities, performance status, and socioeconomic status) imbalance, and minimize patient and physician preferences from biasing treatment selection. Furthermore, RCTs generally control the quality of treatment delivery (dose, duration, and frequency of therapy) to prevent variations in practice from impacting results. Finally, prespecified follow-up criteria and systematic methods for evaluating disease outcomes ensures fair treatment comparisons and minimizes biases due to nonrandom missing data. While an improperly conducted RCT is still subject to all of these factors that can induce bias and confounding, these factors are not controlled for and difficult to even capture in observational research. Thus, RCTs can have flaws from improper conduct and misinterpretation of results, observational research intrinsically, and unavoidably has flaws when it comes to CER.

*Radiation Oncology Service, Hunter Holmes McGuire VA Medical Center, Richmond, VA
†Department of Radiation Oncology, University of Michigan, Ann Arbor, MI
Funding: None
Conflict of interest: None.
Address reprint requests to Payal D. Soni, MD. Hunter Holmes McGuire VAMC, Radiation Oncology Service, 1201 Broad Rock Blvd, Richmond, VA 23249. E-mail: Payal.soni@va.gov

https://doi.org/10.1016/j.semradonc.2019.05.011
1053-4296/Published by Elsevier Inc.

Table 1 Reasons Why RCTs are Superior to Population-Based Observational Research

Bias	Randomization superior to statistical methods to adjust for bias. Observational research inherently impacted by selection bias.
Confounding	Randomization superior to statistical methods to adjust for confounding. Observational research unable to address unmeasured confounders.
Accounting for unmeasured/unknown variables	Randomization inherently favors balance between groups in both measured and unmeasured variables. Propensity-matching of measured variables can induce greater imbalance in unmeasured variables.
Data accuracy	In RCTs data are often centrally collected with methods of quality control. Observational research has variable quality control often with substantial errors in data elements.
Intended use	RCTs by design are intended for research and a specific research question and collect data to answer it. Population-registries often are not created for research, or for specific research questions. Variables collected in observational research are what are readily available.
Treatment information	RCTs assign and collect specific types, doses, and frequency of treatment delivery. Observational research often does not collect this level of detail and simply has the binary receipt or no receipt of treatment.
Treatment intent	RCTs are designed to measure the effect of a treatment with a single intent. Observational research can deliver the same treatment, but it may be for palliative or curative intent purposes.
Side effects/patient reported outcomes	RCTs routinely collect granular information on either physician and/or patient reported outcomes. Observational research rarely has this available.
Quality of information	RCTs often have pathology or imaging centrally reviewed and treatment plans centrally reviewed. Observational research is subject to highly heterogeneous quality.
Causality	RCTs can assign a causal relationship with a treatment effect. Observational research cannot.
Methods	RCTs often have numerous oversight bodies of experts to aid in the prespecified design and endpoints. Observational research rarely has any prespecified design or endpoints.
Development of biomarkers	RCTs are the only method to identify a predictive biomarker. Observational research by definition cannot do this.

Some have raised concerns that the eligibility criteria of RCTs are narrow, and the highly controlled environment of RCTs limit the generalizability of their findings to the real world. However, Unger et al demonstrates that these concerns are greatly exaggerated[1] (Table 2). Unger et al compared overall survival for patients enrolled in the standard treatment arm of 21 national RCTs to matching populations in the Surveillance, Epidemiology, and End Results program (SEER). For cancers with a relatively good prognosis, there was no difference in survival outcomes between the 2 cohorts (mean of adjusted hazard ratios [aHRs] = 0.96; 95% confidence interval [CI] = 0.92-1.01; $P = 0.12$). For cancers with relatively poor prognosis, survival amongst trial patients was better than survival amongst SEER patients in the first year of follow up (mean of aHRs = 0.74; 95% CI = 0.64-0.84; $P < 0.001$), likely due to the selection of healthier patients in trials, but these differences disappeared after the first year (mean of aHRs = 1.05; 95% CI = 0.95-1.15; $P = 0.27$). These data are reassuring, and furthermore, the National Cancer Institute is working toward broadening eligibility criteria for its clinical trials to further ensure a multitude of patients and their outcomes are well represented in clinical trials.

