Journal of Medical Economics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ijme20

Using electronic health record data to identify comparator populations for comparative effectiveness research

Scott D. Ramsey, Blythe J. Adamson, Xiaoliang Wang, Danielle Bargo, Shrujal S. Baxi, Shuhag Ghosh & Neal J. Meropol

To cite this article: Scott D. Ramsey, Blythe J. Adamson, Xiaoliang Wang, Danielle Bargo, Shrujal S. Baxi, Shuhag Ghosh & Neal J. Meropol (2020) Using electronic health record data to identify comparator populations for comparative effectiveness research, Journal of Medical Economics, 23:12, 1618-1622, DOI: 10.1080/13696998.2020.1840113

To link to this article: https://doi.org/10.1080/13696998.2020.1840113

Published online: 10 Nov 2020.

JOURNAL OF MEDICAL ECONOMICS
2020, VOL. 23, NO. 12, 1618–1622
https://doi.org/10.1080/13696998.2020.1840113
Article 0163-RT.R1/1840113

Taylor & Francis
Taylor & Francis Group

ARTICLE COMMENTARY

Using electronic health record data to identify comparator populations for comparative effectiveness research

Scott D. Ramsey[a], Blythe J. Adamson[b,c], Xiaoliang Wang[b], Danielle Bargo[b], Shrujal S. Baxi[b], Shuhag Ghosh[b] and Neal J. Meropol[b]

[a]Fred Hutchinson Cancer Center, Hutchinson Institute for Cancer Outcomes Research, Seattle, WA, USA; [b]Flatiron Health, Inc, New York, NY, USA; [c]The Comparative Health Outcomes, Policy, and Economics (CHOICE) Institute, University of Washington, Seattle, WA, USA

ABSTRACT

Electronic health records (EHRs) can define real world patient populations with high levels of clinical specificity, potentially addressing some of the shortcomings of other types of real world data (RWD) when informing decisions about the comparative effectiveness of medical technologies. An important but under-recognized concern for EHR-derived RWD, however, is that the rich clinical data permits creation of very homogenous subpopulations from the larger group of eligible patients, thereby reducing the representativeness of the cohort relative to clinical practice. In this article, we discuss the tradeoffs between choosing clinical specificity versus representativeness in population sampling for comparative effectiveness research. Using EHR-derived RWD, we provide an example in non-small cell lung cancer to illustrate the concepts, showing wide variation in outcomes among potential comparator cohorts. We close with several recommendations for selecting comparator populations from EHRs that address the balance between matching clinical guidelines and capturing practice variability in comparative effectiveness research.

ARTICLE HISTORY
Received 14 August 2020
Revised 2 October 2020
Accepted 14 October 2020

KEYWORDS
Electronic health records; comparators; health technology assessment; Real World Data

JEL CLASSIFICATION CODES
I11; C49

Introduction

One of the most compelling aspects of electronic health records (EHRs) is their ability to define real world patient populations with high degrees of clinical specificity relative to other real-world databases such as insurance claims or registries. Because EHRs in theory capture nearly all of the major clinical factors that clinicians consider when making therapy choices in practice, when available at scale, they can identify appropriate comparator populations for comparative effectiveness research with high degrees of clinical detail as well as sufficient sample size for evaluating outcomes that are meaningful as population studies[1–3]. These characteristics have the potential to make EHR-based studies highly valuable for comparative effectiveness researchers and health technology assessment (HTA) decision-makers.

It is important to acknowledge, however, that the same feature that represents the principal benefit of EHR-derived data also creates opportunities to *reduce* the representativeness of the comparator population relative to clinical practice. This can happen when the rich clinical data in EHRs are used to reduce practice variability and potential confounding factors by limiting the analysis to very homogenous subpopulations. One example is selecting patients who very closely resemble clinical trial enrollees – both in terms of eligibility criteria and care – in essence simulating an efficacy study (i.e. a clinical trial) versus a study of use in real world practice. Another example is restricting the comparator to patients who received the "right" treatment; that is, those whose care is consistent with clinical practice guideline recommendations. Both approaches skew analyses towards evaluating comparative *efficacy* (impact in optimized situations) versus comparative *effectiveness* (impact in real world settings).

