



Single-arm oncology trials and the nature of external controls arms

Mustafa Hashmi¹ , Jeremy Rassen²  & Sebastian Schneeweiss^{*,1,2,3} 

¹Department of Medicine, Division of Pharmacoepidemiology & Pharmacoeconomics, Brigham & Women's Hospital, Boston, MA 02120, USA

²Aetion, Inc., Boston, MA 02109, USA

³Harvard Medical School, Boston, MA 02115, USA

*Author for correspondence: Tel.: +1 617 278 0930; schneeweiss@post.harvard.edu

Journal of **Comparative Effectiveness Research**

Aim: Single-arm trials with external control arms (ECAs) have gained popularity in oncology. ECAs may consist of primary data from previous trials, electronic health records (EHRs) or aggregate data from the literature. We sought to provide a description of how such studies achieve similarity of patients, comparability of data quality and outcome assessment. **Materials & methods:** In a stratified convenience sample of 15 studies, five used primary data from trials as ECAs, five used secondary data from EHRs and five used aggregate data from the literature. Data were collected from the published literature and public web resources, blinded to the eventual approval decision. **Results:** Studies using ECAs from primary data and EHR data displayed methods to achieve comparability of information, including matched baseline characteristics. Aggregate data from published studies did not attempt to match covariates. The EHR controls often showed calendar time overlap for collecting information while trial data were mostly historic. Outcome data were not consistently reported across studies. US FDA approval was only seen when primary data from trials or EHR data were used as the ECA, however no ECA in this sample directly contributed to approval. **Discussion:** In this nonsystematic review of ECAs for single-arm trials, the ECAs derived from primary data collected by other trials or EHRs show patterns of patient comparability, time overlap, and realistic methodological approaches to achieving balance between treatment arms. They are often submitted to regulators while literature-derived aggregate findings as ECA may serve as benchmarks for pipeline decisions.

First draft submitted: 5 January 2021; Accepted for publication: 28 May 2021; Published online: 22 June 2021

Keywords: adaptive clinical trials • clinical trials • nonrandomized trials • oncology • real-world evidence • trial design

Parallel-group randomized controlled trials (RCTs) are cornerstones to establish causal treatment effects in drug development and approval of oncology products. The 21st Century Cures Act and the Cancer Moonshot initiative highlighted the importance of effectively accelerating the process of evaluating and approving cancer treatments [1]. RCTs typically take substantial time to accrue a patient population, require a meaningfully sized unexposed comparison group to come to robust conclusions, and need to justify ethical questions before randomizing patients to a placebo group [2]. Single-arm trials with external control arms (ECAs) have been used for decades as an alternative, largely in orphan drugs treating rare debilitating disease and now increasingly in highly targeted oncology drugs [3].

The function of an ECA is to provide data on the counterfactual experience had the patients in the experimental arm not been treated. ECAs, therefore, will need to emulate the experimental arm as closely as possible in all its aspects except the drug of interest. The lack of baseline randomization in single-arm trials with ECAs makes this design a nonrandomized cohort study and thus critical choices in the design of the ECAs are needed to ensure comparability of measurement and comparability of patients' baseline risk levels [4].

ECAs are typically based on either primary data from other clinical trial arms, secondary data assembled from electronic health records (EHRs) or aggregate effectiveness data from published literature, and can be formed either

from entire groups ('benchmark comparators') or by selecting individual patients ('individual-level comparators'). Biases can occur if the patient populations in the two groups are not comparable in their outcome risk at baseline, which can happen if the information content and quality is differential by treatment arm; or if the statistical process of making treatment groups comparable regarding their pre-exposure data are insufficient; or if important confounding factors are not observable; or if patients are recruited from different treatment eras where treatment choices differed. Bias can further occur if the two groups are not comparable in their outcome assessment, which can happen if the outcomes are assessed differently; or if the surveillance for the end point differs; or if (2c) patients are recruited from different treatment eras or treatment centers where generally outcome rates differ [4].

This review sought to provide an empirical description of how the various ECA types (primary data, secondary data and aggregate data) achieve similarity of patients, comparability of data quality and outcome assessment.

Materials & methods

Several search strategies were used during the literature search process to establish a convenience sample. Fifteen examples of single-arm trials using ECAs were identified, in which five used primary data from other trials, five used EHR data and five used aggregate results from the literature. These examples were all selected for review before knowing whether the ECA analysis contributed to a regulatory approval decision. Websites including PubMed.gov, clinicaltrials.gov, fda.gov, drugs.com, flatiron.com, cancer.gov and CenterWatch.com were searched for single-arm trials that use ECAs. First authors were contacted to obtain full-length posters and presentations if possible, in cases where we only found an abstract. Studies were then evaluated for comparability of data quality, and appropriateness of statistical methods to achieve balance. They were also examined in order to extract information on cancer type, US FDA regulation status and data collection timeframes for each experimental treatment and ECA.

The methods involved in obtaining data collection timeframes varied because some study authors had not explicitly listed timeframes publicly. Therefore, in an effort to accurately compare ECA structures, data collection timeframes were searched for on clinicaltrials.gov when not made available in the literature. Both the primary completion date and the study completion date were analyzed before choosing one for the study end date. Clinicaltrials.gov defines the primary completion date as the final date subjects were given treatment or examined for primary outcome measures. In contrast, the study completion date is defined as the final date patients were examined or given treatment for primary outcomes, secondary outcomes and adverse events. If the study was published prior to the study completion date, and after the primary completion date, then the primary completion date was used as the end date.

Search strategy

We searched the noted websites using the following search terms and/or following sections of the sites.

PubMed: 'single arm comparator', 'historical control single arm', 'single arm trial oncology', 'single arm trial electronic health record', 'single arm trial historical comparator', 'single arm trial external control', 'single arm oncology external comparison', 'comparative effectiveness single arm trial', 'comparative effectiveness single arm oncology', 'single arm trial real world data', 'single arm trial real world evidence'.

CenterWatch.com: 'Drug information → FDA approved drugs'.

FDA.gov: 'Home → Drugs → Development & Approval Process (Drugs) → Drug Approvals and Databases → Resources for Information on Approved Drugs'.

'Home → Drugs → Development & Approval Process (Drugs) → Drug Approvals and Databases → Drugs@FDA Search → Search by Drug Name, Active Ingredient or Application Number' → entered drug names for each experimental treatment, and searched summary reviews if there were any.

Flatiron.com: 'Publications'.

Drugs.com: 'Find Drugs and Conditions' → entered drug names for each experimental treatment and ECA to find more FDA approval information.

ClinicalTrials.gov: 'Find a study' → entered trial identification information found in each study for data collection timeframes.

Results

We identified 15 single-arm trials with ECAs from among the searched websites. Five of the studies contained ECAs consisting of primary data from trials [5–9], five contained ECAs based on EHR data [2,10–13] and five contained ECAs from aggregate data from the literature [14–18]. Table 1 shows that studies 1–10 accounted for baseline covariates,

Table 1. Comparing data quality and statistical methods to achieve balance among experimental treatments and their respective controls.

Study number Exp vs Ctrl	Comparability of data quality			Statistical methods to achieve balance			Ref.
	Data source of control [†]	Comparability of information to identify the comparison population	Any overlap in timing of data collection	Study size Exp: ECA	Adjusted variables	Covariate balancing methods	Check for achieved balance and authors' comment on balance quality
Studies 1–5 have external control arms consisting of primary data from trials							
Study 1: ceritinib vs crizotinib	Primary data collection of an external control arm from historical trial data (2 single-arm trials and 1 RCT)	- All trials incorporated into both study arms enrolled adult patients with locally advanced or metastatic ALK+ NSCLC - Inclusion criteria for both arms allowed stable brain metastases, and prior treatment for advanced or metastatic ALK+ NSCLC - Shared inclusion and exclusion analyses were made, in which patient data portrayed previous systemic therapy, and no prior ALK-targeted agent treatment - Authors claim there is potential confounding regarding unadjusted and unobserved differences in treatment groups - Authors claim to use study level data because control patient level data were not made publicly available	Minimal overlap	189: 557	- Age - Sex - Race - ECOG PS - Number of prior regimens - Tumor histologic type in ALK+ NSCLC	- Propensity score weighting balanced differences between the populations - Shared inclusion/exclusion criteria - Chi-square tests to conduct comparisons between baseline variables - Sensitivity analyses for baseline adjustment	[5] - Authors state all baseline characteristics were balanced after matching, however, before matching there were large differences regarding race and number of prior regimens - Authors claim to have balanced the once greater percentage of Asians in the experimental treatment population, and the larger number of pretreated patients in the external control population - Authors state that the designs across all the trials were similar, along with outcomes that were consistently defined
Study 2: blinatumomab vs S.O.C.	Primary data collection of an external control arm from existing clinical databases	- All patients had r/r Ph+ ALL - S.O.C. external control had the same eligibility criteria as the experimental study - Missing information in the external S.O.C. control that was available in the experimental study population included prior-graft-vs-host disease, duration of remission with prior allo-HSCT and ECOG PS	No overlap	45: 55	- Age at diagnosis and treatment - Sex - Time from diagnosis to most recent treatment (months) - Prior allo-HSCT status - Prior salvage therapy status - Number of prior salvage therapies - Geographic region	- Propensity score analysis to adjust for covariate balance differences - IPW methods used - Statistical power was enhanced using Bayesian data augmentation	[6] - Authors state that the study populations were relatively similar, however, there were discrepancies in prior treatments and geographic regions - Authors state that age and sex were the most balanced among the study populations - Authors also state chronological between the populations
Study 3: nilotinib vs dasatinib	Primary data collection of an external control arm from a prior RCT (DASISION)	- All patients had newly diagnosed CML-CP - The experimental treatment population portrayed lower counts for median platelets, higher age and a more favorable pre-match ECOG PS - Authors report missing information regarding adverse event rates due to reporting discrepancies	Large overlap	273: 259	- Gender - Age - ECOG PS - Hematology lab values	Baseline covariate matching through weighting patients in the experimental treatment arm	[7] - Authors state that after matching, all baseline characteristics were balanced across the two treatment groups - Authors state that any baseline covariates that were not measured in either or both trials, were not included in baseline matching

[†]The experimental arm is always primary data collection.[‡]There were three historical controls mentioned in the discussion, and the population size of just the PARAMOUNT experimental treatment arm was used here.

ACA: Adenocarcinoma; ALL: Acute lymphoblastic leukemia; allo-HSCT: allogeneic hematopoietic stem cell transplantation; CML-CP: Chronic myeloid leukemia in the chronic phase; ECA: External control arm; ECOG PS: Eastern Cooperative Oncology Group performance status; EHCC: Extrahepatic cholangiocarcinoma; EHR: Electronic health record; GBC: Gallbladder carcinoma; HR+: Hormone receptor-positive endocrine-refractory; IFN: Interferon-alfa; iHCC: Intrahepatic cholangiocarcinoma; IPTW: Inverse probability of treatment weighting; Ph+: Philadelphia chromosome-positive; mRCC: Metastatic renal cell carcinoma; MGMT: methylguanine-methyltransferase; NSCLC: Non-small-cell lung cancer; PFS: Progression-free survival; RCT: Randomized controlled trial; r/r: Relapsed/refractory; S.O.C.: Standard of care; S.P.B.C.: Standard platinum-based chemotherapy; TCR-MM: Triple class refractory multiple myeloma; TN: Triple negative.

Table 1. Comparing data quality and statistical methods to achieve balance among experimental treatments and their respective controls (cont.).

Study number Exp vs Ctrl	Comparability of data quality			Statistical methods to achieve balance			Ref.	
	Data source of control [†]	Comparability of information to identify the comparison population	Any overlap in timing of data collection	Study size Exp: ECA	Adjusted variables	Covariate balancing methods	Check for achieved balance and authors' comment on balance quality	
Study 4: bevacizumab + low dose IFN vs bevacizumab + IFN	Primary data collection of an external control arm from a prior RCT (AVOREN)	- Both studies involved patients with first-line mRCC - Authors performed across study safety comparisons - Authors state that no statistical comparisons were made with the control trial, however, descriptive and historical comparisons were done	No overlap.	146: 272	- Age - Race - Sex - ECOG PS - Tumor type - Number of metastatic sites - Number of lesions per patient - Location of metastases - Prior therapies	- No formal statistical comparisons were made, just across-study safety end point comparisons - Baseline covariates between the studies were displayed to compare them.	- Author states baseline characteristics were generally similar in both populations - Authors state potential differences between the populations could be due to the lack of data collection overlap, differences in treatment modalities, the BSC at each respective time and possible unobserved variables	[8]
Study 5: erlotinib + first-line pemetrexed chemotherapy vs erlotinib + first-line pemetrexed chemotherapy	Primary data collection of an external validation control from two prior RCTs (NCT00457392, NCT00364351)	- The control arm was used to validate experimental arm data, and it was generated using Project Data Sphere, which contained prior erlotinib trial data in patients with locally advanced or metastatic NSCLC as second-line treatment - Prior to controlling for population differences, the patient populations were similar in terms of smoking status, age and gender. However, the experimental treatment population had a lower average ECOG score, and only contained European patients - Authors state that the experimental treatment population had too small of a sample size	No overlap	54: 99	- Age - Race - Gender - Smoking history - ECOG PS - Disease stage	- Baseline characteristics, along with prior treatment, adverse events, and PFS for erlotinib treatment, were extracted from a total of 972 prior trial patients to create the control population of 99. - Controlled for differences in the patient populations	- Authors state that the external control population closely matched the experimental treatment population - Authors claim to have adjusted the differences in the populations regarding ECOG PS, and the imbalance in Europeans/race	[9]
Studies 6–10 have external control arms from EHRs, i.e., secondary data use								
Study 6: alectinib vs ceritinib	Secondary data collection of external control arm from the Flatiron EHR database	- The patient population consisted of adult patients with locally advanced (stage IIIB) or metastatic (stage IV) NSCLC, ALK rearrangement and progression of disease on crizotinib - Inclusion and exclusion criteria from the experimental arm was used for extracting the comparator arm from the Flatiron Health database - Authors claim key prognostic factors that were similar across populations were age, race, gender, prior lines of therapy, and stage at initial diagnosis	Some overlap	183: 67	- Age - Race - Gender - ACA histology - ECOG PS - CNS metastasis - History of smoking - Prior lines of therapy - Stage at initial diagnosis	- Logit propensity scores were applied to balance baseline covariates - Multiple sensitivity analyses - Naive comparison - GenMatch modeling	- Authors claim that all covariates were balanced after weighting. However, ECOG PS was not available consistently in the Flatiron database - Authors believe that there is a possibility of surveillance bias because of differences in data collection	[2]
†The experimental arm is always primary data collection. ‡There were three historical controls mentioned in the discussion, and the population size of just the PARAMOUNT experimental treatment arm was used here. ACA: Adenocarcinoma; ALL: Acute lymphoblastic leukemia; allo-HSCT: allogeneic hematopoietic stem cell transplantation; CML-CP: Chronic myeloid leukemia in the chronic phase; ECA: External control arm; ECOG PS: Eastern Cooperative Oncology Group performance status; EHCC: Extrahepatic cholangiocarcinoma; EHR: Electronic health record; GBC: Gallbladder carcinoma; HR+: Hormone receptor-positive endocrine-refractory; IFN: Interferon-alfa; IHCC: Intrahepatic cholangiocarcinoma; IPTW: Inverse probability of treatment weighting; Ph+ : Philadelphia chromosome-positive; mRCC: Metastatic renal cell carcinoma; MGMT: methylguanine-methyltransferase; NSCLC: Non-small-cell lung cancer; PFS: Progression-free survival; RCT: Randomized controlled trial; r/r: Relapsed/refractory; S.O.C.: Standard of care; S.P.B.C.: Standard platinum-based chemotherapy; TCR-MM: Triple class refractory multiple myeloma; TN: Triple negative.								

[†]The experimental arm is always primary data collection.

*There were three historical controls mentioned in the discussion, and the population size of just the PARAMOUNT experimental treatment arm was used here.

ACA: Adenocarcinoma; ALL: Acute lymphoblastic leukemia; allo-HSCT: allogeneic hematopoietic stem cell transplantation; CML-CP: Chronic myeloid leukemia in the chronic phase; ECA: External control arm; ECOG PS: Eastern Cooperative Oncology Group performance status; EHCC: Extrahepatic cholangiocarcinoma; EHR: Electronic health record; GBC: Gallbladder carcinoma; HR+: Hormone receptor-positive endocrine-refractory; IFN: Interferon-alfa; IHCC: Intrahepatic cholangiocarcinoma; IPTW: Inverse probability of treatment weighting; Ph+: Philadelphia chromosome-positive; mRCC: Metastatic renal cell carcinoma; MGMT: methylguanine-methyltransferase; NSCLC: Non-small-cell lung cancer; PFS: Progression-free survival; RCT: Randomized controlled trial; r/r: Relapsed/refractory; S.O.C.: Standard of care; S.P.B.C.: Standard platinum-based chemotherapy; TCR-MM: Triple class refractory multiple myeloma; TN: Triple negative.

Table 1. Comparing data quality and statistical methods to achieve balance among experimental treatments and their respective controls (cont.).

Study number Exp vs Ctrl	Comparability of data quality			Statistical methods to achieve balance			Ref.
	Data source of control [†]	Comparability of information to identify the comparison population	Any overlap in timing of data collection	Study size Exp: ECA	Adjusted variables	Covariate balancing methods	
Study 7: sunitinib + erlotinib vs erlotinib	Secondary data collection of an external control from the Flatiron EHR database.	– Patient populations had advanced or metastatic NSCLC with no more than two prior platinum-based chemotherapy treatments – Two groups were made with the EHR data, one that matched the trial data collection period, and one that was data post trial data collection	Overlap in the pre-2013 external control group only	480: 165	– Age – Sex – Race – ECOG – Disease stage – Histology – EGFR status – Smoking status – Prior Bevacizumab use – Number of prior chemotherapy regimens	– Propensity score matching algorithm – Inclusion and exclusion criteria to match populations	[10]
Study 8: entrectinib vs crizotinib	Secondary data collection of external control from the Flatiron EHR database	– Patient populations had ROS1 + NSCLC with no prior TKI inhibitor treatment, however, other forms of prior therapies and CNS metastases were permitted – Experimental arm contained patients with an ECOG PS of 0–2, however, the control arm included patients with any ECOG score or even a missing score	Some overlap	53: 54	– Age – Gender – Race/ethnicity – Smoking status – Clinical practice type – Histology – ECOG PS – Brain metastasis – Previous lines of therapy	– Propensity scores were balanced between arms – Experimental arm inclusion and exclusion criteria were applied to the EHR comparator – IPTW applied	[11]
Study 9: selinexor + dexamethasone vs TCR-MM patients not treated with Sd	Secondary data collection of an external control from the Flatiron EHR database.	– Patients gathered for the control arm mimicked the experimental arm – Patient populations had TCR-MM and were previously treated with carfilzomib, pomalidomide, bortezomib, lenalidomide and daratumumab (penta-exposed)	Some overlap	64: 36	– Age – Sex – Race – Carfilzomib-Pomalidomide-Daratumumab refractory prior to index date – Number of prior regimens – Exposure to anthracyclines or alkylating agents prior to index date – Stem cell transplant prior to index date – Baseline hemoglobin – Baseline platelets	– Experimental population inclusion and exclusion criteria were applied to both populations	[12]
Study 10: nivolumab for third-line or later therapy vs S.O.C.	Secondary data collection of an external control arm from the Flatiron EHR database	– Patient populations had advanced/metastatic gastric or gastroesophageal junction cancer and received either treatment as a third-line or later therapy – All patients had an ECOG score of 0/1, and laboratory values confirming adequate hematological and organ function – All patients met the experimental treatment eligibility criteria	Some overlap	42: 43	– ECOG – ALP and hemoglobin – Sex – Prior surgery – Tumor location	– Sensitivity analyses confirmed the adjustment	[13]

[†]The experimental arm is always primary data collection.

[‡]There were three historical controls mentioned in the discussion, and the population size of just the PARAMOUNT experimental treatment arm was used here.

ACA: Adenocarcinoma; ALL: Acute lymphoblastic leukemia; allo-HSCT: allogeneic hematopoietic stem cell transplantation; CML-CP: Chronic myeloid leukemia in the chronic phase; ECA: External control arm; ECOG PS: Eastern Cooperative Oncology Group performance status; EHCC: Extrahepatic cholangiocarcinoma; EHR: Electronic health record; GBC: Gallbladder carcinoma; HR+: Hormone receptor-positive endocrine-refractory; IFN: Interferon-alfa; IHCC: Intrahepatic cholangiocarcinoma; IPTW: Inverse probability of treatment weighting; Ph+: Philadelphia chromosome-positive; mRCC: Metastatic renal cell carcinoma; MGMT: methylguanine-methyltransferase; NSCLC: Non-small-cell lung cancer; PFS: Progression-free survival; RCT: Randomized controlled trial; r/r: Relapsed/refractory; S.O.C.: Standard of care; S.P.B.C.: Standard platinum-based chemotherapy; TCR-MM: Triple class refractory multiple myeloma; TN: Triple negative.

Table 1. Comparing data quality and statistical methods to achieve balance among experimental treatments and their respective controls (cont.).						
Study number Exp vs Ctrl	Comparability of data quality			Statistical methods to achieve balance		
	Data source of control [†]	Comparability of information to identify the comparison population	Any overlap in timing of data collection	Study size Exp: ECA	Adjusted variables	Covariate balancing methods
Studies 10–15 use aggregate data from the literature as an external control						
Study 11: pembrolizumab + capecitabine vs capecitabine	Aggregate data collection of a Phase III trial from multiple prior RCTs	– Patient populations had TN or HR+ metastatic breast cancer	No overlap	29: 27	– Only experimental treatment baseline demographics listed	– Only aggregate results were used for the control populations, no patient characteristics were compared with the experimental arm
Study 12: RT + TMZ, TMZ + irinotecan vs RT + TMZ, TMZ	Aggregate data collection from a Phase III trial and other trial data from the RTOG GBM database.	– Patient populations had adult patients with supratentorial glioblastoma diagnosed after surgery. – The experimental treatment did not require O6-MGMT analysis, therefore, there is missing information	No overlap	152: 287	– Only experimental treatment baseline demographics listed	– Only aggregate results were used for the control populations, no patient characteristics were compared with the experimental arm
Study 13: nab-paclitaxel + gemcitabine vs cisplatin + gemcitabine-diplatin	Aggregate data collection from prior trial data	– Patients populations contained adult patients (18 years and older) with confirmed IHCC, EHCC or GBC – Metastatic or locally advanced unresectable disease – Authors did not identify all patient characteristics that were used in the historical controls	No overlap	60: 204	– Only experimental treatment baseline demographics listed	– Only aggregate results were used for the control populations, no patient characteristics were compared with the experimental arm
Study 14: metformin + standard platinum-based chemotherapy vs S.P.B.C.	Aggregate data collection from prior trial data	– Patient populations contained nondiabetic adult patients with untreated advanced-stage (IIIB or IV) nonsquamous NSCLC	No overlap	14: 359 [‡]	– Only experimental treatment baseline demographics listed	– Only aggregate results were used for the control populations, no patient characteristics were compared with the experimental arm
Study 15: topotecan + bevacizumab vs topotecan + BSC	Aggregate data collection from a prior RCT	– Both patient populations had relapsed small-cell lung cancer patients with one prior chemotherapy regimen, as well as an ECOG PS of 0/1/2	No overlap	50: 71	– Only experimental treatment baseline demographics listed	– Only aggregate results were used for the control populations, no patient characteristics were compared with the experimental arm
[†] The experimental arm is always primary data collection. [‡] There were three historical controls mentioned in the discussion, and the population size of just the PARAMOUNT experimental treatment arm was used here. ACA: Adenocarcinoma; ALL: Acute lymphoblastic leukemia; allo-HSCT: allogeneic hematopoietic stem cell transplantation; CML-CP: Chronic myeloid leukemia in the chronic phase; ECA: External control arm; ECOG PS: Eastern Cooperative Oncology Group performance status; EHCC: Extrahepatic cholangiocarcinoma; EHR: Electronic health record; GBC: Gallbladder carcinoma; HR+: Hormone receptor-positive endocrine-refractory; IFN: Interferon- α ; IHCC: Intrahepatic cholangiocarcinoma; IPTW: Inverse probability of treatment weighting; Ph+: Philadelphia chromosome-positive; mRCC: Metastatic renal cell carcinoma; MGMT: methylguanine-methyltransferase; NSCLC: Non-small-cell lung cancer; PFS: Progression-free survival; RCT: Randomized controlled trial; r/r: Relapsed/refractory; S.O.C.: Standard of care; S.P.B.C.: Standard platinum-based chemotherapy; TCR-MM: Triple class refractory multiple myeloma; TN: Triple negative.						

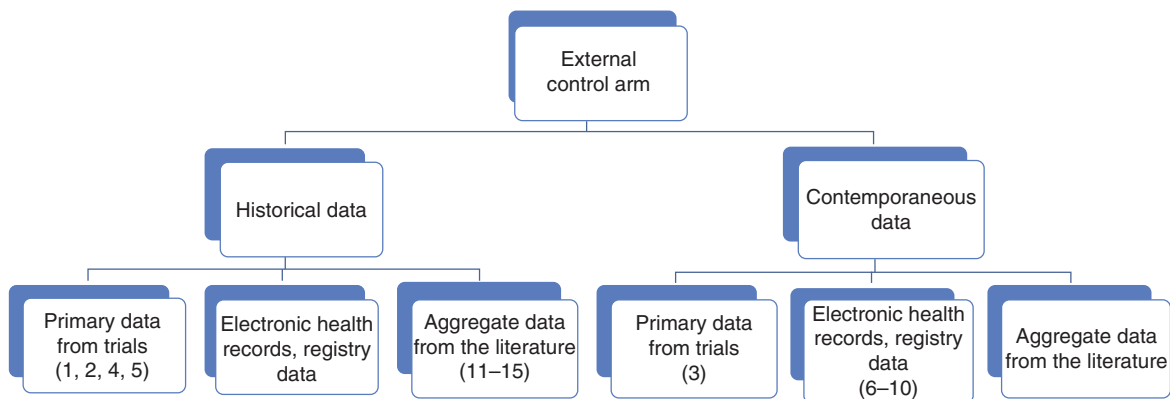


Figure 1. External control arm choices regarding temporality, and data collection mechanism. The numbers in the boxes refer to the example studies described in this article.

whereas studies 11–15 did not. It lists the comparability of data quality of the ECA to the experimental treatment, as well as statistical methods used in each study to achieve balance between the experimental treatment and the ECA. Study authors did not comment on balance quality, nor did they perform statistical tests to achieve balance in any of the aggregate data from the literature ECAs. Besides a few exceptions with studies 4 and 10, both authors' comments on balance quality and statistical tests to achieve balance were seen in studies that used the other two styles of ECAs. Population sizes varied greatly across studies, with some having more patients in the experimental arm and others having more patients in the ECA (Table 1).

Outcome data regarding overall response rate (ORR), median overall survival (OS) and median progression-free survival (PFS) were compared for each study's experimental treatment and ECA in Table 2. Studies varied greatly regarding documentation of each outcome value for the experimental treatment and ECA. Every study that reported outcome data showed improvements in experimental treatment outcome, with the exception of study 4 and study 7. Study 4 had a decrease in median OS and ORR compared with trial external control and study 7 had a decrease in median OS compared with EHR ECA.

Study contribution to regulatory approval and reasons for regulatory approval for the experimental treatment were also assessed. None of the five aggregate data ECAs had any regulatory implications (Table 2). Studies with ECAs from primary data had four experimental treatment approvals, with no documented contributions of the ECA to the approval decision, that is, they were approved based on their experimental arm alone without considering the ECA. Also, studies with ECAs based on EHR data had three experimental treatment approvals. Supplementary Table 1 lists the cancer types by study; non-small-cell lung cancer (NSCLC) was targeted in six studies, whereas all nine other cancers only had one study each that focused on them.

Figure 1 structures the temporal characteristics of ECAs (historical vs concurrent) stratified by data collection methods (primary vs secondary data) and places the 15 examples studies of this review into this framework. This layout shows that historical data are principally used for primary data from trials and aggregate data from the literature ECAs, whereas contemporaneous data are used for EHR and registry data ECAs (Figure 1). Figure 2 displays time overlap for each experimental treatment (blue) and ECA (red). Some experimental treatments and ECAs had multiple trials pooled together, therefore, in these cases all trial data collection timeframes were compiled using the earliest start data and the latest end date. The exact methods used for choosing each start and end date for each study are described in the footnote of Figure 2. Studies that re-used primary data from other trials were usually historical data from completed trials. Like in study 3, there are situations where control arms from an ongoing trial can be repurposed to serve as an ECA for another single-arm trial. This is the case in a well-planned ensemble of trials to maximize the efficiency of the development program. All studies that used EHR data used historical as well as concurrent ECA data. This is not surprising as the most effort in establishing ECAs in EHR data are identifying and accessing a fit-for-purpose data source and bringing data quality and completeness to the desired level. Studies that used published aggregate data on effectiveness as ECAs to benchmark single-arm trials always used historical data, sometimes with substantial lag time with an average of 3+ years.

Table 2. Comparing overall response rate, median overall survival, median progression-free survival and experimental treatment regulatory status between experimental treatments and their respective controls.

Study number	Experimental arm			External control arm			Contributed to approval decision?	Approved?	Reasons for regulatory decision	Ref.
	ORR (%)	Median OS (months)	Median PFS (months)	ORR (%)	Median OS (months)	Median PFS (months)				
Studies 1–5 have external control arms consisting of primary data from trials										
Study 1: ceritinib vs crinoline	68.3	Not reported	13.8	61.2	20.5	8.3	No	Yes	– US FDA granted accelerated approval[†] on 29 April 2014 based on the single-arm ASCEND-1 trial in this study. FDA approval was based on ORR, tumor response rate and duration of response. FDA approval and summary review letters confirmed – FDA granted expanded approval on 26 May 2017 based on the ASCEND-4 RCT which was not used in this study. Approval was based on improved PFS, ORR and treatment response compared with chemotherapy as a control – Therefore, the study that contributed to accelerated approval was used in order to pool patients into study 1	[5]
Study 2: blinatumomab vs S.O.C.	Not reported	7.1	Not reported	Not reported	6.0	Not reported	No	Yes	– On 11 July 2017, the FDA approved blinatumomab, which converted prior accelerated approval to full approval. Approval was based on single-arm ALCANTRA study, which was the trial used in experimental treatment arm of this study. Approval was based on remission rate, complete remission duration, MRD-negative remission numbers and partial hematologic recovery in two cycles with remission – Therefore, the study that contributed to approval was used in study 2 to further assess the results compared with a larger available treatment option external control group	[6]
Study 3: nilotinib vs dasatinib	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	No	Yes	– On 17 June 2010, FDA approved nilotinib for the treatment of adult patients (Ph+ CML) in chronic phase. Approval was based on results of the ENEST and RCT, which used in the experimental arm of this study. Nilotinib showed increased efficacy compared with imatinib, through eliminating Bcr-Abl faster. This led to lower rates of cancer progression in nilotinib – Therefore, the study that contributed to approval was used in study 3 to compare nilotinib to dasatinib	[7]
Study 4: bevacizumab + low dose IFN vs bevacizumab + IFN	28.8	30.7	15.3	35.9	25.8	10.5	No	Yes	– On 3 August 2009 FDA approved bevacizumab + IFN for patients with mRCC, which is the most common form of kidney cancer. Approval is based on the AVOREN RCT, which portrayed a median PFS of 10.2 months compared with 5.4 months in the IFN + placebo arm alone. This is equivalent to an 89% increase in median PFS – Therefore, the study that contributed to approval was used as the control used in study 4	[8]
Study 5: erlotinib + first-line pemetrexed chemotherapy vs erlotinib + first-line pemetrexed chemotherapy	0.0	5.8	1.8	Not reported	Not reported	1.9	No	No	– No approval documents available for this combination for the disease target – Therefore, this study had no contribution to approval	[9]
Studies 6–10 have external control arms from EHR data, i.e., secondary data use										
[†] Accelerated approval occurs when the FDA approves a drug that treats a serious life-threatening disease based on a clinical trial showing a surrogate end point that portrays clinical benefit. The FDA's Accelerated Approval Program allows favorable drugs to reach patients sooner if there is clear indication of effectiveness enhancements over prior treatments. However, a confirmatory study is required to provide evidence of clinical benefit of the treatment, once it is granted accelerated approval. The FDA can then give a treatment expanded approval. BSC: Best supportive care; CML: Chronic myeloid leukemia; EHR: Electronic health record; IFN: Interferon-alfa; NSCLC: Non-small-cell lung cancer; ORR: Overall response rate; OS: Overall survival; PFS: Progression-free survival; RCT: Randomized controlled trial; S.O.C.: Standard of care; S.P.B.C.: Standard platinum-based chemotherapy.										

Table 2. Comparing overall response rate, median overall survival, median progression-free survival and experimental treatment regulatory status between experimental treatments and their respective controls (cont.).

Study number	Experimental arm			External control arm			Contributed to approval decision?	Approved?	Reasons for regulatory decision	Ref.
	ORR (%)	Median OS (months)	Median PFS (months)	ORR (%)	Median OS (months)	Median PFS (months)				
Study 6: alectinib vs ceritinib	Not reported	24	Not reported	Not reported	16	Not reported	No	Yes	– FDA granted accelerated approval [†] of alectinib in December 2015. Approval was based on two single-arm trials which found an ORR of 38% in 87 patients and 44% in 138 patients. Alectinib portrayed clinical benefit through tumor response, and duration of response. FDA granted expanded approval of alectinib on 6 November 2017. Approval was based on the ALEX RCT that portrayed increased PFS compared with crizotinib	[2]
Study 7: sunitinib + erlotinib vs erlotinib	Not reported	8.5	Not reported	Not reported	9.33	Not reported	No	No	– No approval documents available for this combination for the disease target – Therefore, this study had no contribution to approval	[10]
Study 8: entrectinib vs crizotinib	Not reported	Not reported	19	Not reported	18.5	8.8	No	Yes	– FDA granted approval for entrectinib on 15 August 2019 for ROS1 + metastatic NSCLC. Approval was based on ORR and response duration benefits from multiple trials – The trials used in the experimental arm of this study were the studies that were listed as reasons for approval for entrectinib by the FDA (STARTRK-1, STARTRK-2, ALKA-372-001) along with STARTRK-NG	[11]
Study 9: selinexor + dexamethasone vs TCR-MM patients not treated with selinexor + dexamethasone	Not reported	10.4	Not reported	Not reported	5.8	Not reported	No	Yes	– FDA granted approval for selinexor + dexamethasone on 3 July 2019 based on the single-arm STORM trial. Approval was based on 25.3% ORR, median time to first response of 4 weeks and median response duration of 3.8 months – Therefore, the study which was used in the experimental arm of study 9 contributed to approval	[12]
Study 10: nivolumab for third-line or later therapy vs S.O.C.	Not reported	8.97	Not reported	Not reported	5.61	Not reported	No	No	– No approval documents available for this treatment and gastroesophageal junction cancer – Therefore, this study had no contributions to approval	[13]
Studies 10–15 use aggregate data from the literature as an external control										
Study 11: pembrolizumab + capecitabine vs capecitabine	14	15.4	4.0	Not reported	Not reported	3.0	No	No	– No approval documents available for this combination The combination was not found in labels or letters – Therefore, this study had no contribution to approval	[14]
Study 12: RT + TMZ, TMZ + irinotecan vs RT + TMZ, TMZ	Not reported	16.9	Not reported	Not reported	13.7	Not reported	No	No	– No approval documents available for this combination This combination was not found in labels or letters, only for the control – Therefore, this study had no contribution to approval	[15]
Study 13: nab-paclitaxel + gemcitabine-diplatin vs gemcitabine-diplatin	Not reported	19.2	11.8	Not reported	11.7	8	No	No	– No approval documents available for this combination This combination was not found in labels or letters – Therefore, this study had no contribution to approval	[16]
Study 14: metformin + standard platinum-based chemotherapy vs S.P.B.C.	23	11.7	3.9	Not reported	Not reported	Not reported	No	No	– No approval documents available for this combination and disease target – Therefore, this study has no contribution to approval	[17]
Study 15: topotecan + bevacizumab vs topotecan + BSC	16	7.4	4.02	Not reported	Not reported	Not reported	No	No	– No approval documents available for this combination for the disease target – Therefore, this study has no contribution to approval	[18]

[†] Accelerated approval occurs when the FDA approves a drug that treats a serious life-threatening disease based on a clinical trial showing a surrogate end point that portrays clinical benefit. The FDA's Accelerated Approval Program allows favorable drugs to reach patients sooner if there is clear indication of effectiveness enhancements over prior treatments. However, a confirmatory study is required to provide evidence of clinical benefit of the treatment, once it is granted accelerated approval. The FDA can then give a treatment expanded approval.

BSC: Best supportive care; CML: Chronic myeloid leukemia; EHR: Electronic health record; IFN: Interferon- α ; NSCLC: Non-small-cell lung cancer; ORR: Overall response rate; OS: Overall survival; PFS: Progression-free survival; RCT: Randomized controlled trial; S.O.C.: Standard of care; S.P.B.C.: Standard platinum-based chemotherapy.

Discussion

Oncology drug developers often employ single-arm trials with ECAs instead of the parallel-group RCTs to benchmark early studies to inform pipeline decisions and increasingly in regulatory submissions. Key reasons are that highly targeted treatments of advanced cancers will be suitable for only small numbers of patients, and randomizing control patients to standard of care when a promising experimental treatment exists may make study recruitment difficult and/or make the study unethical. Important considerations in planning ECAs include data quality and temporality of the data collection in ECAs relative to the experimental arm.

In our review, we found that those ECAs that borrow control arms or sometimes active arms from other trials have high data granularity that may be used for extensive adjustments to reach balance between treatment groups; usually they are historical data but well-orchestrated evidence development strategies may use recent trial data as ECAs. There is an added efficiency when using existing EHR data in that the time period of information collection is often longer covering historical and concurrent data, and thus, larger ECA sizes can be identified at greater speed. While these studies also address multiple predictor variables to balance treatment groups the difference in data quality between the two arms generally remains a concern that needs to be addressed and refuted in each study. Aggregate findings from previous studies may serve as benchmark for a rapid assessment in pipeline decision making but will have little success with regulators unless effect sizes are astonishing and the treatment progression highly predictable [19]. In our sample, they always rely on historical data, sometimes with substantial lag time.

The combination of limited time overlap and matching on patient predictors can negatively influence inferences drawn from ECA studies [10]. There was large variability in recorded outcomes among the studies, with no clear pattern for any of the ECA formats. Given the select number of studies in this review, we do not report a clear pattern regarding the regulatory impact of ECAs. Therefore, future reviews may focus on selecting a sample of studies that have ECAs with regulatory contributions.

Despite some patterns in this review, an important limitation is the small number of ECA studies examined. Further limiting this review was the inability to clearly compare data collection methods across the studies. Since every study did not list data timeframes distinctly, clinicaltrials.gov was used to fill in the holes for missing data collection timeframes. There seemed to be discrepancies among the different study completion and primary completion dates which could have affected the data collection timeframe results slightly.

It is also important to consider the potential trade-off between achieving contemporaneous data, and complete covariate and outcome data between the arms. Although temporality and comparable date completeness are both important for inter-arm comparisons, future studies may clarify which should be preferred in what circumstances. Due to the small size of our study sample and variability in documentation, it is difficult to see a clear pattern.

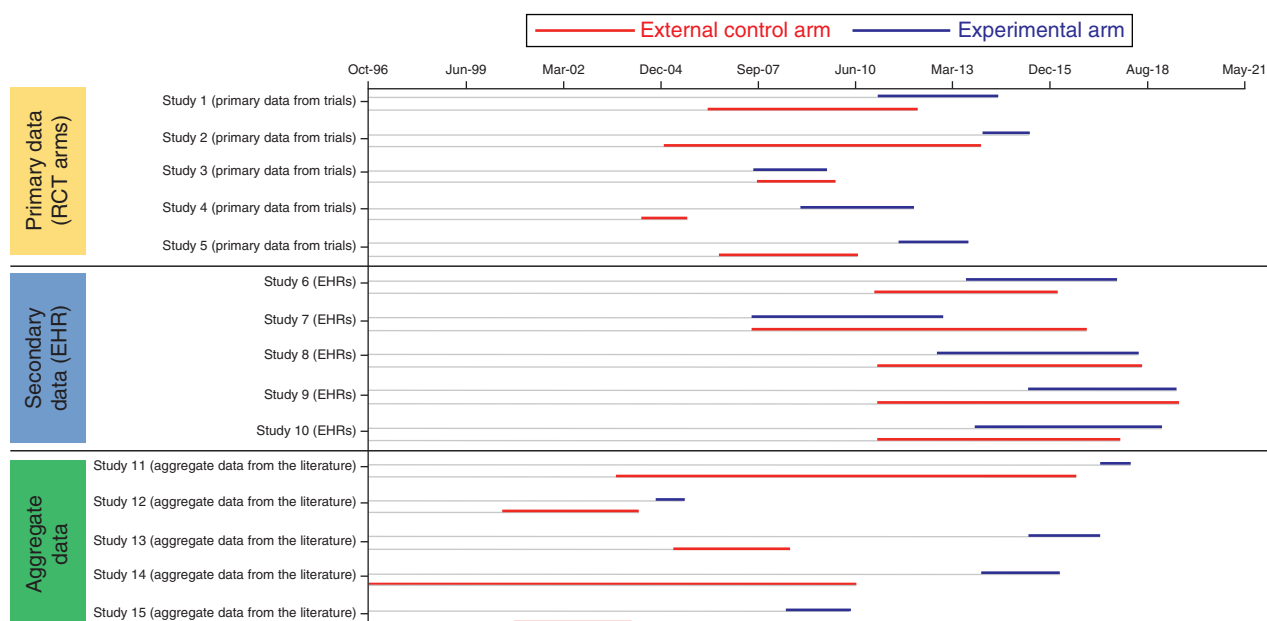


Figure 2. Temporal relationship between experimental arm and external control arm in 15 example studies (cont.).

- 1) Study 1 contained multiple trials that compiled both the experimental treatment arm and the external control containing primary data from trials, thus, leading to multiple study start and end dates. Therefore, the earliest of start dates located on www.clinicaltrials.gov was used for the compiled start date, and the latest of the cut-off dates found in the literature was used for the compiled end date.
 - 2) Study 2 had both the start and end dates for the experimental treatment arm listed in the literature. This study also utilized a clinical trial database for the external control comprising primary data from trials, but they did not specify the data collection timeframe. However, the authors did state that some of the patients were treated 9 years before the start of the experimental treatment start. Therefore, 9 years prior to the start of the experimental treatment start was used as the external control start date, and the start of the experimental treatment time was used as the external control end date.
 - 3) Study 3 data collection timeframes were found using the trial start dates and primary study completion dates on www.clinicaltrials.gov for both the experimental treatment and the external control incorporating primary data from trials. The primary study completion dates were analyzed instead of the study completion dates because study 3 was released after the primary completion and before the study completion of the experimental treatment trial.
 - 4) Study 4 had experimental treatment and external control containing primary data from trials data collection timeframes all found in the literature.
 - 5) Study 5 had experimental treatment data collection timeframes found in the literature. Whereas, the earliest study start date and the primary completion date for the two trials that formed the external control composed of primary data from trials arm were taken from www.clinicaltrials.gov. The study completion date listed on the website for the external control containing primary data from trials was not used because it went past the experimental treatment end date listed in the literature.
 - 6) Study 6 provided data collection timeframes for the EHR external control, however, not for the experimental treatment. Therefore, www.clinicaltrials.gov was used to find the earliest study start date and the latest study completion date of the two trials that comprised the experimental treatment arm. The primary completion date listed was not used because then the external control arm would have a larger data collection timeframe.
 - 7) Study 7 had two categories for the EHR external control arm, a pre-January 1st, 2013 and a post 1 January 2013 group. The start date of the pre-2013 group and the end date of the post-2013 were not specified. To fill in the missing start date of the pre-2013 group, the experimental arm start date from www.clinicaltrials.gov was used. The missing end date of the post 2013 group on the other hand was filled through the first day of the year the paper was published. The study start date and the study completion date of the experimental arm trial were used from www.clinicaltrials.gov because they were not located in the literature. Also, the actual completion date was used instead of the primary completion date because the pre-2013 control group was made in order to match the experimental arm, which had a study completion date in December 2012.
 - 8) Study 8 had experimental and EHR external control data collection timeframes all found in the literature.
 - 9) Study 9 only listed the start date for the EHR external control in the literature. The experimental treatment study start date and completion date (same as primary completion date) were found on www.clinicaltrials.gov. The study completion date of the experimental treatment was used to define the completion date for the EHR external control.
 - 10) Study 10 did not list data collection timeframes in the publicly available abstract, however, a paper was sent directly by the authors containing the EHR external control timeframes. The experimental treatment study timeframes were found on www.clinicaltrials.gov, and the primary completion date was used because the study completion date is in 2022.
 - 11) Study 11 only listed experimental treatment data collection timeframes. The external control comprised aggregate data from the literature timeframes were not listed clearly in the literature, therefore, each cited external control including aggregate data from the literature was searched for on www.clinicaltrials.gov for study start dates and primary completion dates. The study completion dates were not used because the latest end date was past the end date of the experimental treatment.
 - 12) Study 12 had the experimental arm data collection timeframes listed in the literature. The cut-off date for the external control containing aggregate data from the literature was found in the cited literature, and the start date was found by inputting the trial identification into www.clinicaltrials.gov.
 - 13) Study 13 had the experimental arm data collection timeframes listed in the literature. For the external control incorporating aggregate data from the literature, the trial identification from the cited literature was placed into www.clinicaltrials.gov in order to collect the study start date and the primary completion date. The study completion date was not listed, therefore, the primary completion was used.
 - 14) Study 14 had the experimental arm data collection timeframes listed in the literature. The three external control involving aggregate data from the literature timeframes listed in the cited literature of study 14 were compiled into one group with the earliest start date and latest completion date.
 - 15) Study 15 did not list any data collection timeframes in the literature. The experimental treatment arm trial identification placed into www.clinicaltrials.gov in order to receive the study start date and study completion date which was the same as the primary completion date. The external control comprised aggregate data from the literature timeframes were listed in the cited literature of study 15.
- EHR: Electronic health record; RCT: Randomized controlled trials.

Overall, the published literature on ECAs in modern oncology products is thin and often lacks details making it difficult to assess the validity of study design and findings. Regulatory submissions seem to favor reusing control arms from other trials, which increases comparability because of the primary data collection leading to more similar data granularity but they are often hard to find or have meaningful lag time, as well as EHR data that are available in large quantity in no or limited time lag but raise concerns about data quality.

Conclusion

This review of single-arm trials in oncology with focus on the type and composition of ECAs, illustrates that single-arm trials that use ECAs consisting of primary data from prior trials or secondary data from EHRs show patterns of patient comparability, calendar time overlap (especially in EHR ECAs) and statistical methods to achieve balance through baseline covariate matching. At the same time, ECAs based on aggregate data from the literature seem less concerned with achieving balance in baseline patient characteristics. This could raise concerns about the ability to draw concrete conclusions for treatment efficacy based on outcome improvement from these studies.

Future perspective

In terms of how the field will evolve over the coming years, the expansion of available literature on single-arm trials with ECAs will certainly create the potential for larger reviews and individual contributions of ECA studies to approval. Furthermore, patterns may begin to develop with regard to prioritizing temporality, complete outcome data or matched patient populations when creating an effective ECA. Along with a wider scope in available literature comes the potential for evaluating whether single-arm trials with ECAs could be refined or verified by prospective RCTs to reproduce the results of the ECA study. If prospective studies confirm viability, this would further encourage the utilization of ECAs when favorable, which could also lead to more contributions of ECA studies to approval.

Summary points

- Single-arm trials with external control arms (ECAs) may provide an alternative to randomized controlled trials due to randomized controlled trial population size, time of patient accrual and ethical questions that arise when assigning patients to the placebo group.
- The purpose of an ECA is to emulate the patient population of the experimental arm with all aspects except for the drug of interest, however, it must do so without baseline randomization.
- ECAs are usually formed using either primary data from other clinical trial arms, secondary data assembled from electronic health records (EHRs) or aggregate effectiveness data from published literature.
- This nonsystematic review sought to provide a description of how five examples of three respective ECA types (primary data from prior trials, secondary data from EHRs and aggregate data) achieve similarity of patients, comparability of data quality and outcome assessment.
- Temporal characteristics of ECAs show the principal use of historical data for primary data from trials and aggregate data ECAs, and contemporaneous data for secondary data from EHR ECAs.
- Single-arm trials with ECAs from prior trials and EHRs accounted for baseline covariates through statistical methods to achieve balance and patient comparability, whereas aggregate data ECAs did not.
- Using EHR data to create an ECA increases efficiency in that the time period of information collection is often longer covering historical and concurrent data, and thus, larger ECA sizes can be identified at greater speed.
- Outcome data were not consistently reported in a manner in which a clear pattern could be seen with this sample of ECA studies.
- There were no direct contributions to approval in any of the ECA studies, therefore, future reviews can focus on selecting a sample of studies that have ECAs with regulatory contributions.
- Despite some patterns in this review, important limitations include the small number of ECA studies examined, and the inability to clearly compare data collection methods across the studies.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/ce-2021-0003

Financial & competing interests disclosure

This study was funded by the Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, MA, USA. S Schneeweiss is participating in investigator-initiated grants to the Brigham

and Women's Hospital from Bayer, Vertex and Boehringer Ingelheim unrelated to the topic of this study. He is a consultant to Aetion Inc., a real-world evidence software company in which he owns equity. His interests were declared, reviewed and approved by the Brigham and Women's Hospital and Partners HealthCare System in accordance with their institutional compliance policies. J Rassen is an employee of Aetion Inc., a software manufacturer of which he owns equity. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as: ● of interest; ●● of considerable interest

- Carrigan G, Whipple S, Capra WB *et al.* Using electronic health records to derive control arms for early phase single-arm lung cancer trials: proof-of-concept in randomized controlled trials. *Clin. Pharmacol. Ther.* 107(2), 369–377 (2019).
- **Presents a proof of concept for the implementation of a single-arm trial with an external control arm using contemporaneous electronic health record data.**
- Davies J, Martinec M, Delmar P *et al.* Comparative effectiveness from a single-arm trial and real-world data: alectinib versus ceritinib. *J. Comp. Eff. Res.* 7(9), 855–865 (2018).
- **Prepares a proof of concept for the generation of a comparator arm using real world data (RWD) in a rare population for which an randomized controlled trial would not be feasible, by comparing alectinib and ceritinib.**
- Sasinowski FJ, Panico EB, Valentine JE. Quantum of effectiveness evidence in FDA's approval of orphan drugs: update, July 2010 to June 2014. *Ther. Innov. Regul. Sci.* 49(5), 680–697 (2015).
- Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and external controls in clinical trials – a primer for researchers. *Clin. Epidemiol.* 12, 457–467 (2020).
- **Provides an assessment and critique of the utilization of synthetic control methods in the published literature.**
- Tan DSW, Araújo A, Zhang J *et al.* Comparative efficacy of ceritinib and crizotinib as initial *ALK*-targeted therapies in previously treated advanced NSCLC: an adjusted comparison with external controls. *J. Thorac. Oncol.* 11(9), 1550–1557 (2016).
- **Utilizes an external control group using primary data from trials to compare efficacy outcomes between ceritinib and crizotinib for patients with ALP-positive non-small-cell lung cancer.**
- Rambaldi Am Ribera JM, Kantarjian HM *et al.* Blinatumomab compared with standard of care for the treatment of adult patients with relapsed/refractory Philadelphia chromosome-positive B-precursor acute lymphoblastic leukemia. *Cancer* 126(2), 304–310 (2019).
- **Creates an external control cohort using existing primary data from clinical databases in order to compare blinatumomab and standard of care treatments.**
- Signorovitch JE, Wu EQ, Betts KA *et al.* Comparative efficacy of nilotinib and dasatinib in newly diagnosed chronic myeloid leukemia: a matching-adjusted indirect comparison of randomized trials. *Curr. Med. Res. Opin.* 27(6), 1263–1271 (2011).
- Melichar B, Bracarda S, Matveev V *et al.* A multinational Phase II trial of bevacizumab with low-dose interferon- α 2a as first-line treatment of metastatic renal cell carcinoma: BEVLIN. *Ann. Oncol.* 24(9), 2396–2402 (2013).
- Germonpré P, Wyngaert TVD. Second-line erlotinib after failure of pemetrexed-containing chemotherapy in advanced non-small cell lung cancer (NSCLC): real-world effectiveness, safety and tolerability. *PLoS ONE* 14(4), e0215135 (2019).
- Patel K, Ouwens M, Shire N, Kholsa S. The application of electronic medical records (EMRs) as a virtual comparator arm in a lung cancer clinical trial: a case study. *J. Clin. Oncol.* 35(15), e18098 (2017).
- **Illustrates the importance of time point for electronic health record data used as an external control when comparing overall survival data between treatments.**
- Doebele RC, Perez L, Trinh H *et al.* Time-to-treatment discontinuation (TTD) and real-world progression-free survival (rwPFS) as endpoints for comparative efficacy analysis between entrectinib trial and crizotinib real-world *ROS1* fusion-positive (*ROS1*+) NSCLC patients. *J. Clin. Oncol.* 9070–9070 (2019).
- Richardson PG, Jagannath S, Chari A *et al.* Overall survival (OS) with oral selinexor plus low dose dexamethasone (Sd) in patients with triple class refractory-multiple myeloma (TCR-MM). *J. Clin. Oncol.* 37(15), 8014 (2019).
- Chau I, Ayers D, Goring S *et al.* Comparative effectiveness of nivolumab (NIVO) relative to standard of care (SOC) for advanced/metastatic (adv/met) gastric or gastroesophageal junction cancer (GC/GEJC): a simulated treatment comparison (STC). *J. Clin. Onco.* 37(4), 128 (2019).
- Shah AN, Flaum LE, Rademaker A *et al.* A Phase II study of pembrolizumab and capecitabine for triple-negative (TN) and hormone receptor-positive, HER2-negative endocrine-refractory metastatic breast cancer (MBC). *J. Clin. Oncol.* 8(1), 1096 (2019).
- **Produces an example of a single-arm trial utilizing an external control from aggregate data from the literature comparing pembrolizumab and capecitabine to capecitabine alone.**
- Lieberman FS, Wang M, Robins HI *et al.* Phase II study of radiation therapy plus low-dose temozolomide followed by temozolomide and irinotecan for glioblastoma: NRG Oncology RTOG trial 0420. *Int. J. Radiat. Oncol. Bio. Phys.* 103(4), 878–886 (2019).

●● **Formulates a comparison between two treatments for glioblastoma using an external control from aggregate data from the literature.**

16. Shroff RT, Javle MM, Xiao L *et al.* Gemcitabine, cisplatin, and nab-paclitaxel for the treatment of advanced biliary tract cancers: a Phase II clinical trial. *JAMA Oncol.* 5(6), 824–830 (2019).
17. Parikh AB, Kozuch P, Rohs N, Becker DJ, Levy BP. Metformin as a repurposed therapy in advanced non-small cell lung cancer (NSCLC): results of a Phase II trial. *Invest. New Drugs* 35(6), 813–819 (2017).
18. Spigel DR, Waterhouse DM, Lane S, Legenne P, Bhatt K. Efficacy and safety of oral topotecan and bevacizumab combination as second-line treatment for relapsed small-cell lung cancer: an open-label multicenter single-arm Phase II study. *Clin. Lung Cancer.* 14(4), 356–363 (2013).
19. Eichler HG, Oye K, Baird LG *et al.* Adaptive licensing: taking the next step in the evolution of drug approval. *Clin. Pharmacol. Ther.* 91(3), 426–437 (2012).