Effectiveness in the Absence of Efficacy: Cautionary Tales From Real-World Evidence

Safiya Karim, MD[1] and Christopher M. Booth, MD[2]

Background

Real-world data (RWD) are collected outside of traditional clinical trials and may include electronic health records, patient registries, and administrative health care claims.[1,2] Real-world evidence (RWE) is derived from analysis and aggregation of these data. Published reports of RWE have increased substantially in recent years. A PubMed search on the terms "real world data," "real world evidence," and "registry" showed a 600% increase in citations during the period 2002 to 2016 (from 2,435 citations per year to 14,956 citations per year). RWD have historically been used to answer questions related to trends in cancer incidence and mortality; quality of and access to care delivered in routine practice; outcomes of rare cancers; and understanding the incidence of rare events and toxicities in the general population. There is growing interest in the use of RWD to study effectiveness of treatments in the real world (ie, comparative effectiveness research [CER]). Although other reports have highlighted the benefits and some pitfalls of RWE,[3-7] in this commentary, we highlight a specific scenario in which readers of CER should be cautious in their interpretation of reported results. We wrote this commentary because such studies are increasingly common and have the potential for patient harm if therapies are adopted solely on the basis of analyses of RWD. This is particularly important because there are signals that regulatory agencies may begin approving drugs on the basis of observational data.[8]

What Is Comparative Effectiveness Research?

CER in this context can be defined as observational studies that compare the outcomes of two or more treatments in a real-world population of patients not randomly assigned to a treatment. Although these studies can offer important insights into outcomes achieved in routine practice, it is essential that they are carefully designed and considered in the context of existing evidence. Their interpretation should be held to the same level of rigor as that of a clinical trial. The most important methodologic limitation of these of studies is selection bias. Treatment selection in routine practice is strongly influenced by a patient's baseline characteristics; in turn, this results in an inherent imbalance between treatment groups, which in itself can lead to very different outcomes. Although several statistical methods can mitigate this bias (eg, multivariate regression analysis, propensity score analysis), these can only be applied to variables that are known and measurable. Residual confounding, therefore, remains a major limitation and can lead to significant overestimates of effect size.[9] In light of these methodologic shortcomings, we propose a hierarchy of scenarios in which to consider studies of comparative effectiveness.

Placing CER in Context With Existing Evidence

At the top of this hierarchy is the setting in which a randomized controlled trial (RCT) has shown efficacy and RWD are being used to determine if this efficacy is translated into effectiveness in routine practice. RCTs are the gold standard for evaluation of new drugs and are at the top of the evidence-based medicine hierarchy.[10] The strength of RCTs is the high degree of internal validity that is achieved through randomization. However, most RCTs have stringent eligibility criteria that may exclude several important subsets of the population (eg, elderly patients and those with comorbidity[11]), which may result in limited external validity (ie, generalizability), making it difficult to determine which patients will benefit in a heterogeneous real-world population. The use of RWE in this context, therefore, can augment the results of a clinical trial. For example, after several pivotal RCTs, our group explored uptake of adjuvant chemotherapy (ACT) for non–small-cell lung cancer in routine practice. Data from the Ontario population showed an improvement in overall survival (OS) in the era of ACT that was comparable to the results of RCTs.[12] This study used time as an instrumental variable (ie, comparison of the outcome of all surgical cases before and after publication of data from the pivotal RCTs), which mitigates some common forms of bias, including selection bias and immortal time bias. Meyerhardt et al[13] conducted an analysis using the SEER database to determine the effectiveness of bevacizumab with first-line combination chemotherapy for patients with stage IV colorectal cancer. They also found evidence of effectiveness that was consistent with efficacy reported in RCTs.

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on February 20, 2019 and published at jco.org on March 21, 2019: DOI https://doi.org/10.1200/JCO.18.02105

Studies using RWD may also find lack of effectiveness despite efficacy, as was the case in the analysis by Sanoff et al[14] of sorafenib for hepatocellular carcinoma. These discordant findings are not unexpected when one considers the outcomes of delivering a toxic therapy to fit patients in an RCT compared with treatment of older and sicker patients in routine practice. These "negative" effectiveness studies are equally important as those that confirm efficacy and should lead to important questions about the extent to which a novel therapy offers benefit in the real world.

