Article

Feasibility of Using Oncology-Specific Electronic Health Record (EHR) Data to Emulate Clinical Trial Eligibility Criteria [†]

Thomas W. Wilson * [ID]**, Joseph T. Dye, Sarah Spark, Nicholas J. Robert, Janet L. Espirito and E. Susan Amirian**

Real-World Research, Ontada, Irving, TX 75039, USA
* Correspondence: thomas.wilson@mckesson.com; Tel.: +513-349-5846
† An earlier version of this research was presented at the ISPOR conference on 17 May 2022, in Washington DC: Wilson T, Dye J, Spark S et al. P31 Feasibility of Using Oncology Specific Electronic Health Records (EHR) Data to Emulate Clinical Trial Inclusion and Exclusion Criteria. Value in Health, Volume 25, Issue 7, S293. Accessed on 28 January 2022. https://www.valueinhealthjournal.com/article/S1098-3015(22)00241-8/fulltext.

Abstract: We examined eligibility criteria from recent oncology clinical trials to see whether real-world data (RWD) from electronic health records (EHRs) could be used to create external control groups for clinical trials. Trials were identified from the Aggregate Analysis of ClinicalTrials.gov database; the selected trials were for oncology drugs approved by the FDA in 2020. Verbatim text from trial inclusion and exclusion criteria was qualitatively assessed by an expert panel to determine if criteria could be ascertained from structured and unstructured EHR data. Identified criteria were categorized (cancer-related, comorbidity-related, demographic, functional status, and trial operations) and subcategorized. Among 53 identified trials, 20 met the requirements for study inclusion, which included 463 eligibility criteria. Percentages of criteria by category were as follows: cancer-related factors (46%), comorbidities (20%), functional status (18%), trial operations (14%), and demographics (2%). For 18 of the 20 trials, 80% of the eligibility criteria could be ascertained with RWD; for 4 of the 20, it was 100%. When trial operation-specific criteria were excluded, all 20 met the 100% threshold. Our study indicates that both structured and unstructured data from community-based oncology-specific EHRs can be used for determining patient eligibility for external control arms for clinical trials.

Keywords: clinical trial emulation; electronic health records; observational studies; eligibility criteria

1. Introduction

The 21st Century Cures Act of 2016 established the legislative mandate to use real-world evidence (RWE) for regulatory decision making [1,2]. Real-world data (RWD), which are used to generate RWE, are often defined as data collected in routine healthcare delivery versus data collected specifically for research purposes. The use of RWD vs. traditional clinical trials can provide faster, usually more generalizable, and lower-cost assessments of the utility and risk–benefit profile of a medical intervention, potentially enabling accelerated medical product development and advances to patients. However, because RWD collection is not typically customized for any specific research question, there are nuances related to how these data are best used for decision making.

One promising application of RWD involves the construction of external control groups for clinical trials [3]. While prior studies have reported challenges related to the feasibility of emulating trial eligibility criteria with "claims and/or structured" RWD data [4], few, if any, have focused on assessing the utility of accessing data from oncology-specific electronic health records (EHRs) using both structured (e.g., standardized fields) and unstructured (e.g., free text physician notes and radiographic reports) data. Although structured data in EHR systems can be readily extracted for research, richer clinical data relevant to external control group construction often exist in the unstructured data fields

from the EHR. These data can be curated using natural language processing technologies or manual chart abstraction and therefore should be viewed as a significant potential source of deeper RWD even though more resources are required for curation.

The purpose of this study was to examine the specific use case of eligibility criteria from recent oncology clinical trials to assess the degree to which RWD from an oncology-specific EHR in a community setting can be reasonably used to retrospectively access the information needed for emulation of inclusion/exclusion (I/E) eligibility criteria for external control arms in oncology clinical trials. The study only reviews I/E criteria to specifically address the conclusions of Wallach et al. [4]. Future research should examine the ability to emulate other aspects of the RCT, such as clinical outcomes, using RWD.

2. Results

Among the 53 trials identified, 20 met the requirements for inclusion in this study. These trials investigated 13 cancer types. The most common were identified as lung cancer (25%), leukemia (15%), breast cancer (10%), and multiple myeloma (10%). Of the 463 eligibility criteria identified in the trials, the median per study was 21 (range 7–51) total criteria, 8.5 (range 5–31) inclusion, and 12 (range 2–25) exclusion.

