© Copyright 2022

Woojung Lee

# Investing in Clinical Trials for Older Adults:

# The Value and Challenges of Older Adult-Specific Clinical Trials

Woojung Lee

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

David Veenstra, Chair

Joshua Carlson

Anirban Basu

Program Authorized to Offer Degree:

Health Economics & Outcomes Research

University of Washington

# Abstract

## Investing in Clinical Trials for Older Adults: The Value and Challenges of Older Adult-Specific Clinical Trials

Woojung Lee

Chair of the Supervisory Committee:
David Veenstra
Department of Pharmacy

Conducting clinical trials for older adults can address the long-standing issue of the underrepresentation of older adults and the lack of clinical evidence for them. Examining the challenges and the value of such trials could facilitate discussion among trial sponsors on the investment in and the prioritization of older adult-specific trials. This dissertation examines the risk of failure and the real-world value of older adult-specific clinical trials. In Chapter 1, we developed a model that predicts the risk of failure of older adult-specific cancer clinical trials, using trial-level factors. The best-performing machine learning model had an acceptable performance and included nine trial-level factors that can be measured using the trial protocol.

The model can aid in the design and prioritization of future older adult-specific clinical trials. Future works examining the causal relationship between the important factors the trial failure can help develop strategies to reduce the risk of failure of cancer trials for older adults. In Chapter 2, we estimated the impact of cumulative evidence from older adult-specific trials on the prescribing patterns, using phase 3 clinical trials for post-lumpectomy irradiation in early-stage breast cancer (ESBC) as a case study. The difference-in-differences analysis showed that older adult-specific trial results led to a significant decrease in irradiation use over time among older adults and the rate of decrease was significantly accelerated by longer-term follow-up results. These findings confirm the potential of new evidence for older adults to make a substantial change in providers' prescription patterns and the importance of evidence accumulation in driving their behavioral change. Chapter 3 builds on this analysis to quantify the downstream clinical and economic benefits of older adult-specific trial results, using a health-transition model. We found that the older adult-specific trial results on the post-lumpectomy irradiation in ESBC resulted in a substantial cost-saving in the US society, without significantly changing clinical outcomes. These results can inform the discussion around whether to invest in older adult-specific clinical trials. More case studies for trials with different types of results are warranted to have a more comprehensive understanding of how different types of older adult-specific trials generate value.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGEMENT

It is such a blessing that I was able to work with Dr. David Veenstra. Words cannot express how grateful I am for his guidance and mentorship. Your guidance allowed me to grow as an independent researcher and a good scientist. Also, you gave me a lot of research opportunities where I was able to learn interesting research topics and collaborate with many different people, which I believe was a great foundation for my next step. You also have encouraged me when I was down and frustrated, which kept me motivated. I am truly grateful for how you deeply care about your students and help them flourish.

I would also like to thank my dissertation committee members, Josh Carlson and Anirban Basu for their support and intellectual contributions to my growth. I am grateful for your not hesitating to spend time for student mentoring and discussion. I was able to overcome many of the challenges and learn invaluable knowledge and skills thanks to your help.

Also, I would like to sincerely thank all other faculties who I have collaborated or interacted with: Zach, Lou, Beth, Ryan, Aasthaa, Doug, and Shelly. Thank you for sharing with me your tremendous expertise and mentoring for research and career. Your advice and support have shaped what I am now.

I am also very grateful that I met so many good friends and colleagues during my PhD journey. Especially, Nathaniel, Meng, and Kangho: thank you for always making some time for me to answer my dumb and random questions and giving me lots of emotional support. You are my

role models. Also, many thanks to all my friends at CHOICE for serving as sounding boards for my thoughts, questions, and sometimes, frustrations. It has been a pleasure exploring nice places and restaurants in Seattle with you and sharing the ups and downs of our lives.

I also would like to extend gratitude to my family for their unwavering support and prayer. Mom and Dad, thanks for your limitless love and always being supportive of whatever I choose and do. Your dedication and prayer are what give me strength each day. To my husband, Yang Joon: thank you for always believing in me, being my cheerleader, and always trying to boost my self-esteem. You are one of the most precious gifts that God gave me.

Lastly, I thank God for all the blessings and everything you have done for me. I am excited to see what you will do for me throughout the rest of my life.

# Chapter 1. Introduction

The percentage of the population aged 65 or older in the United States (US) has been increasing in the past decade and will continue to rise due to improved life expectancy and aging of the baby boomer generation. The year 2030 will be an important demographic turning point in the history of the United States, when the number of older adults is projected to outnumber children, based on the US Census Bureau's National Population Projections.[1]

Despite this demographic shift that signals the need for a change in clinical evidence generation, older adults have been underrepresented in clinical trials.[2-6] Most trials have failed to collect information about drug efficacy and safety specific to older patients and to measure endpoints that matter the most to them, leading to a paradoxical situation where the outcomes of a drug has not been adequately assessed in the subgroup that accounts for the majority of the patient population. The lack of evidence for older adults forces providers to rely on suboptimal evidence, potentially leading to clinical harm and low-value care, and ultimately a waste of healthcare resources.

The US Food and Drug Administration (FDA) made several recommendations in its recent guidance for industry, including enrolling more older adults in pivotal randomized trials, collecting unique information for older adults, and conducting additional post-market trials.[7] The Association for Clinical Oncology (ASCO) and the International Society for Geriatric Oncology (SIOG) also recently have called for increased efforts to conduct more studies in older adults, highlighting the fact that there is limited evidence on how to treat older patients with cancer.[8, 9] In Europe, the European Medicines Agencies (EMA) adopted its Geriatric Medicine Strategy to undertake specific efforts to ensure that the needs of older people are taken

into account in the evaluation of new medicines.[10, 11] Together, these recommendations underscore the demand for better evidence on the use of drugs in older patients.

Despite the increasing interest in conducting clinical trials in older adults, the challenges of conducting such trials have not been well studied. Conducting clinical trials for older adults could be difficult, leading to early termination and a waste of resources without creating any new scientific progress. Considering the unique characteristics and treatment goals of older adults, unique trial-level factors may play a role in predicting the failure of trials in older adults. Identifying trial-level factors that are predictive of failure of clinical trials for older adults can inform the design and the prioritization of future older adult-specific clinical trials, by helping trial sponsors and researchers assess a trial's likelihood of successful completion. Efforts to understand the potential benefits of older adult clinical trials has also been limited. Considering the limited resources, it is important to assess whether trials in older adults could generate significant value to society. Understanding the clinical and economic benefits such trials can generate in the real world may facilitate discussion on the investment in older adult-specific clinical trials among trial sponsors such as the National Institute of Health (NIH) and industry.

The overall goal of my dissertation was to develop a model that predicts the failure of older adult-specific clinical trials using trial-level factors and to quantify the real-world value of older adult-specific trials that were successfully completed. The definition of older adult-specific trials can vary but it was defined as trials with a lower age limit of 60 years or older based on previous literatures.[12][13] This dissertation was focused on cancer because it is one of the disease areas where treatment decision-making for older adults is particularly challenging due to the vulnerability to adverse drug events, different treatment goals, and the drug costs. The results of the proposed work can facilitate investment and design decisions by public and private

sponsors of clinical trials in older adults by providing a tool to assess the risk of trial failure and highlighting the potential value of older adult-specific trials.

# Chapter 2. Can We Predict Trial Failure among Older Adult-specific Clinical Trials using Trial-level Factors?

## 2.1 Abstract

Objectives: Conducting older adult-specific clinical trials can help overcome the lack of clinical evidence for older adults due to their underrepresentation in clinical trials. Understanding factors contributing to the successful completion of such trials can help trial sponsors and researchers prioritize studies and optimize study design. We aimed to develop a model that predicts trial failure among older adult-specific cancer clinical trials using trial-level factors.

Methods: We identified phase 2-4 interventional cancer clinical trials that ended between 2008 and 2019 and had the minimum age limit of 60 years old or older using Aggregate Analysis of ClinicalTrials.gov data. We defined trial failure as closed early for reasons other than interim results or toxicity or completed with a sample of less than 85% of the targeted size. Candidate trial-level predictors were identified from a literature review. We evaluated eight types of machine learning (ML) algorithms to find the best model. Model fitting and testing were performed using 5-fold nested cross-validation. We evaluated the model performance using the area under receiver operating characteristic curve (AUROC).

Results: Of 209 older adult-specific clinical trials, 87 (42%) were failed trials per the definition of trial failure. The model with the highest AUROC in the validation set was the least absolute shrinkage and selection operator (LASSO) (AUROC in the test set = 0.70; 95% CI: 0.53, 0.86). Nine trial-level factors included in the best model were the study sponsor, the number of participating centers, the number of modalities, the level of restriction on performance score,

study location, the number of arms, life expectancy restriction, and the number of target size. Results were robust to different definitions of trial failure.

Conclusion: We identified multiple trial-level factors predictive of trial failure among older adult-specific clinical trials and developed a prediction model that can help estimate the risk of failure before a study is conducted. The study findings could aid in the design and prioritization of future older adult-specific clinical trials.

## 2.2 Introduction

Although the aging population signals the need for a change in clinical evidence generation, older adults have been underrepresented in clinical trials.[14, 15] The US Food and Drug Administration (FDA) made several recommendations in its recent guidance for industry to address the lack of clinical evidence in older adults, including enrolling more older adults in pivotal randomized trials or conducting additional post-market trials.[16] The Association for Clinical Oncology (ASCO) and the International Society for Geriatric Oncology (SIOG) also have called for increased efforts to conduct more studies in older adults, highlighting the fact that there is limited evidence on how to treat older patients with cancer specifically.[8, 9]

Conducting clinical studies, however, is often time-consuming and expensive. Furthermore, trial failures can arise for various reasons, such as administrative, logistical, and patient recruitment-related issues.[17][18, 19] Unsuccessful completion of a clinical trial represents a waste of resources and time invested in all prior trials and the loss of benefits from alternative studies that could have been pursued. Therefore, it is important for researchers and trial sponsors to ensure that conducted clinical trials can generate meaningful clinical knowledge that can potentially inform clinical practice in older adults.

Considering the increasing demand for older adult-specific clinical evidence, understanding the factors attributable to unsuccessful completion of clinical trials in older adults will become more crucial in the coming years. Several studies have investigated the associations between the trial-level factors and outcomes related to unsuccessful completion of clinical trials (e.g., early termination or low accrual), although they have focused on clinical trials in general patients.[20-32] These studies found that unsuccessful completion of a trial is associated with several trial-level factors such as background, study design-related, disease-related, and

intervention-related factors, suggesting that the probability of trial failure can be estimated before a study is launched.[18]

Older adults often have unique characteristics or treatment goals compared to younger adults, and thus unique trial-level factors may play a role in predicting the failure of trials specific to older adults. For example, older adults' participation may be more challenging and affected by different factors than those affecting the recruitment of general patients.[33] Also, older adult-specific trials may suffer from different logistic or administrative issues (e.g., the lack of funding or interest of sponsors). Although one study examined factors associated with the accrual of older adults in cancer clinical trials, the study mainly examined exclusion criteria-related factors and did not focus on trials specific to older adults.[33] Additional studies specific to clinical trials in older adults and investigating a broader set of trial-level factors could help develop a more comprehensive understanding of how trial-level factors predict the failure of trials among older adults. Ultimately, the study findings would help researchers and trial sponsors better allocate their research budgets to generate evidence in older adults and help design studies that are less likely to fail.

The aim of this study was to identify trial-level risk factors for trial failure in older adult-specific clinical trials. We focused on clinical trials in cancer because it is a disease area where the underrepresentation of older adults has been an important issue.[14][16] Also, previous studies examining similar questions among trials in the general patient population have been done in oncology.[20-32] Therefore, focusing on cancer trials will make it more feasible to compare the findings between trials in older versus general patients.

## 2.3 Methods

*Data sources*

We used data from the 2021 Aggregate Analysis of ClinicalTrials.gov (AACT) database, a publicly available database that contains information about every study registered in ClinicalTrials.gov. Content is downloaded from ClinicalTrials.gov daily and loaded into AACT.[34] ClinicalTrials.gov became available in 2000 and has been maintained by the US National Library of Medicine (NLM) at the National Institute of Health (NIH). Registration is required for studies that meet the definition of an "applicable clinical trial" and either were initiated after September 27, 2007, or initiated on or before that date and were still ongoing as of December 26, 2007.[35] As of March 2021, ClinicalTrials.gov lists 405,147 studies with locations in all 50 States and in 220 countries. Seventy-seven percent of the registered studies are interventional studies for intervention types, including drug or biologic, behavioral, surgical procedure, device, etc., and the other 23% are observational studies.[36]

Information on the registered studies is provided and updated by the sponsor or principal investigator of a study. The data source includes information about clinical studies such as primary purpose (e.g., basic science, treatment, and prevention), study designs (e.g., phase, randomized, and masking), conditions (i.e., type(s) of disease studied), interventions (e.g., intervention type and timing), endpoints (e.g., outcomes and time frame), facilities (e.g., countries and study sites), eligibilities (i.e., inclusion and exclusion criteria), and sponsors. ClinicalTrials.gov also provides information on the current status of a study as follows: not yet recruiting, recruiting, enrolling by invitation, active but not recruiting, suspended, terminated, completed, withdrawn, and unknown. For trials completed or closed early (i.e., terminated,

withdrawn, or suspended), data on both targeted and actual sample sizes are available, and, for closed trials, reasons why the study stopped are provided for most of them.

*Trial inclusion criteria*

Studies that were (a) interventional, (b) phase 2-4, (c) in cancer, (d) ended (closed early or completed) between 2008- 2019, and (e) had a lower bound of age limit of 60 years old or older were included. Studies with a status of "unknown" were excluded. We excluded phase 1 studies from the analysis because the recruited participants in phase 1 trials are often healthy volunteers. Trials that ended before 2008 were excluded because the registration was not mandatory for trials that ended before December 2007, and thus the data is likely to be incomplete. Those that ended after December 2019 were also excluded because ongoing trials in 2020-2021 were likely affected by COVID19, an unusual event. We used the age of 60 as a cut-off for defining older adult-specific clinical trials because: (1) although 65 years of age is the commonly-used age cut-off for defining older adults, this cut-off is arbitrary, meaning that different cut-offs (e.g., 60 or 70) can be used in different settings, (2) the sample size reduces significantly if the cut-off is increased to 65 years of age, which leads to smaller sample size to train and test a model, and (3) 60 years of age has been used as a cut-off for older adult-specific clinical trials in previous studies.[12, 13]

*Outcome variable (trial failure)*

Trial failure was defined based on the review of existing literature on successful trial completion as (1) closed (i.e., terminated, withdrawn, or suspended) early for reasons other than interim results or toxicity issues (i.e., non-informational closure) or (2) completed but with an

actual sample size of less than 85% of the targeted sample size.[18, 37] For closed trials, we reviewed the reasons why trials stopped to categorize them into successful and failed trials. We considered trials ended early due to interim results or toxicity as successful studies because they were able to provide some meaningful scientific knowledge. Trials closed for reasons related to poor accrual, logistics (e.g., cancellation by a sponsor, inadequate budget, or departure of a principal investigator from the institution), and others (e.g., no longer necessary and none given) were considered as non-informational closure. The 85% was chosen as a cut-off for successful vs. failed trials for the completed trials, based on existing literature showing that the statistical power for the primary endpoint becomes seriously compromised when the actual accrual is less than approximately 85% of the target population (we varied this cut-off in the sensitivity analysis).[37-39] Completed trials with missingness with either targeted or actual sample size were excluded from the analysis.

*Candidate predictors*

We identified potential trial-level predictors for trial failure based primarily on a targeted review of the existing literature. Specifically, we reviewed studies that examined trial-level factors associated with or predictive of low accrual or unsuccessful trial completion.[18, 20-32] We also included several factors that may be more relevant to older adult-specific clinical trials and that have not been studied. Predictors included in the analysis are described in the results section.