The unique advantages and potential concerns regarding use of the EHR for comparative effectiveness research have not been widely discussed. Here, we describe a process for using EHR-derived data to define real-world comparator populations to better understand the comparative effectiveness of new therapies. We begin by describing factors that distinguish efficacy versus effectiveness as products move from clinical trials to clinical use in routine care, and how these factors translate into choices for specifying comparator populations from EHRs. We use common scenarios in cancer to illustrate general issues and provide an example in non-small cell lung cancer to demonstrate how modifying the definition of a comparator can impact survival estimates for a comparator. Finally, we provide practical suggestions for specifying comparator populations to generate estimates of comparative efficacy and effectiveness.

CONTACT Scott D. Ramsey sramsey@fredhutch.org Fred Hutchinson Cancer Research Center, Hutchinson Institute for Cancer Outcomes Research, M3-B232, Seattle, WA 98109, USA

www.tandfonline.com/ijme

Conceptual issues

Recommended versus observed care

Because the clinical breadth of the EHR includes nearly all of the clinical factors that are listed in product labels and practice guidelines, identifying populations for comparative effectiveness research that match them is, in theory, feasible and efficient. Focusing on matching this way, however, ignores an important issue that can jeopardize the external generalizability of EHR-derived RWE studies: clinical practice deviates from recommendations. Scores of studies have shown that some patients who receive cancer drugs do not meet criteria described in practice guidelines, and some who don't are in fact eligible for those treatments. The observed variations can be due to knowledge gaps, particularly when prescribing information is changing rapidly, or because clinicians sometimes consider factors that are not included in product labels or practice guidelines – for example, comorbidities, patient preferences, and availability of family support – when making treatment choices.

"Optimized" versus "realized" real world comparators

Most EHR's contain sufficient clinical detail to allow wide latitude when constructing a comparator group. For example, consider a comparative effectiveness study that seeks to use EHR-derived RWD to compare patients who receive a well-established, guideline recommended therapy against a very newly-approved treatment that is an alternative for patients who are eligible for the established therapy. At one extreme, eligibility criteria from the guideline (e.g. tumor morphology, genomic criteria, contraindications) can be applied to select patients from a larger population, with further restriction to select those who actually receive the treatment(s) at the recommended doses and duration. We describe this population as an "optimized" real-world comparator[4]. In cases where 'best supportive care' may be an acceptable alternative, these patients should be included in the optimized comparator population, with an option to exclude them in a subanalysis of treated patients. In contrast to optimized real-world comparators, "realized" real-world comparators might include multiple populations that do not fully meet eligibility criteria, for example: (1) patients who meet eligibility criteria but don't receive the recommended treatment(s); (2) those who receive recommended treatment(s) yet do not fully meet eligibility criteria, and; (3) those who receive the recommended treatment, but with nonstandard dosing or treatment durations. Optimized and realized real-world comparators will differ when observed treatment patterns vary substantially from guidelines or when there are multiple guidelines with somewhat different recommendations. For example, studies have shown that between 55% and 95% of prescribing of colony stimulating factors is inconsistent with expert guidelines[5–9]. In such cases, constructing an optimized (guideline-informed) cohort will produce a markedly different comparator population versus other sampling approaches, the former pointing towards a comparative *efficacy* study and the others more relevant for understanding comparative *effectiveness*.

Understanding the representativeness of EHR-derived comparator populations

Available EHR-derived population versus the larger patient experience

Beyond optimized and realized populations, it is important to consider whether the cohort that forms the EHR-derived database is representative of the population that is the focus of decision-making. Care patterns may differ between the cohort and the intended comparator population. EHR databases representing single centers or academic institutions, for example, could suffer both problems: a non-representative patient population and patterns of treatment that do not represent the larger oncology community.

In some cases, the differences between the available cohort and the larger population are known, in others, there is limited ability to know whether the available cohort is representative of the larger whole. If feasible, population comparability can be partially addressed by sampling from both cohorts to examine sociodemographic and clinical characteristics of both populations. Patterns of care may be evaluated by sampling and evaluation of key care components (e.g. first line chemotherapy rates).

