# Leveraging existing data to contextualize phase II clinical trial findings in oncology

More than 250 000 women are diagnosed with early-stage breast cancer (EBC) in the USA each year.[1] Of these, up to 30% have amplification of the human epidermal growth factor 2 (HER2).[2] The current standard of care for HER2-positive (HER2+) EBC is chemotherapy plus HER2-directed therapy to complete 1 year of treatment.[3] There is growing interest in determining which patients with HER2+ tumors could achieve favorable outcomes with less chemotherapy through better HER2-targeting. The phase II Adjuvant Paclitaxel and Trastuzumab (APT) trial by Tolaney and colleagues[4] provided compelling evidence that patients with small HER2+ tumors without nodal macrometastases can achieve highly favorable outcomes with paclitaxel and trastuzumab alone (TH), avoiding the toxicity associated with multi-agent chemotherapy regimens. Use of TH in this context has been widely adopted in the clinical setting despite the lack of a confirmatory phase III trial.

Phase II clinical trials have well-described limitations for determining new standards of care, including lack of a comparison arm or lack of sufficient power, even if randomization is carried out, to enable firm conclusions regarding relative efficacy. Phase III randomized trials, though they provide stronger evidence regarding the comparative efficacy of treatment alternatives, have problems of their own, including limited feasibility and affordability in clinical contexts with favorable prognoses and low event rates. Added to this is the larger size and design complexity of noninferiority trials traditionally used to test therapeutic minimalization. In small, node-negative, HER2+ breast cancers, where recurrence rates with conventional therapy are low, conducting a prospective, randomized trial would require high patient accruals and long follow-up times, resulting in a study that is prohibitively expensive and takes many years to obtain results. Such study designs in this clinical context do not adequately serve the needs of the oncologic community and the patients whom we serve, due to their inability to provide timely information to guide patient care.

In a study published in this issue of *Annals of Oncology*, Amiri-Kordestani and colleagues[5] took a novel approach to addressing the limitations of the phase II APT trial by using pooled historical clinical trial data as an external control. In doing so, they constructed a multi-agent chemotherapy arm to which TH could be compared and selected patients with similar demographic and clinical characteristics to those in the experimental arm using propensity score matching to address some of the problems with cross-trial comparisons. In their analysis, the authors used existing patient-level data from five clinical trials supporting drug approval by the US Food and Drug Administration (FDA) for HER2+ EBC.[6-9] They pooled data from these trials and selected patients with low-risk EBC. Patients from the historical trials with node-negative disease, tumors $<3$ cm (as required in the APT trial), and treated with trastuzumab and either anthracycline/cyclophosphamide/taxane (ACTH) or taxane/carboplatin (TCH) were matched 1 : 1 to patients treated with TH from the APT trial. To evaluate the impact of HER2-directed therapy, they also matched and compared patients treated with anthracycline/cyclophosphamide/taxane without trastuzumab (ACT) to patients treated with TH. They found similar invasive disease-free survival (iDFS) and overall survival (OS) at 3 and 5 years between the ACTH/TCH and TH arms. They also found improved iDFS and OS at 3 and 5 years when comparing TH with ACT.[5]

Cross-trial comparisons are fraught with limitations, in part due to different study populations and lack of randomization, introducing measured and unmeasured confounders. In this case, the attempt to find comparable patients for comparison across trials is particularly challenging because of inherent differences in the target patient populations for the ACTH and TCH registration trials and the APT trial. The former set of trials deliberately targeted higher risk patients, including those with node-positive, estrogen receptor (ER)-negative, higher grade disease and/or very young age, in order to obtain an event rate sufficient to power a comparison between arms. In contrast, the APT trial tested a de-escalation approach, and thus appropriately limited its patient population to those with relatively lower risk features, including tumors under 3 cm and those without nodal macrometastases, and was conducted in an era where the core populations of the earlier registration trials would have been largely treated in a neoadjuvant fashion. The limited overlap of the trial populations is illustrated by the fact that only 9% of the registration trial patients were included in the external control arm for this analysis, and 96 of 401 APT patients could not be matched to a comparator patient from the registration trial pool. Nevertheless, the authors in this study made appropriate efforts to address measurable confounders through propensity score matching according to age (in years), tumor size (in granular increments of Tmi/T1a, T1b, T1c, and T2), ER status, progesterone receptor status, and histologic grade.[5] The subsets selected for

comparison appear to be substantively similar on measured prognostic features, with the exception of slightly more grade 3 tumors (62.3% versus 58.7%) among TCH/ACTH patients compared with TH patients. While they could not eliminate the possibility of selection bias due to unmeasured confounders, their reasonable and thoughtful approach to homogenize the study groups addresses these concerns as best as can be done outside of a prospective, randomized trial, and sets a hopeful precedent for future use of external control arms, where randomized comparisons are infeasible.