We explored different sets of candidate predictors using three different strategies: (a) all predictors, (b) predictors selected based on the clinical rationale, and (c) the best subset. With the selection strategy ‘a’, we included all trial-level factors available and measurable using the

AACT dataset and with ‘b’, we reduced the number of candidate predictors in ‘a’ based on the clinical rationale. With the strategy ‘c’, we explored subsets (i.e., different combinations) of all candidate predictors. Since the number of different subsets is overwhelming (if we have 10 predictors, for example, there will be 1,024 possible subsets), we started from the predictors selected in ‘b’ and either added one or more predictors with a good clinical rationale or excluded one or more predictors with a less strong rationale. We explored different forms of predictors (i.e., continuous, binary, and categorical) for the predictors that can be potentially converted into different forms. Missing values were addressed using multivariate Imputation by chained equations with 20 iterations.

*Machine learning (ML) algorithms*

We used eight types of machine learning (ML) algorithms: least absolute shrinkage and selection operator (LASSO), Elastic Net, ridge, random forest, AdaBoost classification trees (AdaBoost), boosted generalized linear regression learner (GLMboost), Bayesian additive regression tree (BART), and super learner. Super learner included regularization methods (LASSO, Elastic Net, and ridge), random forest, and BART.

Ridge, LASSO, and Elastic Net are forms of regularized linear techniques that reduce the variance of a model by preventing the parameter estimates from becoming too large.[40] LASSO and Elastic Net could shrink the parameters toward zero, creating a parsimonious model that is considered practical.[40] For Elastic Net, we set alpha, a hyperparameter that is used to assign how much weight will be given to each of the Lasso penalty (alpha = 1) vs. ridge penalty (alpha = 0), at 0.50. We compared these parametric ML models with non-parametric ones. Ensemble tree models and super learner were chosen among the non-parametric learners. We

chose ensemble tree models because it is generally more interpretable than the other learners. Considering the target audience (i.e., trial sponsors and researchers) and implication of this study, models that show at least some predictive properties of each predictor are more likely to be appealing for practical use.[40] Albeit less interpretable, we explored super learner, an ensemble ML algorithm, because it could give a predictive performance better than any single ML model.

*Model fitting and testing*

We used five-fold nested cross-validation (CV) for both model fitting (i.e., choosing optimal model tuning parameters) and selection (i.e., assessing the predictive performance).[41] Finding model tuning parameters using CV was utilized to guard against model over-fitting. Cross-validation was chosen over data splitting because it allows using all the data and is thus beneficial when the data is not expected to be large.[42] Specifically, the dataset was split into training and test datasets, and the training dataset was further split into training and validation datasets. The inner (i.e., nested) CV loop was performed on the training and validation dataset to find an optimal set of model tuning parameters. We used the area under the receiver operating characteristic (AUROC) measured in the validation dataset (i.e., validation AUROC) as the selection criterion for the hyperparameter tuning in the inner loop. AUROC is a performance measurement for the classification problems showing how much the model is capable of distinguishing between classes. Once the optimal hyperparameter was found through the inner CV loop, the model was fitted on the complete training and validation data sets with the optimal hyperparameter. The generalizable AUROC was then calculated in the held-out test dataset, using the outer CV loop (i.e., test AUROC).

The best model was primarily chosen based on the validation AUROC. We also considered the model interpretability (i.e., the degree on which we can obtain information on the predictive properties of each predictor) and model complexity (i.e., single model was preferred than ensemble model such as super learner), in a case when the validation AUROC was not significantly different. For the best model, we calculated AUROC, sensitivity (i.e., the ability to designate failed trials correctly), specificity (i.e., the ability to designate successful trials correctly), positive predictive value (PPV) (i.e., the probability that trials that were predicted to be fail actually fail), and negative predictive value (NPV) (i.e., the probability that trials that were predicted to successful actually success) in the test set.

We weighted the estimated impact of the false positive and false negative results to determine the probability cut-off. False positives (i.e., a model falsely predicts a study to be failed; 1-sensitivity) will result in the loss of benefit from the study that could have been conducted if the model prediction had been correct. False negatives (i.e., a model falsely predicts a study to be successful; 1-specificity) will lead to a waste of money from the failed trials and also the loss of potential benefits from the other studies that would have otherwise been conducted if the resources had not been wasted. We determined that the magnitude of the impact of false positive and false negative would vary greatly by multiple factors, and thus, one is not necessarily greater than the other. Therefore, we chose a cut-off that gives balanced sensitivity and specificity, which was 0.4. Data processing was done using STATA/MP 16.1, and analysis was done using RStudio 2021.09.1.

*Post-hoc association analysis for statistical inference*

As an exploratory analysis, we performed a logistic regression including all predictors to examine the statistical inference of the relationship between the important trial-level factors and trial failure. Odds ratio, 95% confidence interval (CI), and p-value were examined for each of the trial-level factors.

*Sensitivity analyses*

Since the cut-off of 85% that was used to classify successful vs. failed trials among the completed trials is somewhat arbitrary, we varied the cut-off by +/- 10 percentage points in a sensitivity analysis. We repeated the best-performing model in the base case with a different cut-off and examined the model performance and important predictors.

## 2.4 Results

*Analytic sample*

A total of 209 older adult-specific clinical trials were included (Table 1). Among the included older adult-specific clinical trials, 54 were closed early due to the following reasons: poor accrual (21/54, 38%), logistics (17/54, 31%), interim results or toxicity (8/54, 15%), and others (8/54, 16%). There were 155 completed trials included, among which 114 (64%) had an actual sample size equal to or greater than 85% of the targeted sample size and 41 (23%) with less than 85% of the target. Based on the definition of successful vs. failed trials, 122 and 87 were categorized into successful and failed trials, respectively.

About half of the included studies were done in the non-US countries only, and about a third in the US only (Table 2). The mean number of countries and participating centers was 2.1 (standard deviation (SD) = 3.5) and 22.9 (SD = 45.6), respectively. Approximately 45% (95/209) of all trials were sponsored by industry and 10% (21/209) by the public sector. Seventy-one percent of all trials (149/209) were phase 2, and 48.3% (101/209) were randomized studies. There were 4.8% (10/209) of blinded studies. Thirty-seven (77/209) percent of the included trials studied common solid cancer (i.e., breast, colorectal, lung, prostate, and endometrial cancer), and 31.1% (65/209) were in metastatic settings. About a third of the trials were biomarker-specific (69/209, 33.0%). Studies were evenly distributed in terms of the level of restriction by renal dysfunction: 38.2% (80/209) of all included trials had no restriction, 30.1% (63/209) had moderate restriction, and the rest (66/209, 31.6%) had a strict restriction (the definition of moderate and strict restriction is provided in the next section). Thirty-eight percent (80/209) restricted the eligibility based on the life expectancy. The maximum possible performance score

(PS) was 2 in the majority of the included studies (124/209, 59.3%), followed by higher than 3 or non-specific (53/209, 25.4%).

*Included candidate predictors*

Based on the targeted literature search and the consideration of older adult-specific trial-level factors, we included a total of 37 candidate predictors in all-predictor models (Table 2). Candidate predictors included background, study design-related, disease-related, endpoint-related, intervention-related, and eligibility-related factors.

Background factors included study location (i.e., US only, non-US only, and both), the number of participating countries, the number of participating centers, and types of sponsors (i.e., industry, public or government, and others). Study design-related factors included study phase (i.e., 2, 3, and 4; phase 2/3 was treated as phase 3), randomization, blindness, the number of arms, first-line setting, and target sample size.

Disease-related factors included studies on common solid cancer (i.e., breast, colorectal, lung and bronchus, prostate, and endometrial cancer), breast cancer studies, metastatic setting (i.e., metastatic only, non-metastatic only, and non-specific), whether multiple conditions were studied, and whether a study was biomarker-specific.

In terms of endpoint-related factors, we generated four indicators of whether each of the following endpoints was included as a primary endpoint: survival, progression, response, and toxicity. Furthermore, we included whether geriatric assessment (GA) was used to measure any of the endpoints.

Intervention-related factors included whether studied interventions include targeted or immunotherapies, whether a control arm received a placebo, and the number of modalities (ranging from 1 to 4; treatment modalities include drugs, procedure or surgery, radiation, and others).

Eligibility-related factors included the level of restriction on renal dysfunction (i.e., strict, moderate, and none; strict if the participation was restricted to those with the creatinine level of equal to or less than 1.5 times the upper limit of normal and moderate if it is restricted to the creatinine level of greater than 1.5 times the upper limit of normal), whether eligibility was restricted by the life expectancy, the maximum possible PS (i.e., 1, 2, and 3 plus or non-specific), whether a caregiver was needed, whether frailty was assessed (using GA, comorbidity scale, or physicians' assessment), whether the eligibility criteria was restricted by the frailty level (i.e., fit only, frail only, and not-specific), and whether the eligibility was restricted by the tolerability to the standard of care.

Among the 37 candidate predictors, we chose 15 predictors based on a clinical rationale for the selected-predictor models: study location (i.e., US only, non-US only, and both), the number of participating centers, types of sponsor (i.e., industry, public, and others), the number of arms, target sample size, whether a study was biomarker-specific, the number of modalities, the level of restriction based on renal dysfunction (i.e., strict, moderate, and none), whether the eligibility was restricted by life expectancy, maximum possible PS (i.e., 1, 2, and 3 plus or non-specific), whether frailty was assessed.

Based on the exploration of the subsets, the subset with the highest validation AUROC had the following 12 candidate predictors: study location (i.e., US only, non-US only, and both), the number of participating centers, types of sponsor (i.e., industry, public, and others), the

number of modalities, the number of arms, target sample size, whether the eligibility was restricted by life expectancy, maximum possible PS (i.e., 1, 2, and 3 plus or non-specific), and metastatic setting.

*Prediction performance*

When all candidate predictors were included, the validation AUROC obtained using the regularization models were 0.58 (95% CI: 0.49, 0.67), 0.58 (95% CI: 0.49, 0.67), and 0.56 (95% CI: 0.47, 0.65), for the LASSO, Elastic Net, and ridge, respectively. With tree-based models, the validation AUROC was 0.63 (95% CI: 0.54, 0.72), 0.60 (95% CI: 0.51, 0.69), 0.59 (95% CI: 0.50, 0.68), and 0.41 (95% CI: 0.32, 0.49), for random forest, adaBoost, GLMboost, and BART, respectively. The validation AUROC with super learner was 0.63 (95% CI: 0.54, 0.71) (Table 3).

The prediction performance was generally higher when candidate predictors were selected based on the clinical rationale, with the validation AUROC of 0.65 (95% CI: 0.56, 0.73), 0.65 (95% CI: 0.56, 0.74), and 0.65 (95% CI: 0.56, 0.73) for LASSO, Elastic Net, and Ridge, respectively. Tree-based models, except for BART, had a similar validation AUROC to the regularization models: 0.65 (95% CI: 0.56, 0.73), 0.64 (95% CI: 0.55, 0.72), 0.65 (95% CI: 0.57, 0.73), 0.34 (95% CI: 0.28, 0.42) for random forest, adaBoost, GLMboost, and BART, respectively. Super learner had a validation AUROC of 0.66 (95% CI: 0.58, 0.74).

When the best subset of the candidate predictors were used, the validation AUROC with regularization techniques was 0.66 (0.57, 0.75), 0.66 (0.58, 0.74), and 0.66 (0.57, 0.75), for LASSO, Elastic Net, and ridge, respectively. The performance with ensemble trees was a little lower with a validation AUROC of 0.63 (0.55, 0.72), 0.61 (0.52, 0.70), 0.66 (0.58, 0.75), 0.34

(0.26, 0.42) for random forest, adaBoost, GLMboost, and BART, respectively. Super learner had a validation AUROC of 0.66 (0.58, 0.75).

Overall, the validation AUROC was not significantly different across the examined models (i.e., 95% confidence intervals overlap). Especially for the models with selected predictors and with the best subset, the difference in the validation AUROC was within 0.02 across all models, except for BART.

The two models with the highest validation AUROC were LASSO and super learner with the best subset candidate predictors. Since LASSO gives more information on the predictive properties for each candidate predictor and is less complex than the super learner, we chose LASSO with the best subset as the best model.

When tested on the held-out test sample through the outer loop of the nested cross-validation, the best model had the test AUROC of 0.70 (95% CI: 0.53, 0.86). The sensitivity, specificity, PPV, and NPV in the test set were 0.70 (95% CI: 0.28, 0.91), 0.66 (95% CI: 0.35, 0.90), 0.56 (95% CI: 0.41, 0.71), and 0.76 (95% CI: 0.65, 0.87), respectively.

*Importance level of predictors included in the best model*

Among 12 candidate predictors included in the best model (i.e., LASSO with the best subset), the coefficients for types of sponsor being others vs. industry, maximum possible PS of 2 vs. 1, and metastatic setting were shrunk to zero and thus excluded from the final set of predictors (Table 4).

The other nine trial-level factors that remained as final predictors were: study sponsor being NIH vs. industry, the number of centers, the number of treatment modalities, maximum possible PS being 3 plus or non-specific vs. 1, study location both US and non-US vs. US only,

study location being in non-US vs. US only, the number of arms, life expectancy restriction, and the number of target sample size.

Based on the LASSO coefficient, the followings of the nine final factors were positively predictive of the trial failure: study sponsor being NIH vs. industry (OR = 1.25), the number of treatment modalities (OR = 1.51), maximum possible PS being 3 plus or not specific vs. 1 (OR = 1.62), study location being both in the US and non-US vs. US only (OR = 2.25), life expectancy restriction (OR = 1.94) and the number of target sample size (OR = 1.01). The factors that were negatively predictive of the trial failure were: the number of centers (OR = 0.88 per 10 additional centers), study location being in the non-US only vs. US only (OR = 0.41), the number of arms (OR = 0.77, per one more arm).

*Importance level of predictors excluded from the best model (those included in the all-predictor model only)*

The all predictor model with the highest validation AUROC (i.e., random forest) showed that, among the candidate predictors excluded from the best subset, the following ones had an importance level of below zero: targeted or immunotherapies studied, placebo control, trial phase being 4 vs. 2, the renal restriction being moderate vs. strict, frailty level being non-specific vs. fit, frailty level being frail vs. fit, restricted by drug tolerability, whether frailty was assessed, multiple conditions had negative importance (Figure S2.1). The negative importance level suggests that the contribution to the model accuracy is worse than the random permutation.

The rest of the candidate predictors had a positive importance level. The renal restriction was none vs. strict, and the breast cancer trial had a significantly higher importance level than the

others. The direction of prediction of the included candidate predictors was not shown in the random forest model.

*The statistical association between the trial-level factors and the trial failure*

The post hoc logistic regression model showed that, among all candidate predictors, only the number of centers, study being non-US only vs. US only, and life expectancy restriction were significantly associated with the trial failure at a significance level of 5% (Table 4). Having ten additional study centers was associated with the odds of the trial failure lower by a factor of 0.83 (95% CI: 0.71, 0.94), and studies conducted in the non-US countries only had the odds of trial failure that is lower by 0.32 times than those conducted in the US countries only (95% CI: 0.12, 0.82). Studies that restricted the eligibility based on the life expectancy had the odds of trial failure higher by 2.17 times than the studies that do not (95% CI: 1.04, 4.73).

*Sensitivity analyses*

The validation and test AUROC of the best model were similar when alternative cut-offs for the trial failure were used. The validation AUROC was slightly higher than the base-case when the cut-off of 95% was used (validation AUROC: 0.69 (95% CI: 0.61, 0.77); test AUROC: 0.71 (95% CI: 0.40, 0.95)) and slightly lower for 75% (validation AUROC: 0.67 (95% CI: 0.58, 0.76); test AUROC: 0.69 (95% CI: 0.52, 0.86))

## 2.5 Discussion

We identified trial-level factors predictive of the failure of older adult-specific clinical trials in cancer and used them to develop ML models. The best model had an acceptable level of discrimination performance with a generalizable AUROC of 0.70. A total of nine trial-level factors contributed to the prediction of the trial failure in the best model, all of which can be easily measured based on the study protocol by researchers and trial sponsors before conducting a study. The study findings could aid in the design and study prioritization of future older adult-specific clinical trials.