The second context in which CER using RWD can be potentially helpful is when there is some (but not level 1) evidence of efficacy (ie, underpowered RCTs). In this case, RWE can be a useful addition to the knowledge base but should not be considered a substitute for a large phase III RCT. For example, the results of several small RCTs of ACT for bladder cancer have indicated improved outcomes, but definitive evidence was lacking. In this context, Galsky et al[15] conducted an observational study of 5,653 patients with pathologic T3/T4 and/or node-positive bladder cancer using the National Cancer Database. Using propensity scores and landmark analysis, they found that ACT was associated with improved OS. Furthermore, this benefit was seen across all subgroups, including patients older than 70 years, that are often underrepresented in RCTs.[15] This and similar studies were a valuable addition to the existing literature and provided further evidence that ACT is a reasonable standard of care.[16]

The third scenario is the most problematic context in which to use RWD for the purpose of CER: when previous RCTs have shown lack of efficacy. In these cases, we believe the a priori risk of a spurious result is so high that these studies should not be performed. It is difficult to imagine a cancer treatment that does not work under the ideal circumstances of an RCT but that does benefit patients in the general population. Bayesian logic suggests that results demonstrating effectiveness in this setting are far more likely to be artifact compared with data showing effectiveness in a scenario where preexisting RCT evidence has established efficacy. The following two studies in colorectal cancer illustrate what can go wrong when investigators look for effectiveness when RCTs have found no evidence of efficacy.

Effectiveness in CER But Lack of Efficacy in RCTS

Casadaban et al[17] reported a population-based study in which they evaluated the effectiveness of ACT for stage II colon cancer. Previous RCTs[18-21] and a large meta-analysis[22] did not find a statistically significant OS benefit for ACT in stage II colon cancer. The IMPACT B2 (International Multicentre Pooled Analysis of B2 Colon Cancer Trials) meta-analysis included 1,016 patients with stage II disease and reported comparable survival outcomes at 5 years (82% with ACT and 80% with observation; $P = .57$).[34] Although it is possible that prior RCTs were underpowered to detect a statistically significant difference, the existing data suggest that if there is a survival benefit with ACT, it is certain to be small. However, in a retrospective study of 153,110 patients from the National Cancer Database, the authors reported that ACT was associated with an astounding 18% absolute improvement in OS at 5 years (hazard ratio, 0.71; $P < .001$).[17] Moreover, this degree of benefit was observed in all subgroups, including those with no high-risk features. Despite the authors' use of standard statistical approaches (ie, multivariate regression analysis and propensity score analysis), the large effect size strongly hints of residual confounding. On careful review of the study methodology, it appears the authors did not adjust their primary analysis for the effect of comorbidity, and even when it was later added, these data were missing for 60% of patients. Moreover, this study is vulnerable to immortal time bias, whereby patients could not be included in the ACT treatment group unless they survived long enough to get ACT.[23] Conversely, patients who did not receive ACT were not subject to this restriction. Although the authors partially mitigated this risk by excluding patients from the study who died within 30 days of surgery, existing literature shows that the vast majority of patients in routine practice start ACT more than 30 days after surgery. This means there would be a disproportionate number of early deaths (ie, between 30 days postsurgery and the start of ACT) among the no-ACT group. These fundamental limitations, together with the fact that prior RCTs have failed to show a survival benefit with ACT, makes it far more likely that the difference in outcome between the ACT and non-ACT treatment groups reflects residual confounding and other forms of bias than a true treatment effect.

In another recent study, Freischlag et al[24] used the NCDB to evaluate whether incomplete delivery of neoadjuvant radiotherapy (RT) was associated with inferior survival among patients with locally advanced rectal cancer treated during 2006 to 2012. Once again, the research question is problematic. RCTs have clearly demonstrated that neoadjuvant RT for localized rectal cancer improves local control rates and reduces need for colostomy.[25-28] However, in the era of total mesorectal excision (TME), current evidence suggests that preoperative RT offers no (or very small) survival benefit.[29] Results of older meta-analyses reported conflicting results.[25-27,30] However, the studies included in these meta-analyses predated adoption of TME. The Dutch Colorectal Cancer Group, reporting on its pivotal RCT, said that in the era of TME, preoperative RT was associated with reduced local recurrence but not with OS.[31] Thus, the existing evidence suggests that if there is any association between preoperative RT and OS, it is likely to be small.