In contrast, several limitations of population-based observational research studies prevent them from reliably demonstrating causal relationships between treatment and survival in the way that RCTs can. Population registries are created retrospectively from chart review and data may be missing or incorrect.[2,3] To assess the completeness of patient records and accuracy of radiation treatment records in the National Cancer Database and its impact on overall survival, Jacobs et al[4] reviewed patient and treatment details for women with node positive endometrial cancer who underwent hysterectomy within 6 months of diagnosis and adjuvant radiation. Of the 14,298 women included in the study, 54% had at least one missing or unknown variable, and 16% had at least 1 anomalous radiation therapy related data entry. Overall survival was significantly worse amongst those patients with missing/unknown data without anomalous data (HR = 1.29, 95% CI = 1.22-1.37, $P < 0.001$) and with (HR = 1.47, 95% CI = 1.36-1.59, $P < 0.001$) compared to those with complete, nonanomalous treatment records. Others have demonstrated that population registries fail to capture radiation receipt altogether for some patients. Jagsi et al[3] compared SEER reports of radiation receipt to patient self-reports of radiation receipt in a cohort of over 2000 women with ductal

Table 2 Commonly Used Flawed Logic to Favor Observation Research Over RCTs for Comparative Effectiveness Research

Trials take too long to do	This does not address that observational research is subject to more limitations and can lead to erroneous results and conclusions. It simply means it is quicker to do observational research.
Trials are too expensive	This does not address that observational research is subject to more limitations and can lead to erroneous results and conclusions. It simply means it is cheaper to do observational research.
Trials are too hard to do	This does not address that observational research is subject to more limitations and can lead to erroneous results and conclusions. It simply means it is easier to do observational research.
Trials are not generalizable	For this to matter, the treatment effect must have a differential benefit on trial eligible vs ineligible patients. While survival may be better in trial eligible patients, there are almost no treatments that have a significant interaction test by trial eligibility. Long-term survival has been shown to be similar in RCTs compared to SEER data.
Difficult in rare diseases	Difficulty does not mean we should perform easier forms of observational research that may have erroneous results. Hundreds of RCTs have been conducted in rare diseases and it is absolutely possible: Wilms tumors (500 cases per year), osteosarcoma (800 cases per year), Ewings sarcoma (200 cases per year), retinoblastoma (200 cases per year), ocular melanoma (2000 cases per year), mesothelioma (2500 cases per year).
Difficult/unethical to randomize	If a new treatment or avoidance of treatment is suggested to be superior or non-inferior there is rarely a case where an RCT is unethical if there is uncertainty to the experimental treatment. Trials have successfully randomized patients to treatment vs no treatment. RCTs have randomized patients between interventional and noninterventional treatments.
Sample size of trials is too small	RCTs have readily achieved sample sizes >1000 patients and are powered usually to detect an absolute difference in the primary endpoint of ≥2%-5%. Effect sizes smaller than this have been questioned to be of unclear clinical relevance. Sample sizes are only required to be large if the difference of interest is very small. Detecting differences smaller than 2% have been done, but are more challenging.

carcinoma in situ or invasive breast cancer. Of 1292 women who reported having received radiation, 21% (n = 273) were coded as never having received radiation in SEER.

Furthermore, insight into the physician's intent, an important potential confounder, is unavailable. The impact of this and other confounding factors on treatment decision is difficult to measure and adjust for and can inadvertently associate treatment with likelihood of survival. Several statistical strategies exist that aim to adjust for the effects of confounding factors on survival, including multivariable analyses, propensity based matching methods, instrumental variable analyses, and sensitivity analyses. However, there is increasing evidence that population-based studies frequently offer false associations between treatment and survival despite the use of sophisticated statistical methodologies. For example, Giordano et al performed a SEER-Medicare analysis comparing survival amongst men undergoing observation, surgery, and radiation for early stage prostate cancer.[5] Despite adjustments for age, stage, comorbidities, race and other demographics, they found that patients undergoing surgery not only had better survival compared to patients undergoing observation and radiation, which is in contradiction to randomized data,[6] but also found that patients undergoing surgery had better survival compared to a matched population of patients without cancer suggesting residual bias. Pearlstein et al similarly compared prostate cancer treatment outcomes using SEER data linked with Medicare Health Outcomes Survey data that is enriched with prospectively collected patient reported comorbidity and functional status information. Even when using such robust population data, they found implausible differences in patient survival between treatment cohorts.[7] Similar disparities have also been identified between SEER data and randomized trial data in breast cancer.[8]