Several trial-level factors played a role in predicting the failure of older adult-specific trials, including background factors (e.g., study location, the number of centers, and study sponsor), study design-related factors (e.g., the number of study arms, the number of treatment modalities, target sample size), and eligibility-related factors (e.g., life expectancy restriction and maximum possible PS). Specifically, studies performed outside of the US were less likely to fail than those including US study sites. The reasons for a higher likelihood of failure in the US could be due to stricter regulations, longer study protocol approval time, higher study costs, and higher patient recruitment and retainment fees in the US.[43] We also found that studies with a less number of centers are more likely to fail, which is likely because of recruitment challenges (e.g., slow accrual and recruited participants not representative of the whole study population) that could lead to concerns about scientific rigor and external validity.[44] Also, NIH-sponsored trials were more likely to fail than industry-sponsored trials. This finding is likely because the industry generally had more financial resources to recruit and manage the studies.[18] Furthermore, the number of study arms was negatively associated with the trial failure, which could be explained by a higher chance of a patient to be randomized to an experimental treatment

with multiple arms, which could help recruit patients.[45][46, 47] Multiple treatment modalities, on the other hand, were predictive of trial failure, likely due to higher complexity of study procedures and patient burden. [48] Additionally, studies with a higher target sample size were more likely to fail. Achieving a high targeted sample size may require more resources and/or a longer time to recruitment, increasing the chance of early termination or insufficient sample size.[18] Restriction based on the life expectancy was also predictive of a trial failure. The reduced life expectancy is inevitable with aging and other age-related comorbidities, and thus restricting the participation based on the life expectancy may result in the exclusion of a significant proportion of older adults. Lastly, we found that the maximum possible PS being 3+ or non-specific is related to a higher chance of trail failure than the maximum possible PS of 1. This finding was consistent with studies that showed patients enrolled in trials with specific PS eligibility requirements were more likely to be elderly compared to when there were no specific criteria for PS. [33] [49] Without a clear specification of certain eligibility criteria, older adults may not know the potential risks of participating in a trial, making them reluctant to enroll in the study. The information on all these predictors is generally available in the trial protocol, which enables trial sponsors to assess the risk of failure before making study investment and prioritization decisions.

Our study findings also showed trial-level factors that have little contribution to predicting the failure of older adult-specific clinical trials: trial phase 4 vs. 2, metastatic setting, targeted or immunotherapies studied, placebo control, and several eligibility criteria-related factors, including maximum possible PS of 2 vs. 1, the renal restriction being moderate vs. strict, whether frailty was assessed, restriction based on the frailty level, and restriction based on the drug tolerability. Despite the concerns that strict trial eligibility criteria may lead to challenges in

recruiting older adults, several of the eligibility criteria-related factors were not important predictors of the trial failure.[50][51] Specifically, the maximum possible PS of 2 vs. 1 and renal restriction being strict vs. moderate did not matter much in determining the risk of trial failure, suggesting that small differences in the restriction level may not significantly affect the recruitment of older adults. Moreover, restricting the study population based on frailty or drug tolerability had a minimal role in predicting trial failure. These factors can be potential considerations for researchers who try to tailor the trial design to older adults (i.e., geriatricize the trial design). [51] For example, measuring frailty upon the study entry enables researchers to measure treatment efficacy and safety in relation to characteristics unique to older adults (e.g., comorbidities and physical and mental disabilities). Our finding suggests that such study modifications toward older adult-specific trial design may not affect the chance of trial failure.

Cancer clinical trials for older adults and general adults shared similar trial-level predictors of the trial failure. Based on the previous studies focusing on cancer trials among general patients, the following predictors were found to be predictive of trial failure across multiple studies: government-sponsorship, study location being in the US only, small number of centers, multimodalities, and non-breast cancer trials.[18, 20-32] All of these factors, except for breast cancer trials, contributed to predicting the failure of older adult-specific trials in our best model and had a consistent prediction direction. Although breast cancer trial was not one of the final predictors in our best model, it showed a relatively higher importance level in the best model with all predictors (i.e., random forest). Also, the multivariate logistic regression showed a negative association between breast cancer trial and the trial failure, which was consistent with the prediction direction among general cancer trials. One difference between trials in older vs. genera adults was the importance of the maximum possible PS. While maximum possible PS

was one of the important predictors for older adult-specific trials, it was not for trials among general patients. [27] Considering younger patients generally have good PS and thus less sensitive to the restriction based on PS, it makes sense that PS being unimportant in predicting the failure of trials for younger patients. Life expectancy restriction was not considered in any of the studies for general trials, thus comparison was not possible. For many of the trial-level factors that we found unimportant among older adult-specific trials, there have been mixed findings across the existing studies. Existing studies were done in several different settings (e.g., National Cancer Institute's Cooperative Group Trials, Canadian Cancer Trials, etc.) and for different tumor types (e.g., multiple, lung, breast, etc.), which could be the reason we observe different predictive properties for some trial-level factors. Although there was one study that comprehensively included all cancer clinical trials, a limited set of trial-level factors was examined. [25] Additional studies with a representative sample of cancer clinical trials and a more extensive set of trial-level factors could enable a more direct comparison between the results from trials for older vs. general patients.

Our study findings can be applied by trial sponsors and researchers in a couple of ways. First, this study provides a parsimonious set of predictors that are potentially associated with the risk of trial failure. If confirmed by inference studies showing how these predictors are statistically associated with the trial failure among older adult-specific clinical trials, our results can help researchers modify trial-level factors in a way to reduce the risk of study failure.[25] Furthermore, the models developed in this study can be useful in aiding trial sponsors in prioritizing clinical trials by providing a quantitative measure for the likelihood of trial success. The best-performing model had a NPV of 76%, meaning that the proportion of failed trials among those assumed to be successful would be around 24%. Depending on the error rate with

the current system, the model can be used by trial sponsors or researchers to reduce the costs arising from falsely investing in the trials that will end up failing.

## 2.6 Limitations

This study is not without limitations. First, some key potential predictors of trial failure, such as clinical importance of a research question and deviation from the standard of care, were not included due to the data unavailability or difficulties in operationalizing. Other trial-level factors may need to be considered when predicting the failure of older adult-specific cancer trials.

Furthermore, the sample size was small because we focused on older adult-specific clinical trials that are relatively uncommon, and thus there may be a relatively higher chance of overfitting. Several rules (e.g., one in ten rule, meaning that one predictive variable can be estimated for every ten events) recommend reducing the number of candidate predictors to reduce the risk of overfitting. Therefore, to alleviate the risk, we explored different number of candidate predictors by reducing the number based on the clinical rationale.

Furthermore, 8 out of 54 closed trials did not specify the reason for closing. Although we assumed that trials closed early without any specified reasons are less likely to be successful (e.g., informational termination), there might have been misclassification due to the missingness. However, the impact is unlikely to be substantial because they account for less than 4% (8/209) of the total sample.

Moreover, the cut-off of 85% to divide successful vs. unsuccessful clinical trials is arbitrary. Although we chose the one within the range of values used in previous studies, there is still a risk of misclassification. We varied this cut-off in a sensitivity analysis to alleviate the concern about finding significantly different results with a different cut-off.

Finally, although the ML approach allowed us to examine the importance of predictors, it is challenging to make a clear distinction between important versus unimportant predictors of

trial failure. Some may interpret the model results differently, especially for the middle-ranked factors.

## 2.7 Conclusion

We identified trial-level factors that are predictive of failure of cancer clinical trials among older adults and developed an ML model that can be used to predict the risk of failure of older adult-specific clinical trials. The identified trial-level factors and a prediction model based on these factors can aid in the design and prioritization of future older adult-specific clinical trials. Future works examining the causal mechanism behind the associations between the important factors and the trial failure can help develop strategies to reduce the risk of failure of cancer trials for older adults.

## 2.8 Tables & Figures

| Table 2. 1. Summary of outcome classification, stratified by trial status (n=209) | | |
|---|---|---|
| | Successful trial (n=122) | Failed trial (n=87) |
| Closed trials (n=54) | Due to interim results or toxicity (n=8) | Due to reasons other than interim results or toxicity (n=46) |
| Completed trials (n=155) | Actual sample size of ≥ 85% of target (n=114) | Actual sample size of < 85% of target (n=41) |

| Table 2. 2. Summary of trial characteristics stratified by outcome classification (n=209) | | | |
|---|---|---|---|
| | Successful trials (n=122) | Failed trials (n=87) | p-value[e] |
| | Number (%[d]) or mean (SD) | | |
| Background factors | | | |
| Location[a] | | | |
| US only[b] | 33 (27.1%) | 36 (42.9%) | 0.01 |
| Non-US only | 81 (66.4%) | 38 (45.2%) | |
| Both | 8 (6.6%) | 10 (11.9%) | |
| Number of countries | 2.3 (0.4) | 1.8 (0.3) | 0.38 |
| Number of facilities[e] | 28.9 (4.4) | 15.2 (4.5) | 0.04 |
| Sponsor[e] | | | |
| Industry[b] | 55 (45.1%) | 40 (46%) | 0.26 |
| NIH | 9 (7.4%) | 12 (13.8%) | |
| Others | 58 (47.5%) | 35 (40.2%) | |
| Study design-related factors | | | |
| Phase | | | |
| 2[b] | 87 (71.3%) | 62 (71.3%) | 0.84 |
| 3 | 30 (24.6%) | 20 (23.0%) | |
| 4 | 5 (4.1%) | 5 (5.8%) | |
| Randomized | 63 (51.6%) | 38 (43.7%) | 0.26 |
| Blinding | 9 (7.4%) | 1 (1.2%) | 0.04 |
| Number of arms[a] | 1.8 (0.1) | 1.5 (0.1) | 0.07 |
| First-line setting | 85 (69.7%) | 60 (70.0%) | 0.91 |
| Target sample size[a] | 183.5 (21.4%) | 188.5 (31.5) | 0.89 |
| Disease-related factors | | | |
| Common solid cancer | 46 (37.7%) | 31 (35.6%) | 0.76 |
| Breast cancer | 16 (13.1%) | 12 (13.8%) | 0.89 |
| Metastatic setting | | | |
| Metastatic only[b] | 40 (32.8%) | 25 (28.7%) | 0.24 |
| Non-metastatic only | 2 (1.6%) | 5 (5.8%) | |
| Non-specific | 80 (65.6%) | 57 (65.5%) | |
| Multiple conditions | 13 (10.7%) | 11 (12.6%) | 0.66 |
| Biomarker-specific[a] | 36 (29.5%) | 33 (37.9%) | 0.20 |
| Endpoint-related factors | | | |
| Survival[c] | 52 (42.6%) | 34 (39.1%) | 0.61 |
| Progression[c] | 23 (18.9%) | 13 (14.9%) | 0.46 |
| Response[c] | 36 (29.5%) | 28 (32.2%) | 0.68 |
| Toxicity[c] | 9 (7.4%) | 8 (8.2%) | 0.64 |
| GA was used to measure outcomes | 12 (9.8%) | 8 (9.2%) | 0.88 |
| Intervention-related factors | | | |
| Targeted or immunotherapies studied | 42 (34.4%) | 39 (45.4%) | 0.11 |
| Placebo control | 5 (4.1%) | 1 (1.2%) | 0.21 |
| Number of modalities[a] | 1.1 (0.4) | 1.2 (0.5) | 0.10 |
| Eligibility-related factors | | | |

| | | | |
|---|---|---|---|
| Restricted by renal dysfunction[a] | | | |
| Strict[b] | 34 (27.9%) | 32 (36.8%) | 0.39 |
| Moderate | 39 (32.0%) | 24 (27.6%) | |
| None | 49 (40.2%) | 31 (35.6%) | |
| Restricted by life expectancy[a] | 42 (34.4%) | 39 (45.4%) | 0.11 |
| Maximum performance status allowed[e] | | | |
| 1[b] | 21 (17.2%) | 11 (12.6%) | 0.25 |
| 2 | 75 (61.5%) | 49 (56.3%) | |
| 3 plus or non-specific | 26 (21.3%) | 27 (31.0%) | |
| Caregiver needed | 1 (0.8%) | 2 (2.3%) | 0.38 |
| Frailty assessed[a] | 28 (23.0%) | 23 (26.4%) | 0.56 |
| Restricted by frailty | | | |
| Fit only[b] | 8 (6.6%) | 5 (5.8%) | 0.63 |
| Frail only | 19 (15.6%) | 18 (20.7%) | |
| Non-specific | 95 (77.9%) | 64 (73.6%) | |
| Restricted by drug tolerability | 15 (12.3%) | 15 (17.2%) | 0.32 |

[a] Included in the selected-variable-model; [b] Used as a reference group; [c] Whether it was used as a primary endpoint or not; [d] All percentages are calculated by column; [e] Chi-squared test and t-test were used for categorical and continuous variables, respectively; GA = geriatric assessment; all of the variables in this table were included in the all-variable-model.

| Table 2. 3. Area Under the Receiver Operating Characteristics (AUROC) in the Validation Set | | | |
|---|---|---|---|
| | Predictor selection strategies | | |
| | All predictors (n=37) | Selected predictors based on the clinical rationale (n=15) | Best subset (n=12)[a] |
| Machine learning model | AUROC (95% confidence interval) | | |
| LASSO | 0.58 (0.49, 0.67) | 0.65 (0.56, 0.73) | 0.66 (0.57, 0.75) |
| Elastic Net | 0.58 (0.49, 0.67) | 0.65 (0.56, 0.74) | 0.66 (0.58, 0.74) |
| Ridge | 0.56 (0.47, 0.65) | 0.65 (0.56, 0.73) | 0.66 (0.57, 0.75) |
| Random Forest | 0.63 (0.54, 0.72) | 0.65 (0.56, 0.73) | 0.63 (0.55, 0.72) |
| AdaBoost | 0.60 (0.51, 0.69) | 0.64 (0.55, 0.72) | 0.61 (0.52, 0.70) |
| GLMboost | 0.59 (0.50, 0.68) | 0.65 (0.57, 0.73) | 0.66 (0.58, 0.75) |
| BART | 0.41 (0.32, 0.49) | 0.34 (0.28, 0.42) | 0.34 (0.26, 0.42) |
| Super Learner[b] | 0.63 (0.54, 0.71) | 0.66 (0.58, 0.74) | 0.66 (0.58, 0.75) |

[a] We explored several subsets of all candidate predictors to find a model that has a higher validation AUROC than the best-performing model with the 15 selected predictors; [b] Super learner included regularization methods (LASSO, Elastic Net, and ridge), random forest, and BART; AdaBoost: Classification Trees; GLMboost: Boosted Generalized Linear Regression Learner; BART: Bayesian Additive Regression Tree.