Very small populations

In some cases, there are too few comparator patients available in real-world samples to reliably estimate comparative effectiveness. For example, developing a comparator population for the neurotrophic tropomyosin receptor kinase (NTRK) fusion-targeted drug larotrectinib (which received licensing approval after a single-arm study) at product launch would require sampling from a population of patients with tumors identified as having NTRK gene fusions who were not treated with larotrectinib. Because routine NTRK fusion testing was uncommon prior to the launch of larotrectinib, identifying a comparator population will be difficult[10].

Evolving practice patterns

Another problematic situation for selecting a comparator population is when patterns of treatment are evolving rapidly; for example, when a newly introduced treatment is rapidly replacing existing treatments. In this case, as with rare cancers, retrospective real-world data might have insufficient numbers of patients that represent contemporary practice. In general, because practice continually evolves, databases become less representative of the current patient experience as the gap between the last observed patient and the present time grows.

Considerations when using EHR-derived data to select real-world comparators

When selecting appropriate real-world comparator population(s) for comparative effectiveness research, it is important to consider many issues, including the larger purpose of the analysis, the decision makers' perspectives, and the primary

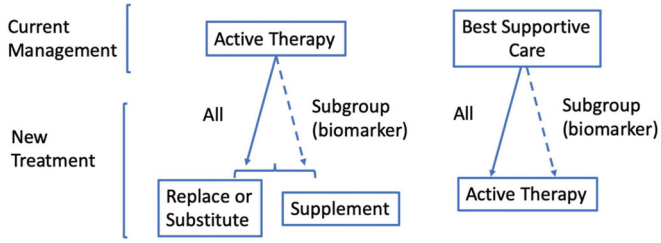

Figure 1. Clinical context for adoption of a new therapy.

question of interest (e.g. is technology A safer and more effective than technology B?)[11]. Below we describe three salient features: (1) clinical context; (2) the objectives of the comparison, and; (3) the availability of factors used to select the comparator population, particularly potential confounders.

Clinical context

Because of the high level of clinical specificity afforded by the EHR, it is useful to first consider the clinical context; that is, the relevant clinical setting for the new treatment and comparator. Relevant issues for clinical context include the patient population, indication and contraindications to the alternative treatments (which, for the great majority of HTA submissions, are new treatments) and intended use of the new therapy relative to the comparator (Figure 1). In some cases, the alternative is intended to be a *substitute or replacement* for existing treatments, intended for *all* patients receiving an existing therapy, or for a subgroup, defined by clinical or biological characteristics (e.g. a genomic variant). Some new therapies are intended to *supplement* an existing treatment, either provided concurrently or prior to use of an existing therapy. Lastly, an alternative treatment can be intended for those who would otherwise receive supportive care only; for example, in patients whose disease has progressed after receiving all recommended treatments.

Because of the heterogeneity of trials and the fact that a product label is a synthesis of evidence from trials, trial-based comparators have significant limitations for understanding effectiveness of a new product in clinical practice. To address this limitation, real world comparators from EHR-derived databases that meet eligibility and treatment recommendations that are consistent with practice guidelines may be a valuable adjunct for decision-makers. While such comparators are closer to clinical practice than trial-based meta-analyses (and thus may have superior internal representativeness), they can ignore important factors that influence practice guidelines. This is a particular problem if guidelines recommend alternative therapies within a class of drugs of which the comparator is one option, because the alternative drugs have differential benefit over the comparator. Selecting "weak comparator" drugs that guidelines recommend against limits external representativeness of these studies.

Objective of the comparative effectiveness study: optimized or realized populations?