Two systematic analyses of observational studies published prior to 2000 had suggested reasonable agreement with matching randomized trials.[9,10] However, less than 15% of the studies included in both analyses were performed using population-based registries, and less than 20% were focused on a survival endpoint. This is important, as especially in cancers where patients may die of noncancer causes, survival is a composite of both cancer-associated and noncancer associated mortality. Furthermore, both analyses were done on a select group of observational studies from select journals, and oncology studies were poorly represented. Therefore, the findings from these 2 analyses cannot be applied to the current population-based CER sensation in cancer research. More recently, Soni et al performed a comprehensive systematic analysis of all population based CER studies in all forms of cancer performed using SEER, SEER-Medicare, and the National Cancer Database between 2000 and 2016, and found no agreement between these studies and matching randomized trials beyond what would be expected by chance alone (k = 0.037 95% CI = −0.027 to 0.1). Survival results from 350 registry-based treatment comparisons were compared to corresponding randomized treatment comparisons, and only 40% reported congruent findings. Furthermore, there was no correlation between

survival HRs (concordance correlation coefficient = 0.083 [95% CI = −0.068 to 0.230]). No modifiable study design factors including the database used, variable adjustments, use of instrumental variable or sensitivity analyses, or degree of similarity in populations and eligibility criteria could improve agreement.[11]

It is often argued that observational studies are a necessary component of CER despite their limitations as RCTs cannot be done in many scenarios. We would challenge this notion given that even rare diseases, such as ocular melanoma,[12] mesothelioma,[13] and pediatric cancers,[14,15] among many others, have not just one randomized trial, but often multiple trials that have helped define the standards of care. Furthermore, the notion that patients cannot be randomized between interventional and noninterventional treatments is also overly used, as these types of trials has been effectively done in prostate cancer,[6,16-20] lung cancer,[21,22] cervical cancer,[23] ocular melanoma,[12] and brain and spine metastases.[24,25]

Population-based CER studies are just as likely to be false as they are to be true with no reliable way of distinguishing one from the other, which can be dangerous. The data can be misinterpreted and misused by patients, physicians, and policy makers and thereby lead to patient and societal harm.[26] These registries and analytical methods need to continue to improve; until then, readers should interpret their findings with caution. Just because RCTs can be subjected to limitations, does not mean that an even more flawed method of research, population-registry research, should be preferred. Rather, we would suggest that improved methods of conducting RCTs should be sought to expand eligibility criteria and minimize the obstacles of conducting RCTs.