## 2.9 Supplement

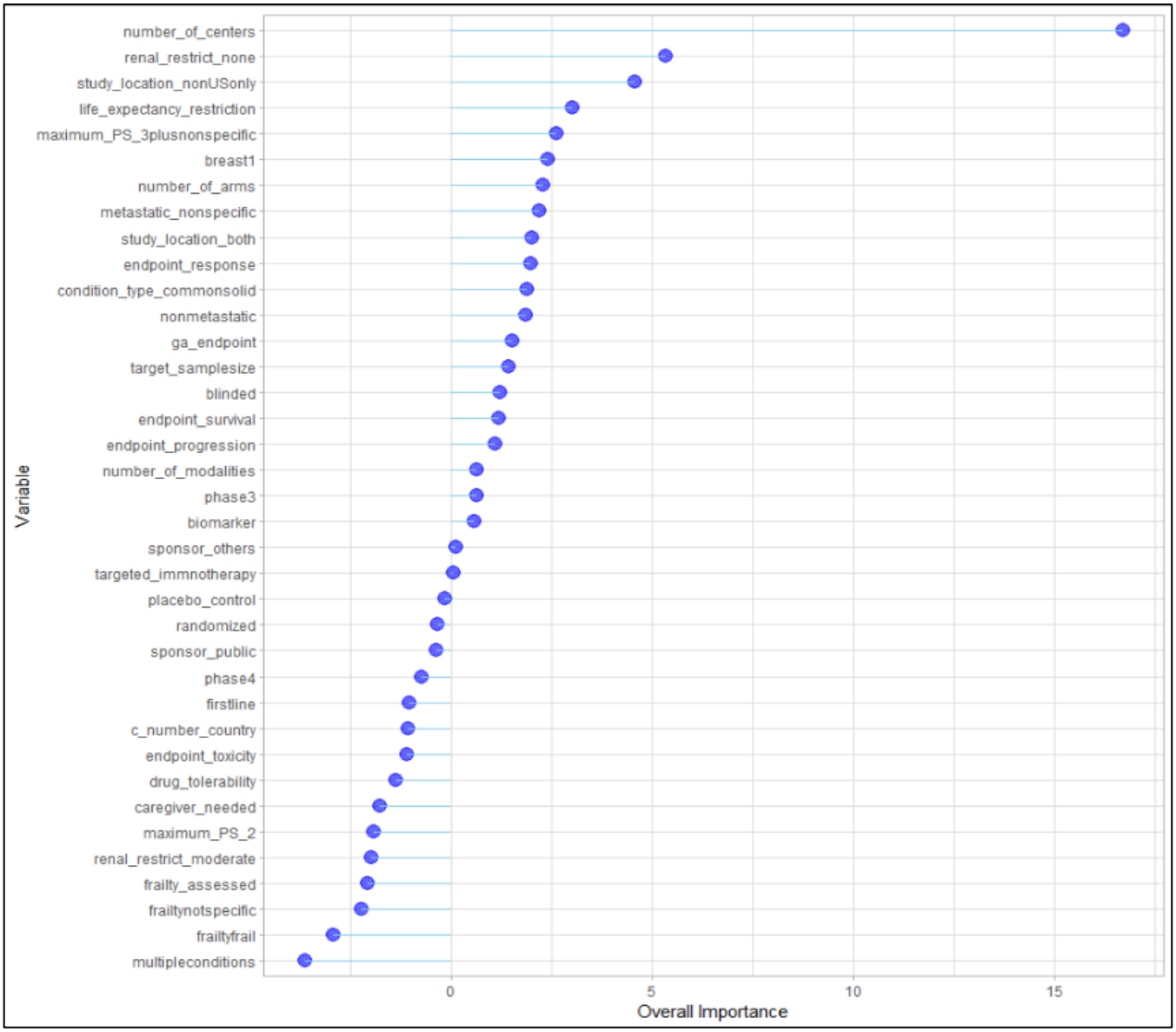


Figure S2. 1. Mean decrease accuracy of trial-level factors obtained from random forest model with all predictors (n=37)

# Chapter 3. How Does Cumulative Evidence from Older Adult-specific Trials Influence Clinical Practice? A Difference-in-Differences Analysis of Irradiation in Early-stage Breast Cancer

## 3.1 Abstract

Background: Despite increasing focus on conducting cancer clinical trials in older adults, it is unclear whether such evidence influences practice patterns. We aimed to estimate the impact of cumulative evidence from older adult-specific trial results from the CALGB 9343 and PRIME II trials that found post-lumpectomy irradiation has little benefit among older adults with early-stage breast cancer (ESBC).

Methods: Patients diagnosed with ESBC between 2000 and 2018 were identified from the SEER registry data. We examined the incremental immediate effect, incremental average yearly effect, and cumulative effect of a series of CALGB 9343 and PRIME II results on the utilization level of post-lumpectomy irradiation. We conducted difference-in-differences analyses, comparing those aged 70 or older vs. less than 65 years old.

Results: The initial 5-year CALGB 9343 results in 2004 led to a significant immediate (-0.038, 95% CI: -0.064, -0.012) and average yearly decrease (-0.008, 95% CI: -0.013, -0.003) in the probability of irradiation use among those aged 70 or older compared to those below 65 years of age. 11-year CALGB 9343 results in 2010 significantly accelerated the average yearly effect by 1.7 percentage points (95% CI: -0.030, -0.004). The other later results did not significantly

change the time trend. The cumulative effect of all results between 2004 to 2018 was -26.3 percentage points (95% CI: -0.29, -0.24).

Conclusion: Cumulative evidence from older adult-specific trials in ESBC led to decreasing use of irradiation over time among elderly patients. The rate of decrease after the initial results was accelerated by long-term follow-up results.

## 3.2 Introduction

Older adults have been underrepresented in cancer clinical trials.[14, 52] A lack of drug efficacy and safety data specific to older adults forces providers to rely on suboptimal evidence, potentially leading to low-value care. Acknowledging the growing importance of clinical evidence for older adults, several organizations, such as the US Food and Drug Administration and the Association for Clinical Oncology (ASCO), have called for more studies among older adults.[8, 9, 16, 53] Given the increasing attention to clinical trials in older adults, assessing the impact of such trials may inform and facilitate discussion around prioritizing and implementing older adult-specific trials.

Post-lumpectomy irradiation in early-stage breast cancer (ESBC) is one of the clinical areas where evidence in older adults has been generated. Clinical trials among general patients in the 1980s showed that post-lumpectomy irradiation provides the same overall outcomes as mastectomy, leading to the use of irradiation as an adjuvant therapy among ESBC patients.[54] Starting from 2004, however, the results of phase 3 clinical trials in older adults were published, demonstrating the minimal benefit of post-lumpectomy irradiation among older women with ESBC. Specifically, Cancer and Leukemia Group B (CALGB) 9343 examined the benefit of post-lumpectomy irradiation therapy in women aged 70 years or older with estrogen receptor (ER)-positive stage 1 (i.e., T1N0M0) breast cancer in the United States. Its 5-year results published in 2004 demonstrated little benefit of irradiation in recurrence and survival.[55] The longer-term results presented in ASCO 2010 meeting and published in 2013 confirmed the findings of the earlier reports.[56, 57] Subsequently, the results from another older adult-specific trial named Post-operative Radiotherapy In Minimum-risk Elderly (PRIME) II were disclosed in the 2013 San Antonio breast cancer symposium and were published in 2015.[58, 59] PRIME II

investigated the same research question with slightly broader eligibility criteria (i.e., those with ESBC that is hormone receptor-positive, T1-T2, N0, and M0 and were 65 years or older) in European countries. Investigators found consistent results as the findings from CALGB 9343, concluding that post-lumpectomy irradiation can be omitted among older adults with ESBC.

Previous studies have investigated the impact of CALGB 9343 and/or PRIME II studies on the level of post-lumpectomy irradiation use. These studies consistently found that there has been a decrease in the level of post-lumpectomy irradiation use among older women with ESBC.[60-67] However, none of these studies adjusted for time-varying covariates, making it difficult to ascertain whether the decrease in the irradiation use is attributable to the older adult-specific trial results or other time-varying factors during the study period. For example, advances in surgical techniques and therapeutic options might have lowered the level of irradiation use.[68, 69] Conversely, the improvements in the field of radiation oncology could have led to increasing irradiation use over time, which would have resulted in the underestimation of the effects of the trial results.[70] Furthermore, none of the previous studies examined the impact of additional results from the trials. More specifically, there have been a series of disclosures of the trial results over time since 2004, and previous studies only examined the aggregated effect of multiple disclosures.[55-59] Disentangling the contribution of each disclosure of trial results may help strategize how to generate and release the older adult-specific trial results.

The goal of this study was to examine the impact of each of a series of CALGB 9343 and PRIME II results on the post-lumpectomy irradiation use, controlling for time-varying unobserved variables.

## 3.3 Methods

*Study design*

We performed a difference-in-differences (DiD) analysis to estimate the change in the level of post-lumpectomy irradiation use before and after the disclosures of a series of older adult-specific clinical trial results among patients older than 70 years old. DiD analysis can facilitate causal inference by comparing the results in an exposed group to the pre-post changes in a control group that was not exposed to the intervention in question but otherwise was supposed to have the same trends in outcomes as the exposed group.[71]

The interventions of interest were the disclosures of CALGB 9343 and PRIME II results as either abstract or journal publications (Figure 1): (a) 5-year results CALGB 9343 published in New England Journal of Medicine in September 2004, (b) 11-year results of CALGB 9343 presented in ASCO in May 2010, (c) 13-year results of CALGB 9343 published in Journal of Clinical Oncology in May 2013, (d) 5-year results of PRIME II presented in San Antonio BC Symposium in December 2013, and (e) 5-year results of PRIME II study published in Lancet in January 2015. Since interventions 'c' and 'd' happened in the same year, we combined them and examined their aggregated effect (Note that the month of diagnosis was not available in the data we used).

Based on the disclosure year of each intervention, we divided the whole study period into the following five periods: pre-period (2000–2003; before intervention 'a'), period 1 (2004–2009; between 'a' and 'b'), period 2 (2010–2012; between 'b' and 'c/d'), period 3 (2013–2014; between 'c/d' and 'e'), and period 4 (2015–2018; after 'e') (Figure 1). Each patient was assigned to each of these periods based on their year of diagnosis. More rationale for the pre-and post-periods is provided in Supplement 1.

We included those aged 70 or older at the time of diagnosis in the treatment group (hereafter referred to as 'older group') since they were affected by both studies. We included those younger than 65 years old at the time of diagnosis as a control group (hereafter referred to as 'younger group') to adjust for time-varying factors (Supplement 2 for the choice of the control group). Patients between 65 and 69 were excluded from the analysis because they were partially affected by the interventions.

*Data source and patient population*

We used the Surveillance, Epidemiology, and End Results Program (SEER) Database, a nationally representative, population-based data that includes all cancer cases within specific US geographic regions (Supplement 3).[72, 73]

The study period spanned from January 2000 to December 2018, corresponding to the dates for which data for all 18 SEER registries were available at the time of study design. Consistent with the eligibility criteria of the CALGB 9343 trial, we included patients who were (1) female, (2) diagnosed with first or only stage 1 (T1N0M0 and tumor size $\leq$ 2 cm) ER-positive breast cancer, (3) diagnosed between 2000 and 2018, (4) underwent a lumpectomy, and (5) aged 70 years or older at the time of diagnosis in the older group.[55] Those who satisfied inclusion criteria 1-4 and were younger than 65 years old were included in the younger group as a control.

*Statistical Analysis*

The association between a series of trial results and the level of post-lumpectomy irradiation use was examined using a DiD analysis. The outcome variable was a binary indicator of whether a patient received irradiation after lumpectomy during the first course of therapy,

which was ascertained using surgery/irradiation sequence records (Supplement 4). We estimated a linear model with the outcome variable instead of a logistic model because linear combination of parameters, which we used to estimate measurements of interest, is more feasible with a linear model. We used a sensitivity analysis to compare linear vs. logistic models.[74, 75] As independent variables, we included a binary variable of whether a patient is in the older vs. younger groups (hereafter referred to as 'group indicator'), year of diagnosis dummy variables with the last year of the pre-period, 2003, as a reference, and interaction terms of the group indicator and year of diagnosis dummy variables (Supplement 5). The model was adjusted for patient characteristics available in the dataset, including age at diagnosis, race, ethnicity, and marital status.

Parallel trend assumption (i.e., assumption that outcomes in the intervention group would have continued along the same trajectories as the control group in the absence of an intervention) was tested by examining whether the DiD estimates during the pre-period (2000–2003) are jointly equal to zero, using a joint F test.[76] All tests of statistical significance were 2-sided with a significance level of 5%.

*Measurements*

We estimated the incremental immediate effect and incremental yearly average effect of each intervention. The incremental immediate effect of each intervention was defined as the immediate drop in the probability of post-lumpectomy irradiation use between periods, incremental to the average yearly change (i.e., slope) in the previous period. We defined the incremental average yearly effect as the average yearly change in the probability of the post-

lumpectomy irradiation use in a given period, incremental to the average yearly change in the previous period.

We also estimated the cumulative effect of a series of interventions. Specifically, we estimated the total amount of decrement in the probability of post-lumpectomy irradiation use in response to intervention(s) ‘a’, ‘a’ and ‘b’, ‘a’ through ‘d’ and ‘a’ through ‘e’ (Supplement 6 for the details of the measurements).

*Sensitivity analyses*

Since the outcome of interest was binary, we used a logistic regression model to examine how DiD estimates differ from when linear regression was used (Supplement 7 for the details). We compared the cumulative effect of the disclosures with the base-case results. The incremental effects of each disclosure were not estimated because of the challenges of linearly combining the odds ratios.

## 3.4 Results

*Patient characteristics*

Our study sample consisted of 206,564 patients, of which 36.4% (n=75,091) were aged 70 years old or older (Table 1; online only). Seventy-one percent and 61% of the older group received post-lumpectomy irradiation during the pre- (2000-2003) and post-period (2004-2018), respectively. The mean age was 76.9 (standard deviation (SD) = 5.4) and 53.9 (SD = 7.3) years old in the older and the younger groups, respectively. The majority of patients were non-Hispanic white in both groups.

*Unadjusted temporal trends of the post-lumpectomy irradiation*

Unadjusted trends of the post-lumpectomy irradiation use showed that the fraction of patients receiving irradiation decreased over time in the older group after initial publication of older adult-specific trial results in 2004 (Figure 2). On the other hand, there was a slightly increasing trend over time in the younger group.

*Parallel trend assumption*

The DiD estimates during the pre-period (2000-2003) were jointly equal to zero ($F = 0.99$, $p = 0.40$), suggesting that the pre-period trend was not significantly different between the older and younger groups.

*Incremental immediate effect and incremental average yearly effect of each intervention*

The probability of the post-lumpectomy irradiation use significantly dropped by 3.8 percentage points (95% CI: -0.064, -0.012) in 2004 in the older group, compared to the younger group, showing a significant immediate effect of intervention 'a' (Table 2). The additional interventions 'b', 'c/d', and 'e' did not have a significant incremental immediate effect. In other words, the decrement of the probability of post-lumpectomy irradiation use between periods was not significantly different from the average yearly decrement in the previous period.

Intervention 'a' had a significant incremental average yearly effect of -0.008 (95% CI: -0.013, -0.003), meaning that the probability of post-lumpectomy irradiation use decreased by 0.8 percentage points per year on average during period 1, compared to the pre-period. The intervention 'b' was associated with a significant change in the average yearly effect by -0.017 (95% CI: -0.030, -0.004) during period 2 as compared to period 1, significantly accelerating the rate of decrease. The later interventions 'c/d' and 'e' did not have a significant incremental average yearly effect, indicating that the average yearly effect during periods 3 and 4 stayed constant as that of period 2.

*The cumulative effect of a series of interventions*

The probability of post-lumpectomy irradiation use among older adults lowered by 8.0 percentage points (95% CI: -0.104, -0.055) at the end of period 1 (Figure 3, Table 2). At the end of period 2, the cumulative effect 'a' and 'b' reached -15.4 percentage points (95% CI: -0.178, -0.130). and it became -18.7 percentage points (95% CI: -0.021, -0.016) with the additional interventions 'c' and 'd' in 2013.The cumulative effect of all intervention was - 26.3 percentage points (95% CI: -0.287, -0.239) over the whole post periods.

*Sensitivity analysis*

The cumulative effect was smaller with a logistic regression by approximately 2 percentage points compared to a linear regression. It was -0.060 (95% CI: -0.084, -0.038), -0.131 (95% CI: -0.153, -0.108), -0.164 (95% CI: -0.188, -0.143), and -0.243 (95% CI: -0.265, -0.221) for periods 1, 2, 3, and 4, respectively (Supplement 7). However, the confidence interval of these estimates overlapped between the two models.

## 3.5 Discussion

We examined the incremental and cumulative effect of a series of the results of CALGB 9343 and PRIME II, phase 3 older adult-specific trials that found post-lumpectomy irradiation use in ESBC patients aged 70 or older has little benefit. The initial 5-year follow-up results of CALGB 9343 were associated with a significant immediate and yearly decrease in the probability of irradiation use, and a follow-up disclosure of longer-term CALGB 9343 results further accelerated the rate of decrease. The cumulative evidence from the trials led to a significant decrease in the probability of post-lumpectomy irradiation use by approximately 25 percentage points between 2004 and 2018.

Our study builds on a body of literature on the impact of CALGB 9343 and/or PRIME II and ascertains the impact of the older adult-specific trials over time, accounting for time-varying factors. Earlier studies compared the level of irradiation omission before and after the results of these trials in older patients, showing a short-term and longer-term decrease after the disclosure of the trial results.[60-67] None of these studies, however, accounted for a counterfactual scenario–what the level of post-lumpectomy irradiation use would have been without the CALGB 9343 and/or PRIME II results–in the analysis. Considering that many factors are changing over time, potentially affecting the irradiation use in breast cancer care, ignoring those factors may yield misleading results either by underestimating or exaggerating the actual impact of the trial results. In this study, we observed a slightly increasing trend of irradiation use in younger patients, suggesting that estimating the impact of the trials based on trends only in older adults might have led to an underestimation of their true effect. In this study, the increasing trend among younger adults was not prominent, and thus it did not change the conclusion drawn from

the previous studies. However, we believe continued efforts are needed in future time-trend research to appropriately account for the time trend due to other factors.