Freischlag et al[24] explored this issue in a retrospective study of 17,600 patients treated during 2006 to 2012. They reported that complete delivery of neoadjuvant radiation was associated with a 10% improvement in OS at 5 years compared with incomplete RT (73% *v* 63%; $P < .01$). Moreover, in contrast to findings of previous RCTs,[32,33] they did not find any difference in resection margin positivity or

permanent colostomy rate.[24] Is it plausible that despite the results from a large RCT with long follow-up times, this study found a 10% improvement in OS but no benefit in local control? Again in this study, the large "survival benefit" observed is more likely due to residual confounding from patient characteristics that allowed complete delivery of RT rather than the RT itself.

Conclusion

Although RWD can provide valuable insight into the benefit of treatments in the real world, there are inherent limitations to this study design. Studies of comparative effectiveness are ideally performed with a multidisciplinary team involving clinicians, epidemiologists, and biostatisticians. These studies are best suited for settings in which there is existing evidence to believe that a given treatment is efficacious (ie, to understand if efficacy translates to effectiveness). In settings where RCTs do not exist or may not be feasible, RWD can be informative; however, these studies should be interpreted with caution. Clinicians should not adopt new therapies on the basis of RWE in isolation. This is particularly true when RCTs have revealed no evidence of treatment benefit; reports of "effectiveness" in this setting are more likely artifact and may be misleading. Journal editors and clinicians should be critical of studies that report effectiveness in the absence of efficacy and should question the plausibility of such findings.

AFFILIATIONS

[1]University of Calgary, Calgary, Alberta, Canada
[2]Queen's University Cancer Research Institute, Kingston, Ontario, Canada

CORRESPONDING AUTHOR

Christopher M. Booth, MD, Division of Cancer Care and Epidemiology, Queen's University Cancer Research Institute, 10 Stuart St, Kingston, Ontario, K7L 3N6, Canada; e-mail: booth@queensu.ca.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST AND DATA AVAILABILITY STATEMENT

Disclosures provided by the authors and data availability statement (if applicable) are available with this article at DOI https://doi.org/10.1200/JCO.18.02105.

AUTHOR CONTRIBUTIONS

Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

ACKNOWLEDGMENT

C.M.B. is funded as the Canada Research Chair in Population Cancer Care.

REFERENCES

1. Makady A, de Boer A, Hillege H, et al: What is real-world data? A review of definitions based on literature and stakeholder interviews. Value Health 20:858-865, 2017
2. Garrison LP Jr, Neumann PJ, Erickson P, et al: Using real-world data for coverage and payment decisions: The ISPOR Real-World Data Task Force report. Value Health 10:326-335, 2007
3. Booth CM, Karim S, Peng Y, et al: Radical treatment of the primary tumor in metastatic bladder cancer: Potentially dangerous findings from observational data. J Clin Oncol 36:533-535, 2018
4. Chang GJ: Is there validity in propensity score-matched estimates of adjuvant chemotherapy effects for patients with rectal cancer? JAMA Oncol 4:921-923, 2018
5. Choudhury A, Hoskin PJ: Bladder cancer and the National Cancer Data Base: New insight or misinformation? Cancer 124:1105-1107, 2018
6. Goodwin PJ, Ballman KV, Small EJ, et al: Evaluation of treatment benefit in *Journal of Clinical Oncology*. J Clin Oncol 31:1123-1124, 2013
7. Massarweh NN, Chang GJ: Translating research findings into practice-the unfulfilled and unclear mission of observational data. JAMA Surg https://doi.org/10.1001/jamasurg.2018.3118 [epub ahead of print on October 24, 2018]
8. Corrigan-Curay J, Sacks L, Woodcock J: Real-world evidence and real-world data for evaluating drug safety and effectiveness. JAMA 320:867-868, 2018
9. Lee J, Bosco F, Geiger AM, , et al A most stubborn bias: Confounding by indication in observational studies. J Clin Epidemiol 63:64-74, 2011
10. Sackett DL: Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest 95:2S-4S, 1989 (2 suppl)
11. Booth CM, Tannock IF: Randomised controlled trials and population-based observational research: Partners in the evolution of medical evidence. Br J Cancer 110:551-555, 2014
12. Booth CM, Shepherd FA, Peng Y, et al: Adoption of adjuvant chemotherapy for non-small-cell lung cancer: A population-based outcomes study. J Clin Oncol 28:3472-3478, 2010
13. Meyerhardt JA, Li L, Sanoff HK, et al: Effectiveness of bevacizumab with first-line combination chemotherapy for Medicare patients with stage IV colorectal cancer. J Clin Oncol 30:608-615, 2012
14. Sanoff HK, Chang Y, Lund JL, et al: Sorafenib effectiveness in advanced hepatocellular carcinoma. Oncologist 21:1113-1120, 2016
15. Galsky MD, Stensland KD, Moshier E, et al: Effectiveness of adjuvant chemotherapy for locally advanced bladder cancer. J Clin Oncol 34:825-832, 2016
16. Booth CM, Tannock IF: Benefits of adjuvant chemotherapy for bladder cancer. JAMA Oncol 1:727-728, 2015
17. Casadaban L, Rauscher G, Aklilu M, et al: Adjuvant chemotherapy is associated with improved survival in patients with stage II colon cancer. Cancer 122:3277-3287, 2016