Given the amount of discretion researchers have in selecting patient populations from EHRs, it is worth considering the implications of choosing optimized versus realized populations. An optimized population may provide useful information if the purpose of the study is to understand incremental benefits and harms of a newly approved therapy in a best-case scenario. This may be particularly helpful for treatments that have been approved without randomized controlled trials–an increasingly common issue for cancer therapies[12]. On the other hand, restricting real-world populations in this way runs the risk of creating false expectations regarding the incremental benefit of new treatment in practice. When there is little difference between optimized and realized reference populations the distinction is minimized. If, however, the EHR reveals that substantial proportions of patients who either meet criteria and are not treated or are treated but only partially meet criteria, then the choice between optimized and realized populations takes on more significance, because the heterogeneity of both the patient populations and treatment effects will differ between groups.

Availability and completeness of relevant clinical factors in the EHR

Incomplete records include those where the factor of interest wasn't recorded in the EHR at the time it was used to make a decision or when it was not considered at all. For example, studies show that functional status (e.g. ECOG performance status), while commonly accepted as a quality metric, is recorded for only about 60% of the cancer patients who receive systemic therapy[13]. If lack of recording of performance status reflects systematic bias (e.g. clinicians less likely to record scores for patients with better performance status), ignoring this issue when selecting patients based on this factor will result in a biased comparator population relative to the larger real world population. Prior evaluation of completeness of key components of the record that are used to select the comparator population and tests for whether the missing data are likely to be "at random" or correlated with other factors is critical to address potential concerns about incompleteness of records.

While the EHR is a uniquely comprehensive record of clinical factors that influence treatment choice and outcomes, it is important to consider whether other patient factors that are not commonly included in the EHR may be influencing decision-making and/or observed outcomes. As noted, financial factors, availability of family support, or stated preferences of family members may influence observed choices and outcomes. Such unmeasured confounding factors can threaten representativeness of findings. Discussions with providers about the relevance of these issues, as well as use of statistical techniques to address unmeasured confounding (example: instrumental variables) may be necessary to address these concerns.

Case study: an EHR-based comparator in advanced non-small cell lung cancer

To illustrate the impact of alternative approaches to defining the comparator population, we used data from the US nationwide Flatiron Health EHR-derived de-identified database[14] to construct comparator populations with varying degrees of similarity to the OAK trial, an open-label, randomised, phase 3 clinical trial of atezolizumab ($n = 425$) or docetaxel (n-425) in patients with advanced (stage IIIb or IV) non-small cell lung cancer (NSCLC) who had previously received treatment with one to two cytotoxic chemotherapy regimens[15]. The database includes data from over 280 cancer clinics (~800 sites of care), representing more than 2.4 million patients with cancer in the United States.

Using the EHR-derived data, we generated three comparator populations: (1) a population that includes real-world patients who meet inclusion and exclusion criteria for the reference population in the OAK trial[15] and received docetaxel (optimized); (2) patients who meet inclusion/exclusion criteria and received any therapy except atezolizumab (realized 1); (3) a population with fewer restrictions on inclusion/exclusion criteria (primary age 19 and higher and receipt of one or more prior platinum-based therapies) and received the comparator treatment (realized 2). Details of the inclusion/exclusion criteria definitions are provided in Table 1.

We adopted a method of moments approach, analogous to propensity scores, that weights patients in each cohort to achieve cohort-level balance for baseline characteristics in the clinical trial[16]. Specifically, patients were weighted on trial median age ($\leq$64 vs >64 years), gender (male, female), race (white vs other), smoking history (ever, never), and histology (squamous, non-squamous). Median overall survival in months and corresponding 95% confidence interval (95% CI), were estimated using the non-parametric Kaplan-Meier method, in each cohort separately. Restricted time mean overall survival was also estimated using Kaplan-Meier method, with two different time windows: (1) censoring patients at the end of trial follow-up (31 July 2016), and (2) all patients followed up to 28 February 2020.

The case study found that overall survival was longer among patients who met clinical trial inclusion/exclusion criteria (optimized and realized cohort 1) compared to patients who only partially met trial selection criteria (realized cohort 2) (Table 2). An implication for researchers is that realized cohorts may result in shorter overall survival estimates than optimized cohorts. For payers and HTA decision makers, this could result in more realistic estimates of clinical effectiveness and is a great option for understanding disease burden. Optimized cohorts may be preferred for use as indirect comparisons to new therapies studied only in trials, because realized cohorts could lead to misleading conclusions and overestimates of the effectiveness of newer therapies studied in the trial.

