Mimicking Clinical Trials Using Real-World Data – A Novel Method and Applications

Wei-Jhih Wang

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Anirban Basu, Chair

Aasthaa Bansal

Caroline S. Bennette

Program Authorized to Offer Degree:

Pharmacy

University of Washington

Abstract

Mimicking Clinical Trials Using Real-World Data – A Novel Method and Applications

Wei-Jhih Wang

Chair of the Supervisory Committee:
Professor, Anirban Basu
Department of Pharmacy

Background

Real-world data (RWD) have been widely used to evaluate the effectiveness and long-term safety of new treatments in practice to support evidence from clinical trials. Yet, the use of RWD would be limited when new treatments were slowly adopted in practice. Since clinical trial evidence may not be generalizable, it is important to know when and for whom the physicians would adopt the new treatments post clinical trials. Therefore, the first aim of this dissertation was to identify whether physician decisions on the adoption of treatment were influenced by the similarity of patient characteristics between trial participants and general target population.

Additionally, RWD could be used to assess the generalizability of trials and possibly generate an external control group for a single arm study along with the propensity score-based approaches, which require individual-level trial data. When only summary data from trials are available, clinical trial samples can be simulated by using the observed correlations among baseline patient

characteristics in the RWD. However, correlations found in RWD may not apply to trial data, and a different correlation structure among covariates would potentially bias the outcomes, especially if the outcomes data generating process includes effect modifiers or non-linear models. Therefore, using trial summary data and RWD, we aimed to develop an algorithm to approximate the true correlations of participants' characteristics in a trial and generate trial data accordingly.

Methods

In Aim 1, we conducted an exploratory analysis using individual-level trial data and the SEER-Medicare linked database. The trial aimed to evaluate the efficacy of panitumumab plus FOLFIRI for treated metastatic colorectal cancer (mCRC), initially published in 2010. Using a propensity score approach, we estimated the likelihood of being enrolled in the trial given patient characteristics. Then, we observed the pattern of adopting any panitumumab-FOLFIRI regimen as the second-line therapy for the elderly patients with mCRC in practice, by the quartile-based groups determined by the probability of being enrolled in a trial. We also examined whether the pattern of adopting the regimen changed before and after 2010. The 95% confidence intervals (CI) of adopting the panitumumab-FOLFIRI regimen were obtained using the bootstrap method.

In Aim 2, we developed an iterative algorithm using copula and resampling, which are based on the estimated propensity score for likelihood of enrollment in a trial given participants' characteristics. Validation was performed using Monte-Carlo simulations under different scenarios where the marginal and joint distributions of covariates differ between trial samples and RWD. We also illustrated this method with two applications using an actual colorectal

cancer trial and the SEER-Medicare linked database. First, we applied existing methods (i.e. standardized mean difference, B-index) to assess the generalizability of the trial. Then, we explored the feasibility of using RWD to generate an external control group by applying a parametric Weibull model trained in RWD to predict overall survival in the simulated trial cohort. The real-world target population were selected based on the eligibility trial criteria if applicable.

Results

In Aim1, among 2,815 treated metastatic colorectal cancer patients in the SEER-Medicare linked database, only 41 (1.5%) patients used any panitumumab-FOLFIRI regimen from 2008 to 2016. A positive association was observed between the probability of receiving the panitumumab-FOLFIRI regimen and the probability of being enrolled in the trial. After the trial results published in 2010, the probability of receiving any panitumumab-FOLFIRI regimen was 3% (95% CI: 0.011, 0.044) for patients who had higher probabilities to be enrolled in a trial (above the third quartile), whereas it was only 1 % (0.004, 0.016) for those who had lower probabilities (below the first quartile). By contrast, fewer than 1 % patients received any panitumumab-FOLFIRI regimen before 2010 regardless of the likelihood of being in a trial. The pattern of adopting the panitumumab-FOLFIRI regimen significantly differed before and after 2010 among patients whose probabilities of being enrolled in a trial were above the median. The mean difference in the probability of adopting the regimen before and after 2010 were 0.01 (0.002, 0.019) and 0.027 (0.015, 0.042) in the highest two quartiles respectively. Yet, the difference was very small because only few patients used the regimen in the real world. Additionally, we found

that the treatment was adopted broadly many years later to patients whose characteristics were less similar with trial participants.

In Aim 2, across all of the simulation scenarios, this iterative algorithm could successfully approximate correlations of covariates in a trial, which were closer to the true correlations than the correlations in RWD. The algorithm could also successfully reproduce the joint distribution of baseline characteristics for an actual cancer trial using summary data from the trial and the real-world cohort from the SEER-Medicare linked database. Therefore, we can obtain the similar estimates of standardized mean difference and B-index using simulated trial data (1.25 and 0.77 respectively) to assess the generalizability of the trial, comparing to the value of using the individual-level trial data (1.33 and 0.78 respectively). Additionally, no difference in approximated correlations was observed when we applied different eligibility trial criteria to select the target population. Lastly, we found that the adjusted survival estimates among the simulated trial population were close to the actual trial Kaplan-Meier (K-M) estimates, where the 95% confidence intervals of these survival curves were overlapped.

Conclusions

In this dissertation, we provided an approach for exploring how physicians would adopt the new treatment in practice post clinical trials using a specific colorectal cancer trial as an example in Aim1. Our study also implied that slowly adopting a new treatment in practice would lead to insufficient sample sizes and limit the use of RWD. Additionally, we found that our algorithm could be a feasible way to simulate individual-level clinical trial data when only summary data from the trials are available, which could be used to assess the generalizability of clinical trials

given participants' characteristics and inform decision making around the applicability of trial results to a real-world population.

TABLE OF CONTENTS

List of Figures

List of Tables

ACKNOWLEDGEMENTS

I have been at the University of Washington seven years since I studied my MS in Health Services, I am so grateful to have great mentors and amazing friends here and have a good time over years. I do appreciate any opportunities and assistance I received, and also all support and encouragements I got when I have a hard time.

I would like to sincerely thank my dissertation committee: Anirban Basu, Aasthaa Bansal, Carrie Bennette, and Yvonne Lin, for their mentorship and guidance. During this process, I was guided to think critically, to deliver complicated technical information efficiently, and to overcome the challenges in academia. This experience is definitely invaluable in my life.

Special thanks to Beth Devine, David Veenstra, and Ryan Hansen. Thank Beth for giving me many opportunities over years and always being very supportive for international students, particularly in helping us to correct our program category to a STEM (science, technology, engineering and math) degree; thank Dave for encouraging me and giving me suggestions to explore my capability and value; and thank Ryan for giving me research opportunities and also helping me when I was looking for a summer internship in industry.

Additionally, thanks to other faculty members for their instructions through the courses, and thanks to my cohort (especially for Nathaniel Hendrix), the group meeting members, and all friends I met in the CHOICE Institute for their company over five years. Lastly, many thanks to my parents and sisters for their understanding and support over these years.

Chapter 1: Introduction

Sound evidence-based medicine is essential for health policy makers to adopt treatments, design insurance coverage plan and efficiently allocate scarce resources. A decade ago, the Institute of Medicine's Roundtable on Evidence-Based Medicine set a goal that "By the year 2020, 90 percent of clinical decisions will be supported by accurate, timely, and up-to-date clinical information and will reflect the best available evidence."[1,2] Since high burden of cancer remains over years, the Institute of Medicine (IOM), now National Academy of Medicine (NAM), released recommendations on high-quality cancer care in 2013. In addition to deliver patient-centered care, IOM suggested developing evidenced-based cancer care and informing clinical decisions with scientific research (e.g. clinical trials, comparative effectiveness research).[3]

A well-designed randomized controlled trial (RCT) is usually considered as the gold standard for investigating the safety and efficacy of new treatments. However, it is infeasible to gather evidence through RCT due to ethical issues or rare outcomes. Additionally, the most critical pitfall of the clinical trials is the generalizability for several reasons. In oncology, fewer than two percent of total cancer patients are represented in cancer trials, where the elderly and racial minorities are much more likely to be underrepresented in the trials.[4,5] On the other hand, trial participants required to meet restrictive eligibility criteria tend to be healthier and less diverse than the general population. Yet even after applying same eligibility criteria to select the real-world target population for a comparison, the discrepancy between trial participants and general population cannot be fully eliminated, indicating the limits of external validity of the clinical trials.[6]

Therefore, using real-world data (RWD) is a feasible and essential approach to provide invaluable evidence that complements clinical trials data to inform business strategies, clinical practice, coverage and payment decisions.[7,8] In a broad sense, any information of patient health care collected outside traditional clinical trials could be considered as RWD. More precisely, the US Food and Drug Administration (FDA) defines RWD as any data reflect patient health status and the delivery of health care and are routinely collected from a range of sources, including but not limited to electronic health records (EHRs), claims and billing data, product and disease registries, and patient-generated information.[9] Along with appropriate study design and statistical methods, RWD have been widely used to understand the burden of illness and treatment patterns, and to evaluate the effectiveness and long-term safety of the treatments in the practice setting.[10,11] Moreover, RWD can be used to assess the generalizability of trials by evaluating the similarities between trials participants and target population, and then project potential treatment effects with weighting to target population.[12-14]

In 2016, the 21st Century Cures Act highlights the importance of real-world evidence (RWE), which is defined by FDA as the clinical evidence reveals the usage and potential benefits/risks of medical products derived from RWD.[15,16] Nevertheless, it could be challenging to use real-world evidence for decisions making, particularly in regulatory decisions.[17,18] Most common RWD, such as EHRs and administrative claims data, mainly reflect the adoption of treatments in practice depending on physician judgment and patient preferences as well as insurance coverage, implying potential selection bias and unobserved confounders. Appropriate study design and analysis are critical to ensure the credibility of evidence from observational RWD studies.

Consequently, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the International Society for Pharmacoepidemiology (ISPE) special task force made recommendations regarding good procedural practices for RWD studies to improve the quality and reproducibility of RWE, which may help to enhance confidence in RWE for decision making.[19,20] Additionally, the use of RWD is driven by data availability, completeness and accuracy. Clearly, using RWD is infeasible to evaluate unapproved treatments, and it is insufficient to provide evidence before the treatments were used widely in the general population either. Also, incomplete clinical information (e.g. disease severity, imaging and laboratory testing results) and miscoding data could potentially influence the reliability of RWE and would limit the possibility of RWD to replicate clinical trial findings. A recent study showed that only 15% of clinical trials published in 2017 in the United states can be replicated using sufficient data from claims or EHRs to identify intervention, indication, eligibility criteria and primary end points in the trials.[21]

Committing an act of incorporating RWE in decision making, FDA has released a new framework for FDA's Real-World Evidence Program in the end of 2018, which emphasizes the feasibility of using RWD in regulatory decisions.[22] One of the strategies proposed for study designs is to explore the potential use of RWD-based external control for single arm studies. Compared to a historical control derived from past clinical trials, using RWD-based external control may be a better option because the reliability of evidence using historical control depends on the comparability of population at different time points.[23,24]

Along with propensity score-based approaches, recent studies used EHRs and individual-level cancer trial data to demonstrate the possibility of using an external RWD control group.[25,26] The purpose of using propensity score (PS) methods is to compare and balance observed patient characteristics between trial and RWD population and allow us to apply PS weighting for adjustment. Yet, it is difficult to use this approach when individual-level trial data are not available. Although data sharing of clinical trials is recommended to maximize the value of clinical trials, it is still difficult to access to most individual-level clinical trial data, especially outside the industry. Fortunately, while only published summary trial data are available, advanced statistical methods (e.g. matching adjusted indirect comparison (MAIC), copulas methods) could be potentially useful to address this issue and fill the gap. [27,28]

Clearly, RWD are not advocated as a substitute of RCTs, but afford the potential opportunities to increase efficiency with innovative approaches for future decision making. To get a better understanding of RWD, it is important to explore how new technologies have been adopted in the general population. Are treatment recipients different from RCTs participants? Are physician decisions associated with patient characteristics? Additionally, it is necessary to develop a novel approach to overcome the challenges when individual-level trial data are not available.

In this dissertation, an actual trial data and the SEER-Medicare linked database were used to answer the above questions. In the Aim 1, we explored physician decisions on adaptation of treatments among the elderly colorectal cancer patients, focusing on the similarity of patient characteristics between treatment recipients and RCTs participants. In the Aim 2, we developed an iterative algorithm incorporating propensity score methods with copulas, which could be

potentially used in two applications: (1) to assess the generalizability of trials, and (2) to assess the feasibility of generating an external control using published summary trial data and RWD. We first demonstrated and validated the algorithm through Monte Carlo simulations, and also applied the algorithm to an empirical example. The actual individual-level trial data were only used for validation in the Aim 2.

References

1. *Institute of Medicine, Evidence-Based Medicine and the Changing Nature of Health Care: 2007 IOM Annual Meeting Summary.* Washington, DC: The National Academies Press; 2008.
2. *Institute of Medicine, Leadership Commitments to Improve Value in Health Care: Finding Common Ground: Workshop Summary.* Washington, DC: The National Academies Press; 2009.
3. Ganz PA. Institute of Medicine report on delivery of high-quality cancer care. *J Oncol Pract.* 2014;10(3):193-195.
4. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: Race-, sex-, and age-based disparities. *JAMA.* 2004;291(22):2720-2726.
5. *Institute of Medicine, Forum on Drug Discovery, Development, and Translation. Transforming Clinical Research in the United States: Challenges and Opportunities: Workshop Summary.* Washington DC: The National Academies Press; 2010.
6. Steventon A, Grieve R, Bardsley M. An Approach to Assess Generalizability in Comparative Effectiveness Research: A Case Study of the Whole Systems Demonstrator Cluster Randomized Trial Comparing Telehealth with Usual Care for Patients with Chronic Health Conditions. *Medical Decision Making.* 2015;35(8):1023-1036.
7. Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. *British Journal Of Cancer.* 2014;110:551.
8. Garrison LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report. *Value in Health.* 2007;10(5):326-335.
9. US Food and Drug Administration, Real-World Evidence, https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence. 2020.
10. Basu A, Axelsen K, Grabowski DC, et al. Real-World Data: Policy Issues Regarding their Access and Use. *Medical Care.* 2016;54(12):1038-1044.
11. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence — What Is It and What Can It Tell Us? *New England Journal of Medicine.* 2016;375(23):2293-2297.
12. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society).* 2010;174(2):369-386.
13. Cole SR, Stuart EA. Generalizing Evidence From Randomized Clinical Trials to Target Populations: The ACTG 320 Trial. *American Journal of Epidemiology.* 2010;172(1):107-115.
14. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of Trial Results Using Inverse Odds of Sampling Weights. *Am J Epidemiol.* 2017;186(8):1010-1014.
15. 21st Century Cures Act , https://www.congress.gov/114/bills/hr34/BILLS-114hr34enr.pdf. 2016.
16. Barlas S. The 21st Century Cures Act: FDA Implementation One Year Later: Some Action, Some Results, Some Questions. *P & T : a peer-reviewed journal for formulary management.* 2018;43(3):149-179.
17. Beaulieu-Jones BK, Finlayson SG, Yuan W, et al. Examining the Use of Real-World Evidence in the Regulatory Process. *Clin Pharmacol Ther.* 2020;107(4):843-852.

18. Hampson G, Towse A, Dreitlein WB, Henshall C, Pearson SD. Real-world evidence for coverage decisions: opportunities and challenges. *Journal of Comparative Effectiveness Research.* 2018;7(12):1133-1143.
19. Berger ML, Sox H, Willke RJ, et al. Good Practices for Real-World Data Studies of Treatment and/or Comparative Effectiveness: Recommendations from the Joint ISPOR-ISPE Special Task Force on Real-World Evidence in Health Care Decision Making. *Value Health.* 2017;20(8):1003-1008.
20. Wang SV, Schneeweiss S, Berger ML, et al. Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0. *Pharmacoepidemiol Drug Saf.* 2017;26(9):1018-1032.
21. Bartlett VL, Dhruva SS, Shah ND, Ryan P, Ross JS. Feasibility of Using Real-World Data to Replicate Clinical Trial Evidence. *JAMA Netw Open.* 2019;2(10):e1912869.
22. U.S. Food and Drug Administration, Framework for FDA's Real-World Evidence Program, https://www.fda.gov/scienceresearch/specialtopics/realworldevidence/default.htm. 2018.
23. Viele K, Berry S, Neuenschwander B, et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharm Stat.* 2014;13(1):41-54.
24. Pandis N, Cobourne MT. Clinical trial design for orthodontists. *J Orthod.* 2013;40(2):93-103.
25. 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.
26. Davies J, Martinec M, Delmar P, et al. Comparative effectiveness from a single-arm trial and real-world data: alectinib versus ceritinib. *Journal of Comparative Effectiveness Research.* 2018;7(9):855-865.
27. Signorovitch JE, Wu EQ, Yu AP, et al. Comparative Effectiveness Without Head-to-Head Trials. *PharmacoEconomics.* 2010;28(10):935-945.
28. Signorovitch JE, Sikirica V, Erder MH, et al. Matching-Adjusted Indirect Comparisons: A New Tool for Timely Comparative Effectiveness Research. *Value in Health.* 2012;15(6):940-947.

Chapter 2: An Exploratory Analysis of Adopting a New Cancer Treatment in Practice – a Case Study of the Colorectal Cancer Trial

2.1 ABSTRACT

Background

Real-world data (RWD) have been widely used to evaluate the effectiveness and long-term safety of new treatments in practice to support evidence from clinical trials. Yet, the use of RWD would be limited when new treatments were slowly adopted in practice. Since clinical trial evidence may not be generalizable, it is important to know when and for whom the physicians would adopt the new treatments post clinical trials. Therefore, the aim of the study was to identify whether physician decisions on the adoption of treatment were influenced by the similarity of patient characteristics between trial participants and general target population.

Methods

We conducted an exploratory analysis using individual-level trial data and the SEER-Medicare linked database. The trial aimed to evaluate the efficacy of panitumumab plus FOLFIRI for treated metastatic colorectal cancer (mCRC), initially published in 2010. Using a propensity score approach, we estimated the likelihood of being enrolled in the trial given patient characteristics. Then, we observed the pattern of adopting any panitumumab-FOLFIRI regimen as the second-line therapy for the elderly patients with mCRC in practice, by the quartile-based groups determined by the probability of being enrolled in a trial. We also examined whether the pattern of adopting the regimen changed before and after 2010. The 95% confidence intervals (CI) of adopting the panitumumab-FOLFIRI regimen were obtained using the bootstrap method.

Results

Among 2,815 treated metastatic colorectal cancer patients in the SEER-Medicare linked database, only 41 (1.5%) patients used any panitumumab-FOLFIRI regimen from 2008 to 2016. A positive association was observed between the probability of receiving the panitumumab-FOLFIRI regimen and the probability of being enrolled in the trial. After the trial results published in 2010, the probability of receiving any panitumumab-FOLFIRI regimen was 3% (95% CI: 0.011, 0.044) for patients who had higher probabilities to be enrolled in a trial (above the third quartile), whereas it was only 1 % (0.004, 0.016) for those who had lower probabilities (below the first quartile). By contrast, fewer than 1 % patients received any panitumumab-FOLFIRI regimen before 2010 regardless of the likelihood of being in a trial. The pattern of adopting the panitumumab-FOLFIRI regimen significantly differed before and after 2010 among patients whose probabilities of being enrolled in a trial were above the median. The mean difference in the probability of adopting the regimen before and after 2010 were 0.01 (0.002, 0.019) and 0.027 (0.015, 0.042) in the highest two quartiles respectively. Yet, the difference was very small because only few patients used the regimen in the real world. Additionally, we found that the treatment was adopted broadly many years later to patients whose characteristics were less similar with trial participants.

Conclusions

Using a specific colorectal cancer trial as an example, our study may provide an approach for exploring how physicians would adopt the new treatment in practice post clinical trials. Our study also implied that slowly adopting a new treatment in practice would lead to insufficient sample sizes and limit the use of RWD.

2.2 Introduction

In 2013, the Institute of Medicine (IOM), now National Academy of Medicine (NAM), released recommendations on high-quality cancer care that emphasized the development of evidence-based practice guidelines and informing clinical decisions with scientific research.[1] While randomized controlled trials (RCTs) are considered as the gold standard for evaluating the efficacy and safety of new treatments, however, RCTs have limited generalizability. Therefore, real-world data (RWD) have been widely used to understand the long-term benefits and harms of new treatments in routine practice to complement the clinical trial findings. [2-4] Yet, challenge arises when the new treatments were slowly adopted in practice, implying sufficient RWD would not be available in terms of electronic health records (EHRs) and insurance claims data.

In practice, physicians play a crucial role in clinical decisions by assessing available treatment options for improving patient health outcomes. Physician prescribing decisions are possibly influenced by a variety of factors, including the benefits, risks, and costs of drugs; patient / physician preferences; peer influence; and marketing by the pharmaceutical industry.[5,6] Several surveys have been conducted to understand patient or physician preferences for cancer treatments. In oncology, the benefits and toxicities of treatments are the key factor affecting clinical decisions.[7] When adopting new treatments, physicians tend to use information from scientific publications and medical congress announcements to justify their prescribing decisions, which could be also affected by their peer physicians.[6,8] Additionally, most patient requests or demands for a specific cancer treatment were found to be clinically appropriate based on physician judgments.[9]

Yet, when new evidence becomes publicly available, it is unclear how physicians would translate scientific evidence into clinical practice. Since clinical trial evidence may not be generalizable, it is important to understand when and for whom physician would adopt the new treatment in practice. It is possible that physicians may start to adopt the treatment to patients who are more likely to have treatment benefits. Therefore, we aimed to identify whether physician decisions on the adoption of treatment are influenced by the similarity of patient characteristics between trial participants and general target population. Given the availability of individual-level clinical trial data, we used a colorectal cancer trial as a case in this study.

Colorectal cancer is the second leading cause of cancer death in the United States, and the five-year survival of metastatic colorectal cancer is only 14%.[10,11] In the past two decades, potential treatment options including monotherapy and combination therapy for metastatic colorectal cancer were investigated through plenty of clinical trials.[12,13] Panitumumab (Vectibix™), an epidermal growth factor receptor (EGRF) monoclonal antibody, was approved for patients with chemo-refractory metastatic colorectal cancer in 2006.[14] Following the approval, Amgen Inc. conducted a phase 3 randomized multicenter clinical trial to investigate the efficacy of panitumumab in combination with FOLFIRI (Irinotecan plus leucovorin and 5-fluorouracil (5-FU)) as the second-line treatment for metastatic colorectal cancer (NCT00339183, Amgen protocol number: 20050181). The first trial results published in November 2010 showed a significant improvement in progression-free survival (PFS) between the panitumumab-FOLFIRI arm and FOLFIRI alone arm, while the final results were released in January 2014.[15,16] Hence, the aim of this study is to explore how physicians would adopt the panitumumab-FOLFIRI

regimen in practice and whether the pattern of adopting the regimen differs before and after the published trial results.

2.3 Methods

Data sources

We requested the individual-level clinical trial data through the Project Data Sphere Cancer Research Platform, which is operated by CEO Roundtable on Cancer's Life Sciences Consortium (LSC).[17] Thereby, we had access to the trial protocol and individual-level data of the Amgen trial (20050181) which compared the efficacy of panitumumab plus FOLFIRI with the efficacy of using FOLFIRI alone for previously treated metastatic colorectal cancer. Then, we used the Surveillance, Epidemiology and End Results (SEER)-Medicare Linked Database as the real-world data. The SEER-Medicare linked database is composited of cancer registries and Medicare claims, which includes detailed information of cancer status (e.g. stage, histology, etc.) and health services utilizations (e.g. diagnosis, procedure, treatment, etc.).[18] The 2018 linkage files represent cancer patients diagnosed before 2015 and include Medicare claims from 2007 through the end of 2016.

Real-world target population

Since metastasis occurring after initial diagnosis is not available in the SEER-Medicare linked database, we defined our target population as treated metastatic colorectal cancer patients who were initial diagnosed from 2008 to 2015. We first identified colorectal cancer patients with adenocarcinoma using the International Classification of Diseases for Oncology, third revision (ICD-O-3) site codes (C180-C189, C199, C209) and histology codes (8140, 8480, 8490, 8210,

8220, 8261), and then defined their cancer stage using the American Joint Committee on Cancer (AJCC) staging system.[19] We excluded patients with other primary cancers in this study. The patients' medical history related to the eligibility criteria of the trial was identified using International Classification of Diseases, Ninth/tenth Revision, Clinical Modification (ICD-9-CM/ICD-10-CM), ICD-9-CM procedure codes, ICD-10 Procedure Coding System (ICD-10-PCS), Healthcare Common Procedure Coding System (HCPCS) codes, and Current Procedural Terminology (CPT) codes.

Lines of therapy

We focused on common treatments that were approved for metastatic colorectal cancer before 2015, including 5-FU, leucovorin, oxaliplatin, irinotecan, capecitabine, cetuximab, bevacizumab, panitumumab (Vectibix), ziv-aflibercept (Zaltrap), regorafenib (Stivarga), trifluridine-tipiracil (LONSURF), ramucirumab (Cyramza).[20] Then, we used Healthcare Common Procedure Coding System code (HCPCS) and National Drug Code (NDC) to identify Intravenous (IV) and oral chemotherapy drugs respectively. The following criteria was used to determine the lines of therapy, which were consistent with the approaches in the previous treatment pattern studies for metastatic colorectal cancer.[21-26] Since the treatments could be used as a single-agent or used in combination, we considered all agents that were prescribed during the first 36 days after the initial treatment as the first-line treatment. A new line of therapy was defined as treatment switching after first-line treatment except maintenance therapy (e.g. adding capecitabine / bevacizumab to the previous chemotherapy regimens) or discontinuing one agent from the combination of chemotherapy. Additionally, a treatment was considered as the next line when the gap between two treatments was more than 90 days regardless of the reasons. This study was

reviewed and approved by the Institutional Review Board (IRB) in the University of Washington.

Exploratory analysis

Only patients aged 65 years and older were included in our analysis because of data availability. Descriptive statistics was used to examine the pattern of adopting any panitumumab-FOLFIRI or FOLFIRI regimens for the second-line therapy before and after 2010 when the first trial results were published. Note that the panitumumab-FOLFIRI or FOLFIRI regimens were potentially used in combination with any other agents in practice. Additionally, we explored and compared the characteristics of patients who received the panitumumab-FOLFIRI regimen to those who used other treatments as the second line treatment.

To identify whether physicians were more likely to adopt the panitumumab-FOLFIRI regimen for patients who had similar characteristics with trial participants, we first combined the AMGEN trial data (Group=1) with the SEER-Medicare data (Group=0) and used a logistic regression to fit a saturated propensity score model (i.e. including all covariates and possible interaction terms) to estimate the probability of being enrolled in a trial given patient characteristics at their start date of second-line treatment. The baseline patient characteristics included age, gender, cancer subtype (primary tumor site located in colon or rectum), and prior treatments (oxaliplatin or bevacizumab). Then, we modeled the probability of receiving the panitumumab-FOLFIRI regimen on the probability of being enrolled in the trial using a logistic regression. The bootstrap method was used to obtain 95% confidence intervals of the probability of receiving the panitumumab-FOLFIRI regimen and of the mean difference in the probability of

receiving the panitumumab-FOLFIRI regimen before and after 2010. All analyses were conducted using SAS version 9.4 (SAS Institute, Inc., Cary, NC, USA) and RStudio 3.6.0 (RStudio Inc., Boston, MA).

2.4 Results

In our analysis, we included 276 trial participants who aged 65 years and older from the Amgen trial, and identified a total of 2,815 treated metastatic colorectal cancer patients from 2008 to 2015 in the SEER-Medicare linked database. Table 2.1 shows the distribution of patient characteristics between trial participants and the real-world target population. We found that trial participants were slightly younger and had smaller variation in age than the real-world population. The trial cohort included more male participants (71% vs. 55%) and fewer patients with primary tumor site located in colon (67% vs. 82%) than the real-world cohort. For the prior treatment, both the trial cohort and the real-world cohort included about 70% patients treated with oxaliplatin, whereas the percentage of patients using bevacizumab was much lower in the trial cohort than in the real-world cohort (16% vs 61%).

Of total 2,815 metastatic colorectal cancer patients aged 65 years and older, 725 (26%) patients received the second-line treatment from 2008 to 2010, and the rest of them received treatments after 2010. From 2008 to 2016, we found that few patients received the panitumumab-FOLFIRI regimen for the second-line therapy; only 41 (1.5%) patients used panitumumab plus FOLFIRI in combination with any other drugs, including 29 patients who only used panitumumab plus FOLFIRI without other agents (Table 2.2). Overall, only 153 (5.5%) patients used panitumumab-related therapy, and 1,559 (55.4%) of patients received treatments related to neither

panitumumab nor FOLFIRI. The percentage of patients using panitumumab was 6% after 2010, which was slightly higher than that before 2010.

Owing to a small sample size, patients using any treatments in combination with panitumumab plus FOLFIRI were considered as receiving the panitumumab-FOLFIRI regimen in the following analysis. Using a saturated propensity score model, the average probability of being enrolled in a trial for the real-world population was 0.08 given their baseline characteristics, where the maximum value was only 0.44.

Figure 2.1 shows the average predicted probability of receiving panitumumab plus FOLFIRI among the quartile-based groups determined by the probability of being enrolled in a trial. Overall, we observed a positive association between the probability of receiving any panitumumab-FOLFIRI regimen in practice and of being enrolled in trial from 2008 to 2016. Interestingly, the opposite patterns of the association were found before and after 2010. Following the first trial results published in 2010, the probability of receiving any panitumumab-FOLFIRI regimen was slightly increasing from 1 % to 3 % for patients whose probabilities of being enrolled in a trial were increasing from the lowest quartile (below the first quartile) to the highest quartile (above the third quartile). By contrast, fewer than 1 % patients received any panitumumab-FOLFIRI regimen before 2010 regardless of the likelihood of being in a trial. A significant difference in the probability of being treated with any panitumumab-FOLFIRI regimen was found before and after 2010 among patients who were more likely to be enrolled in a trial (i.e. above the median probability of enrollment), shown in Table 2.3. Additionally, we also observed that the probability of receiving any panitumumab-FOLFIRI regimen slightly

increased for those who were less likely to be enrolled in a trial after 2014, when the final trial results were published.

While comparing patient characteristics of treatment recipients to non-recipients, we found that recipients were slightly younger than non-recipients and the majority were male (Table 2.4). None of the treatment recipients received hospice care before their second-line therapy. The percentage of receiving oxaliplatin (81%) or panitumumab (10%) in prior treatment were also higher among 41 treatment recipients than those who did not receive the panitumumab-FOLFIRI regimen. Most importantly, we found that only 39 % of treatment recipients would meet the eligibility criteria of the trial in terms of medical history.

2.5 Discussion

In this study, using a colorectal cancer trial as a case study, we explored how physicians would adopt the new treatment in practice when the trial evidence was available and whether the decisions would be associated with patient characteristics. Therefore, we examined the pattern of using any panitumumab-FOLFIRI regimen among the elderly patients and estimated the likelihood of receiving the treatment and of being enrolled in the trial given available patient characteristics. Unfortunately, only few patients used any panitumumab-FOLFIRI regimen in our real-world cohort. From our exploratory analysis, we observed that physician decisions on the adoption of treatment in practice might be positively associated with the similarity of patient characteristics between trial participants and the real-world population after the first trial results published. Also, the treatment may be adopted more broadly many years later to patients whose characteristics were less similar with trial participants.

As the target population are different from the trial participants, it is important to know when and for whom the physicians would adopt the new treatment post clinical trials, which might help to optimize the use of real-world data. Ideally, RWD are expected to provide more representative evidence to support findings from the clinical trials. By comparing the patient characteristics between the trial and the real-world population, we can examine whether physicians adopted the new regimens to the broader population, and this indicator may help us to determine when to use RWD regarding the health outcomes among diverse populations. However, a new treatment might be slowly adopted in practice after the trial evidence available, which would lead to insufficient sample sizes and limit the use of RWD.

Several reasons may explain why only 1.5% patients received any panitumumab-FOLFIRI regimen as their second-line treatment in our selected target population. First, our real-world cohort only included patients aged 65 years and older. It is possible that the elderly may be more vulnerable and less likely to tolerate the toxicities of using multiple drugs in combination. Physicians may tend to adopt new treatments to the younger and healthier population after evaluating the potential benefits and risks of the treatments. Second, the trial evidence only showed using panitumumab plus FOLFIRI as the second-line treatment can significantly improve progression-free survival among patients with wild-type KRAS tumors. However, approximately 30-60% of colorectal cancer patients may carry KRAS mutate tumors.[27,28] It is very likely that physicians would not adopt the panitumumab-FOLFIRI regimen to patients with mutant KRAS tumors since no evidence of clinical benefits was found among these patients. Even for patients with wild-type KRAS tumors, physicians may consider and adopt other

regimens as the second-line treatment because there was no significant improvement in overall survival when patients used the panitumumab-FOLFIRI regimen. In practice, the most common second-line regimens for metastatic colorectal cancer patients included FOLFIRI / FOLFOX (with or without bevacizumab) or cetuximab-based regimens (with or without FOLFIRI/irinotecan), and all of these treatments were approved more than fifteen years ago.[29-31] Since the published clinical trial results of a new treatment may be the only available evidence in the earlier years, the slow adoption of new treatments may imply that physicians would need more evidence to inform clinical decisions. Also, incorporating scientific evidence into clinical guidelines may help to speed the adoption of new treatments in practice.

The key limitations of this study are data availability. Surprisingly, we only found few metastatic colorectal cancer patients using the panitumumab-FOLFIRI regimen from the SEER-Medicare linked database, which are not sufficient to examine the trend of adopting the new treatment by years or to perform a statistical hypothesis testing. The number of patients receiving the panitumumab-FOLFIRI regimen may be underestimated in our analysis because the metastatic colorectal cancer patients were selected based on their initial diagnosis and must be age 65 years and older. Nevertheless, a previous study using the Flatiron Health electronic medical records database also showed that only 2.5% metastatic colorectal cancer patients received the panitumumab-FOLFIRI regimen from 2013 to 2018, indicating the regimen may not be frequently adopted in general.[32] Additionally, some patient characteristics related to the likelihood of being enrolled in a trial were not available in the SEER-Medicare linked database, including Eastern Cooperative Oncology Group (ECOG) performance status and KRAS status. Therefore, the similarity of patient characteristics between trial participants and the rea-world

population was only evaluated given the common variables available in both the AMGEN trial and SEER-Medicare linked database. Lastly, we only used one cancer trial to explore whether physicians tended to adopt the panitumumab-FOLFIRI regimen to metastatic colorectal cancer patients whose characteristics were more similar with the trial participants; thus, our findings would not be applicable for other trials.

2.6 Conclusions

Despite these limitations, our study may provide an approach for exploring how physicians would adopt the new treatment in practice post clinical trials using a specific colorectal cancer trial as an example. Further work is needed to explore when the new treatment would be adopted to broader population for other treatments and cancer types. Also, data sharing of clinical trials is essential since it would be challenging to conduct such studies when individual level trial data are not available.

Acknowledgments

- "This study is based on research using information obtained from www.projectdatasphere.org, which is maintained by Project Data Sphere. Neither Project Data Sphere nor the owner(s) of any information from the web site have contributed to, approved or are in any way responsible for the contents of this study."
- "This study used the linked SEER-Medicare database. The interpretation and reporting of these data are the sole responsibility of the authors. The authors acknowledge the efforts of the National Cancer Institute; the Office of Research, Development and Information, CMS; Information Management Services (IMS), Inc.; and the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER-Medicare database."
- "This work was facilitated though the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington"

2.7 Tables and Figures

Table 2.1. Patient Characteristics Between Trial Participants and the Real-World Population

Characteristics	The real-world cohort (n=2815)	The trial cohort (n=276)
Age, mean (SD)	72.59 (5.98)	70.74 (3.91)
Male, n (%)	1539 (54.7)	195 (70.7)
Prior treatment – oxaliplatin, n (%)	1990 (70.7)	191 (69.2)
Prior treatment – bevacizumab, n (%)	1710 (60.7)	44 (15.9)
Primary tumor site located in colon, n (%)	2317 (82.3)	186 (67.4)

Table 2.2. The Second-line Treatment Patterns Among Real-world population

Treatments for the second-line therapy	Total (N)	Before 2010 (2008~2010) n (%)	After 2010 (2011~2016) n (%)
Overall	2815	725	2090
Any combination with panitumumab	153	28 (0.04)	125 (0.06)
Panitumumab alone	48	10 (0.01)	38 (0.02)
Any panitumumab-FOLFIRI	**41**	5 (0.01)	36 (0.02)
Only panitumumab-FOLFIRI	**29**	4 (0.01)	25 (0.01)
Any combination with FOLFIRI*	1144	280 (0.39)	864 (0.41)
FOLFIRI plus capecitabine/bevacizumab	630	160 (0.22)	470 (0.22)
FOLFIRI alone	276	69 (0.10)	207 (0.10)
Other drugs	1559	422 (0.58)	1137 (0.54)

*included any panitumumab-FOLFIRI regimen

Table 2.3. The Probability of Adopting the Regimen by the Probability of Being in a Trial

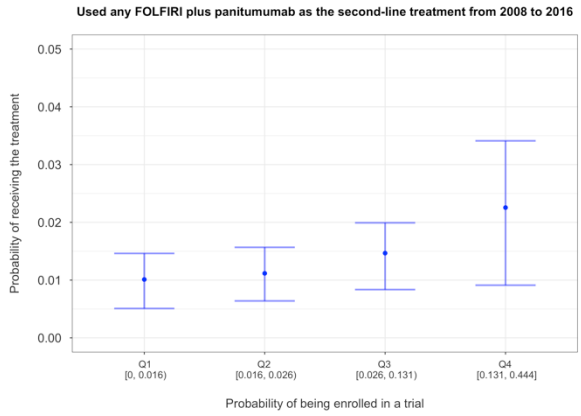
Quartile-based groups*	Probability of adopting the panitumumab-FOLFIRI regimen mean (Bootstrapped 95% CI)			
	Overall (2008 to 2016)	Before 2010 (2008~2010)	After 2010 (2011~2016)	Mean difference before and after 2010
First quartile	0.01 (0.005, 0.015)	0.009 (0, 0.017)	0.011 (0.004, 0.016)	(-0.009, 0.011)
Second quartile	0.011 (0.006, 0.016)	0.009 (0, 0.016)	0.012 (0.006, 0.017)	(-0.007, 0.012)
Third quartile	0.015 (0.008, 0.02)	0.007 (0.001, 0.012)	0.017 (0.009, 0.023)	(0.002, 0.019)
Fourth quartile	0.023 (0.009, 0.034)	0.002 (0.001, 0.002)	0.029 (0.011, 0.044)	(0.015, 0.042)

*group by the probability of being in a trial for each time period

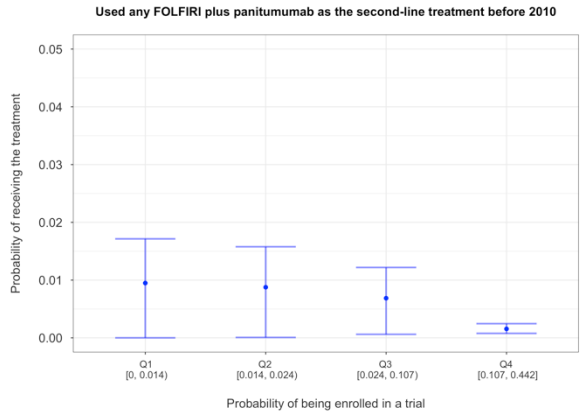
Table 2.4. Patient Characteristics Between Treatment Recipients and non-Recipients

Characteristics	Received any panitumumab-FOLFIRI regimen	
	Yes (n=41)	No (n=2774)
Age, mean (SD)	69.66 (4.09)	72.63 (5.99)
Male, n (%)	29 (70.7)	1510 (54.4)
Prior treatment – oxaliplatin, n (%)	33 (80.5)	1957 (70.5)
Prior treatment – bevacizumab, n (%)	24 (58.5)	1686 (60.8)
Primary tumor site located in colon, n (%)	34 (82.9)	2283 (82.3)
Used panitumumab as the first-line therapy (%)	4 (9.8)	25 (0.9)
Any hospice before the second-line therapy (%)	0 (0.0)	3 (0.1)
Met criteria regarding medical history (%)	16 (39.0)	1118 (40.3)

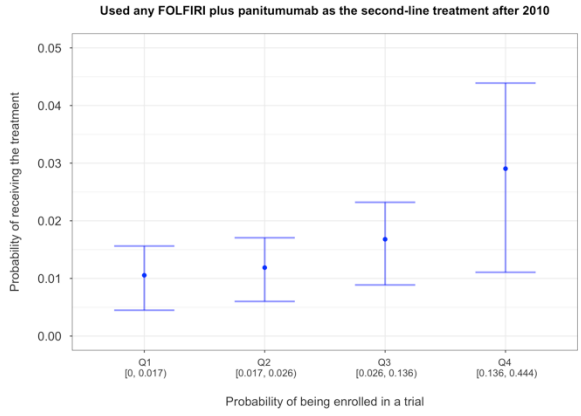
(A)



(B)



(C)



(D)

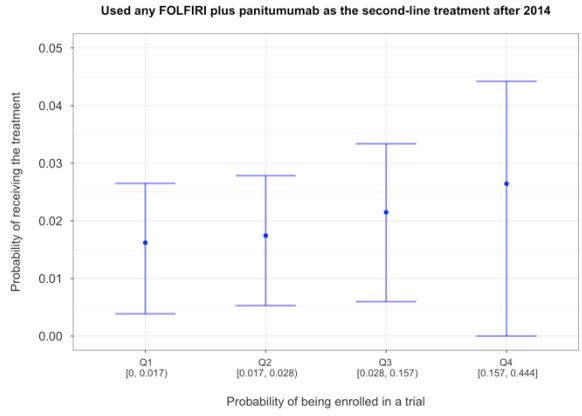


Figure 2.1. the Trend of Adopting the panitumumab-FOLFIRI Regimen in Different Time Periods: (A) from 2008 to 2016 (B) before 2010 (C) after 2010 (2011-2016) (D) after 2014 (2014-2016).

(x-axis: quartile-based groups by the probability of being enrolled in a trial, ranged from the minimum value to the first quartile (Q1), from Q1 to the median (Q2), from Q2 to the third quartile (Q3), and from Q3 to the maximin value for each group respectively; y-axis: the average predicted probability of received the regimen)

2.8 References

1. Ganz PA. Institute of Medicine report on delivery of high-quality cancer care. *J Oncol Pract.* 2014;10(3):193-195.
2. Garrison LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report. *Value in Health.* 2007;10(5):326-335.
3. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence — What Is It and What Can It Tell Us? *New England Journal of Medicine.* 2016;375(23):2293-2297.
4. Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. *British Journal Of Cancer.* 2014;110:551.
5. Ali Murshid M MZ. Models and theories of prescribing decisions: A review and suggested a new model. *Pharmacy Practice.* 2017;15(2):990.
6. Keating NL, O'Malley AJ, Onnela JP, Gray SW, Landon BE. Association of Physician Peer Influence With Subsequent Physician Adoption and Use of Bevacizumab. *JAMA Netw Open.* 2020;3(1):e1918586.
7. González JM, Ogale S, Morlock R, et al. Patient and physician preferences for anticancer drugs for the treatment of metastatic colorectal cancer: a discrete-choice experiment. *Cancer management and research.* 2017;9:149-158.
8. Theodorou M TV, Pavlakis A, Maniadakis N, Fragoulakis V, Pavi E, Kyriopoulos J. Factors influencing prescribing behaviour of physicians in Greece and Cyprus: results from a questionnaire based survey. *BMC Health Services Research.* 2009;9:150.
9. Gogineni K, Shuman KL, Chinn D, Gabler NB, Emanuel EJ. Patient Demands and Requests for Cancer Tests and Treatments. *JAMA Oncology.* 2015;1(1).
10. Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin.* 2020.
11. Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975–2017, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/statfacts/html/colorect.html, based on November 2019 SEER data submission. 2020.
12. Modest DP, Pant S, Sartore-Bianchi A. Treatment sequencing in metastatic colorectal cancer. *Eur J Cancer.* 2019;109:70-83.
13. Yazdi MH, Faramarzi MA, Nikfar S, Abdollahi M. A Comprehensive Review of Clinical Trials on EGFR Inhibitors Such as Cetuximab and Panitumumab as Monotherapy and in Combination for Treatment of Metastatic Colorectal Cancer. *Avicenna J Med Biotechnol.* 2015;7(4):134-144.
14. Van Cutsem E, Peeters M, Siena S, et al. Open-Label Phase III Trial of Panitumumab Plus Best Supportive Care Compared With Best Supportive Care Alone in Patients With Chemotherapy-Refractory Metastatic Colorectal Cancer. *Journal of Clinical Oncology.* 2007;25(13):1658-1664.
15. Peeters M, Price TJ, Cervantes A, et al. Randomized phase III study of panitumumab with fluorouracil, leucovorin, and irinotecan (FOLFIRI) compared with FOLFIRI alone as second-line treatment in patients with metastatic colorectal cancer. *J Clin Oncol.* 2010;28(31):4706-4713.

16. Peeters M, Price TJ, Cervantes A, et al. Final results from a randomized phase 3 study of FOLFIRI {+/-} panitumumab for second-line treatment of metastatic colorectal cancer. *Ann Oncol.* 2014;25(1):107-116.
17. CEO Roundtable on Cancer's Life Sciences Consortium (LSC), Project Data Sphere platform, https://projectdatasphere.org/projectdatasphere/html/home. 2018.
18. National Cancer Institute, The SEER-Medicare Linked Database, 2007-2015.
19. World Health Organization, International classification of diseases for oncology (ICD-O) – 3rd edition, 1st revision. In: 3rd ed. ed. Geneva2013: https://apps.who.int/iris/handle/10665/96612. Accessed 2013.
20. U.S. Food and Drug Administration, Hematology/Oncology (Cancer) Approvals & Safety Notifications, https://www.fda.gov/drugs/resources-information-approved-drugs/hematologyoncology-cancer-approvals-safety-notifications. 2020.
21. Parikh RC, Du XL, Morgan RO, Lairson DR. Patterns of Treatment Sequences in Chemotherapy and Targeted Biologics for Metastatic Colorectal Cancer: Findings from a Large Community-Based Cohort of Elderly Patients. *Drugs Real World Outcomes.* 2016;3(1):69-82.
22. McLean J, Rho YS, Kuruba G, et al. Clinical Practice Patterns in Chemotherapeutic Treatment Regimens for Metastatic Colorectal Cancer. *Clin Colorectal Cancer.* 2016;15(2):135-140.
23. Abrams TA, Meyer G, Schrag D, Meyerhardt JA, Moloney J, Fuchs CS. Chemotherapy usage patterns in a US-wide cohort of patients with metastatic colorectal cancer. *J Natl Cancer Inst.* 2014;106(2):djt371.
24. Seal BS, Sullivan SD, Ramsey SD, et al. Systemic therapy for colorectal cancer: patterns of chemotherapy and biologic therapy use in nationally representative US claims database. *BioDrugs.* 2014;28(2):229-236.
25. Sonbol MB, Mountjoy LJ, Firwana B, et al. The Role of Maintenance Strategies in Metastatic Colorectal Cancer: A Systematic Review and Network Meta-analysis of Randomized Clinical Trials. *JAMA Oncol.* 2019:e194489.
26. Hess GP, Wang PF, Quach D, Barber B, Zhao Z. Systemic Therapy for Metastatic Colorectal Cancer: Patterns of Chemotherapy and Biologic Therapy Use in US Medical Oncology Practice. *Journal of oncology practice.* 2010;6(6):301-307.
27. Wilson PM, LaBonte MJ, Lenz H-J. Molecular Markers in the Treatment of Metastatic Colorectal Cancer. *The Cancer Journal.* 2010;16(3).
28. Dinu D, Dobre M, Panaitescu E, et al. Prognostic significance of KRAS gene mutations in colorectal cancer--preliminary study. *J Med Life.* 2014;7(4):581-587.
29. Ikoma N, Raghav K, Chang G. An Update on Randomized Clinical Trials in Metastatic Colorectal Carcinoma. *Surg Oncol Clin N Am.* 2017;26(4):667-687.
30. Cunningham D, Humblet Y, Siena S, et al. Cetuximab Monotherapy and Cetuximab plus Irinotecan in Irinotecan-Refractory Metastatic Colorectal Cancer. *New England Journal of Medicine.* 2004;351(4):337-345.
31. Guglielmi AP, Sobrero AF. Second-line therapy for advanced colorectal cancer. *Gastrointest Cancer Res.* 2007;1(2):57-63.
32. Hess LM, Cui ZL, Mytelka DS, Han Y, Goodloe R, Schelman W. Treatment patterns and survival outcomes for patients receiving second-line treatment for metastatic colorectal cancer in the USA. *International Journal of Colorectal Disease.* 2019;34(4):581-588.

Chapter 3: Develop a Novel Iterative Algorithm to Simulate Individual-Level Trial Data Using Published Summary Data of Trials and Real-World Data

3.1 Abstract

Introduction

Real-world data (RWD) could be used to assess the generalizability of trials and possibly generate an external control group for a single arm study along with the propensity score-based approaches, which require individual-level trial data. When only summary data from trials are available, clinical trial samples can be simulated by using the observed correlations among baseline patient characteristics in the RWD. However, correlations found in RWD may not apply to trial data, and a different correlation structure among covariates would potentially bias the outcomes, especially if the outcomes data generating process includes effect modifiers or non-linear models. Therefore, using trial summary data and RWD, we aimed to develop an algorithm to approximate the true correlations of participants' characteristics in a trial and generate trial data accordingly.

Methods

We developed an iterative algorithm using copula and resampling, which are based on the estimated propensity score for likelihood of enrollment in a trial given participants' characteristics. Validation was performed using Monte-Carlo simulations under different scenarios where the marginal and joint distributions of covariates differ between trial samples and RWD. We also illustrated this method with two applications using an actual colorectal

cancer trial and the SEER-Medicare linked database. First, we applied existing methods (i.e. standardized mean difference, B-index) to assess the generalizability of the trial. Then, we explored the feasibility of using RWD to generate an external control group by applying a parametric Weibull model trained in RWD to predict overall survival in the simulated trial cohort. The real-world target population were selected based on the eligibility trial criteria if applicable.

Results

Across all of the simulation scenarios, this iterative algorithm could successfully approximate correlations of covariates in a trial, which were closer to the true correlations than the correlations in RWD. The algorithm could also successfully reproduce the joint distribution of baseline characteristics for an actual cancer trial using summary data from the trial and the real-world cohort from the SEER-Medicare linked database. Therefore, we can obtain the similar estimates of standardized mean difference and B-index using simulated trial data (1.25 and 0.77 respectively) to assess the generalizability of the trial, comparing to the value of using the individual-level trial data (1.33 and 0.78 respectively). Additionally, no difference in approximated correlations was observed when we applied different eligibility trial criteria to select the target population. Lastly, we found that the adjusted survival estimates among the simulated trial population were close to the actual trial Kaplan-Meier (K-M) estimates, where the 95% confidence intervals of these survival curves were overlapped.

Conclusions

Our algorithm could be a feasible way to simulate individual-level clinical trial data when only summary data from the trials are available, which could be used to assess the generalizability of clinical trials given participants' characteristics and inform decision making around the applicability of trial results to a real-world population. Yet, the correlations of patient characteristics were not too different between the trial participants and the target population in our empirical example. Furthermore, we only explored and illustrated the ideas of using RWD to predict outcomes in the simulated trial data since an outcome prediction model must be developed in RWD with sufficiently large sample sizes. Further research is needed to apply this algorithm to other empirical examples which reflect a significant difference in correlations among covariates between the trial cohort and the real-world cohort and perform model selection to build an outcome prediction model when larger data are available.

3.2 Introduction

Existing methods was developed to assess the generalizability of a trial and project treatment effect from a trial to the general population using the propensity score-based approaches.[1-3] However, it is challenging to apply these methods when either individual-level trial data or real-world data (RWD) are not available. While only trial summary data are available, matching-adjusted indirect comparisons (MAIC) can be used to adjust treatment effect applying propensity score-based approaches with the generalized method of moments (GMM) estimates.[4,5] Yet, these approaches assume that the correlations among covariates are the same among two data sources (e.g. trials vs. real-world data, among two trials etc.) or an outcome model does not depend on the correlations among covariates, which may not be hold. However, correlations among covariates could potentially bias the expected value and variances of predicted outcomes, especially for a non-linear regression model.[6-8] Even for a linear model, covariances should be considered to obtain unbiased expected value of the outcomes when effect modifiers exist in the outcome models. To explain the importance of correlation structure among covariates, we provided a simple numeric example in the appendix to demonstrate the impact of correction structure in covariates on the predicted outcomes given linear and or non-linear outcome models (Appendix 1). The example shows that correlations of covariates must be considered in the data generating process to ensure the accuracy of predictions from the outcome models.

However, challenge arises when correlations among covariates are unknown owing to a lack of individual-level data, particularly for clinical trial data. Therefore, we developed an iterative algorithm to approximate the correlations of covariates in a trial using published trial summary data and individual-level real-world data. Using approximated correlations, we expected to

reproduce trial population and explored two potential applications. Instead of adjusting treatment effects, we aimed to combine the simulated individual-level trial data with real-world data to assess the generalizability of clinical trials and to assess the feasibility of generating external control groups which was proposed in the Framework for FDA's Real-World Evidence Program.[9]

In the following sections, we first demonstrated a conceptual framework of the algorithm using copulas and propensity score-based approaches. Then, we illustrated and validated the algorithm through Monte-Carlo simulations, including designs of data generating process. Finally, we applied this algorithm to an actual colorectal cancer trial and real-world data, where we had individual-level trial data for validation.

3.3 Conceptual framework

We developed an iterative algorithm to approximate a true correlation structure in a trial using copulas and propensity score approaches. Copulas allow us to determine a joint distribution of multiple variables when only marginal distributions of each variable are available (Eq 1), which have been used widely in risk management in finance.[10]

$$F(x_1, \ldots\ldots, x_m) = C\big(F_1(x_1), \ldots\ldots, F_m(x_m)\big) \quad (1)$$

where $F(x_1, \ldots\ldots, x_m)$ represent a joint distribution function with m margins $F_1, \ldots F_m$.

In this study, we applied a Gaussian copula (Eq 2) in the algorithm that was flexible and robust using only mean and standard deviation to form the marginal distributions.[11]

$$C(u) = \Phi\big(\Phi^{-1}(u_1), \ldots\ldots, \Phi^{-1}(u_m)\big) \quad (2)$$

Let S denote a trial sample of patients from a population denoted by D. Based on the property of copulas, we can form the joint distribution of a trial sample (S) and population (D) respectively as the following:

$$h_j(\boldsymbol{X}) = C_j(\boldsymbol{U}) \cdot \prod_{l=1}^{k} f(X_l), \quad j = S, D \qquad (3)$$

where $f()$ denote the marginal density of each covariate, $h()$ denote their joint density function, and $C_j(\boldsymbol{U})$ is a copula of uniform distributions that captures the dependence structure of patient characteristics $\boldsymbol{X}$=(X_1, X_2…X_k). Note that both the copula and the density functions are indexed by S and D. From Eq (3), we can also write:

$$Pr(\boldsymbol{X}|S) = C_s(\boldsymbol{U}) \cdot \prod_{l=1}^{k} Pr\,(X_l|S), \text{ and}$$

$$Pr(\boldsymbol{X}|D) = C_D(\boldsymbol{U}) \cdot \prod_{l=1}^{k} Pr(X_l|D) \qquad (4)$$

Since a trial sample (S) is not a random sample from the population (D), we defined that a trial sample was selected from the target population with non-random sampling weight $\theta(\boldsymbol{X})$. Therefore,

$$Pr(\boldsymbol{X}|S) = \theta(\boldsymbol{X}) \cdot Pr(\boldsymbol{X}|D)$$

Using Bayes' theorem, we derived $\theta(\boldsymbol{X})$ as the odds of sampling weights

$$\theta(\boldsymbol{X}) = \frac{Pr(\boldsymbol{X}|S)}{Pr(\boldsymbol{X}|D)} = \frac{Pr(S|\boldsymbol{X}) \cdot Pr(D)}{Pr(D|\boldsymbol{X}) \cdot Pr(S)} = \frac{Pr(S|\boldsymbol{X})}{1 - Pr(S|\boldsymbol{X})} \cdot \frac{Pr(D)}{Pr(S)} \qquad (5)$$

where $0 < Pr(S|\boldsymbol{X}) < 1$.

Using Eq (4) and Eq (5), the propensity score-based weights can be shown as

$$\theta(\boldsymbol{X}) = \frac{Pr(\boldsymbol{X}|S)}{Pr(\boldsymbol{X}|D)} = \frac{C_s(\boldsymbol{U}) \cdot \prod_{l=1}^{k} \Pr\,(X_l|S)}{C_D(\boldsymbol{U}) \cdot \prod_{l=1}^{k} \Pr(X_l|D)}$$

Let $\theta^*(\mathbf{X})$ represent the true weighting function, which when applied to the population data can reproduce an observed sample. If $\theta^*(\mathbf{X})$ exists, then by definition, the dependence structure

observed in the weighted population data should reflect the dependence structure in the sample. That is:

$$C_s(\boldsymbol{U}) \cdot \prod_{l=1}^{k} Pr(X_l|S) = C_{S^*}(\boldsymbol{U}) \cdot \prod_{l=1}^{k} Pr(X_l|\theta^*(\boldsymbol{X}), D) \quad (6)$$

However, since the estimation of $\theta^*(\mathbf{X})$ would require inidviudal-level data from both the sample and the population, we followed an iterative approach to approximate the true $\mathrm{C_S}(\mathbf{U})$.

The iterative algorithm includes several steps (Figure 3.1). First, we used a copula with the correlations of covariates from real-world data and published trial summary data to simulate an initial trial sample in the algorithm. Second, we applied the propensity score methods to the combined data (the initial trial sample and real-world data) to estimate the probability of being enrolled in a clinical trial given individuals' characteristics. Since the trial sample was not randomly selected from the target population in the real world, therefore, we resampled real-world data using the odds of sampling weights that was a function of the propensity score. Individuals were supposed to have more weighting when they had similar characteristics as those who were enrolled in the clinical trial. Note that the range of propensity score was set between 0.05 and 0.95 to avoid extreme weighting.

While repeating above process, the correlations of individuals' characteristics in the resampled real-world data changed over iterations to more closely resemble the correlations of participants' characteristics in a trial. The iterations were stopped when the coefficients remain stable, which was defined as maximum absolute changes in coefficients of the propensity score model to be less than a fraction of standard deviation of the coefficients. We set $c = 1$ and used one standard deviation as a stop criterion (i.e. SD*1).

$$Max\{|(\boldsymbol{\beta}^{i+1} - \boldsymbol{\beta}^{i}) - SD^{i} * c|\} \leq 0 \quad (7)$$

where $\boldsymbol{\beta}^{i}$ was the vector of coefficients of the logistic regression in the i-th iteration. This indicated that resampled real-world data were drawn based on similar odds of sampling weights given small changes in the coefficients among iterations, and implied that the correlations of covariates in the resampled real-world data should reflect the correlation structure in a trial. To demonstrate and validate this algorithm, we first presented Monte-Carlo simulations and then applied this algorithm to an actual trial and real-world data in this study.

3.4 Monte-Carlo Simulation Designs

Data Generating Process

We designed a Monte Carlo simulation along with the assumption that a trial sample was a part of the real-world population who had non-zero probability to be enrolled in a trial. We illustrated this iterative algorithm using two continuous variables (age and logarithm of income) that were associated with the enrollment of clinical trials. Both age and logarithm of income were assumed to follow a normal distribution. We first simulated real-world data (N=50,000) using a multivariate normal distribution of age (X_1) and logarithm of income (X_2) with a correlation (ρ). We let

$$X_1 \sim N(45, 5^2)$$

$$X_2 \sim N(11, 1^1)$$

and $\rho_{X_1X_2} = 0.04$,

which can be estimated from Medical Expenditure Panel Survey (MEPS) data.

To generate a trial sample, we designed a propensity score model which was modeling the probability of being in a trial given individuals' characteristics and applied to the simulated real-world data. The individuals in the real world must have a non-zero probability to be enrolled in a trial ($0 < p < 1$), and the probability of being in a trial was changed among target population given individuals' characteristics. Then, a trial sample (n=2,000) was drawn from the simulated real-world data using the odds as weighting. Different propensity score models were used in each scenario.

A general form of the propensity score models can be shown as the following.

$$logit(p) = -2.5 + ln(0.8) \cdot Age + \ln(1.1) \cdot log\ (income) + \beta_3 \cdot Age \cdot log\ (income)$$

We expected to see the difference in correlation of age and log income between a trial and the real-world data, which was associated with the effect of the interaction terms in a model. Therefore, we fixed the intercept and coefficients of age and log income but changed the coefficient of the interatom term (β_3) from ln(0.9) to ln(1.1) to show different magnitude of the correlations between covariates in different scenarios. About 11 % of the population would be enrolled in a trial given these models among all of the 21 scenarios. The coefficients of age and log income were arbitrary, however, the rationale behind the choice of coefficients was based on the association between these variables and the enrollment of trials. That is, the elderly and low income population were less likely to be enrolled in a trial than younger population and higher income population.[12,13] We also centered covariates to the mean in order to interpret the intercept, which meant the probability of being in a trial is 0.08 for the individuals who are 45 years old and have income around $59,874.

Sampling Process

In each scenario, we repeated sampling process m times to draw m trial samples from the real-world data using the same propensity score model to capture the uncertainty of sampling. Then, we ran the algorithm k times using the marginal distribution of covariates from each trial sample. That is, we obtained k approximated correlations and took average over them for each sample. Finally, we took average again over m average approximated correlations from all trial samples. The overview of the validation process is shown in Figure 3.2.

3.5 Simulation Results

In our simulations, we drew 50 trial samples ($m = 50$) from the simulated real-world data and ran algorithm 50 times ($k = 50$) using one standard deviation as a stop criterion in each scenario. The details of each scenarios are listed in the Table 3.1, including the distribution of true correlations and covariates (age and log income) over 50 trial samples, the number of iterations and the approximated correlation in each scenario. Based on our data generating process, the age distribution in each trial sample was slightly younger than those in the simulated real-world data, and the distribution of log income varied depending on the correlation between age and log income. The average iterations were about seven to nine times to get an approximated correlation from the algorithm each time, and run time was about three to four hours to repeat the sampling process (m=50, k=50) across all scenarios. All simulations were executed on a laptop with a 2.7 GHz processor and 8GB RAM at 1333MHz.

To evaluate the performance of the algorithm, we examined the distance between the true correlation and the approximated correlation derived from the algorithm in each scenario, shown

in Figure 3.3A. For comparison, we also demonstrated the difference in correlations between a true trial and simulated real-world data in Figure 3.3A. While using a propensity score model without an interaction term (ln β_3=1), same correlations of covariates were observed in true trial samples and simulated real-world data ($\rho = 0.04$). From other scenarios, we observed that a trial sample would have either stronger positive or negative correlation between age and log income than that in the simulated real-world data when the effect of interaction term was far from zero. Most importantly, we found that the iterative algorithm could successfully change the dependence structure of covariates from the simulated real-world data and move towards the true correlation in a trial - the approximated correlation was close to the correlation in a true trial across all of the scenarios. Additionally, we tested the algorithm using a different stop criterion in nine scenarios. Instead of using one standard deviation as the stop point, we let the algorithm to be stopped when maximum absolute changes in coefficients was less than a half standard deviation (i.e. SD*0.5).

We observed that approximated correlations were similar regardless of using one or half standard deviation as the stop point (Figure 3.3B). Not surprisingly, more iterations and longer run time were needed for the algorithm to converge when we used more restricted stop criterion. The average iterations were about 30 times (range from 4 to 255) to reach the converge point each time, and run time were about 14 hours to complete the repeated sampling process. Our findings implied that using one standard deviation as the stop criterion would be more feasible to use this algorithm in practice, where more coefficients would be included in a model.

3.6 An empirical example

Data Sources

Individual-Level Cancer Clinical Trial

To validate the algorithm, we also applied the algorithm to an actual cancer trial that were conducted by Amgen Inc. in 2006. This phase 3 randomized multicenter clinical trial aimed to investigate the efficacy of panitumumab (Vectibix™) in combination with FOLFIRI (Irinotecan plus leucovorin and 5-fluorouracil (5-FU)) as second line treatment for metastatic colorectal cancer (NCT00339183, Amgen protocol number: 20050181). We requested and had access to the individual-level trial data including the protocol through the Project Data Sphere Cancer Research Platform, which is operated by CEO Roundtable on Cancer's Life Sciences Consortium (LSC).[14] Therefore, we were able to obtain the correlations of selected variables using the individual-level trial data and used bootstrap to estimate the 95% confidence intervals of the correlations in the trial for validation purpose. We also restricted population who aged 65 years and older because we used the SEER-Medicare linked database as the real-world data.

Total 276 patients who aged 65 years and older were selected from the AMGEN trial, and the summary of patient characteristics among this population are shown in Table 3.2. We selected patient characteristics from the published trial summary table, including age, gender, cancer subtype (primary tumor site located in colon or rectum), and prior treatments (oxaliplatin/bevacizumab). Unfortunately, we were not able to include Eastern Cooperative Oncology Group (ECOG) performance status and sites of metastatic disease because of data availability in the SEER-Medicare linked database. Race were also excluded because most trial participants were white (99%). We assumed that patients' age follows a normal distribution and

let binary variables follow a latent normal distribution while using the gaussian copula and convert them to binary variables afterwards.

Real-World Data

The target population were defined as treated metastatic colorectal cancer patients who would meet the eligibility criteria of the Amgen trial. Metastatic colorectal cancer patients were selected from the SEER-Medicare linked database based on their initial diagnosis from 2008 to 2015. To select our target population, we used International Classification of Diseases for Oncology, third revision (ICD-O-3) site codes (C180-C189, C199, C209) and histology codes (8140, 8480, 8490, 8210, 8220, 8261) to identify colorectal cancer patients with adenocarcinoma, and then we used American Joint Committee on Cancer (AJCC) staging system to select metastatic colorectal cancer patients. All patients with other primary cancers were excluded in this study.

To determine whether patients received the second-line treatment, we first identified common Intravenous (IV) or oral chemotherapy drugs that were approved for metastatic colorectal cancer before 2015 using Healthcare Common Procedure Coding System (HCPCS) codes or National Drug Code (NDC). The common treatments Include 5-FU, leucovorin, oxaliplatin, irinotecan, capecitabine, bevacizumab, cetuximab (Erbitux), panitumumab (Vectibix), ziv-aflibercept (Zaltrap), regorafenib (Stivarga), trifluridine-tipiracil (LONSURF), ramucirumab (Cyramza), which could be used as a single-agent or use in combination, e.g. FOLFOX (oxaliplatin plus leucovorin and 5-FU), FOLFIRI (irinotecan plus leucovorin and 5-FU), and CAPOX (oxaliplatin plus capecitabine) etc.

We examined the treatment patterns after the initial diagnosis to determine the lines of therapy. The first-line treatment was defined as any drugs that were used during the first 36 days after the initial treatment and used at least one or more cycles. Patients were considered to receive the second-line treatment when they switched to another treatment after their first-line treatment. Extra agents for maintenance therapy (e.g. adding capecitabine/bevacizumab to the previous chemotherapy regimens) or removing one agent from the combination of chemotherapy were not considered as a new line. If the time period between two treatments was longer than 90 days, then the treatment was considered as the next line regardless of the reasons. These criteria were consistent with the definition of treatment patterns in metastatic colorectal cancer from the previous studies.[15-20]

After determining the lines of therapy, we followed the clinical trial protocol to select patients who received one and only one previous fluoropyrimidine-based chemotherapy (e.g. capecitabine and 5-FU) and were not treated with irinotecan, cetuximab, or erlotinib. Then, using the start date of the second-line treatment as an index date, we screened patients' history of diagnosis and procedures using International Classification of Diseases, Ninth/tenth Revision, Clinical Modification (ICD-9-CM/ICD-10-CM), ICD-9-CM procedure codes, ICD-10 Procedure Coding System (ICD-10-PCS), HCPCS codes, and Current Procedural Terminology (CPT) codes. Applying the clinical trial's exclusion criteria, we excluded patients who had any history of central nervous system (CNS) metastases, pneumonitis or pulmonary fibrosis, inflammatory bowel disease, human immunodeficiency virus (HIV) infection, or hepatitis B/C infection etc. We also excluded patients if they had significant cardiovascular diseases within one year, surgical procedures with anesthesia within 28 days, radiotherapy or any active infection within

14 days before the index date. Finally, we obtained the correlation matrix of selected patient characteristics which were listed in the Table 3.2. This study was reviewed and approved by the Institutional Review Board (IRB) in the University of Washington.

Setting

A total of 3,879 metastatic colorectal cancer patients were found to use any second-line treatment from 2008 to 2015. Table 3.3 shows the sample sizes of population after applying different criteria. In order to have a larger sample size to be re-sampled in the iterative algorithm, our final target population (n=990) was defined as metastatic colorectal cancer patients who met most criteria but were allowed to have chemotherapy within 30 days before the index date. We also performed a sensitivity analysis to examine the impact of using different target population in the algorithm. Using these target population with the summary data from the Amgen trial to initial the iterative algorithm, we let a simulated trial sample size to be 500 in each iteration and repeated the algorithm 100 times to obtain the average approximated correlations.

Assess the generalizability

Two existing metrics were considered to assess the generalizability, including standardized mean difference and B-index.[1,21] The standardized mean difference was defined as the difference in average propensity scores between a trial sample and the population divided by pooled standard deviation of propensity scores. Although no specific criterion was set to dichotomize whether a trial was generalizable or not, a smaller value of standardized mean difference was preferred. On the other hand, the B-index ranged from zero to one was developed by comparing the density of propensity score quantitatively, which higher value indicated better generalizability.

External control group

To illustrate the potential use of RWD to generate an external control group, we first simulated trial data using the approximated correlations of patient characteristics for those who used FOLFIRI in the trial from the algorithm. Then, we used a parametric Weibull model to fit the overall survival among patients who used FOLFIRI alone or maintenance therapy with FOLFIRI as the second-line treatment in the real-world. To generate a comparable cohort, patients who received any third-line treatments were excluded from our real-world population. Finally, we applied the outcome model to predict survival among the simulated trial population given their characteristics.

3.7 Empirical Results

Comparison of the simulated trial data and the actual trial data

To evaluate the performance of the algorithm, we compared the approximated correlation to the true correlation in the AMGEN trial for each pair of covariates, and also compared the join distribution of simulated data using approximated correlations and true correlations. We found that 9 of 10 approximated correlations were close to the true correlations and fell into the 95% confidence intervals (95% CI) except the correlation between using oxaliplatin and using bevacizumab in the prior therapy (Table 3.4.). Nevertheless, the joint distributions of selected patient characteristics were very similar when we simulated data using approximated correlations or true correlations, which implied the iterative algorithm should be able to approximate correlations in a true trial and generate similar trial population given selected variables (Figure 3.4A). While comparing the joint distributions using an existing metric, high similarity (B-

index=0.99) was found between the simulated data using approximated correlations and the individual-level trial data. From the sensitive analysis, we noticed that those approximated correlations were consistent among different scenarios in terms of applying the most or the less restrictive criteria to the target population.

Comparison of the simulated trial data and the real-world data

Additionally, we estimated the propensity scores of being in a trial using simulated data and individual-level real-world data. Similar distributions of propensity score were found regardless of using either approximated correlations or true trial correlations to simulate data (Figure 3.4B). While assessing the generalizability using existing metrics, we obtained 1.25 and 0.77 of the standardized mean difference and B-index respectively. Comparing to the gold standard which were calculated using individual-level trial data and the real-world data, we found that the estimates were very similar to the value of the standardized mean difference and B-index, which were 1.33 and 0.78 respectively.

Comparison of the actual trial and the real-world data

We also examined whether a dependence structure of covariates differed between actual trial participants and real-world target population. Among ten correlations of paired covariates from the real-world target population, we found that eight of them fell outside the 95% CI of the true correlations in a trial. Yet, some of these correlations were very close to the bound of the 95% CI. Only four of them would fall outside the 99% CI of the true correlations, which indicating that the dependence structure of covariates among the target population may not be too different from that in the true trial. Therefore, we observed similar join distribution of patient

characteristics in the simulated data regardless of using the approximated correlations derived from the algorithm or the correlations from the real-world target population with the same marginal distributions of covariates from the true trial.

Predicted survival using the simulated data

Of 990 metastatic colorectal cancer patients who met trial eligibility criteria, only 172 patients received FOLFIRI alone or maintenance therapy with FOLFIRI as the second-line therapy. Therefore, we can only define an arbitrary outcome model with interaction terms in this example because of a small sample size. The model included age, age squared, and three interaction terms (i.e. male and oxaliplatin, oxaliplatin and tumor type, age and tumor type). After simulating 100 trial samples, we applied the Weibull outcome model fitted using RWD and predicted the survival estimates given patient characteristics for each trial sample. Then, we obtained an average adjusted survival curves among simulated trial samples. We found that the adjusted survival estimates among the simulated trial population using either approximated correlations or true trial correlations were close to the actual RCT Kaplan-Meier (K-M) estimates. The overlapping 95% CI between these survival curves are shown in the Figure 3.5.

3.8 Discussion

We developed an iterative algorithm to approximate correlations of covariates in a trial using trial summary data and individual-level real-world data. Through simulations, we found that effect modifiers in propensity score models changed the correlations of covariates in a trial, where stronger effect leads to a bigger difference in correlations between a trial and real-world data. Using the algorithm, we could approximate correlations of covariates in a trial which

reduced the difference in correlations between true trials and real-world data. Furthermore, we demonstrated that the algorithm could successfully reproduce the joint distribution of patient characteristics in an actual caner trial.

The iterative algorithm could be used in different ways to support existing approaches when there is lack of individual-level trial data. First, the algorithm could be a potential tool to fill the gap in existing methods to assess the generalizability of a trial. The existing approaches aimed to compare the distribution of individuals' characteristics in a trial to those in the real world, where the join distribution of covariates would be changed by the correlations among covariates. Thus, we believed that using the approximated correlations of covariates from the algorithm to simulate individual-level trial data would be more appropriate than using the correlations from the real-world to do so. Our approach would allow researchers to evaluate the discrepancies between clinical trials and clinical practice in any disease area, which could help to inform better clinical or coverage decisions in different population. Second, the algorithm may help to examine the assumption of same correlations among different data sources. When a similar dependence structure was observed between a trial and real-world data, this finding could support the evidence derived from the methods under this assumption. Even when the assumption was not held, our approach may be applied to estimate propensity score-based weights rather than using generalized method of moments (GMM) for matching-adjusted indirect comparisons (MAIC). Additionally, while we trained an outcome model using RWD, using the algorithm could potentially provide an opportunity to obtain adjusted survival estimates among trial population given their characteristics.

Several limitations should be mentioned in this study. First, applying the algorithm to approximate correlations in an actual trial were constrained by the availability of common variables between the AMGEN trial and SEER-Medicare linked database. Ideally, we preferred to include all variables that were listed in the published trial summary table to build the join distribution of patient characteristics. However, some of them were not available in the SEER-Medicare linked database, including ECOG performance status, KRAS status and sites of metastatic disease. Thus, the joint distribution of patient characteristics was determined only using five available covariates in our empirical example, which may explain the similarity between the trial and the target population in this study. Note that the joint distribution of patient characteristics was only for the elderly who aged 65 years or older in this study. The impact of unobserved covariates would depend on the difference in these covariates between the trial population and the target population in the real world. The unobserved covariates might be an issue of using the propensity score approach for causal inference, however, our goal was only to compare the patient characteristics between two populations rather than making the causal inference in this study.

Second, our target population were not representative of metastatic colorectal cancer patients who used second-line treatment in the real world. Using the SEER-Medicare linked database, we only included patients who were initially diagnosed as metastatic colorectal cancer because disease progression after the initial diagnosis was unavailable in the cancer registry data. Therefore, the sample size of our target population was smaller than we expected. A larger sample size of the target population would potentially increase the variations in patient characteristics, and significant difference may be observed between the trial and the target

population. Likewise, owing to an insufficient sample size, we can only use an arbitrary outcome model to illustrate the ideas of using RWD to generate an external control group. Since the survival prediction model is crucial to obtain credible and robust predictions, the model should be trained and determined using the sufficiently large sample size. Therefore, further research is needed to validate our approach and evaluate the potential use of RWD for clinical trial design.

Additionally, we were not able to apply all trial eligibility criteria to identify target population in the SEER-Medicare linked database because of data availability. The unavailable information included patients' ECOG performance status, laboratory data and radiographically documented disease progression using RECIST (Response evaluation criteria in solid tumors) criteria. Nevertheless, it may be acceptable to select target population only using main inclusion and exclusion criteria based our findings in the sensitivity analysis. No difference in approximated correlations was observed when we applied different criteria to the target population.

Finally, we assumed linear correlations among covariables and only considered Gaussian copula in the algorithm. However, Pearson correlation is not applicable for multinomial variables. In our empirical example, most variables were binary which can be converted from a latent continuous distribution, and the only multinomial variable, sites of metastatic disease, was excluded because of data availability in the SEER-Medicare linked database. In addition, appropriate assumptions of marginal distributions were also required to form the joint distribution of covariates using copulas. The algorithm may not be able to successfully reproduce the joint distribution of patient characteristics among trial population if users applied inappropriate assumptions. Additional simulations could be considered to explore the impact of including polynomial terms in the

simulation model, and other types of copulas would be needed to model non-linear correlations of covariates.

3.9 Conclusions

In sum, the iterative algorithm may help to approximate correlations in a trial and be adopted to apply existing methods. Yet, we only validated the algorithm in one actual trial and limited to the elderly because of data availability. Data sharing of clinical trials are highly recommended and actual individual-level trial data are preferred to inform decisions than a model-based approach with assumptions. Further research is needed to apply this algorithm to other empirical examples which reflect a significant difference in correlations among covariates between the trial cohort and the real-world cohort, to perform model selection to build an outcome prediction model when larger data are available, to explore the potential use in microsimulations for economic evaluation, and to develop a test to identify whether the correlations are different between RWD and the trial before applying the algorithm.

Acknowledgments

- "This study is based on research using information obtained from www.projectdatasphere.org, which is maintained by Project Data Sphere. Neither Project Data Sphere nor the owner(s) of any information from the web site have contributed to, approved or are in any way responsible for the contents of this study."
- "This study used the linked SEER-Medicare database. The interpretation and reporting of these data are the sole responsibility of the authors. The authors acknowledge the efforts of the National Cancer Institute; the Office of Research, Development and Information, CMS; Information Management Services (IMS), Inc.; and the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER-Medicare database."
- "This work was facilitated though the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington"

3.10 Tables and Figures

Table 3.1. Description of each scenario (using one standard deviation as the stop criterion)

Scenarios	Effect of interaction term ($\log \beta_3$)	Average over m samples ($m = 50$)			Average over m samples and k times algorithm ($m = k = 50$)		
		True correlation mean (95% CI)	Age Mean (SD)	Log income Mean (SD)	Iterations mean [range]	Approximated correlation	Run time (hours)
S1	0.9	-0.48 (-0.52, -0.44)	37.62 (5.49)	11.78 (5.49)	8.64 [4, 46]	-0.32 (-0.34, -0.30)	4.06
S2	0.91	-0.44 (-0.48, -0.39)	38.07 (5.38)	11.67 (5.38)	8.64 [4, 36]	-0.29 (-0.31, -0.27)	4.02
S3	0.92	-0.38 (-0.42, -0.34)	38.43 (5.26)	11.56 (5.26)	8.78 [4, 32]	-0.27 (-0.28, -0.25)	4.02
S4	0.93	-0.32 (-0.37, -0.27)	38.72 (5.17)	11.48 (5.17)	8.61 [4, 30]	-0.24 (-0.26, -0.22)	4.09
S5	0.94	-0.27 (-0.32, -0.23)	38.95 (5.12)	11.40 (5.12)	8.57 [4, 50]	-0.20 (-0.23, -0.17)	4.03
S6	0.95	-0.22 (-0.27, -0.18)	39.08 (5.02)	11.33 (5.02)	8.42 [4, 33]	-0.17 (-0.19, -0.14)	3.74
S7	0.96	-0.17 (-0.21, -0.13)	39.26 (4.99)	11.27 (4.99)	8.16 [4, 38]	-0.13 (-0.16, -0.10)	3.73
S8	0.97	-0.12 (-0.16, -0.07)	39.38 (4.95)	11.21 (4.95)	8.17 [3, 36]	-0.10 (-0.12, -0.07)	3.85
S9	0.98	-0.06 (-0.11, -0.02)	39.43 (4.90)	11.15 (4.90)	7.93 [3, 39]	-0.06 (-0.10, -0.03)	3.57
S10	0.99	-0.01 (-0.07, 0.04)	39.49 (4.89)	11.10 (4.89)	7.74 [3, 43]	-0.03 (-0.06, 0.00)	3.53
S11	1	0.04 (0.00, 0.07)	39.52 (4.90)	11.04 (4.90)	7.22 [3, 41]	0 (-0.02, 0.03)	3.39
S12	1.01	0.09 (0.04, 0.13)	39.54 (4.89)	10.99 (4.89)	6.87 [3, 42]	0.03 (0.01, 0.05)	3.21
S13	1.02	0.13 (0.09, 0.17)	39.50 (4.92)	10.94 (4.92)	6.89 [3, 44]	0.06 (0.04, 0.08)	3.06
S14	1.03	0.17 (0.13, 0.22)	39.42 (4.93)	10.88 (4.93)	7.39 [3, 43]	0.09 (0.07, 0.12)	3.25
S15	1.04	0.22 (0.17, 0.27)	39.35 (4.97)	10.83 (4.97)	7.69 [3, 44]	0.12 (0.09, 0.15)	3.54
S16	1.05	0.27 (0.23, 0.30)	39.22 (5.01)	10.76 (5.01)	8.05 [3, 47]	0.15 (0.13, 0.18)	3.75
S17	1.06	0.31 (0.27, 0.34)	39.13 (5.07)	10.71 (5.07)	8.22 [3, 46]	0.18 (0.16, 0.21)	3.73
S18	1.07	0.34 (0.30, 0.38)	38.95 (5.10)	10.64 (5.10)	8.33 [3, 36]	0.21 (0.19, 0.24)	3.92
S19	1.08	0.39 (0.35, 0.42)	38.76 (5.13)	10.57 (5.13)	8.35 [4, 38]	0.25 (0.23, 0.26)	3.82
S20	1.09	0.42 (0.38, 0.45)	38.59 (5.20)	10.51 (5.2)	8.51 [4, 36]	0.27 (0.25, 0.29)	3.99
S21	1.1	0.45 (0.42, 0.49)	38.34 (5.23)	10.42 (5.23)	8.49 [4, 38]	0.30 (0.28, 0.31)	3.97

Table 3.2. Summary of Patient Characteristics

Variables		Mean (SD)	Distribution
Continuous	Age	70.74 (3.91)	Normal distribution
Binary	Male =1	0.7 (0.46)	Latent normal distribution & convert it to binary
	Primary tumor (colon=1, rectal=0)	0.67 (0.47)	
	Prior treatment - oxaliplatin =1	0.69 (0.46)	
	Prior treatment - bevacizumab =1	0.16 (0.37)	

Table 3.3. Sample Sizes of the Real-World Population Selected with Different Criteria

Population	N
Metastatic colorectal cancer patients who used any second-line treatments	3879
Age >=65	2815
Previous fluoropyrimidine-based chemotherapy & without prior anti-EGFr antibody therapy or irinotecan therapy	2433
No hospice care before the index date	2430
Met other criteria (e.g. medical history)	990
No chemotherapy within 30 days before the index date	546

Table 3.4. Comparing true correlations to approximated correlations using different RWD

Variables	True correlation in a trial (95%CI)	Approximated correlation Using RWD n=990	Sensitive analysis Approximated correlation using different RWD		
			RWD n=546	RWD n=2430	RWD n=2815
Age*Male	0.1 (-0.02, 0.21)	0.03 (-0.03, 0.10)	0.03 (-0.06, 0.13)	0.03 (-0.04, 0.09)	0.03 (-0.03, 0.09)
Age*Oxali	0.03 (-0.10, 0.15)	-0.01 (-0.08, 0.06)	-0.01 (-0.12, 0.09)	0.01 (-0.05, 0.07)	0.02 (-0.04, 0.08)
Age*Bevac	-0.01 (-0.11, 0.10)	-0.02 (-0.09, 0.04)	-0.01 (-0.11, 0.09)	-0.04 (-0.10, 0.02)	-0.04 (-0.10, 0.02)
Age*Type	0.05 (-0.06, 0.17)	0 (-0.08, 0.07)	0 (-0.11, 0.10)	-0.02 (-0.08, 0.04)	-0.02 (-0.07, 0.04)
Male*Oxali	-0.05 (-0.17, 0.07)	0 (-0.12, 0.11)	-0.01 (-0.16, 0.15)	-0.02 (-0.11, 0.08)	0 (-0.09, 0.08)
Male*Bevac	-0.09 (-0.21, 0.04)	-0.01 (-0.11, 0.08)	-0.02 (-0.17, 0.13)	-0.05 (-0.14, 0.04)	-0.06 (-0.15, 0.02)
Male*Type	-0.09 (-0.21, 0.02)	-0.02 (-0.14, 0.10)	-0.01 (-0.21, 0.18)	-0.02 (-0.13, 0.08)	-0.04 (-0.12, 0.05)
Oxali*Bevac	0.18 (0.09, 0.27)	0.01 (-0.07, 0.10)	0.01 (-0.13, 0.15)	0.05 (-0.04, 0.14)	0.03 (-0.05, 0.10)
Oxali*Type	0.02 (-0.10, 0.14)	0.05 (-0.06, 0.16)	0.03 (-0.16, 0.22)	0.03 (-0.08, 0.13)	0.02 (-0.08, 0.12)
Bevac*Type	-0.08 (-0.20, 0.05)	0.02 (-0.07, 0.11)	0.02 (-0.12, 0.17)	0.05 (-0.04, 0.13)	0.05 (-0.03, 0.13)

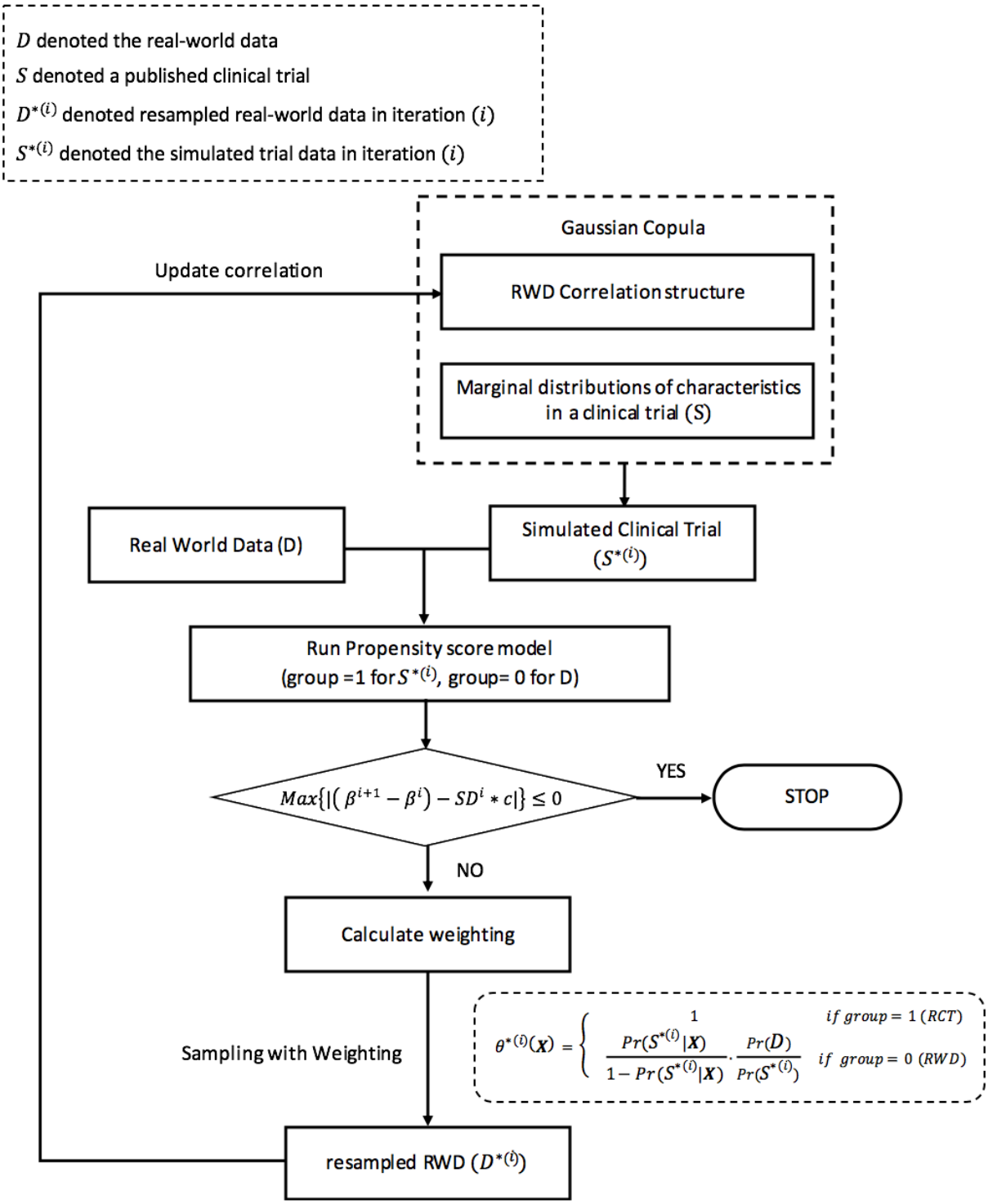


Figure 3.1. The iteration process

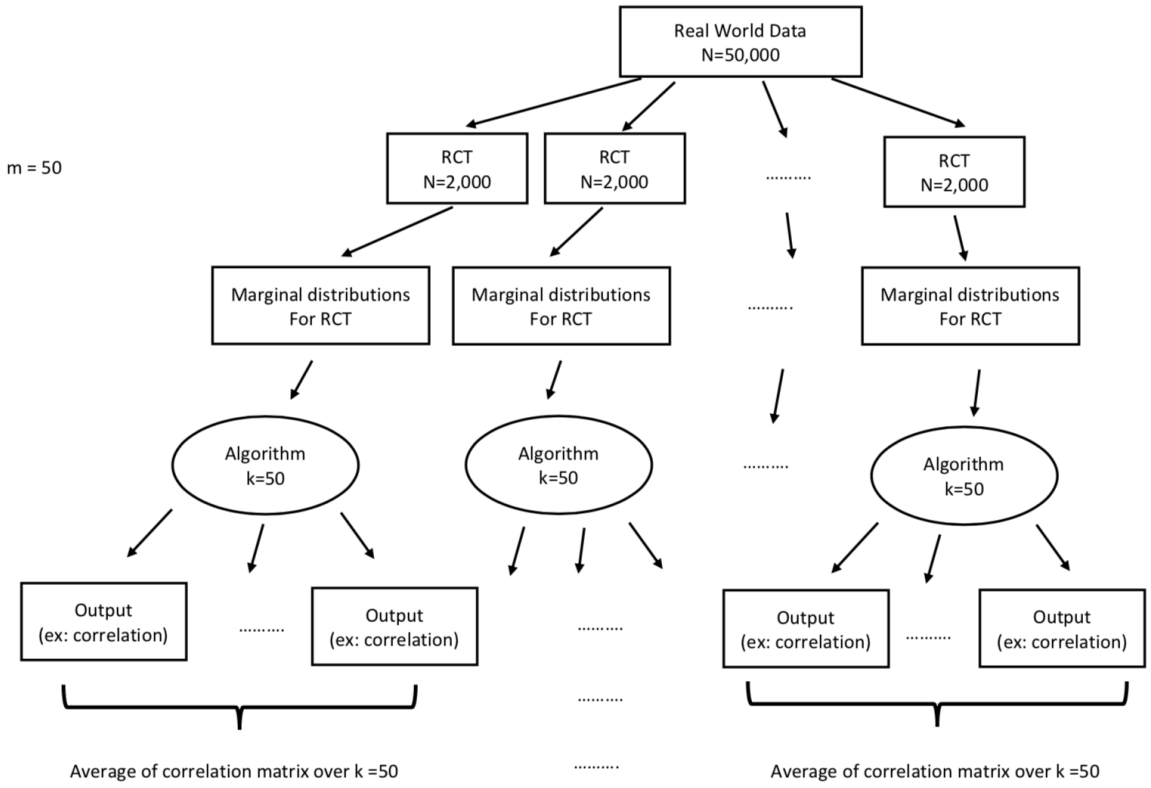
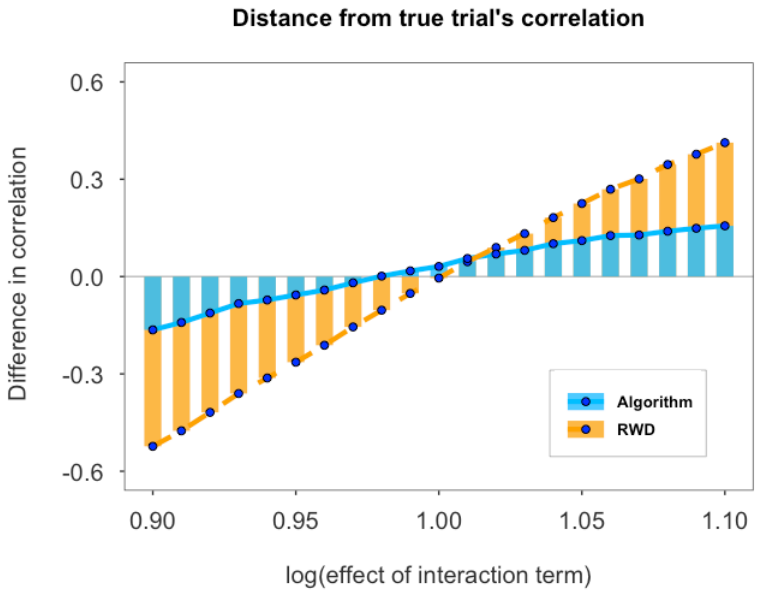


Figure 3.2. Overview of the Validation Process (An example for m =50, k=50)

(A)



(B)

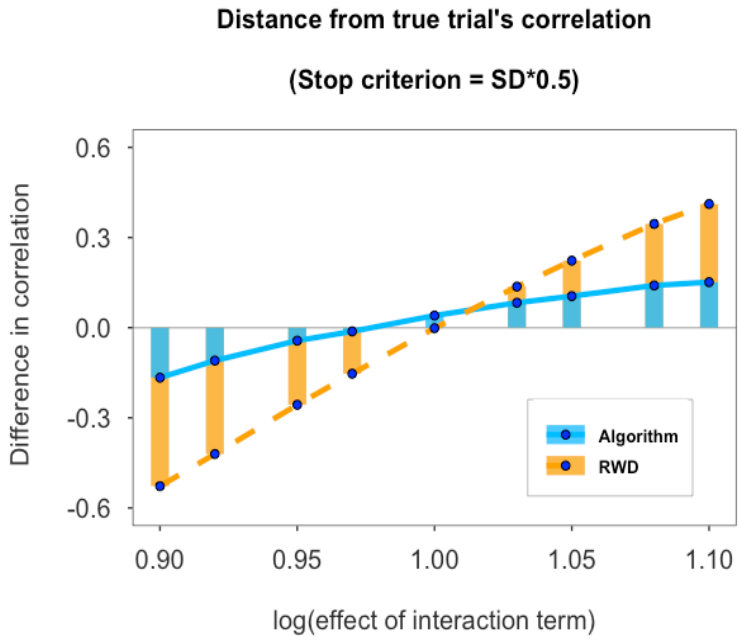
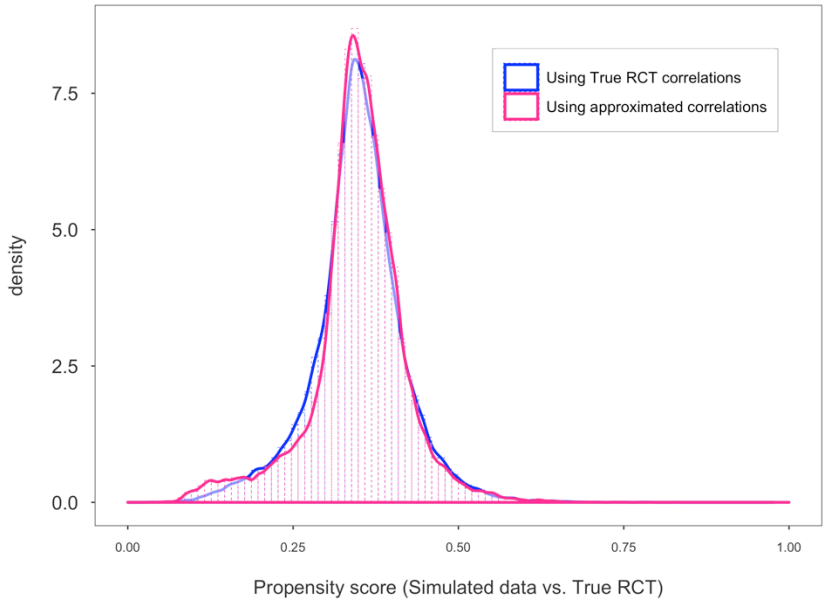


Figure 3.3. The performance of the algorithm with a different stop criterion (A) one standard deviation (B) half standard deviation

(x-axis: different scenarios; y-axis: difference in correlations between the true trial correlations and the approximated correlations derived from the algorithm (or the correlations in the RWD))

(A)



(B)

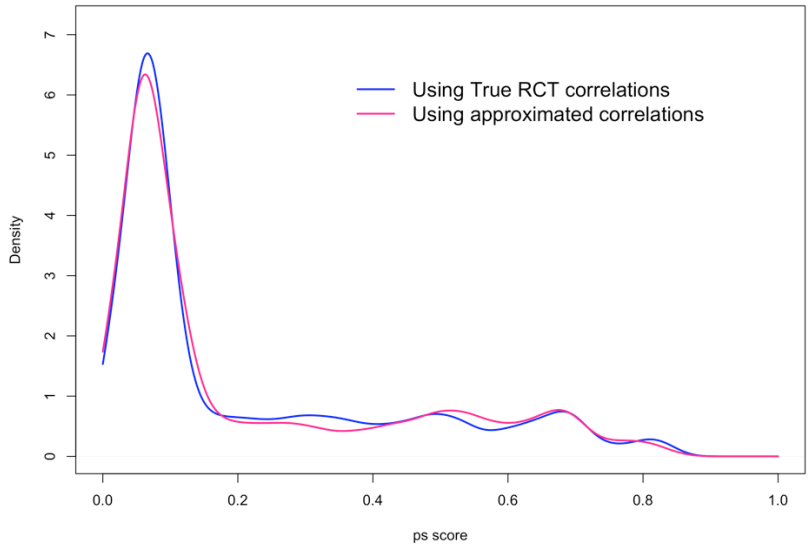


Figure 3.4. The PS distribution (A) simulated trial data vs. true trial (B) simulated data vs. RWD

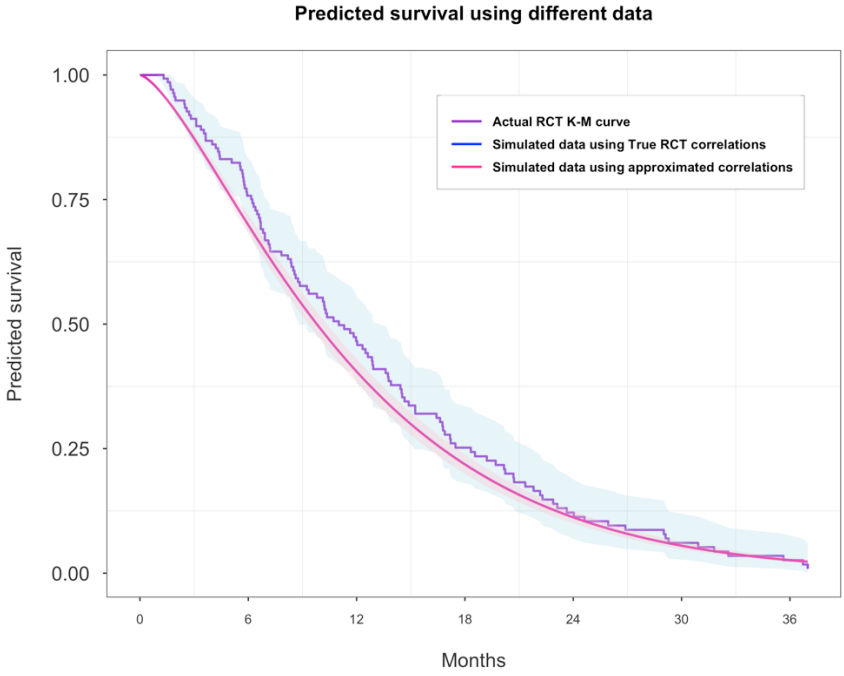


Figure 3.5. Comparison of the Actual RCT Kaplan-Meier Curve and the Adjusted Survival Curves using simulated data
(with either approximated correlations or true trial correlations of patient characteristics)

3.11 References

1. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society).* 2010;174(2):369-386.
2. Cole SR, Stuart EA. Generalizing Evidence From Randomized Clinical Trials to Target Populations: The ACTG 320 Trial. *American Journal of Epidemiology.* 2010;172(1):107-115.
3. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of Trial Results Using Inverse Odds of Sampling Weights. *Am J Epidemiol.* 2017;186(8):1010-1014.
4. Signorovitch JE, Sikirica V, Erder MH, et al. Matching-Adjusted Indirect Comparisons: A New Tool for Timely Comparative Effectiveness Research. *Value in Health.* 2012;15(6):940-947.
5. Signorovitch JE, Wu EQ, Yu AP, et al. Comparative Effectiveness Without Head-to-Head Trials. *PharmacoEconomics.* 2010;28(10):935-945.
6. Pingel R, Waernbaum I. Correlation and efficiency of propensity score-based estimators for average causal effects. *Communications in Statistics - Simulation and Computation.* 2017;46(5):3458-3478.
7. Pingel R, Waernbaum I. Effects of correlated covariates on the asymptotic efficiency of matching and inverse probability weighting estimators for causal inference. *Statistics.* 2015;49(4):795-814.
8. Basu A, Polsky D, Manning WG. Use of Propensity Scores in Non-Linear Response Models: The Case for Health Care Expenditures. *National Bureau of Economic Research Working Paper Series.* 2008;No. 14086.
9. U.S. Food and Drug Administration, Framework for FDA's Real-World Evidence Program, https://www.fda.gov/scienceresearch/specialtopics/realworldevidence/default.htm. 2018.
10. McNeil AJ, Frey R, Embrechts P. *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton University Press; 2015.
11. Rowan Iskandar FA-E, Thomas A. Trikalinos. Abstract 1B-3 On Simulating Multivariate Distributions with Arbitrary Marginal Distributions and Dependence Structures. *39th Annual Meeting of the Society for Medical Decision Making, Pittsburgh, Pennsylvania.* 2017.
12. Unger JM, Gralow JR, Albain KS, Ramsey SD, Hershman DL. Patient Income Level and Cancer Clinical Trial Participation: A Prospective Survey Study. *JAMA Oncology.* 2016;2(1):137-139.
13. Unger JM, Hershman DL, Albain KS, et al. Patient Income Level and Cancer Clinical Trial Participation. *Journal of Clinical Oncology.* 2013;31(5):536-542.
14. CEO Roundtable on Cancer's Life Sciences Consortium (LSC), Project Data Sphere platform, https://projectdatasphere.org/projectdatasphere/html/home. 2018.
15. Parikh RC, Du XL, Morgan RO, Lairson DR. Patterns of Treatment Sequences in Chemotherapy and Targeted Biologics for Metastatic Colorectal Cancer: Findings from a Large Community-Based Cohort of Elderly Patients. *Drugs Real World Outcomes.* 2016;3(1):69-82.

16. McLean J, Rho YS, Kuruba G, et al. Clinical Practice Patterns in Chemotherapeutic Treatment Regimens for Metastatic Colorectal Cancer. *Clin Colorectal Cancer.* 2016;15(2):135-140.
17. Abrams TA, Meyer G, Schrag D, Meyerhardt JA, Moloney J, Fuchs CS. Chemotherapy usage patterns in a US-wide cohort of patients with metastatic colorectal cancer. *J Natl Cancer Inst.* 2014;106(2):djt371.
18. Seal BS, Sullivan SD, Ramsey SD, et al. Systemic therapy for colorectal cancer: patterns of chemotherapy and biologic therapy use in nationally representative US claims database. *BioDrugs.* 2014;28(2):229-236.
19. Hess GP, Wang Pf Fau - Quach D, Quach D Fau - Barber B, Barber B Fau - Zhao Z, Zhao Z. Systemic Therapy for Metastatic Colorectal Cancer: Patterns of Chemotherapy and Biologic Therapy Use in US Medical Oncology Practice. 2010.
20. Sonbol MB, Mountjoy LJ, Firwana B, et al. The Role of Maintenance Strategies in Metastatic Colorectal Cancer: A Systematic Review and Network Meta-analysis of Randomized Clinical Trials. *JAMA Oncol.* 2019:e194489.
21. Tipton E. How Generalizable Is Your Experiment? An Index for Comparing Experimental Samples and Populations. *Journal of Educational and Behavioral Statistics.* 2014;39(6):478-501.

3.12 Appendix

Linear model

Assume that we have a linear outcome model $Y = X_1 + X_2$ without an interaction term. According to the properties of random variables, we can simply derive the expected value and variance of sum of multiple random variables by the following equations.

$$E(Y) = E(X_1 + X_2) = E(X_1) + E(X_2) \quad (8)$$

$$Var(Y) = Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2cov(X_1, X_2) \quad (9)$$

$$cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2) = \rho_{x_1x_2} \cdot \sqrt{Var(X_1)Var(X_2)} \quad (10)$$

If X_1 and X_2 are independent, the covariance will be zero because of no correction among two variables. In other words, Equations (9) and (10) show that the correlation $(\rho_{x_1x_2})$ between X_1 and X_2 should be included to get correct variance of the outcome (Y) if X_1 and X_2 are not independent. Additionally, Equations (10) and (11) indicated that the covariance should be considered to get correct expected value of the outcome if we have an interaction term in the linear models.

$$E(Y) = E(X_1 + X_2 - X_1 \cdot X_2) = E(X_1) + E(X_2) - E(X_1 \cdot X_2) \quad (11)$$

Non-linear model

Similarly, the correlation among covariates should be considered to obtain correct expected value and variance of the outcome when we have non-linear outcome models. Let our non-linear outcome model as $Y = e^{X_1+X_2}$, and assume that both X_1 and X_2 follow standard normal distribution N(0,1). Then, the expected value and variance of the outcome can be shown in Equations (12) and (13).

$$E(Y) = E(e^{X_1+X_2}) = E(e^Z) = e^{\mu+\frac{\sigma^2}{2}}, \qquad where\ Z \sim N(\mu, \sigma^2) \tag{12}$$

$$Var(Y) = e^{2\mu+\sigma^2} \cdot (e^{\sigma^2} - 1) \tag{13}$$

Note that $\sigma^2 = {\sigma_{x_1}}^2 + {\sigma_{x_2}}^2 + 2 \cdot \rho \cdot \sigma_{x_1} \cdot \sigma_{x_2}$. Thus, we can see that correlation between X_1 and X_2 will affect both expected value and variance of $e^{X_1+X_2}$.

To have better understanding on how the correlation among covariates would affect the mean and standard deviation of the outcomes, we also provide a numerical example through simulations.

Example

We first created true data and calculated the mean and standard deviation of the outcome as the gold standard. Then, we simulated data with and without correlation and compared the results. As we expected, we can get more accurate mean and standard deviation of the outcome when we simulated data with correlation than without correlation, especially when the outcome models are a linear model with effect modifiers or non-linear models.

Outcome Models - Mean, (SD)		Summary Data of Outcome Y		
		Actual Data	Simulated Data **with correlation**	Simulated Data **without correlation**
M1	Linear model without interaction	-0.06 (2.26)	0 (2.24)	-0.01 (1.73)
M2	Linear model with interaction	-0.45 (2.12)	-0.41 (2.00)	0 (1.73)
M3	Non-linear model without interaction	11.17 (85.85)	12.36 (87.59)	4.43 (16.32)
M4	Non-linear model with interaction	1.89 (3.75)	1.79 (3.16)	4.27 (41.67)

Linear models: M1 = $Y = X_1 + X_2 + X_3$; M2 = $Y = X_1 + X_2 - X_1 \cdot X_2$

Non-linear models: M3 =$Y = e^{X_1+X_2+X_3}$; M4 =$Y = e^{X_1+X_2-X_1 \cdot X_2}$

Where X_1, X_2, X_3 have correlation $\rho_{x_1x_2} = 0.4$, $\rho_{x_1x_3} = 0.25$, $\rho_{x_2x_2} = 0.3$, and all $\boldsymbol{X} \sim N(0,1)$

Chapter 4: Conclusions

In this dissertation, we developed an iterative algorithm to approximated correlations of covariates in a trial using published trial summary data and individual-level real-world data. When trial population are very different from the real-world population, we believe that using approximated correlations of covariates from the algorithm would be a better way to mimic individual-level trial data than using RWD correlation to do so.

The simulated individual-level trial data can be used to compare the distributions of patient characteristics among diffident populations and assess the generalizability of trials using existing methods, which may help to inform clinical or coverage decisions. In the same manner, when the individual-level trial data are not available, we can use the simulated trial data instead in Aim 1 to assess the similarity of patient characteristics between trial participants and real-world population, and to explore how and for whom physicians would adopt treatments in practice post clinical trials.

Further research is needed to apply this algorithm to other empirical examples which reflect a significant difference in correlations among covariates between the trial cohort and the real-world cohort, to perform model selection to build an outcome prediction model when larger data are available, to explore the potential use in microsimulations for economic evaluation, and to develop a test to identify whether the correlations are different between RWD and the trial before applying the algorithm.

VITA

Wei-Jhih Wang is a fifth-year PhD candidate in Pharmaceutical Outcomes Research and Policy at The Comparative Health Outcomes, Policy, and Economics (CHOICE) Institute at the University of Washington. Prior to pursuing her PhD, Wei-Jhih received her Masters in Health Services from the University of Washington in 2015 and Masters in Biostatistics from National Taiwan University in 2008. Wei-Jhih's research focuses on the applications of Data Sciences and Health Economics, including methodological topics, comparative effectiveness research and outcomes research.