References

1. Unger JM, Barlow WE, Martin DP, et al: Comparison of survival outcomes among cancer patients treated in and out of clinical trials. J Natl Cancer Inst 106: 2014. dju002
2. Park HS, Lloyd S, Decker RH, et al: Limitations and biases of the surveillance, epidemiology, and end results database. Curr Probl Cancer 36:216-224, 2012
3. Jagsi R, Abrahamse P, Hawley ST, et al: Underascertainment of radiotherapy receipt in surveillance, epidemiology, and end results registry data. Cancer 118:333-341, 2012
4. Jacobs CD, Carpenter DJ, Hong JC, et al: Radiation records in the national cancer database: Variations in coding and/or practice can significantly alter survival results. JCO Clin Cancer Inform 3:1-9, 2019
5. Giordano SH, Kuo YF, Duan Z, et al: Limits of observational data in determining outcomes from cancer therapy. Cancer 112:2456-2466, 2008
6. Hamdy FC, Donovan JL, Lane JA, et al: 10-year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. N Engl J Med 375:1415-1424, 2016
7. Pearlstein KA, Basak R, Chen RC: Comparative effectiveness of prostate cancer treatment options: Limitations of retrospective analysis of cancer registry data. Int J Radiat Oncol Biol Phys 103:1053-1057, 2019
8. McGale P, Cutter D, Darby SC, et al: Can observational data replace randomized trials? J Clin Oncol 34:3355-3357, 2016
9. Concato J, Shah N, Horwitz RI: Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 342:1887-1892, 2000
10. Benson K, Hartz AJ: A comparison of observational studies and randomized, controlled trials. N Engl J Med 342:1878-1886, 2000
11. Soni PD, Hartman HE, Dess RT, et al: Comparison of population-based observational studies with randomized trials in oncology. J Clin Oncol 4:1209-1216, 2019
12. Jampol LM, Moy CS, Murray TG, et al: The COMS randomized trial of iodine 125 brachytherapy for choroidal melanoma: IV. Local treatment failure and enucleation in the first 5 years after brachytherapy. COMS report no. 19. Ophthalmology 109:2197-2206, 2002
13. Zalcman G, Mazieres J, Margery J, et al: Bevacizumab for newly diagnosed pleural mesothelioma in the Mesothelioma Avastin Cisplatin Pemetrexed Study (MAPS): a randomised, controlled, open-label, phase 3 trial. Lancet 387:1405-1414, 2016
14. Evans AE, Jenkin RD, Sposto R, et al: The treatment of medulloblastoma. Results of a prospective randomized trial of radiation therapy with and without CCNU, vincristine, and prednisone. J Neurosurg 72:572-582, 1990
15. Packer RJ, Gajjar A, Vezina G, et al: Phase III study of craniospinal radiation therapy followed by adjuvant chemotherapy for newly diagnosed average-risk medulloblastoma. J Clin Oncol 24:4202-4208, 2006
16. Sathya JR, Davis IR, Julian JA, et al: Randomized trial comparing iridium implant plus external-beam radiation therapy with external-beam radiation therapy alone in node-negative locally advanced cancer of the prostate. J Clin Oncol 23:1192-1199, 2005
17. Hoskin PJ, Rojas AM, Bownes PJ, et al: Randomised trial of external beam radiotherapy alone or combined with high-dose-rate brachytherapy boost for localised prostate cancer. Radiother Oncol 103:217-222, 2012
18. Morris WJ, Tyldesley S, Rodda S, et al: Androgen Suppression Combined with Elective Nodal and Dose Escalated Radiation Therapy (the ASCENDE-RT Trial): An Analysis of survival endpoints for a randomized trial comparing a low-dose-rate brachytherapy boost to a dose-escalated external beam boost for high- and intermediate-risk prostate cancer. Int J Radiat Oncol Biol Phys 98:275-285, 2017
19. Bill-Axelson A, Holmberg L, Garmo H, et al: Radical prostatectomy or watchful waiting in early prostate cancer. N Engl J Med 370:932-942, 2014
20. Wilt TJ, Jones KM, Barry MJ, et al: Follow-up of prostatectomy versus observation for early prostate cancer. N Engl J Med 377:132-142, 2017
21. Albain KS, Swann RS, Rusch VW, et al: Radiotherapy plus chemotherapy with or without surgical resection for stage III non-small-cell lung cancer: A phase III randomised controlled trial. Lancet 374:379-386, 2009
22. Moghanaki D. SVL. Veterans Administration Lung Cancer Surgery or Stereotactic Radiotherapy Trial (VALOR). Available at: http://www.clinicaltrials.gov/ct2/show/NCT02984761. Accessed February 27, 2019.
23. Landoni F, Colombo A, Milani R, et al: Randomized study between radical surgery and radiotherapy for the treatment of stage IB-IIA cervical cancer: 20-Year update. J Gynecol Oncol 28:e34, 2017
24. Patchell RA, Tibbs PA, Walsh JW, et al: A randomized trial of surgery in the treatment of single metastases to the brain. N Engl J Med 322:494-500, 1990
25. Patchell RA, Tibbs PA, Regine WF, et al: Direct decompressive surgical resection in the treatment of spinal cord compression caused by metastatic cancer: A randomised trial. Lancet 366:643-648, 2005
26. Eggertson L: Lancet retracts 12-year-old article linking autism to MMR vaccines. CMAJ 182:E199-E200, 2010