The categorization of each criterion into categories and subcategories was complex; actual examples of each selected individual criterion and their placement into categories and subcategories by the panel are detailed in Table 1.

Table 1. Process overview for determining likelihood of being recorded in the EHR using the verbatim text of trial criteria, performed by an expert panel. *n* = 1 example per category/subcategory.

Example Verbatim Trial Criteria [a]	Inclusion (I) orExclusion (E) [b]	Category [c]	Subcategory [d]	Likelihood of Presence in EHR
Participants must have PD-L1 IHC testing with results performed by a central laboratory during the screening period	I	Cancer-Related	Biomarker or Genetic Marker	Yes
Histologically confirmed MBC, current stage IV	I	Cancer-Related	Cancer Stage	Yes
Newly diagnosed acute myeloid leukemia (AML)	I	Cancer-Related	Cancer Type	Yes
At least one measurable untreated lesion	I	Cancer-Related	Measurable Disease	Yes
Active or untreated central nervous system (CNS) metastases as determined by computed tomography (CT) or magnetic resonance imaging (MRI) evaluation	E	Cancer-Related	Metastasis	Yes
Isolated myeloid sarcoma (i.e., myeloblastoma or chloroma) allowed regardless of bone marrow results	I	Cancer-Related	Other Cancer	Yes
Intrathecal [drug name] [e] given at diagnosis allowed	I	Cancer-Related	Prior Cancer Treatments/Procedures	Yes
Progression of disease after 1 or 2 prior regimens in the metastatic setting	I	Cancer-Related	Progressive Disease	Yes
Significant history of cardiovascular disease	E	Comorbidity-Related	Cardiovascular Status	Yes
Patients with Down syndrome over 4 years of age are eligible	I	Comorbidity-Related	Comorbidity	Yes
Topical or inhalation steroids for other conditions allowed	I	Comorbidity-Related	Concomitant Medications	Yes
Known contraindication to receive [drug name][e] at the planned doses	E	Comorbidity-Related	Contraindication	Yes
Participant has received strong or moderate cytochrome P450 3A4 (CYP3A) inducers 7 days prior to the initiation of study treatment	E	Comorbidity-Related	Drug–Drug/Food Interaction	Yes
Have adequate organ function, as defined by the study protocol	I	Comorbidity-Related	Organ Function	Yes

Table 1. *Cont.*

Example Verbatim Trial Criteria [a]	Inclusion (I) or Exclusion (E) [b]	Category [c]	Subcategory [d]	Likelihood of Presence in EHR
History of idiopathic pulmonary fibrosis, organizing pneumonia, drug induced pneumonitis, idiopathic pneumonitis, or evidence of active pneumonitis on screening chest CT scan. History of radiation pneumonitis in the radiation field (fibrosis) is permitted	E	Comorbidity-Related	Pulmonary Status	Yes
Male or female participants ≥ 55 years of age	I	Demographic	Demographic	Yes
Adequate hematologic and end-organ function	I	Functional Status	Hematologic Status	Yes
History of hepatic encephalopathy	I	Functional Status	Hepatic Status	Yes
Positive test for Human Immunodeficiency Virus (HIV)	E	Functional Status	HIV Status	Yes
Participant has a known hypersensitivity to the components of [drug name] [e] or its excipients	E	Functional Status	Hypersensitivity	Yes
Active or history of autoimmune disease or immune deficiency	E	Functional Status	Immune Status	Yes
Active hepatitis B or hepatitis C	E	Functional Status	Infection	Yes
The participant has any known significant ophthalmologic abnormalities of the surface of the eye	E	Functional Status	Ocular Status	Yes
Eastern Cooperative Oncology Group (ECOG) performance status—0, 1, 2, 3	I	Functional Status	Performance Status	Yes
Creatinine clearance inferior to 30 mL per minute (Modification of Diet in Renal Disease (MDRD) Formula)	E	Functional Status	Renal Status	Yes
Not pregnant or nursing	I	Functional Status	Reproductive Status	Yes
Matched family donor criteria (for patients with intermediate-risk or high-risk disease)	I	Trial-Specific	Donor Matching	No
Life expectancy of at least 3 months	I	Trial-Specific	Life Expectancy	No
Negative pregnancy test	I	Trial-Specific	Trial Operations	No

[a] Verbatim text of criteria from 20 selected studies. One example for each of the 29 subcategories developed (n = 463 total criteria); [b] designated in the protocol; [c] criterion classified into five (5) broad categories; [d] criterion further classified into twenty-nine (29) subcategories; [e] all proprietary drug names removed and replaced with [drug name].

Percentages of criteria by category were cancer-related factors (46%), comorbidities (20%), demographics (2%), functional status (18%), and trial operations (14%) (Table 2) [5,6].

Eighteen of the 20 trials met the 80% threshold for eligibility criteria likely to be ascertainable with RWD, while 4 of 20 trials met the 100% threshold when all criteria were considered. Trial-specific criteria are criteria that may be essential to the operation of the clinical trial, but not necessarily data collected within the context of routine care and documentation in the real-world setting. For example, "signed informed consent" to participate in the clinical trial would not be a part of routine care and therefore would not be in the EHR. Documentation of the use of multiple birth control methods for female patients of child-bearing age may be required to ensure patient and fetal safety in the clinical trial setting. However, documentation of this in the real-world setting is not necessarily required. When trial-specific criteria were removed from the assessment, all 20 trials met the 100% threshold for ascertainable trial criteria (Figure 1).

Table 2. Number and percentage of criteria per category (*n* = 5) and subcategory (*n* = 29).

Criteria Category			Criteria Subcategory		
Name	n of Criteria Found across all 20 Trials	% of the Total Criteria across All 20 Trials	Name	n	%
Cancer-Related	212	46	Biomarker or Genetic Marker	21	10
			Cancer Stage	10	5
			Cancer Type	35	17
			Measurable Disease	12	6
			Metastasis	7	3
			Other Cancer	51	24
			Prior Cancer Treatments/Procedures	71	33
			Progressive Disease	5	2
Comorbidity-Related	92	20	Cardiovascular Status	14	15
			Comorbidity	48	52
			Concomitant Medications	4	4
			Contraindication	1	1
			Drug–Drug/Food Interaction	5	5
			Organ Function	15	16
			Pulmonary Status	5	5
Demographic	11	2	Age and Sex	11	100
Functional Status	82	18	Hematological Status	19	23
			Hepatic Status	6	7
			HIV Status	6	7
			Hypersensitivity	7	9
			Immune Status	4	5
			Infection	12	15
			Ocular Status	3	4
			Performance Status	17	21
			Renal Status	3	4
			Reproductive Status	5	6
Trial-Specific	66	14	Donor Status	8	12
			Life Expectancy	4	6
			Trial Operations	54	82
Total	463	100			

HIV, human immunodeficiency virus.
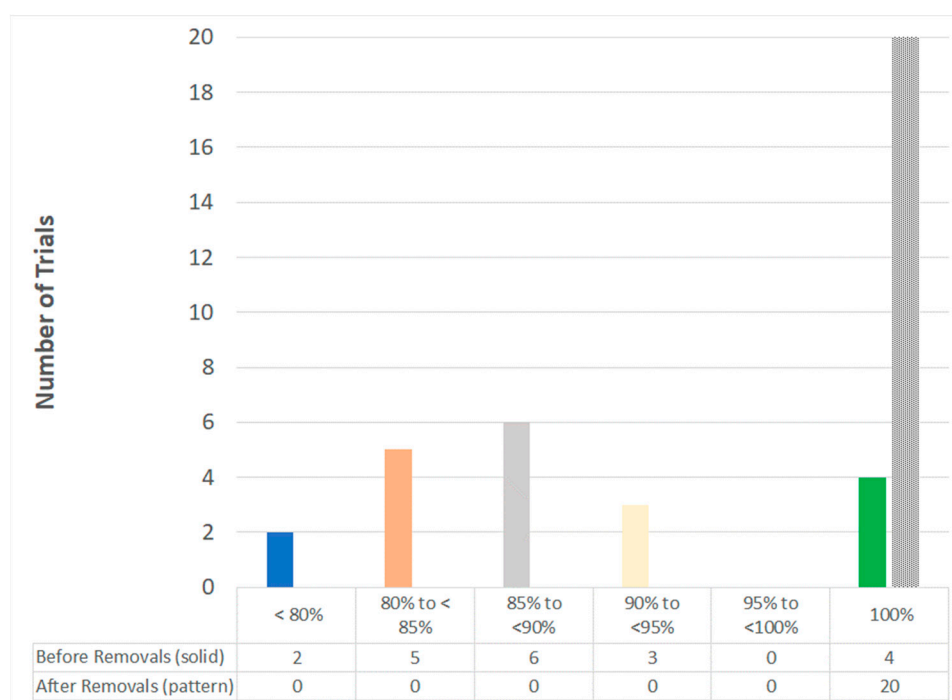
Figure 1. Trial emulation rate before and after removing trial-specific criteria. Trial-specific criteria are criteria that may be essential to the operation of the clinical trial, but not necessarily data collected within the context of routine care and documentation in the real-world setting. For example, "signed informed consent" to participate in the clinical trial would not be a part of routine care and therefore would not be in the EHR. Other examples include life expectancy, agreement to be abstinent, and donor status.

3. Discussion

Building on previously cited reports [4], the inclusion of unstructured data along with structured data from oncology-specific EHRs vastly improves the proportion of eligibility criteria that are likely to be ascertainable for clinical trial emulation. Rich unstructured data are accessible in patient charts and should not be neglected as a potential source of information for evaluating the eligibility of patients, especially as opportunities for implementing scalable technological solutions for abstracting data become more tangible. Our study indicates that RWD-based external controls constructed from oncology-specific EHR data are a conceivable solution.

In the context of phase 3 clinical trials, when external control arms are being designed for regulatory purposes, sponsors and regulators should perhaps carefully consider crafting the inclusion and exclusion criteria to be applicable, as much as possible, in both the clinical trial setting and in real-world practice. We would recommend starting with data available in RWD and, when applicable, using these data to generate inclusion/exclusion criteria. Further, specific operational criteria that do not impact study validity should be identified and clearly noted when external control arms are envisioned. The judicious use of eligibility criteria to identify real-world external controls would maximize external validity while preserving internal validity. Inclusivity in clinical trials is a major topic in the current literature [7], and the use of RWD to construct external control arms can aid in improving inclusivity and generalizability. Indeed, restrictive clinical trial eligibility criteria have been cited as one of the major barriers to the participation of a more diverse population in trials, and "revised criteria may improve participant diversity, without compromising safety or study results" [8]. Therefore, prioritization of key criteria and the relaxation of any criteria that are largely unrelated to internal validity would optimize opportunities for inclusion. As stated earlier, variables found in structured vs. unstructured data will vary from EHR system to EHR system, and structured data often need to be supplemented

with other data sources [9]. However, as unstructured data are the "gold standard," the best way to emulate an RCT is through structured data supplemented with unstructured data. The level of agreement between these two sources of data would use the unstructured data as the gold standard. The value of unstructured data is an under-appreciated and poorly researched resource. The actual empirical trial emulation test of inclusion/exclusion criteria (as well as trial process and outcome measures) would be an obvious next step in research.

4. Methods

In accordance with 45 CFR §46, institutional review board approval was not required for this study because public information was used (no patient data were utilized and informed consent was not required).

FDA approvals of oncology drugs in 2020 were identified and matched to trial data from the Aggregate Analysis of ClinicalTrials.gov (AACT) database (accessed 17 December 2021). Trial data from phase 3 clinical trials with outcome information were included. The verbatim text from the reported inclusion and exclusion eligibility criteria in these trials was downloaded from AACT for evaluation. Reported criteria that contained multiple components were separated into discrete, individual criteria. For example, if a criterion was stated as "women over age 35", then the components were listed as separate entries for sex (women) and age (over age 35).

These were qualitatively assessed by an expert panel (the authors) representing expertise in medicine, pharmacy, epidemiology, nursing, and chart abstraction. The panel has expertise and experience accessing RWD from an oncology EHR (iKnowMed) for research purposes; iKnowMed is the EHR for The US Oncology Network, a network of community oncology practices covering 40 states of the United States with 1.2 million newly diagnosed cancer patients yearly; iKnowMed is used by over 2700 oncology providers [10]. There are many EHR systems available, and the degree to which specific data are included in structured versus unstructured fields will vary from system to system and even within different implementations of the same system. The panel assessed each criterion for its inclusion in the medical record as part of routine patient care (versus clinical trial purposes), and each criterion was scored as "likely" or "not likely" to be ascertainable in structured and/or unstructured EHR data. Any discordance in scores was adjudicated through discussion.

The resulting tabulation of criteria was then organized by the panel into categories and subcategories to facilitate assessment. Many criteria could be worded as a positive (e.g., adequate hepatic function) or a negative (inadequate hepatic function), affecting placement as an inclusion or exclusion criterion. Either distinction would be classified as a single subcategory by the panel.

The criteria were classified into 5 different categories and 29 subcategories as follows:

1. Cancer-Related:
   - Biomarker or Genetic Marker, Cancer Stage, Cancer Type, Measurable Disease, Metastasis, Other Cancer, Prior Cancer Treatments/Procedures, Progressive Disease.

2. Comorbidity-Related:
   - Cardiovascular Status, Comorbidity, Concomitant Medications, Contraindication, Drug–Drug/Food Interaction, Organ Function, Pulmonary Status.

3. Demographic:
   - Age and Sex.

4. Functional Status:
   - Hematological Status, Hepatic Status, HIV Status, Hypersensitivity, Immune Status, Infection, Ocular Status, Performance Status, Renal Status, Reproductive Status.

5. Trial-Specific:

○   Donor Status, Life Expectancy, Trial Operations.

Under the trial-specific category, a subcategory of trial operations was created for those criteria related to logistical and operational elements of running a clinical trial such as informed consent, pre-randomization pregnancy testing, and behavioral restrictions (such as promise to abstain from sex). The proportion of trials with at least 80% and those with 100% of eligibility criteria judged to be likely ascertainable with RWD [4] was determined. The data analysis for this paper was generated using SAS software v9.4 (SAS Institute Inc., Cary, NC, USA. Accessed on 1 January 2022 to 21 March 2023. https://www.sas.com/en_us/legal/editorial-guidelines.html).

References

1. US Food and Drug Administration. Real-World Evidence. Real-World Data (RWD) and Real-World Evidence (RWE) Are Playing an Increasing Role in Health Care Decisions. Available online: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence (accessed on 5 July 2022).
2. International Society for Pharmacoeconomics and Outcomes Research. Top 10 HEOR Trends. Available online: https://www.ispor.org/heor-resources/about-heor/top-10-heor-trends/ (accessed on 5 July 2022).
3. Thorlund, K.; Dron, L.; Park, J.J.H.; Mills, E.J. Synthetic and external controls in clinical trials-a primer for researchers. *Clin. Epidemiol.* **2020**, *12*, 457–467. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7218288/ (accessed on 5 July 2022). [CrossRef] [PubMed]
4. Wallach, J.D.; Zhang, A.D.; Skydel, J.J.; Bartlett, V.L.; Dhuva, S.S.; Shah, N.D.; Ross, J.S. Feasibility of using real-world data to emulate post-approval confirmatory clinical trials of therapeutic agents granted US Food and Drug Administration accelerated approval. *JAMA Netw. Open* **2021**, *4*, e2133667. Available online: https://pubmed.ncbi.nlm.nih.gov/34751763/ (accessed on 5 July 2022). [CrossRef] [PubMed]
5. American Society of Clinical Oncology. The ASCO Post. 2020 FDA Approvals of Drugs for Cancer Treatment. 2020. Available online: https://ascopost.com/issues/december-25-2020/2020-fda-approvals-of-drugs-for-cancer-treatment/ (accessed on 5 July 2022).
6. Clinical Trials Transformation Initiative. Download AACT. Available online: https://aact.ctti-clinicaltrials.org/download (accessed on 17 December 2021).
7. US Food and Drug Administration. Enhancing the Diversity of Clinical Trial Populations—Eligibility Criteria, Enrollment Practices, and Trial Designs Guidance for Industry. 2020. Available online: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enhancing-diversity-clinical-trial-populations-eligibility-criteria-enrollment-practices-and-trial (accessed on 28 January 2023).

8. Riner, A.N.; Girma, S.; Vudatha, V.; Mukhopadhyay, N.; Skoro, N.; Gal, T.S.; Freudenberger, D.C.; Herremans, K.M.; George, T.J.; Trevino, J.G. Eligibility criteria perpetuate disparities in enrollment and participation of Black patients in pancreatic cancer clinical trials. *J. Clin. Oncol.* **2022**, *40*, 2193–2202. Available online: https://pubmed.ncbi.nlm.nih.gov/35316089/ (accessed on 28 January 2023). [CrossRef] [PubMed]
9. Von Lucadou, M.; Ganslandt, T.; Prokosch, H.U.; Toddenroth, D. Feasibility analysis of conducting observational studies with the electronic health record. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 202. Available online: https://pubmed.ncbi.nlm.nih.gov/31660955/ (accessed on 2 May 2023). [CrossRef] [PubMed]
10. Ontada. Available online: https://www.ontada.com/ (accessed on 1 May 2023).