Our overall approach to estimating survival follows commonly used recommendations when individual-level survival data are available, including consideration of a range of functional forms, evaluating model fit and clinical plausibility of the survival function (i.e. assessing internal and external validity)[17–19]. Our focus on choices that can be made for selecting the comparator population from real world data – particularly when clinically detailed, patient level data are available – has received less discussion. Developing methodological standards for selecting real world comparators will become important as the use of EHR-derived datasets for health technology assessments grows over time.

Table 1. Summary table for cohort selection criteria.

I/E criteria	Optimized	Realized 1	Realized 2
aNSCLC, ≥18 year old	✓	✓	✓
Prior platinum-based combination therapy	✓	✓	✓
No prior docetaxel, anti-CTLA-4 or PD-L1 therapies	✓	✓	
Prior TKI therapy if EGFR/ALK mutated	✓	✓	
ECOG score ≤ 1 or missing	✓	✓	
Adequate organ function (defined by clinical trial criteria)	✓	✓	
No HIV, hepatitis B or hepatitis C	✓	✓	
Meet 90-day gap rule	✓	✓	✓
2L/3L Treatment	Docetaxel	Any except atezolizumab	Docetaxel
Weighting	✓	✓	✓
Treatment start window	11 March 2014–29 April 2015		
Follow-up duration	a. From treatment start date to trial end date (6 July 2016) b. From treatment start date to 29 February 2020		

Abbreviations. aNSCLC, advanced non-small cell lung cancer; CTLA-4, cytotoxic T-lymphocyte associated protein-4; PD-L1, programmed cell death-ligand 1; TKI, tyrosine kinase inhibitor; EGFR, epidermal growth factor receptor; ALK, anaplastic lymphoma kinase; HIV, human immunodeficiency virus.

Table 2. Survival analysis results for the four comparator cohorts.

Cohort	*N*	*N* (weighted)	Median OS, months (censored[d])	Mean OS, months (censored)	Mean OS, months
Optimized Cohort[a]	126	116.86	7.2 (6.0 – 8.9)	10.2 (8.5 – 11.9)	12.3 (9.6 – 15.0)
Realized 1 Cohort[b]	511	477.48	7.8 (7.3 – 9.0)	11.3 (10.5 – 12.2)	14.4 (12.9 – 15.9)
Realized 2 Cohort[c]	282	260.37	5.3 (4.5 – 6.6)	8.6 (7.6 – 9.7)	10.1 (8.5 – 11.7)

[a]Meet inclusion/exclusion criteria for OAK and received comparator (docetaxel).
[b]Meet inclusion/exclusion criteria for OAK and received any therapy except atezolizumab.
[c]Partially meet inclusion/exclusion for OAK and received comparator (docetaxel).
[d]Censored: patients are censored at the end of trial follow-up (July 2016).

Closing: summary considerations for researchers and HTA decision makers

In closing, we believe there are three broad categories where HTAs can use and interpret comparators for decision making. First, evaluation of real world outcomes relative to trial outcomes for matched comparators when a head to head comparison is available from a clinical trial (optimized). This provides the HTA an estimate of benefit that can be achieved in real world settings for a trial-defined subgroup of patients. Second, to understand the factors that may contribute to observed differences in real world outcomes versus trial outcomes for comparators (optimized vs realized). This provides the HTA an estimate of the prevalence and identifies factors that are impacting observed outcomes for the comparator in the real world (e.g. adherence, suboptimal dosing). Lastly, in the absence of direct comparison, EHR data can be used to understand the potential benefit of a new drug vs the comparator: an indirect comparison with clinical matching. If the difference in outcomes between optimized and realized for the comparator is large, this suggests that the new drug might have benefits that are not readily captured by the trial.

Transparency

Declaration of funding

This study was sponsored by Flatiron Health, which is an independent subsidiary of the Roche group.