Another limitation of cross-trial comparisons, as the study authors point out, is time-trend bias, which could not be easily adjusted for using statistical methods. The historical clinical trials used for comparison in this study enrolled the bulk of patients between 2000 and 2005,[6-9] while the APT trial was conducted between 2007 and 2010.[4] While improvements in treatment delivery and supportive care that may impact survival should always be considered, it is unlikely that there was substantial variation in clinical care during such a short gap in time.

Acknowledging that evaluation of phase II study findings with a prospective, randomized, controlled trial is ideal but not always feasible, is there any other way to contextualize phase II findings? Another data source that could be leveraged for comparison—though also fraught with limitations—is electronic health record (EHR)-derived patient-level data. While single-institution databases lack the numbers needed for comparison, and claims databases lack the granular clinical information required to construct a well-matched comparison group, novel datasets such as the American Society of Clinical Oncology's CancerLinQ database that integrate patient-level data from multiple health systems may one day provide the volume of patients and the richness of clinical data to enable such comparisons.[10] The FDA has identified advancing real-world data into regulatory-quality real-world evidence as a key strategic priority.[11] In its present state, due in large part to the limitations of the EHRs from which it is sourced, CancerLinQ and other pooled EHR databases generally lack sufficient completeness of clinical information, such as tumor size, nodal status, histologic grade, and receptor status, necessary to enable matched comparisons of real-world and study data. Perhaps even more importantly, aggregated EHR data rarely capture cancer recurrence and progression events, a key end point of clinical trials, in a complete, accurate, and structured manner. Additionally, the more diverse populations treated in clinical practice, while improving the generalizability of the comparator pool, may also prove challenging to match to highly selected clinical trial participants, predictably resulting in discordance between real-world and trial outcomes.[12] Improving the data quality of aggregate real-world databases requires improving the quality of structured data elements documented in the EHR, as well as additional tools such as natural language processing and human curation.[10,13] These challenges are not insurmountable, but will take time and innovation to reduce reliance on clinician data entry, which is inherently variable and resource intensive, and increase automation, linkage, and curation processes to improve data quality and enhance the interoperability of data across different health systems and records. The Minimal Common Oncology Data Elements (mCODE™) initiative reflects these efforts to leverage real-world data for research purposes. This initiative aims to provide a common standard and language for oncology to standardize and collect data in a computable manner so it can be aggregated and analyzed with data from many other patients, exchanged through interoperable EHR systems, collected in a streamlined manner without burdening clinicians, and protect patient privacy. The rigor of data collection required in the trial setting, and the manpower required to support it, is not the standard in clinical practice, but data collection in clinical practice can evolve to be better adapted to provide real-world evidence, leveraging the power of the EHR and data mining tools.

In the meantime, Amiri-Kordestani and colleagues[5] provide an innovative approach for using historical clinical trial data to contextualize phase II study findings in HER2+ EBC that can be easily applied to other contexts. Their study provides reassuring findings that patient outcomes with TH are similar to those achieved with ACTH or TCH in this patient population with low clinical risk, and therefore de-escalation of therapy for patients with small, node-negative HER2+ breast cancers is appropriate. Their approach is particularly well-suited for studies looking at de-escalation therapies. In breast oncology, a similar approach could be applied to the results of ongoing studies evaluating the impact of HER2-directed therapy, without or with limited chemotherapy, in patients with HER2+ EBC.[14,15] The FDA's Oncology Center of Excellence has longstanding interest in streamlining drug approvals through use of synthetic control arms based on data from prior clinical trials to evaluate the efficacy of new drugs or serve as a comparator arm for prospective clinical trials in rare tumor types.[11,16] Doing so can improve accrual to clinical trials and increase patient access to promising new therapies. This study by Amiri-Kordestani and colleagues, notably led by FDA investigators, provides a promising example of the FDA's commitment to leveraging existing data to advance scientific progress.

Historically, large randomized clinical trials have been the flagships of treatment advancement in oncology. However, those who argue that prospective, randomized trials are the only standard by which to practice clinical medicine will, in many cases, find themselves waiting for a trial that will never materialize, or one that takes so long that it is completed after the standards of care have evolved, making the hypothesis less relevant or difficult to interpret. Comparative studies, such as the one by Amiri-Kordestani and colleagues, are not intended to replace phase III trials. However, in the context of clinical conditions with favorable outcomes and low event rates, study designs that utilize existing comparator data to replace prospectively collected control arms may be more affordable, nimble, and timely, and, when understood for what they are rather than what they are not, serve as a valuable contribution to the body of evidence in clinical

oncology. The authors here demonstrate how, with a thoughtful study design, we can use existing data to advance scientific knowledge in a timely and affordable manner to answer clinically-relevant questions in oncology.

E. M. Ray*, L. A. Carey & K. E. Reeder-Hayes
*Division of Oncology, Department of Medicine, University of North Carolina at Chapel Hill, Lineberger Comprehensive Cancer Center, Chapel Hill, USA*
(*E-mail: emily_ray@med.unc.edu).

Available online 22 September 2020

https://doi.org/10.1016/j.annonc.2020.09.008
DOI of original article: https://doi.org/10.1016/j.annonc.2020.08.2106

## FUNDING

None declared.

## DISCLOSURE

The authors have declared no conflicts of interest.

## REFERENCES

1. U.S. Cancer Statistics. Breast Cancer Stat Bite | CDC. Available at https://www.cdc.gov/cancer/uscs/about/stat-bites/stat-bite-breast.htm. Accessed August 27, 2020.
2. Slamon DJ, Godolphin W, Jones LA, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science*. 1989;244:707-712.
3. National Comprehensive Cancer Network: NCCN Clinical Practice Guidelines in Oncology: Breast Cancer.
4. Tolaney SM, Guo H, Pernas S, et al. Seven-year follow-up analysis of adjuvant paclitaxel and trastuzumab trial for node-negative, human epidermal growth factor receptor 2-positive breast cancer. *J Clin Oncol*. 2019;37:1868-1875.
5. Amiri-Kordestani L, Xie D, Tolaney SM, et al. A Food and Drug Administration analysis of survival outcomes comparing the Adjuvant Paclitaxel and Trastuzumab trial with an external control from historical clinical trials. *Ann Oncol*. 2020;31(12):1704-1708.
6. Slamon D, Eiermann W, Robert N, et al. Adjuvant trastuzumab in HER2-positive breast cancer. *N Engl J Med*. 2011;365:1273-1283.
7. Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med*. 2005;353:1673-1684.
8. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med*. 2005;353:1659-1672.
9. von Minckwitz G, Procter M, de Azambuja E, et al. Adjuvant pertuzumab and trastuzumab in early HER2-positive breast cancer. *N Engl J Med*. 2017;377:122-131.
10. Miller RS, Wong JL. Using oncology real-world evidence for quality improvement and discovery: the case for ASCO's CancerLinQ. *Future Oncol*. 2018;14:5-8.
11. FDA. Breaking down barriers between clinical trials and clinical care: incorporating real world evidence into regulatory decision making. Scott Gottlieb, Food and Drug Administration. Speech given January 28, 2019. Available at https://www.fda.gov/news-events/speeches-fda-officials/breaking-down-barriers-between-clinical-trials-and-clinical-care-incorporating-real-world-evidence. Accessed September 3, 2020.
12. Kumar A, Guss ZD, Courtney PT, et al. Evaluation of the use of cancer registry data for comparative effectiveness research. *JAMA Netw Open*. 2020;3:e2011985.
13. Rubinstein SM, Warner JL. Cancerlinq: origins, implementation, and future directions. *JCO Clin Cancer Inform*. 2018;2:1-7.
14. ATOP TRIAL: T-DM1 in HER2 positive breast cancer - full text view - ClinicalTrials.gov. Available at https://clinicaltrials.gov/ct2/show/NCT03587740. Accessed August 27, 2020.
15. CompassHER2-pCR: Decreasing chemotherapy for breast cancer patients after pre-surgery chemo and targeted therapy - full text view - ClinicalTrials.gov. Available at https://clinicaltrials.gov/ct2/show/NCT04266249. Accessed September 4, 2020.
16. Menefee ME, Gong Y, Mishra-Kalyani PS, et al. Project switch: docetaxel as a potential synthetic control in metastatic non-small cell lung cancer (mNSCLC) trials. *J Clin Oncol*. 2019;37:9105.