18. Matsuda C, Ishiguro M, Teramukai S, et al: A randomised-controlled trial of 1-year adjuvant chemotherapy with oral tegafur-uracil versus surgery alone in stage II colon cancer: SACURA trial. Eur J Cancer 96:54-63, 2018
19. Schippinger W, Samonigg H, Schaberl-Moser R, et al: A prospective randomised phase III trial of adjuvant chemotherapy with 5-fluorouracil and leucovorin in patients with stage II colon cancer. Br J Cancer 97:1021-1027, 2007
20. Gray R, Barnwell J, McConkey C, et al: Adjuvant chemotherapy versus observation in patients with colorectal cancer: A randomised study. Lancet 370: 2020-2029, 2007
21. Moertel CG, Fleming TR, Macdonald JS, et al: Intergroup study of fluorouracil plus levamisole as adjuvant therapy for stage II/Dukes' B2 colon cancer. J Clin Oncol 13:2936-2943, 1995
22. Gill S, Loprinzi CL, Sargent DJ, et al: Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: Who benefits and by how much? J Clin Oncol 22:1797-1806, 2004
23. Suissa S: Immortal time bias in pharmaco-epidemiology. Am J Epidemiol 167:492-499, 2008
24. Freischlag K, Sun Z, Adam MA, et al: Association between incomplete neoadjuvant radiotherapy and survival for patients with locally advanced rectal cancer. JAMA Surg 152:558-564, 2017
25. Glimelius B, Grönberg H, Järhult J, et al: A systematic overview of radiation therapy effects in rectal cancer. Acta Oncol (Madr) 42:476-492, 2003
26. Colorectal Cancer Collaborative Group: Adjuvant radiotherapy for rectal cancer: A systematic overview of 8,507 patients from 22 randomised trials. Lancet 358: 1291-1304, 2001
27. Cammà C, Giunta M, Fiorica F, et al: Preoperative radiotherapy for resectable rectal cancer: A meta-analysis. JAMA 284:1008-1015, 2000
28. Sauer R, Liersch T, Merkel S, et al: Preoperative versus postoperative chemoradiotherapy for locally advanced rectal cancer: Results of the German CAO/ARO/ AIO-94 randomized phase III trial after a median follow-up of 11 years. J Clin Oncol 30:1926-1933, 2012
29. Wong RK, Berry S, Spithoff K, et al: Preoperative or postoperative therapy for stage II or III rectal cancer: An updated practice guideline. Clin Oncol (R Coll Radiol) 22:265-271, 2010
30. Buyse M, Zeleniuch-Jacquotte A, Chalmers TC: Adjuvant therapy of colorectal cancer. Why we still don't know. JAMA 259:3571-3578, 1988
31. Kapiteijn E, Marijnen CA, Nagtegaal ID, et al: Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer. N Engl J Med 345: 638-646, 2001
32. van Gijn W, Marijnen CA, Nagtegaal ID, et al: Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer: 12-Year follow-up of the multicentre, randomised controlled TME trial. Lancet Oncol 12:575-582, 2011
33. Sauer R, Becker H, Hohenberger W, et al: Preoperative versus postoperative chemoradiotherapy for rectal cancer. N Engl J Med 351:1731-1740, 2004
34. IMPACT B2 Investegators: Efficacy of adjuvant fluorouracil and folinic acid in B2 colon cancer: International Multicentre Pooled Analysis of B2 Colon Cancer Trials (IMPACT B2) Investigators. J Clin Oncol 17:1356-1363

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Effectiveness in the Absence of Efficacy: Cautionary Tales From Real-World Evidence

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 ascopubs.org/jco/site/ifc.

Safiya Karim
Honoraria: Sanofi

No other potential conflicts of interest were reported.