Declaration of financial/other relationships

BJA, XW, DB, SB, SG, NJM report employment at Flatiron Health, Inc., which is an independent subsidiary of the Roche group and stock ownership in Roche. SB and NJM report equity ownership in Flatiron Health, Inc. SDR reports consultancy fees and research funding from Flatiron Health; this work was developed during his research sabbatical at Flatiron Health.

JME peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

Author contributions

Concept and design: SDR.
Data collection, analysis and interpretation: SDR, BJA, XW.
Manuscript writing, review and approval: All.

Acknowledgements

None reported.

References

[1] 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. 2020;107(2):369–377.

[2] Chau I, Le DT, Ott PA, et al. Developing real-world comparators for clinical trials in chemotherapy-refractory patients with gastric cancer or gastroesophageal junction cancer. Gastric Cancer. 2020; 23(1):133–141.

[3] Huang Bartlett C, Mardekian J, Cotter MJ, et al. Concordance of real-world versus conventional progression-free survival from a phase 3 trial of endocrine therapy as first-line treatment for metastatic breast cancer. PloS One. 2020;15(4):e0227256.

[4] Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016; 183(8):758–764.

[5] Aapro MS, Bohlius J, Cameron DA, et al. 2010 update of EORTC guidelines for the use of granulocyte-colony stimulating factor to reduce the incidence of chemotherapy-induced febrile neutropenia in adult patients with lymphoproliferative disorders and solid tumours. Eur J Cancer. 2011;47(1):8–32.

[6] Crawford J, Dale DC, Kuderer NM, et al. Risk and timing of neutropenic events in adult cancer patients receiving chemotherapy: the results of a prospective nationwide study of oncology practice. J Natl Compr Canc Netw. 2008;6(2):109–118.

[7] Potosky AL, Malin JL, Kim B, et al. Use of colony-stimulating factors with chemotherapy: opportunities for cost savings and improved outcomes. J Natl Cancer Inst. 2011;103(12):979–982.

[8] Ramsey SD, McCune JS, Blough DK, et al. Colony-stimulating factor prescribing patterns in patients receiving chemotherapy for cancer. Am J Manag Care. 2010;16(9):678–686.

[9] Smith TJ, Khatcheressian J, Lyman GH, et al. 2006 update of recommendations for the use of white blood cell growth factors: an evidence-based clinical practice guideline. J Clin Oncol. 2006; 24(19):3187–3205.

[10] Solomon JP, Linkov I, Rosado A, et al. NTRK fusion detection across multiple assays and 33,997 cases: diagnostic implications and pitfalls. Mod Pathol. 2020;33(1):38–46.

[11] Goodman CS. HTA 101: introduction to health technology assessment. Bethesda (MD): National Library of Medicine; 2014.

[12] Hatswell AJ, Baio G, Berlin JA, et al. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999–2014. BMJ Open. 2016;6(6): e011666.

[13] Backenroth D, Haimson JD, Meropol NJ, et al. Development and initial validation of a prognostic score based on structured data from EHRs. [Abstract]. In: Proceedings of the American Association for Cancer Research Annual Meeting 2019; 2019 Mar 2–Apr 3; Atlanta, GA. Philadelphia (PA): AACR; Cancer Res 2019; 79(13 Suppl):Abstract 3303.

[14] Ma X, Long L, Moon S, et al. Comparison of population characteristics in real-world clinical oncology databases in the US: Flatiron Health, SEER, and NPCR. medRxiv. 2020. DOI:10.1101/2020.03.16.20037143

[15] Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. Lancet. 2017;389(10066):255–265.

[16] Verde PE, Ohmann C. Combining randomized and non-randomized evidence in clinical research: a review of methods and applications. Res Synth Methods. 2015;6(1):45–62.

[17] Jackson C, Stevens J, Ren S, et al. Extrapolating survival from randomized trials using external data: a review of methods. Med Decis Making. 2017;37(4):377–390.

[18] Latimer NR. Survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. Med Decis Making. 2013;33(6): 743–754.

[19] Latimer N. NICE DSU technical support document 14: survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. Decision Support Unit. [June 2011; 2020 Sept 29]. Available from: http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf