Published in final edited form as:

Conf Proc IEEE Eng Med Biol Soc. 2016 August ; 2016: 2970–2973. doi:10.1109/EMBC.2016.7591353.

Assessing the Population Representativeness of Colorectal Cancer Treatment Clinical Trials

Zhe He, IEEE [Member],

School of Information and Institute for Successful Longevity, Florida State University, Tallahassee, FL 32308 USA (phone: 850-644-5775; zhe@fsu.edu)

Zhiwei Chen,

Department of Computer Science at Florida State University, Tallahassee, FL 32308 USA (zc15d@my.fsu.edu)

Thomas J George Jr.,

UF Health Cancer Center, Gainesville, FL 32610 (thom.george@medicine.ufl.edu)

Gloria Lipori, and

UF Health and UF Health Sciences Center, Gainesville, FL 32610 (pflugg@shands.ufl.edu)

Bian Jiang, IEEE [Member]

Department of Health Outcomes and Policy, University of Florida, Gainesville, FL 32610 USA (bianjiang@ufl.edu)

Abstract

The generalizability (external validity) of clinical trials has long been a concern for both clinical research community as well as the general public. Results of trials that do not represent the target population may not be applicable to the broader patient population. In this study, we used a previously published metric Generalizability Index for Study Traits (GIST) to assess the population representativeness of colorectal cancer (CRC) treatment trials. Our analysis showed that the quantitative eligibility criteria of CRC trials are in general not restrictive. However, the qualitative eligibility criteria in these trials are with moderate or strict restrictions, which may impact their population representativeness of the real-world patient population.

I. Introduction

Randomized controlled trials are well regarded as the gold standard for generating medical evidence [1]. To support evidence-based medicine (EDM), clinical trials should have a balanced internal validity (methodological quality) and external validity (generalizability) [2]. However, trials overemphasizing the internal validity are often criticized for the lack of population representativeness (i.e., the study population does not capture the characteristics of the real-world target population) [3]. Various factors could influence the representativeness of trials. Besides real-world issues such as geographical proximity to the trial site, overly restrictive or vague eligibility criteria may also limit the representation of a certain population subgroups (e.g., old adults with multiple comorbidities) [4], yielding low *a priori* generalizability. Traditionally, researchers often compared enrolled patients with the real-world patients to assess *a posteriori* generalizability [5]. However, this cannot be done

before the completion of the trial and the publication of the trial results, thereby delaying the detection of systematic biases in trial design.

To quantify *a priori* generalizability, Weng et al. previously introduced a quantitative metric that compares the study population of the trials, derived from their eligibility criteria, with the real world patients who would benefit from the trial results [6]. The metric, Generalizability Index on Study Traits (GIST), can quantify the population representativeness of a set of clinical trials with respect to a single quantitative variable such as age and BMI. In previous work, we have demonstrated its effectiveness in assessing the population representativeness of clinical trials on Type 2 diabetes at large scale [7, 8].

Cancer increases in incidence with age and is the second leading cause of death in the US following heart disease [9]. Colorectal cancer (CRC) is the second-leading cause of cancer deaths, affects both genders and is associated with advancing age [10]. Elderly patients have been reported to be underrepresented in cancer-treatment trials [11, 12]. However, this analysis was based on a relatively small number of assessed trials. In this study, we aim to investigate the population representativeness issue on the disease domain level by including all the CRC treatment trials registered on [ClinicalTrials.gov](https://clinicaltrials.gov) over the last 10 years. We use the GIST metric to assess the population representativeness, and to identify restrictive quantitative eligibility criteria in these CRC treatment trials. To augment the analysis of quantitative eligibility criteria, we further conduct a preliminary analysis on the scale of restrictive qualitative eligibility criteria in cancer trials [13] using a semi-automated approach. To profile a real-world population of CRC patients (target population), we used the patient data within the University of Florida (UF) Health Integrated Data Repository (IDR). To profile the study population of CRC treatment trials, we used the eligibility criteria of 1,308 such trials with a start date between 01/2004 and 12/2013 in [ClinicalTrials.gov](https://clinicaltrials.gov). This work intends to improve the transparency of the population representativeness of CRC treatment trials and help trial designers make informed decision when designing new trials, thereby optimizing the balance between the internal validity and the external validity.

II. Background

A. [ClinicalTrials.gov](https://clinicaltrials.gov)

[ClinicalTrials.gov](https://clinicaltrials.gov) is the official clinical study and results registry created and maintained by the U.S. National Library of Medicine (NLM). Mandated by the Food and Drug Administration Amendments Act (FDAAA) of 1997 at FDAAA 801, all US-based trials of drugs, device, and biologic interventions other than Phase I trials have to be registered in [ClinicalTrials.gov](https://clinicaltrials.gov). As of March 7, 2016, 209,951 studies with sites in 192 countries are registered. Study summaries in [ClinicalTrials.gov](https://clinicaltrials.gov) include structured study descriptors such as study design, study phase, intervention, as well as unstructured inclusion/exclusion criteria that specify the characteristics of subjects to be included or excluded in the study.

B. UF Health Integrated Data Repository (IDR)

Supported by the UF Clinical and Translational Institute (CTSI) and the UF Health, the UF Health IDR is a secure, clinical data warehouse (CDW) that aggregates data from the university's various clinical and administrative information systems, including the Epic electronic medical record (EMR) system. As of January 2016, the IDR contains data for encounters that occurred after June 2011, with a total of more than 492 million observational facts pertaining to 709,422 patients [14].

III. Methods

We first give two definitions that are used in this paper:

Study population: patients who satisfy the eligibility criteria of a clinical study. In this study, we consider the study population as the overall population of all patients that are eligible for CRC treatment clinical trials.

Target population: patients to whom the results of the clinical studies are intended to be applied. The target population characteristics can be only approximated by with the available patient data. In this study, the target population includes all the CRC patients in UF Health system.

Then we describe the data preparation and data analysis as follows.

A. Data Preparation

We have previously built the COMPACT database, which includes structured meta-data and fine-grained eligibility features from all the trials on ClinicalTrials.gov [15]. Based on the COMPACT database, we have developed a web-based tool VITTA (<http://is.gd/VITTA>), allowing its users to flexibly select a set of trials of the same disease domain and profile their study population with respect to one quantitative eligibility criteria at a time [16]. Using the VITTA tool, we identified 1,308 CRC treatment trials with a start date between 01/2004 and 12/2013. Then, we selected five frequently used quantitative eligibility criteria in CRC trials, including 'age', 'creatinine', 'platelet count', 'alanine transaminase' (ALT), and 'aspartate aminotransferase' (AST). We then generated the study population with respect to each of these variables (i.e., the percentage of trials that permit each value of a variable).

To profile the target population of CRC patients, we identified 3,178 patients with CRC in UF Health IDR using ICD-9-CM codes 153.*, 154.* and ICD-10-CM codes C18.*, C20.*, and C21.*. A patient may have multiple lab values over time. The latest lab values were used in this study. Due to the use of de-identified patient data for secondary analysis, this study is exempted from the ethics review.

B. Data Analysis

1) Calculating GIST scores for quantitative criteria—We calculated the GIST scores of frequently used quantitative eligibility variables, one at a time, in CRC treatment trials to quantify the population representativeness. A GIST score is the sum across all consecutive

non-overlapping value intervals of the percentage of studies that recruit patients in that interval, multiplied by the percentage of patients observed in that interval:

$$\text{GIST} = \sum_{i=1}^N \frac{\sum_{j=1}^T I([i_{\text{low}}, i_{\text{high}}] \subset w_j)}{T} * \frac{\sum_{k=1}^P I(i_{\text{low}} \leq y_k < i_{\text{high}})}{P} \quad (1)$$

where N is the number of distinct value intervals of the quantitative feature, T is the number of trials, P is the number of patients, w_j is the inclusion value interval of the quantitative feature for the j^{th} study, such that indicator I can be defined as j^{th} study interval subsumes the i^{th} interval low and high boundary values, and y_k is the observed value of the quantitative feature for the k^{th} patient such that an indicator I can be defined when k^{th} patient has a value of the quantitative feature falls within the i^{th} interval.

The GIST score ranges between 0 and 1 and characterizes the proportion of patients that would be potentially eligible across trials, with 1 being perfectly generalizable and 0 being completely not generalizable. Note that the formula for calculating the GIST score can also be applied to categorical variables, whereby the values are integers. In a previously published study, we have used a simulation-based approach to evaluate the GIST metric in assessing the population representativeness of clinical trials with respect to a single quantitative eligibility criterion [17].

By ranking the GIST scores of the selected variables, we can easily identify the quantitative eligibility criteria that are more restrictive than others.

2) Identifying restrictive qualitative criteria—Previously, Lewis et al. analyzed the participation of older adults in cancer clinical trials and identified a set of qualitative criteria that have strict restrictions: “no history of congestive heart failure”, “no active angina”, “no myocardial infarction ever”, “no myocardial infarction ever past five years”, “no history of ischemic heart disease”, “no abnormal conduction disease”, “no arrhythmia requiring treatment”, “no history of hypertension”, “no history of peripheral vascular disease”, “no history of stroke”, “no history of transient ischemic attack”, “no thromboembolic disease history”, “no history of chronic cerebrovascular accident”, “no history of deep vein thrombosis”, and “no history of pulmonary embolism” [13]. Meanwhile, Lewis et al. also listed the moderate restrictions for each of them. For example, the moderate restriction version of “no history of hypertension” is “no poorly controlled hypertension”. In this study, we assessed the scale of such eligibility criteria with moderate or strict restrictions in the colorectal cancer treatment clinical trials.

In free-text eligibility criteria, trial designers may use the same criteria in different ways. For example, one may use “no history of hypertension” in the inclusion criteria, while another one may use “history of hypertension” in the exclusion criteria. It is more complex when the temporal constraint is added (e.g., “no history of uncontrolled hypertension in the past 90 days”). To assess the scale of these restrictive qualitative criteria in CRC trials, we first automatically identified the occurrences of the key terms (e.g., ‘congestive heart failure’, and ‘angina’) in the free-text eligibility criteria of all CRC treatment trials. To account for

different strings for the same key term, we identified all the synonyms of the key terms using the Unified Medical Language System (UMLS) [18], which has integrated over 170 source vocabularies and mapped the terms with the same meaning into the same concept. If the key term (e.g., ‘arrhythmia requiring treatment’) is not covered by the UMLS, we just used key term itself for the search. After the criteria were automatically annotated, we randomly selected 100 trials and manually reviewed the annotated criteria to check if they accurately matched the restrictive criteria listed above with respect to negation and temporal information. As such, we identified all the qualitative eligibility criteria with moderate or strict restriction in the random sample. We report our findings in the Results section below.

IV. Results

A. Basic Characteristics

Table 1 shows the characteristics of CRC patients in UF Health IDR. Out of the 3,178 CRC patients, 65.7% are white. The mean age is 65 years old. The standard deviations of the lab tests are large, indicating the diverse health status of the patient population.

B. GIST Scores for Frequently Used Quantitative Variables

Table 2 shows the GIST scores for each of the five variables. Each variable has two GIST scores. The first score only includes the trials that have specified a permissible value range for the variable. The second score includes all the trials regardless of the usage of the variable. The trials without the specific variable were deemed to allow all the possible values for the variable. Age and AST had the highest GIST score (0.91), whereas creatinine has the lowest GIST score (0.78).

C. Visualization of the Study Population and the Target Population

We visualized the study population and the target population of CRC treatment trials with respect to two variables: creatinine (Figure 1) and age (Figure 2). Each dot on the blue solid curve represents the percentage of patients with a specific value, while each dot on the dashed green curve represents the percentage of trials that consider patients with a specific value. As shown in Figure 1, most CRC treatment trials accept patients with creatinine < 1.4 mg/dl, which include 88.3% of CRC patients.

With respect to age, most trials allowed patients above 20 years old. About 20% trials explicitly excluded patients who are older than 80 years old. Note that even though most trials do not explicitly exclude older adults by the age criterion, other restrictive eligibility criteria such as ‘*cognitive impairment*’, or ‘*no history of myocardial infarction*’ may indirectly limit the representation of older adults.

D. Preliminary Analysis on Restrictive Qualitative Criteria

In all 1,308 CRC treatment trials, we identified that 636 trials contain one or more keywords in the restrictive qualitative eligibility criteria as described in [13]. The most frequent keywords are ‘congestive heart failure’ (482 trials), ‘myocardial infarction’ (473 trials), and ‘ischemic heart disease’ (35 trials). Among 100 randomly selected CRC trials, 44 trials had 76 occurrences of the keywords of the restrictive qualitative eligibility criteria. We manually

reviewed these 76 occurrences and found that 8 (10.5%) occurrences were not restrictive criteria (false positives), 58 (76.3%) occurrences were moderately restrictive criteria, and 10 (13.2%) occurrences were strictly restrictive criteria. Table III lists the qualitative eligibility criteria with moderate and strict restrictions that occurred more than twice in the random sample.

V. Discussion

In this work, we assessed the population representativeness of 1,308 CRC treatment trials in [ClinicalTrials.gov](#) with a start date between 01/2004 and 12/2013. According to the GIST scores, five frequently used quantitative eligibility criteria in CRC treatment trials (i.e., age, creatinine, platelet count, ALT, AST) did not exhibit significant restrictions to the target population. However, qualitative eligibility criteria with moderate or strict restrictions were frequently observed in these trials, according to our preliminary analysis. Patients with clinically evident congestive heart failure and/or myocardial infarction in the past were often excluded by these CRC treatment trials.

To ensure internal validity, many trials are justified to investigate certain population subgroups that may not represent the real-world population broadly. Therefore, rather than maximize the external validity of clinical studies, we intend to improve the transparency of *a priori* generalizability of CRC cancer treatment trials, with the ultimate goal of helping trial designers make informed decision for choosing patient selection criteria.

A. Limitations

Some limitations need to be noted in this work. First, the GIST metric does not take geographic location and enrollment number into account. Second, we only used CRC patient data in UF Health IDR to profile the target population. Our study is not generalizable to the national population. Third, we only reviewed a small amount of random sample of qualitative eligibility criteria. Future research is needed to automatically identify the contextual information of complex eligibility criteria. Fourth, the COMPACT database uses the API of [ClinicalTrials.gov](#) for condition indexing. There may be trial indexing errors.

B. Future Work

In future work, we plan to investigate the correlation among eligibility criteria that may contribute to the population representativeness issues of cancer trials with recommendations to better *a priori* align clinical trial eligibility with the target population. We also plan to extend the VITTA tool to enable interactive analysis of the population representativeness of clinical studies.

VI. Conclusion

In this study, we assessed the population representativeness of CRC treatment trials that started between 2004 and 2013 with our previously published methods. This work could inform the CRC clinical research community of population representativeness issues and help future trial developers better balance the internal and external validity required in trial design.

Acknowledgments

The development of the COMPACT database and the VITTA tool was supported by U.S. National Library of Medicine Grant R01LM009886 (PI: Weng) and the U.S. National Center for Advancing Translational Science (NCATS) Award UL1TR000040 (PI: Ginsberg). This work was partially supported by an Amazon Web Service in Education Research Grant Award (PI: He) and NIH/NCATS Clinical and Translational Science Award UL1TR001427 (PIs: Nelson & Shenkman). The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institutes of Health.

References

1. [April 9, 2014] From the NIH Director: The Importance of Clinical Trials. Available from: <http://www.nlm.nih.gov/medlineplus/magazine/issues/summer11/articles/summer11pg2-3.html>
2. Yessaian A, Mendivil AA, Brewster WR. Population characteristics in cervical cancer trials: search for external validity. *Am J Obstet Gynecol*. 2005; 192(2):407–13. [PubMed: 15695979]
3. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet*. 2005; 365(9453):82–93. [PubMed: 15639683]
4. Filion M, Forget G, Brochu O, Provencher L, Desbiens C, Doyle C, Poirier B, DuRocher M, Camden S, Lemieux J. Eligibility criteria in randomized phase II and III adjuvant and neoadjuvant breast cancer trials: not a significant barrier to enrollment. *Clin Trials*. 2012; 9(5):652–659. [PubMed: 23060323]
5. Lee PY, Alexander KP, Hammill BG, Pasquali SK, Peterson ED. Representation of elderly persons and women in published randomized trials of acute coronary syndromes. *JAMA*. 2001; 286(6):708–13. [PubMed: 11495621]
6. Weng C, Li Y, Ryan P, Zhang Y, Gao J, Liu F, Bigger JT, Hripcsak G. A Distribution-based Method for Assessing The Differences between Clinical Trial Target Populations and Patient Populations in Electronic Health Records. *Applied Clinical Informatics*. 2014; 5(2):463–479. [PubMed: 25024761]
7. He Z, Wang S, Bornanian E, Weng C. Assessing the Population Representativeness of Type 2 Diabetes Trials by Combining Public Data from ClinicalTrials.gov and NHANES. *Stud Health Technol Inform*. 2015; 2015(216):569–73.
8. He Z, Ryan P, Hoxha J, Wang S, Carini S, Sim I, Weng C. Multivariate analysis of the population representativeness of related clinical studies. *J Biomed Inform*. 2016; 60:66–76. [PubMed: 26820188]
9. CDC. [2014 March 7, 2016] Deaths: Leading Causes for 2013. Available from: http://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_02.pdf
10. Society, AC. [March 11, 2016] Colorectal Cancer Facts and Figures 2016. Available from: <http://www.cancer.org>
11. Hutchins LF, Unger JM, Crowley JJ, Coltman CA Jr, Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. *N Engl J Med*. 1999; 341(27):2061–7. [PubMed: 10615079]
12. Sorg C, Schmidt J, Buchler MW, Edler L, Marten A. Examination of external validity in randomized controlled trials for adjuvant treatment of pancreatic adenocarcinoma. *Pancreas*. 2009; 38(5):542–50. [PubMed: 19287330]
13. Lewis JH, Kilgore ML, Goldman DP, Trimble EL, Kaplan R, Montello MJ, Housman MG, Escarce JJ. Participation of patients 65 years of age or older in cancer clinical trials. *J Clin Oncol*. 2003; 21(7):1383–9. [PubMed: 12663731]
14. [March 11, 2016] The Statistics of UF Health Integrated Data Repository. Available from: <http://idr.ufhealth.org/2016/01/27/new-i2b2-stats/>
15. He Z, Carini S, Hao T, Sim I, Weng C. A Method for Analyzing Commonalities in Clinical Trial Target Populations. *AMIA Annu Symp Proc*. 2014; 2014:1777–86. [PubMed: 25954450]
16. He Z, Carini S, Sim I, Weng C. Visual aggregate analysis of eligibility features of clinical trials. *J Biomed Inform*. 2015; 54:241–55. [PubMed: 25615940]
17. He Z, Chandar P, Ryan P, Weng C. Simulation-based Evaluation of the Generalizability Index for Study Traits. *AMIA Annu Symp Proc*. 2015; 2015:594–603. [PubMed: 26958194]

18. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res.* 2004; 32(Database issue):D267–70. [PubMed: 14681409]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Visualization of the target population and the study population of colorectal cancer treatment trials with respect to creatinine.

Figure 2.
Visualization of the target population and the study population of colorectal cancer treatment trials with respect to age.

TABLE I

Basic Characteristics of the Patient Cohort

Characteristics	Number
Patients, n	3,178
<i>Gender</i>	
Male (%)	1,601 (50.4%)
Female (%)	1,577 (49.6%)
<i>Race/Ethnicity</i>	
Asian (%)	23 (0.7%)
Black/African American (%)	365 (11.5%)
White (%)	2,089 (65.7%)
Other (%)	701 (22.1%)
Age (mean, SD)	65 ± 14
Creatinine (mg/dl)	1.03 ± 0.88
Platelet count (/mm ³)	236610 ± 112568
ALT (Upper limit of normal)	0.78 ± 2.04
AST (Upper limit of normal)	1.14 ± 5.51

TABLE II

Gist Scores of Five Quantitative Eligibility Criteria in Colorectal Cancer Treatment Trials

Variable	GIST Score (Number of trials)	
	<i>Trials with the variable</i>	<i>All the trials</i>
Age	0.91 (1,308)	0.91 (1,308)
Creatinine	0.78 (422)	0.93 (1,308)
Platelet count	0.88 (332)	0.97 (1,308)
ALT	0.88 (318)	0.97 (1,308)
AST	0.91 (160)	0.99 (1,308)

TABLE III

Frequency of Qualitative Eligibility Criteria with Moderate or Strict Restrictions in the Random Sample

Qualitative Criteria	Level of Restriction	Number of Trials (%)
No clinically evident congestive heart failure	Moderate	30 (68%)
No myocardial infarction past 12 months	Moderate	17 (39%)
No myocardial infarction past 6 months	Moderate	15 (34%)
No history of congestive heart failure	Strict	5 (11.4%)
No myocardial infarction ever	Strict	2 (5%)
No history of stroke	Strict	2 (5%)

^aThe denominator of the percentage values is the total number of trials in the manual review (44)