The study finding also demonstrated the importance of evidence accumulation in prescription patterns, which was not explicitly studied in the previous studies. Specifically, we found that the additional longer-term CALGB 9343 results significantly accelerated the average yearly decrease in the post-lumpectomy irradiation use. Also, this accelerated rate stayed constant throughout the rest of the post-period when additional longer-term CALGB 9343 results and initial PRIME II results were released. Assuming the level of providers' behavioral changes would reach a plateau after a certain period, the rate of change might have been slowed if there were no additional results. These findings suggest that the release of additional results added more certainty to the trial findings, making providers feel more comfortable changing the prescribing patterns. A high level of certainty may be important especially when providers need to change the prescribing patterns that have been maintained for a long time or when trials suggest omitting existing treatments that have minimal benefits (as opposed to adding a new treatment with demonstrated benefits). Our finding suggests that one way to incorporate treatment heterogeneity in older adults into the clinical practice could be a stepwise approach where evidence is accumulated and assured by additional findings. Longer-term follow-up of the same patients or a trial with the same question but in different patient populations would help the affirmation process. One thing to note, however, is that there may be a diminishing return to the additional trial results, and thus it is important to know the optimal level of investment in generating older adult-specific evidence.

## 3.6 Limitations

Our study has several limitations. First, the findings from our study may be generalized to other disease or treatment areas.

Moreover, studies have reported under-ascertainment of radiation therapy in the SEER data, with the sensitivity and positive predicted value of approximately 82% and 98%, respectively for ESBC.[77, 78] Although the fraction of patients who received irradiation might have been underestimated, the aim of this study was to estimate the temporal change rather than the absolute value of the utilization level. Also, we included younger patients who were likely to have a similar under-ascertainment issue as a control group. Therefore, our estimates were unlikely be affected by the under-ascertainment issue.

Furthermore, it was unclear how much of the patients diagnosed in 2004 were exposed to the initial disclosure. Although our assumption that those diagnosed in 2004 were likely to be affected by the initial disclosure was based on the existing studies (it often takes a couple of months for a patient to receive irradiation followed by lumpectomy), we conducted a scenario analysis to address the uncertainty (Supplement 8).[60, 79-81]

Also, despite the attempt to adjust for time-varying factors, there may still be remaining unadjusted confounders. Furthermore, studies published in this clinical area and advances in breast cancer treatment might not have had an equal impact on the irradiation use in the younger and older age groups, potentially preventing estimating the appropriate counterfactual and biasing our estimates in either direction. However, we expect that the disproportionate impact of some time-varying factors on the clinical practice would be smaller than the impact of phase 3 clinical trial results and the guideline changes in response to the results. [82]

## 3.7 Conclusion

We assessed the effect of a series of results from CALGB 9343 and PRIME II, phase 3 trials that showed post-lumpectomy irradiation has minimal benefit among older adults with ESBC. Our study findings suggest that new evidence among older adults can potentially change the providers' prescription patterns, and rate of change can be accelerated if trial results are accumulated. Additional well-controlled studies, especially for positive trials (i.e., trials suggesting adding a new treatment among older adults), may help understand the potential impact of older adult-specific trials more comprehensively.

## 3.8 Tables & Figures

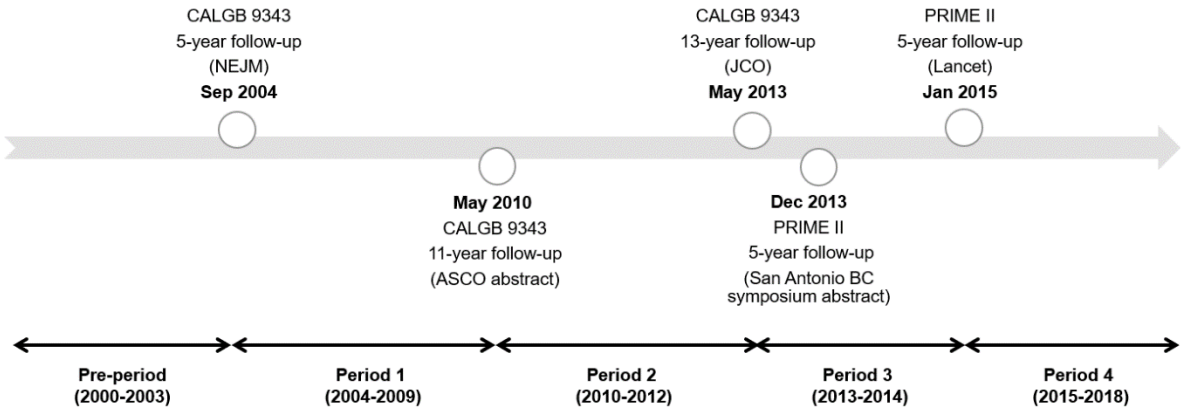


Figure 3. 1. Timeline for the disclosures of CALGB 9343 and PRIME II results
CALGB = Cancer and Leukemia Group B; PRIME = Post-operative Radiotherapy In Minimum-risk Elderly; NEJM = New England Journal of Medicine; ASCO = American Society of Clinical Oncology; JCO = Journal of Clinical Oncology; BC = Breast Cancer. Note: Although shorter-term CALGB 9343 results were released in 2001 as an ASCO abstract, we did not include it as an intervention of interest because the median follow-up was only two years, and thus the effect of post-lumpectomy on survival was still controversial at this point. Our decision was also based on all other existing studies examining the changes in the post-lumpectomy irradiation use after the CALGB 9343 study that consider evidence disclosed in 2004 as the first seminal piece of evidence in older adults.

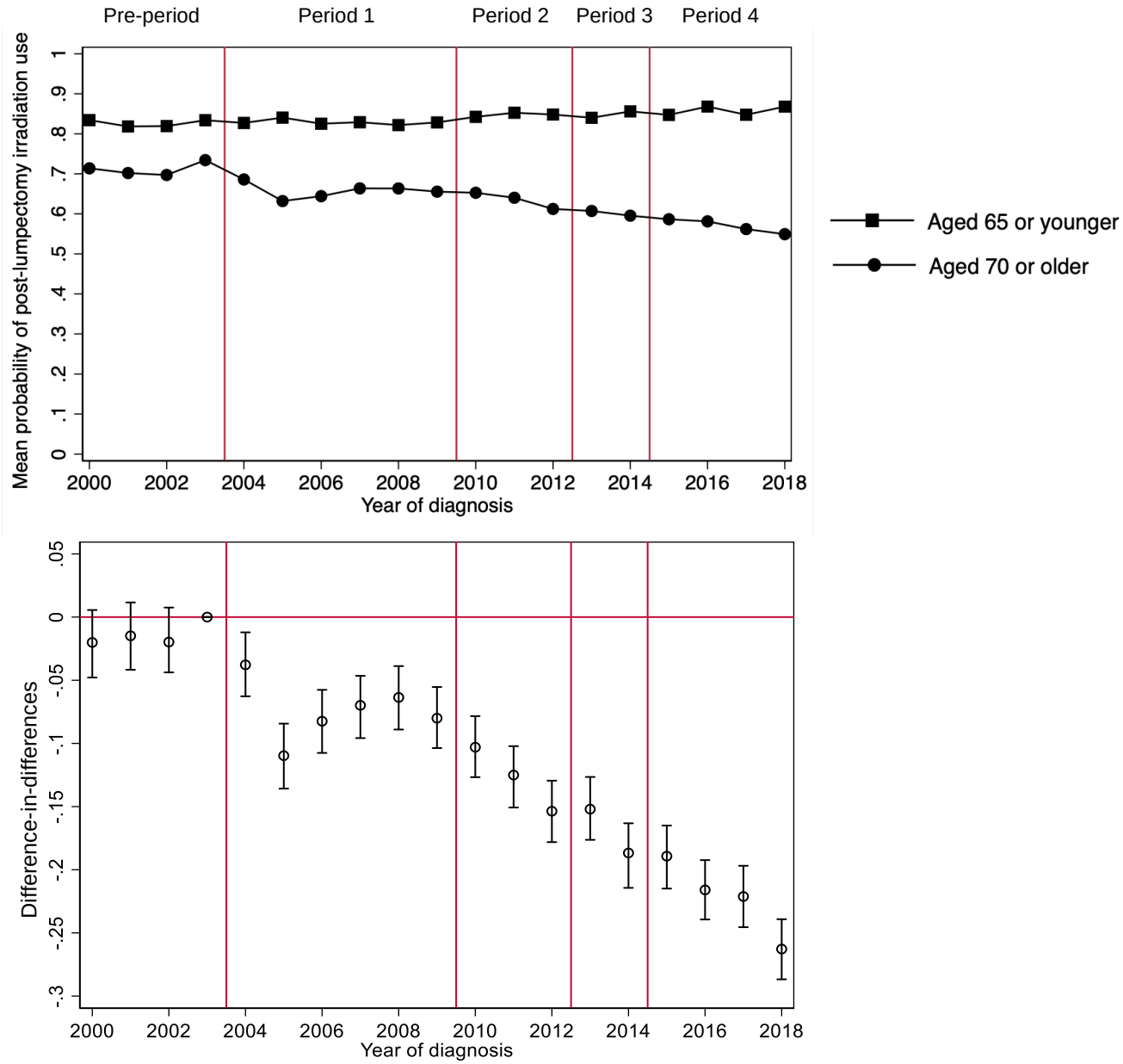


Figure 3. 2. The mean probability of post-lumpectomy irradiation use by year of diagnosis (top) and difference-in-differences estimates of the impact of a series of disclosures of CALGB 9343 and PRIME II results on the probability of post-lumpectomy irradiation use (bottom)

The vertical lines correspond to the time when older adults with early-stage breast cancer started to be affected by the following disclosures: 'a' (5-year follow-up results of CALGB disclosed in 2004), 'b' (11-year follow-up results of CALGB disclosed in 2010), 'c/d' (13-year follow-up

results of CALGB and 5-year follow-up results of PRIME disclosed in 2013), and ‘e’ (5-year follow-up results of PRIME disclosed in 2015).

| Table 3. 1. Characteristics of study participants (n = 206,564) | | | |
|---|---|---|---|
| | Older group (≥70 y/o) (n = 75,091) | Younger group (<65 y/o) (n = 131,473) | p-value |
| | Number (percentage)[a] | Number (percentage)[a] | |
| The probability of receiving of post-lumpectomy irradiation[a] | | | |
| During the pre-period (2000-2003) | 0.71 (0.45) | 0.82 (0.38) | <.001 |
| During the post-period (2004-2018) | 0.61 (0.49) | 0.85 (0.36) | <.001 |
| Age at diagnosis[a] | 76.90 (5.44) | 53.90 (7.30) | <.001 |
| Race/ethnicity | | | <.001 |
| Hispanic | 4,733 (6.30%) | 13,285 (10.10%) | |
| Non-Hispanic white | 61,649 (82.10%) | 95,136 (72.36%) | |
| Non-Hispanic black | 4,306 (5.73%) | 10,239 (7.79%) | |
| Non-Hispanic others | 4,403 (5.86%) | 12,813 (9.75%) | |
| Marital status | | | <.001 |
| Married | 34,339 (45.73%) | 85,395 (64.95%) | |
| Divorced | 6,454 (8.59%) | 14,635 (11.13%) | |
| Separated | 317 (0.42%) | 1,369 (1.04%) | |
| Never married | 5,348 (7.12%) | 19,481 (14.82%) | |
| Unmarried or domestic partner | 72 (0.10%) | 302 (0.23%) | |
| Widowed | 25,310 (33.71%) | 5,158 (3.92%) | |
| Unknown | 3,251 (4.33%) | 5,133 (3.90%) | |
| Year of diagnosis | | | <.001 |
| 2000 | 2,944 (3.92%) | 5,279 (4.02%) | |
| 2001 | 3,151 (4.20%) | 5,650 (4.30%) | |
| 2002 | 3,194 (4.25%) | 5,757 (4.38%) | |
| 2003 | 3,025 (4.03%) | 5,618 (4.27%) | |
| 2004 | 3,326 (4.43%) | 6,062 (4.61%) | |
| 2005 | 3,416 (4.55%) | 6,537 (4.97%) | |
| 2006 | 3,503 (4.67%) | 6,588 (5.01%) | |
| 2007 | 3,500 (4.66%) | 6,815 (5.18%) | |
| 2008 | 3,663 (4.88%) | 6,880 (5.23%) | |
| 2009 | 3,831 (5.10%) | 7,059 (5.37%) | |
| 2010 | 3,768 (5.02%) | 6,851 (5.21%) | |
| 2011 | 4,095 (5.45%) | 7,204 (5.48%) | |
| 2012 | 4,149 (5.53%) | 7,301 (5.55%) | |
| 2013 | 4,358 (5.80%) | 7,392 (5.62%) | |
| 2014 | 4,547 (6.06%) | 7,666 (5.83%) | |
| 2015 | 4,757 (6.33%) | 8.149 (6.20%) | |
| 2016 | 5,133 (6.81%) | 8,258 (6.27%) | |

| | | | |
|---|---|---|---|
| 2017 | 5,201 (6.93%) | 8,289 (6.30%) | |
| 2018 | 5,500 (7.39%) | 8,128 (6.18%) | |

[a] Mean and standard deviation was reported for the probability of receiving of post-lumpectomy irradiation and age at diagnosis; y/o: years old; We calculated p-values using a chi-square test (for categorical variables) or t-test (for continuous variables).

| Table 3. 2. The incremental and cumulative effect of disclosures of CALGB 9343 and PRIME II results | | | |
|---|---|---|---|
| Period (corresponding years)[a] | Incremental immediate effect between periods | Incremental average yearly effect over the period | Cumulative effect[b] |
| | Point estimate (95% CI) | | |
| Period 1 (2004-2009) | -0.038* (-0.064, -0.012) | -0.008* (-0.013, -0.003) | -0.080* (-0.104, -0.055) |
| Period 2 (2010-2012) | -0.015 (-0.042, 0.013) | -0.017* (-0.030, -0.004) | -0.154* (-0.178, -0.130) |
| Period 3 (2013-2014) | 0.027 (-0.004, 0.058) | -0.009 (-0.036, 0.017) | -0.187* (-0.021, -0.016) |
| Period 4 (2015-2018) | 0.032 (-0.008, 0.072) | 0.010 (-0.014, 0.035) | -0.263* (-0.287, -0.239) |
| Periods 2-4 (2010-2018)[c] | N/A | -0.012* (-0.017, -0.006) | N/A |
| Periods 1-4 (2004-2018)[d] | N/A | -0.016* (-0.018, -0.014) | N/A |

[a] Periods 1, 2, 3, and 4 correspond to the disclosures 'a' (5-year follow-up results of CALGB), 'b' (11-year follow-up results of CALGB), 'c/d' (13-year follow-up results of CALGB and 5-year follow-up results of PRIME), and 'e' (5-year follow-up results of PRIME), respectively; CI: confidence interval; [b] The cumulative effect in a certain period includes that of the previous periods; [c] The incremental yearly effect of later disclosures ('b' through 'e') compared to the initial disclosure ('a') was collectively examined; [d] The incremental yearly effect of all disclosures ('a' through 'e') compared to the pre-period was collectively examined. * indicates that the estimate was significantly different from zero at a significance level of 5%.

## 3.9 Supplement

*Supplement 1: The rationale for the choice of the control group*

We thought that younger patients with the same disease would be an appropriate control group because they were likely to be affected by all changes made in breast cancer care similarly to older adults. Also, the overall consensus that post-lumpectomy irradiation is recommended among younger adults has not changed during the study period–while the guideline changed among older adults soon after the 5-year follow-up CALGB 9343 results, the recommendation stayed unchanged among those younger regarding whether to use the post-lumpectomy irradiation.[1] Given that clinical practice is mainly affected by the clinical guidelines, we assumed that there was no significant event specific to younger patients during the study period that substantially affected the standard practice only in them.[2]

*Supplement 2: The rationale for the decision of pre-and post-periods*

In the base-case, we included 2000-2003 in the pre-period and 2004-2018 in the post-period. Although the initial results of CALGB 9343 were not published until September 2004, we assumed that patients diagnosed earlier than September 2004 would have been affected by the study results because it often takes a couple of months for a patient to receive irradiation followed by lumpectomy.[3-6] In other words, for those diagnosed early in 2004, decisions on adjuvant treatments after lumpectomy were likely to be made later in 2004 when the CALGB results were published. In the scenario analysis, we tested an alternative scenario where patients diagnosed in 2004 were not affected by older adult-specific trial results and thus included in the pre-period.

*Supplement 3: Details about the SEER registry data*

SEER has collected cancer incidence data from population-based cancer registries from 18 states covering approximately 34.6 percent of the US population. Trained registrars collect data from all clinical settings that diagnose or treat cancer. Included information are patient demographics, primary tumor site, tumor morphology, tumor properties, stage at diagnosis, and first course of treatment, and vital status. Stage and histologic details are reported for all cancers, allowing specific subpopulations to be studied. SEER data also have the data for the first course of treatment, including surgery, irradiation, or chemotherapy.

*Supplement 4: The ascertainment of the outcome variable*

The receipt of irradiation after lumpectomy was ascertained using surgery/irradiation sequence records. These records show the order in which surgery (i.e., lumpectomy) and radiation therapies were administered for those patients who had surgery and/or irradiation. Among those who meet the eligibility criteria described above, patients with 'radiation after surgery' were assigned 1, and those with 'no radiation' were 0. Patients who received irradiation before or during lumpectomy or had unknown information about the receipt or sequence (i.e., radiation before surgery, radiation both before and after surgery, intraoperative radiation, surgery both before and after radiation, and sequence unknown) were excluded from the analysis.

*Supplement 5: The rationale for the model specification*

By having the interaction term of group indicator and year of diagnosis dummy, the model specification we chose allows separately estimating the DiD for each year (DiD for a given year is equal to the coefficient of the interaction term). In other words, it provides a non-

parametric way of taking the difference in the changes over time between the older and younger group for each year, enabling us to estimate the incremental effect of each disclosure of trial results easily.

*Supplement 6. Estimation of the measurements*

1) Incremental immediate effect

   We first took the difference in the DiD estimates between the first year of a given post-period and the previous year to estimate the total immediate change in the level of post-lumpectomy irradiation use between periods. Then we subtracted the average yearly change in the previous period to estimate the incremental immediate effect, assuming that the immediate effect of an intervention would be the same as the average yearly change during the previous period without a significant incremental immediate effect of additional intervention.

2) Incremental average yearly effect

   We took the difference in the DiD estimates for the last and first years of a given period and divided the difference by the number of years in the period to calculate the yearly average effect of the interventions that were in effect during that period. Then, we subtracted the yearly average effect during the previous period to estimate the incremental yearly average effect of additional intervention.

   In addition to the incremental average yearly effect of each intervention, we estimated the aggregated incremental yearly effect of the later interventions (‘b’ through ‘e’) to examine whether the additional interventions collectively accelerated the average yearly effect of the first intervention. We also estimated the average yearly effect of all

interventions (‘a’ through ‘e’) to examine the yearly change in the level of irradiation use over a whole post-period.

3) Cumulative effect

   We estimated the cumulative effect of a series of interventions as DiD for the last year of each post-period: DiD for 2009, 2012, 2014, and 2018 were estimated for a cumulative effect of intervention(s) ‘a’, ‘a’ and ‘b’, ‘a’ through ‘d’ and ‘a’ through ‘e’, respectively. Note that ‘c’ and ‘d’ were combined because they happened in the same year.

*Supplement 7. Sensitivity analysis*

How we estimated difference-in-differences (DiD) in the logistic regression model: Because the interpretation of DiD estimates (i.e., the coefficients of the interaction terms) is not straightforward in a logistic model, we used a recycling prediction to convert the DiD estimated in odds ratio to the absolute difference. Then, we estimated the 95% CI of the DiD estimates for each year by running 1,000 times of bootstrapping.

Results:

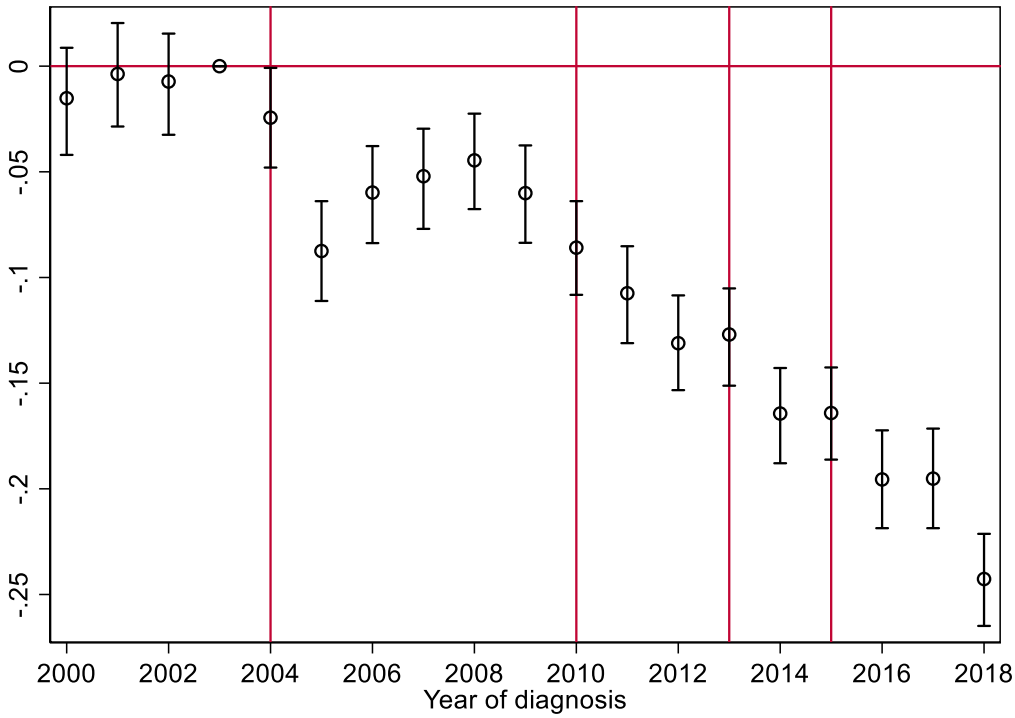


Figure S3. 1. Difference-in-differences estimates of the impact of a series of disclosures of CALGB 9343 and PRIME II results on the probability of post-lumpectomy irradiation use using a logistic regression

The red vertical lines correspond to the time when older adults with early-stage breast cancer started to be affected by the following disclosures: ‘a’ (5-year follow-up results of CALGB disclosed in 2004), ‘b’ (11-year follow-up results of CALGB disclosed in 2010), ‘c/d’ (13-year follow-up results of CALGB and 5-year follow-up results of PRIME disclosed in 2013), and ‘e’ (5-year follow-up results of PRIME disclosed in 2015).

*Supplement 8. Scenario analysis under an assumption that patients diagnosed in 2004 were not affected by the trial results.*

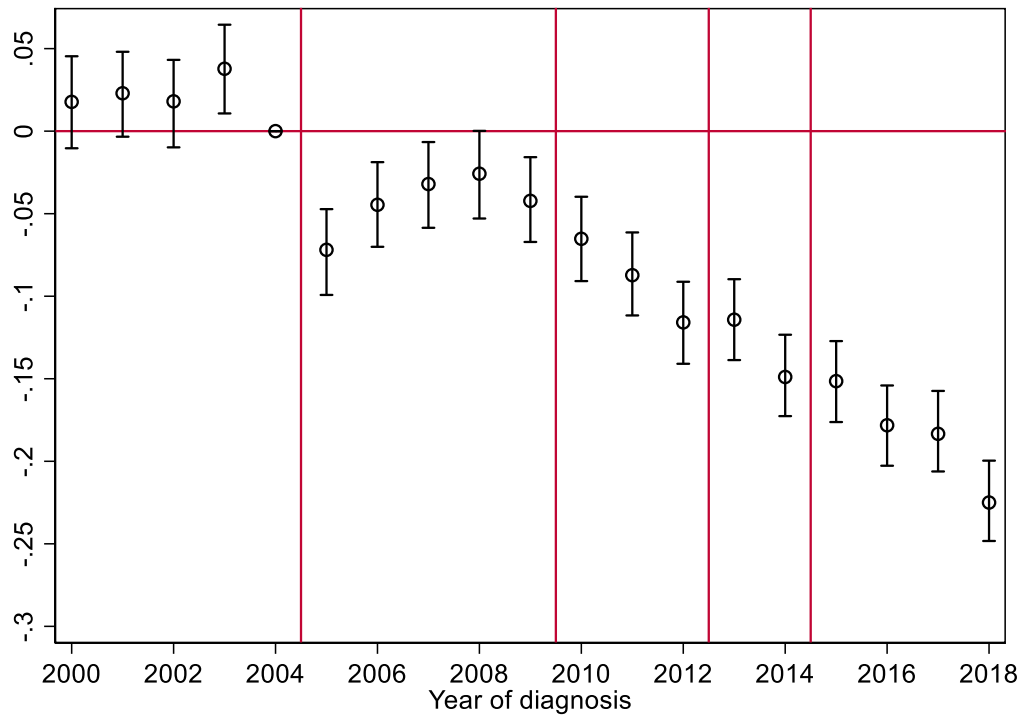


Figure S3. 2. Difference-in-differences estimates of the impact of a series of disclosures of CALGB 9343 and PRIME II results on the probability of post-lumpectomy irradiation use including 2004 in the pre-period

The vertical lines are correspondent to the time when older adults with early-stage breast cancer started to be affected by the following disclosures, respectively: 'a' (5-year follow-up results of CALGB disclosed in 2004), 'b' (11-year follow-up results of CALGB disclosed in 2010), 'c&d' (13-year follow-up results of CALGB and 5-year follow-up results of PRIME disclosed in 2013), and 'e' (5-year follow-up results of PRIME disclosed in 2015).

Table S3. 1. The cumulative, incremental immediate, incremental average yearly effect of disclosures of CALGB 9343 and PRIME II results including 2004 in the pre-period

| Period (corresponding years)[a] | Incremental immediate effect between periods | Incremental average yearly effect over the period | Cumulative effect[b] |
|---|---|---|---|
| | Point estimate (95% CI) | | |
| Period 1 (2005-2009) | -0.071* (-0.098, -0.046) | 0.007* (0.001, 0.014) | -0.042* (-0.067, -0.016) |
| Period 2 (2010-2012) | -0.030* (-0.058, -0.002) | -0.033* (-0.047, -0.019) | -0.116* (-0.141, -0.091) |
| Period 3 (2013-2014) | 0.027 (-0.004, 0.058) | -0.009 (-0.036, 0.017) | -0.149* (-0.173, -0.123) |
| Period 4 (2015-2018) | 0.032 (-0.008, 0.072) | -0.010 (-0.014, 0.035) | -0.225* (-0.248, -0.220) |
| Periods 2-4 (2010-2018) | N/A | -0.027* (-0.034, -0.021) | N/A |
| Periods 1-4 (2010-2018) | N/A | -0.012* (-0.014, -0.010) | N/A |

[a] Periods 1, 2, 3, and 4 correspond to the disclosures 'a' (5-year follow-up results of CALGB disclosed in 2004), 'b' (11-year follow-up results of CALGB disclosed in 2010), 'c&d' (13-year follow-up results of CALGB and 5-year follow-up results of PRIME disclosed in 2013), and 'e' (5-year follow-up results of PRIME disclosed in 2015); [b] The cumulative effect in a certain period includes that of the previous periods; * indicates that the estimate was significantly different from zero at a significance level of 5%.

# Chapter 4. The Real-world Value of Older Adult-specific Clinical Trials: Post-lumpectomy Irradiation among Older Adults with Early-stage Breast Cancer

## 4.1 Abstract

Objectives: Although there is increasing interest in conducting cancer clinical trials in older adults, the benefit of such trials is unclear. We aimed to quantify the real-world clinical and economic effects of two phase 3 trials (CALGB 9343 and PRIME II) that found little clinical benefit of post-lumpectomy radiation therapy (RT) among older women with early-stage breast cancer (ESBC).

Methods: We developed a health-transition model to quantify the incremental clinical and economic outcomes between scenarios with vs. without older adult-specific trial results from a societal perspective between 2004 and 2018. We derived model effectiveness inputs from CALGB 9343. The total number of affected patient population in the US and the change in the probability of omitting post-lumpectomy RT due to the trial results were derived from a retrospective analysis of the SEER registry data for patients with ESBC. Sensitivity analyses were conducted to calculate the 95% credible interval (CR) of the incremental clinical and economic outcomes between the two scenarios.

Results: Between 2004 and 2018, the trial results led to 32,936 (95% CR: 31,442, 34,365) fewer patients receiving post-lumpectomy RT among those aged 70 years or older diagnosed with ESBC in the US and a decrease in the cost of $419M USD (95% CR: -$239M, -$648M). The difference in projected life years (1,083 years, 95% CI: -2,677, 7,463) and QALYs (683 years, 95% CI: -2,797, 7,199) were not significant. At a willingness-to-pay threshold of $100k/QALY,

the probability of older adult-specific trial results generating a positive net monetary benefit was 97.5%.

Conclusion: The CALGB 9343 and PRIME II trial results resulted in a substantial cost-saving in the US society. Our results suggest that older adult-specific clinical trials that demonstrate the negligible benefit of intervention in older people could result in a significant monetary benefit. Further case studies are needed for different types of older adult-specific trials to understand the value of older adult-specific trials comprehensively.

## 4.2 Introduction

Older adults have long been under-represented in cancer clinical trials.[14] The lack of clinical evidence generated in older adults has led to a situation where clinical decision-makers need to rely on suboptimal evidence, such as results from the younger patient population or subgroup analysis. The percentage of older adults has been increasing in the past decade and is expected to rise in the coming years.[1] This demographic shift may signal the need for change in how clinical evidence is being generated in clinical trials and the importance of examining potential treatment heterogeneity among older people.

The US Food and Drug Administration made several recommendations in its recent guidance to encourage the generation of evidence for older adults.[16] The Association for Clinical Oncology and the International Society for Geriatric Oncology also recently have called for increased efforts to conduct more studies in older adults, highlighting that there is limited evidence on how to treat older patients with cancer.[8, 9] Based on these recommendations, one way to generate better clinical evidence for older adults could be to conduct additional clinical trials for older adults, especially in areas where existing trials are mainly focused on younger people.

Although older adult-specific trials can lead to important findings on the treatment heterogeneity between older vs. younger patients, a relatively small number of trials dedicated to older adults have been conducted. [13, 55, 57, 59, 83, 84] Quantifying the value of older adult-specific trials could facilitate discussions of trial sponsors on the prioritization and the investment in older adult-specific trials. One approach to quantifying the value of a clinical trial is to assess the real-world value of a clinical trial, defined as the clinical and economic benefits generated in the real world by the availability of older adult-specific information.[85]

One of the examples where older adult-specific clinical trial was conducted is post-lumpectomy radiation therapy (RT) in early-stage breast cancer (ESBC). Post-lumpectomy RT had been a standard of care for those diagnosed with ESBC since 1980s.[54] However, the results of Cancer and Leukemia Group B (CALGB) 9343, which examined the benefit of post-lumpectomy RT in women aged 70 years or older with ESBC in the United States (US), were published 2004, showing that post-lumpectomy RT has minimal clinical benefit among older adults.[55] The longer-term results were published in 2010 and 2013 with the same conclusion.[57] Subsequently, the Post-operative Radiotherapy in Minimum-risk Elderly (PRIME) II was conducted among those older than 65 years old with ESBC in European countries and showed results consistent with the CALGB 9343.[59] Given the minimal benefits of adjuvant RT, both trials concluded that post-lumpectomy RT could be omitted among older adults with ESBC.

The aim of this was to assess the real-world clinical and economic value of older adult-specific clinical trials using trials of post-lumpectomy RT in ESBC as a case study.

## 4.3 Methods

*General framework*

We calculated the real-world value of CALGB 9343 and PRIME II results as the incremental clinical and economic outcomes among the affected patient population (i.e., those who are eligible for post-lumpectomy RT omission) between scenarios with vs. without the results in 2004-2018 (Figure 1). Hypothetically, CALGB 9343 and PRIME II results change the level of post-lumpectomy RT use in older adults, leading to a difference in clinical and economic outcomes among the affected patient population between the two scenarios.

Let $N$ denote the total number of the affected patient population, $p$ and $q$ the probabilities of omitting post-lumpectomy RT with and without the trial results, $Y_{noRT}$ and $Y_{RT}$ the lifetime per-person clinical and economic outcomes with and without omitting RT, and $i$ calendar years. Then, the incremental clinical and economic outcomes among the affected patient population in the scenarios with and without the trial results during the study period can be calculated as follows (Figure 1):

$$\sum_{i=2004}^{2018} N_i \times (p_i - q_i) \times (Y_{noRT_i} - Y_{RT_i})$$

*Model structure*

We used a hybrid decision tree and Markov model to quantify the real-world value of older adult-specific trial results. In the decision tree, women eligible for omitting irradiation were divided into two branches: using and omitting RT. At the end of the decision tree, women entered the Markov model and were tracked for lifetime clinical and economic outcomes. The Markov model included the following five health states: No recurrence, loco-regional recurrence, remission, distant metastasis, and death (Figure 2). The model was based on a patient of age 75

years old. The clinical and economic outcomes were estimated from a societal perspective and were discounted at 3% per annum.

*Advancement in radiation therapy over time*

With the advancement in RT, different types of RT have been used during the study period. While different types of RT are not significantly different in terms of efficacy, they require a different number of visits and has different adverse event (AE) profiles and costs.[86-92] Therefore, we varied the utilization level of different types of RT by calendar year. We stratified RT by the number of fractions (e.g., whole-breast irradiation (WBI) and hypo-fractionated irradiation (hypoWBI)) and by means of delivery (e.g., three-dimensional conformal radiotherapy (3D-CRT) and intensity-modulated radiotherapy (IMRT)). For simplicity, we included the two most commonly used numbers of fractions and means of delivery, and thus included a total of four different types of RT: WBI-3D CRT, WBI-IMRT, hypoWBI-3D CRT, and hypoWBI-IMRT. The temporal utilization level of each type of RT was derived from existing studies that examined the trend of hypo-WRI and IMRT use among older adults with ESBC in the US and a linear extrapolation.[93, 94]

*Key model assumptions*

We made following assumptions in our model: (1) the risk of loco-regional recurrence and distant metastasis observed in the CALGB 9343 10-year follow-up results would stay constant beyond the trial period in the base case scenario, (2) the risk of distant metastasis is higher in the remission state than in the no-recurrence state by approximately 2.5 times[95], (3) the probability of receiving salvage mastectomy and lumpectomy is approximately the same in

the loco-regional state, (4) the proportion of patients with local (stage 1-2) and regional recurrence (stage 3) among those who experienced loco-regional recurrence was 79% and 21%, respectively[57], and (5) those receiving WBI visited the facility five days a week for six weeks, and those receiving hypoWBI made five visits a week for three weeks.[57, 87]

*Clinical inputs*

We estimated the annual transition probabilities from no-recurrence to loco-regional recurrence and to distant metastasis states without RT, using the 10-year results of CALGB 9343 in the no RT group, and used the hazard ratios (HR) for the RT group (Table 1).[57] The probability of developing distant metastasis in the remission state was estimated by multiplying a factor of 2.5 to that in the no-recurrence state based on our assumption.[95] The probability of death from the distant metastasis was derived from a study estimated the mortality among those who fulfilled CALGB 9343 criteria and had metastatic breast cancer.[96] We assumed that the risk of death from the no recurrence, loco-regional recurrence, and remission states follow the US general mortality and thus used the 2019 Social Security Life Table.[97] We estimated BC-non-specific mortality by subtracting BC-specific mortality.[98]

For AE due to RT, we reviewed the types of AE studied in CALGB 9343 and chose the ones that could be severe enough to impact the health utility and healthcare costs (i.e., grade 3 or 4).[55, 99] As a result, skin-related toxicities were considered in our model. We derived the prevalence of skin-related toxicities from an existing literature that examined the frequency of AE for different types of RT.[99]

*Utility inputs*

We derived utility inputs from existing cost-effectiveness studies in breast cancer (BC) (Table 1).[96, 100-102] The utility for loco-regional recurrent was calculated by subtracting the weighted average of the disutility values for salvage mastectomy (0.16) and lumpectomy (0.10) from the utility for no-recurrence.[101] The disutility of AE was applied to the first year in the no-recurrence state for those who received post-lumpectomy RT.[101] We adjusted utility values for all health states by using an age-specific modifier.[103]

*Economic inputs*

We derived the direct costs for each health state from existing literature. The costs for the first year in the no-recurrence state were stratified by the types of treatment a patient received (Table 1). The difference in medical costs between WBI-3D CRT vs WBI-IMRT and WBI vs hypo-WBI were derived from studies that used the SEER-Medicare data of women diagnosed with stage 1-2 BC.[104, 105] We derived the direct costs for the remaining years in the no-recurrence state from Shih et al., which estimated the continuing costs for older patients with stage 1 BC.[106] For the direct costs in the remission state, we used the continuing costs for local and regional BC patients obtained from the same study. The costs of loco-regional recurrence and distant metastasis were informed by Grady et al. that used the SEER-Medicare data.[107] The costs of loco-regional recurrence were calculated by taking a weighted average of year one costs of local vs. regional recurrence. We used the costs for stage 4 BC from Grady et al. for distant metastasis. Given that RT-related skin toxicity is usually treated with corticosteroid or topical cream, we approximated that the one-off costs for treating AE are $50 USD.

The indirect costs included transportation and time costs to receive RT. The average number of miles of traveling and the number of hours spent for treatments per fraction were

informed by existing literature.[108] We used the Fuel Cost Calculator (assuming gas/fuel price of approximately $4 per gallon and miles per gallon of about 20~40 mpg) to estimate the average cost per mile and the US Bureau of Labor Statistics for the national average hourly wage for women.[109, 110] All costs were presented in 2021 USD.

*The size of the affected population*

The affected patient population in the US was defined as those who meet the eligibility criteria of CALGB 9343 trial and were diagnosed between 2004 and 2018, 15-year time period after the older adult-specific trial results became available (Supplement 1).[55]

We conducted a retrospective analysis of the SEER registry to estimate the number of patients eligible for RT omission for each calendar year from 2004 to 2018 (Figure 3). Since the SEER registry data covers approximately 28 percent of the US population, we adjusted the yearly incidence in the entire US by multiplying a factor of 3.6 (100/28).[72] We took this approach based on a study showing that the incidence of late-stage BC is the same between non-SEER and SEER regions and assumed a similar relationship would apply for ESBC.[111]

*The changes in the level of post-lumpectomy RT utilization due to trial results*

We analyzed the SEER registry data to examine the impact of results of CALGB 9343 and PRIME II on the utilization level of post-lumpectomy RT between 2004 and 2018, using difference-in-differences analysis (unpublished results, Figure 3).

*Study outcomes*

Study outcomes included the incremental number of RT non-users, LYs, QALYs, societal costs, and net monetary benefit (NMB) comparing scenarios with vs. without older adult-specific trial results. The NMB was calculated using a willingness-to-pay per QALY threshold of 100,000 USD.

*Characterizing uncertainty*

We conducted 5,000 Monte Carlo simulations to characterize the uncertainty in the results. We varied model inputs based on 95% confidence interval (CI) or standard errors. When the uncertainty range is not available, model inputs were varied by +/- 20% or 30% of the base case value.

*Scenario analysis*

Studies have suggested that the risk of relapse in treated BC patients is minimal beyond ten years.[112] Therefore, in a scenario analysis, we assumed that no loco-regional recurrence or distant metastasis occurred among women who stayed in the no recurrence state for ten years.

## 4.4 Results

*The number of radiation therapy non-users*

The estimated affected patient population in the US was 224,096 between 2004 and 2018. After we applied the incremental probability of omitting RT due to older adult-specific trial results each calendar year, the total number of patients who omitted RT was 32,936 higher in the scenario with older adult-specific trials vs. without among the affected patient population between 2004 and 2018 (Table 2).

*Clinical outcomes*

The incremental LYs gained for a patient over the lifetime horizon was 0.03 years, comparing no RT vs. RT. When the per-person incremental LYs were applied to the entire affected population, the LYs among the affected population were higher by 1,083 years comparing scenarios with vs. without older adult-specific trial results (Table 2).

The incremental QALYs per patient with no RT vs. RT over time was 0.02 years on average across years (Supplement 2 for the comparison with other studies). Therefore, the incremental QALYs gained among the affected patient population with vs. without older adult-specific trial results was 683 years.

*Economic outcomes*

The per-person cost-saving comparing no RT vs. RT was $12,492 on average across years, which resulted in the incremental cost-saving of $419 million USD among the affected patient population in scenarios with vs. without older adult-specific trial results (Table 2).

The real-world value of trial results in monetized QALYs was $68 million USD with a willingness-to-pay per QALY threshold of $100,000 USD, which led to the incremental NMB of $487 million USD comparing the two scenarios. The incremental NMB had an overall increasing trend from 2004 to 2018 (Figure S4.1).

*Sensitivity analysis*

The probabilistic sensitivity analysis showed that the 95% CR of incremental number of patients who omitted RT in scenarios with vs. without older adult-specific trial results was 31,531 to 34,388. The 95% CR of real-world value of older adult-specific trial results in LYs and QALYs were -2,476 to 7,512 years and -2,797 to 7,199 years, respectively (Figure S4.2). The 95% CR of the societal cost-saving and the incremental NMB due to the older adult-specific trial results was -$671 to - $240 million USD and -$0.2 million to $4.5 billion USD, respectively.

The probability of a scenario with older adult-specific trial results generating a positive incremental NMB, compared to a scenario without was 99.8%, 97.5%, and 93.8% with a willingness-to-pay threshold of $50,000, $100,000 and $150,000, respectively (Figure S4.3).

*Scenario analysis*

When we assumed that those who stayed in the no-recurrence state for more than ten years did not experience loco-regional recurrence or distant metastasis, the realized value of older adult-specific trials in terms of LYs and QALYs was 896 and 543 years, respectively, which was a little lower than the base case. The realized value in terms of costs and NMB was $446 million USD and $500 million USD.

## 4.5 Discussion

We estimated the clinical and economic benefits generated in the real world by older adult-specific trials of post-lumpectomy RT in ESBC between 2004 and 2018. We found that the number of older adults who omitted lumpectomy RT was significantly higher with the older adult-specific trials vs. without. The older adult-specific trial results led to a significant reduction in the societal costs and the impact of the trial results on clinical outcomes was not clear. At a willingness-to-pay per QALY of $100,000, the probability of older adult-specific trial results generating a positive NMB was 98%.

We found that the real-world value of older adult-specific trial results has increased over the 15-year time horizon, mainly due to the increasing trend of prescription pattern change over time. Between 2004 and 2009, when only short-term results of CALGB 9343 were available, the incremental likelihood of omitting RT was relatively small and there was no clear increasing trend. With the subsequent publications of longer-term results of CALGB 9343 and the results of PRIME II after 2010, however, the level of RT use decreased constantly, generating a higher societal value over time. This finding may suggest the importance of the cumulative effect of trial results in generating the value of clinical trials. Especially in situations where trial results recommend changing providers' prescription patterns that have been maintained for a long time, more confirming evidence may be required to derive the change in providers' decisions. Therefore, researchers studying the treatment effect heterogeneity in older patients may need to consider decreasing uncertainty in their findings by, for example, following up the patients for a longer term to further increase the value created by the trial results.

A case study that quantifies the real-world value of older adult-specific trials like ours has implications for decision-making among trial sponsors on whether to invest in older adult-

specific trials. By showing that older adult-specific trials brought substantial benefits to a society, our study may suggest that clinical trials for older adults may be worth the investment. Especially, in situations like ours when older adult-specific trials show negative results, meaning that a drug that works well in younger adults is shown to have no or little benefit among older adults, trial results could be translated into economic benefits without sacrificing clinical outcomes. It should be noted, however, that the size of benefits would vary depending on several factors such as the size of affected patient population, the level of change in prescribing patterns, and efficacy and costs of an intervention. Therefore, a careful consideration of these factors is needed when assessing the size of the value. Furthermore, different types of older adult-specific trial results may generate real-world value differently. If, for example, older adult-specific clinical trials found that intervention with null or negative clinical effect in younger adults has a significant benefit in older people (i.e., positive results), the trial results will likely improve clinical outcomes but also increase the costs. There could also be a case where older adult-specific clinical trials do not find a heterogeneous treatment effect compared to younger patients (i.e., neutral results). More studies are warranted to understand how the value is generated by different types of older adult-specific trial results.

To our knowledge, there has been one study that examined the potential downstream impact of older adult-specific evidence. Greenup et al. examined the potential cost savings associated with an evidence-based RT among older women with ESBC in the US.[113] This study quantified the upfront savings if all older adults in their database omit post-lumpectomy RT and found that per-person cost-saving would be approximately $11,700, which was a little lower than our estimate ($14,500). While this study showed the potential economic impact of older adult-specific trial results, it did not account for the real-world prescription pattern changes

and the actual number of the affected patient population. Thus, the study results are hypothetical. Also, clinical outcomes were not incorporated in assessing the downstream impact of the trial results, and economic outcomes were not tracked for the lifetime horizon, meaning that the estimated benefit is not inclusive of all aspects of the impact.

Moreover, we did not include the caregiver costs arising from post-lumpectomy RT. Given the high level of the frailty of those older than 70 years old, a caregiver may need to accompany a patient to the RT facilities, which results in additional indirect costs of RT and a higher cost-saving from the trial results.

## 4.6 Limitations

Our study has several limitations. First, the findings from our study cannot be generalized to older adult-specific clinical trials in other disease or treatment areas. More case studies may be needed to understand the impact of older adult-specific trials in different situations.

Moreover, we did not take into account the costs of conducting clinical trials. The median cost of conducting phase 3 trials is approximately $21.4 million.[114] However, it varied from approximately $5 million to $75 million USD depending on factors including study design, size, and length. Trials like CALGB 9343 that had more than ten years of follow-up might be more expensive than other trials.[114] Furthermore, trials that recruit older or frail adults may have a higher recruitment cost.[115, 116] Further studies on how the costs of older adult-specific clinical trials are compared to the trials in the general patient population would help determine the net economic return of such trials.

Furthermore, there is uncertainty about the size of the affected patient population. We assumed that patients who meet the eligibility criteria of the CALGB 9343 study are those eligible for the omission of post-lumpectomy RT. However, it is possible that some providers have applied the results to patients more broadly or conservatively. Also, the study period was limited to up until 2018 due to the data limitation. Considering the continuing decreasing trend of the probability of receiving post-lumpectomy RT, our study might not have captured the full value of CALGB 9343 and PRIME II.

## 4.7 Conclusion

We quantified the real-world clinical and economic benefits of results from CALGB 9343 and PRIME II phase 3 trials that demonstrated post-lumpectomy RT has minimal benefit among older adults with ESBC. We found that approximately 33,000 older adults with ESBC omitted RT after lumpectomy in response to the trial results over a 15-year period and reduction in the US societal costs of $419 million USD. While these results can contribute to the discussion around whether and how to invest in older adult-specific clinical trials, more case studies, especially for trials with positive and neutral results, need to be done to have a more comprehensive understanding of how older adult-specific trials generate value.

## 4.8 Tables & Figures

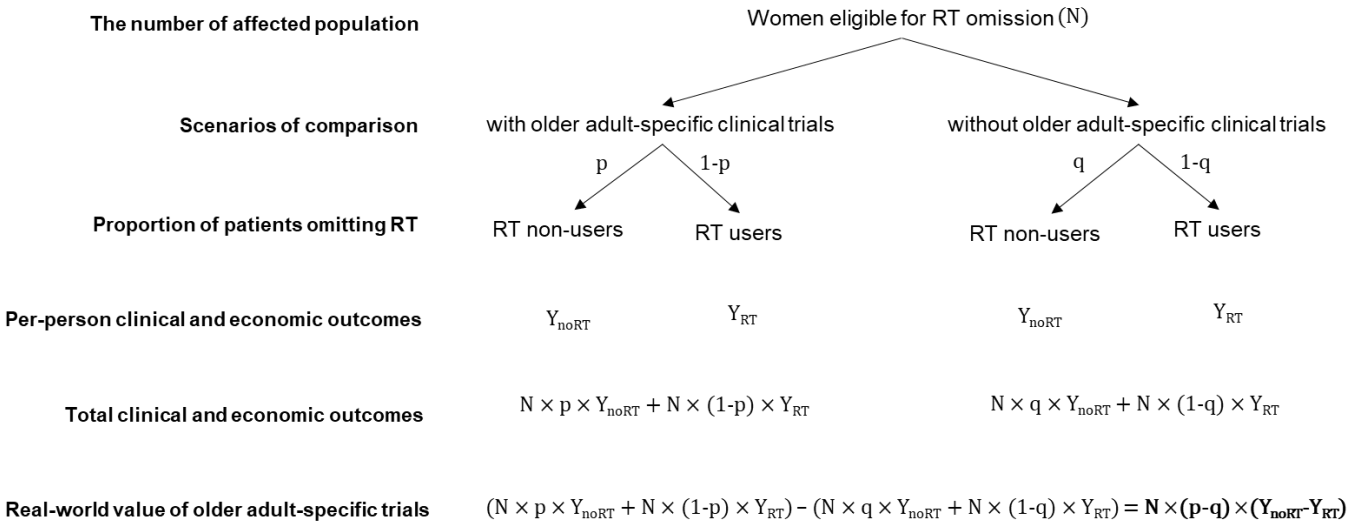


Figure 4. 1. Overall analytic approach

N = the number of the affected population; p = the probability of omitting post-lumpectomy RT in a scenario with older adult-specific clinical trials; q = the probability of omitting post-lumpectomy RT in a scenario without older adult-specific clinical trials; YnoRT = lifetime clinical and economic outcomes (i.e., Quality-adjusted life years, life years, societal costs, and net monetary benefits) without post-lumpectomy RT; YRT = lifetime clinical and economic outcomes without post-lumpectomy RT with post-lumpectomy RT; RT: radiation therapy

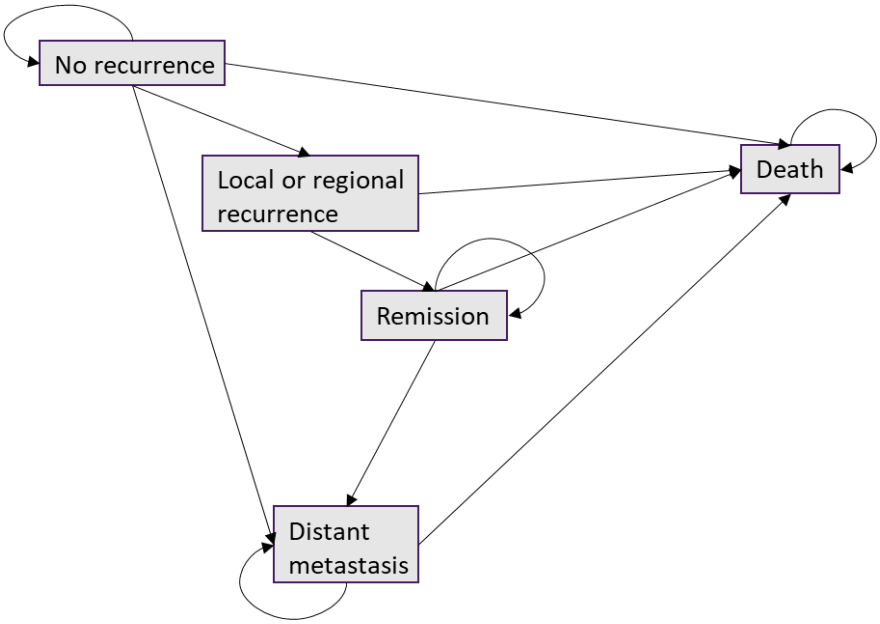


Figure 4. 2. Markov model structure

All patients who underwent lumpectomy after the diagnosis of ESBC started from the no recurrence state where they could remain in the state, transitioned to loco-regional recurrence or distant metastasis, or died from non-BC causes. Once patients developed a loco-regional recurrence, they stayed in that state for one year and then transitioned to remission or died from non-BC causes. Those who moved to the remission state, patients stayed in that state, transitioned to distant metastasis, or died from non-BC causes. In the distant metastasis state, patients could stay in that state or died from BC- or non-BC causes.

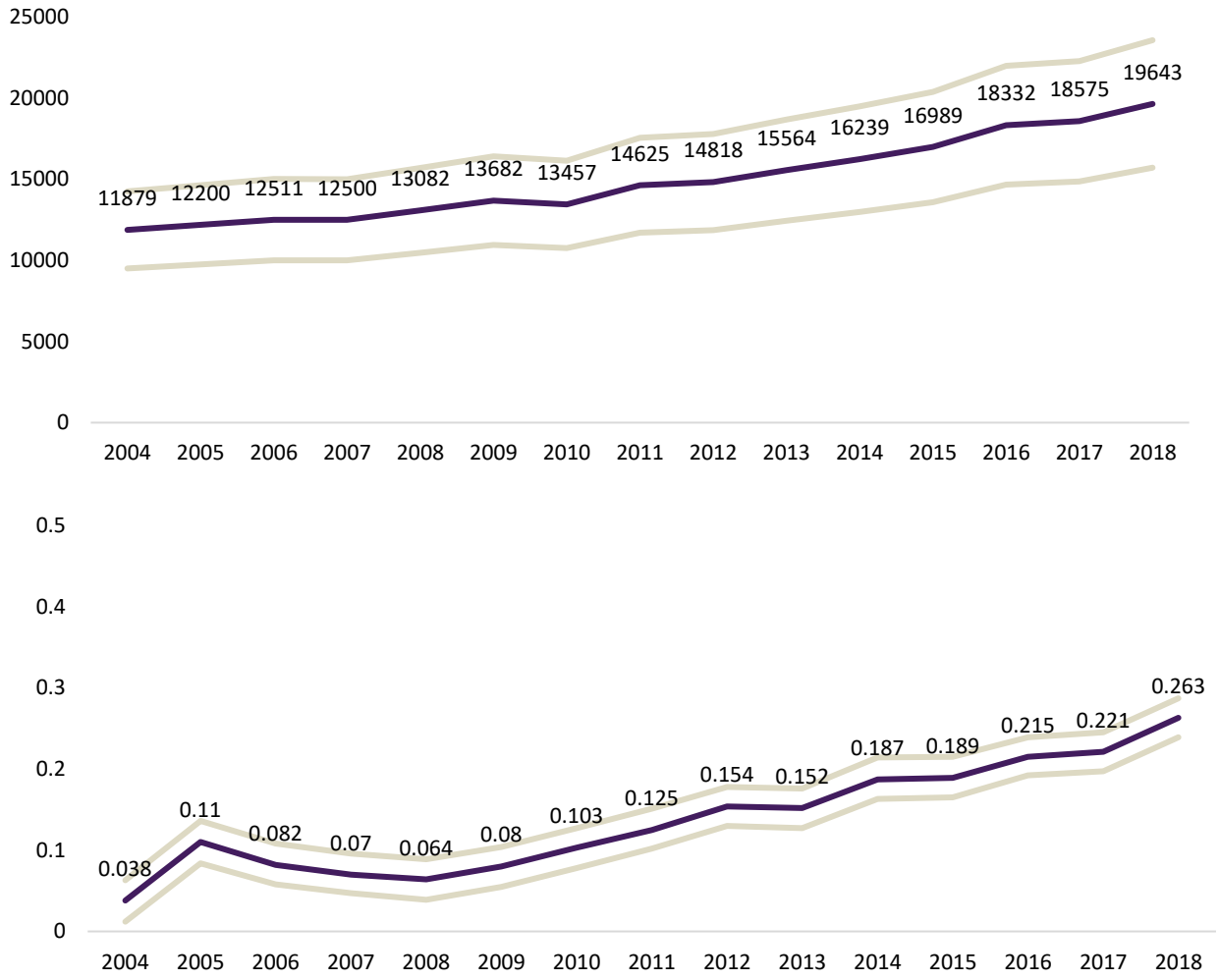


Figure 4. 3. The number of the affected patient population (top) and the incremental probability of patients omitting RT with vs. without trial results (bottom) over time

Gray lines indicate upper and lower bounds. The number of the affected patient population was varied by +/- 20% and the incremental proportion of patients omitting RT was varied based on the 95% confidence interval obtained from the difference-in-differences analysis. Normal distribution was used for both inputs in the probabilistic sensitivity analysis.

| Table 4. 1. Clinical, utility, and economic inputs and ranges for sensitivity analyses | | | | | | |
|---|---|---|---|---|---|---|
| Variables | Deterministic | Lower | Upper | Distribution | Uncertainty range | Source |
| General | | | | | | |
| Age at diagnosis | 75 | N/A | N/A | N/A | N/A | [56] |
| Discount rate | 3% | N/A | N/A | N/A | N/A | N/A |
| Yearly transition probabilities | | | | | | |
| No recurrence to loco-regional recurrence | | | | | | |
| No RT | 0.009 | 0.006 | 0.014 | Normal | 95% CI | [56] |
| Hazard ratio (RT vs. No RT) | 0.18 | 0.07 | 0.42 | Lognormal | 95% CI | [56] |
| No recurrence to distant metastasis | | | | | | |
| No RT | 0.005 | 0.009 | 0.003 | Normal | 95% CI | [56] |
| Hazard ratio (RT vs. No RT) | 1.20 | 0.63 | 2.32 | Lognormal | 95% CI | [56] |
| Ratio of the risk of distant metastasis from no recurrence vs. remission state | 2.50 | 1.75 | 3.25 | Lognormal | +/- 30% | [95] |
| Distant metastasis to death | 0.23 | 0.18 | 0.28 | Normal | +/- 20% | [96] |
| No recurrence to death | Life table | N/A | N/A | N/A | N/A | [97] |
| Remission to death | Life table | N/A | N/A | N/A | N/A | [97] |
| Loco-regional recurrence to death | Life table | N/A | N/A | N/A | N/A | [97] |
| Probability of G3 skin toxicity of RT | | | | | | |
| CF-WBI-3D CRT | 0.024 | 0.02 | 0.03 | Beta | +/- 20% | [99] |
| CF-WBI-IMRT | 0.018 | 0.01 | 0.02 | Beta | +/- 20% | [99] |
| HF-WBI-3D CRT | 0.007 | 0.01 | 0.008 | Beta | +/- 20% | [99] |
| HF-WBI-IMRT | 0.002 | 0.00 | 0.002 | Beta | +/- 20% | [99] |
| Utilities[a] | | | | | | |
| No recurrence (year 1) | 0.71 | 0.57 | 0.85 | Beta | +/- 20% | [96, 100-102] |
| No recurrence (year 2+) | 0.71 | 0.57 | 0.85 | Beta | +/- 20% | [96, 100-102] |
| Loco-regional recurrence | 0.61 | 0.49 | 0.73 | Beta | +/- 20% | [96, 100-102] |

| | | | | | | |
|---|---|---|---|---|---|---|
| Remission | 0.71 | 0.51 | 0.76 | Beta | +/- 20% | [96, 100-102] |
| Distant metastasis | 0.54 | 0.43 | 0.65 | Beta | +/- 20% | [96, 100-102] |
| Disutility of G3 skin toxicity | 0.01 | 0.01 | 0.012 | Beta | +/- 20% | [101] |
| Direct costs (yearly) | | | | | | |
| No recurrence (year 1) | | | | | | |
| No RT | $10,817 | $8,654 | $12,981 | Gamma | +/- 20% | [100, 104] |
| CF-WBI-3D CRT | $23,546 | $18,837 | $28,255 | Gamma | +/- 20% | [100, 104] |
| CF-WBI-IMRT | $32,940 | $26,352 | $39,527 | Gamma | +/- 20% | [100, 104, 105] |
| HF-WBI-3D CRT | $19,126 | $15,301 | $22,951 | Gamma | +/- 20% | [100, 104] |
| HF-WBI-IMRT | $28,519 | $22,815 | $34,223 | Gamma | +/- 20% | [100, 104, 105] |
| No recurrence (year 2+) | $803 | $643 | $964 | Gamma | +/- 20% | [100] |
| Loco-regional recurrence | $55,964 | $44,772 | $67,157 | Gamma | +/- 20% | [107] |
| Remission | $1,499 | $1,199 | $1,798 | Gamma | +/- 20% | [100] |
| Distant metastasis (year 1) | $83,490 | $66,792 | $100,188 | Gamma | +/- 20% | [107] |
| Distant metastasis (year 2+) | $32,564 | $26,051 | $39,076 | Gamma | +/- 20% | [107] |
| G3 skin toxicity | $393 | $314 | $472 | Gamma | +/- 20% | Assumption |
| Indirect costs (year 1 only) | | | | | | |
| CF-WBI | $1,500 | $1,200 | $1,800 | Gamma | +/- 20% | [96, 109, 110] |
| HF-WBI | $750 | $600 | $900 | Gamma | +/- 20% | [96, 109, 110] |

[a] The age-specific modifier of 0.771 was applied

| Table 4. 2. The real-world value of older adult-specific clinical trial results | | |
|---|---|---|
| Real-world value | Base-case | 95% CR[b] |
| Post-lumpectomy RT non-users | 32936 | 31442, 34365 |
| Life years | 1083 years | -2677, 7463 years |
| Quality-adjusted life years | 683 years | -2797, 7199 years |
| Monetized QALYs[a] | $68 million USD | -$28 million, $72 million USD |
| Societal costs | -$419 million USD | -$239 million, -$648 million USD |
| Net monetary benefits[a] | $487 million USD | -$0.2 million, $4.5 billion USD |

[a]We used a willingness-of-pay per QALY threshold of $100,000 USD to calculate monetized QALYs and net monetary benefits; [b]5,000 times of Monte Carlo simulations were used to estimate the 95% credible interval; CR: credible intervals; USD: 2021 US dollars

## 4.9 Supplement

*Supplement 1. Eligibility criteria of CALGB 9343*

Patients who (1) were female, (2) were diagnosed with first or only stage 1 (T1N0M0 according to TNM classification and tumor size ≤ 2 cm) estrogen receptor (ER)-positive breast cancer, (3) underwent lumpectomy (i.e., partial mastectomy or less than total mastectomy), and (4) were aged 70 years or older at the time of diagnosis.

*Supplement 2. Comparing the base-case incremental QALYs per patient with vs. without omitting RT with other cost-effectiveness studies*

The incremental QALYs per patient with vs. without omitting RT in the base-case scenario were different from the existing cost-effectiveness studies using clinical inputs derived from the same older adult-specific trials. While we found the incremental QALYs gained of 0.02 years with omitting RT in the base case, it was -0.2 years in Sen 2014 and -0.03 years in Ward 2020. [1, 2] The 95% CR for the incremental QALYs with vs. without omitting RT crossed zero in our study, but this information was not available in the other two studies. Therefore, whether the QALYs were significantly different between the two arms in the other studies was unclear. The difference in the base-case incremental QALYs gained is likely due to the different model structures. While we had a separate health state for remission for those who recovered from a loco-regional recurrence and used a higher utility value for them, other studies did not. Therefore, patients who experienced the loco-regional recurrence were assumed to have a lower utility even after the recovery, which amplifies the impact of the higher risk of loco-regional recurrence with RT omission. Moreover, Sen 2014 assumed that the utility value is lower for those who omitted RT than those who received it based on Hayman et al. where utilities were

measured among general patients with ESBC using standard gambles. [3] In this study, the utility for not receiving RT was lower than the one for receiving RT due to the fear of experiencing recurrence. However, this study is outdated (conducted before 1997) and was not specific to older adults who have a low risk of loco-regional recurrence. Lastly, other studies did not incorporate the higher risk of distant metastasis with receiving RT, which was shown in the long-term CALGB 9343 results, overestimating the survival with RT.

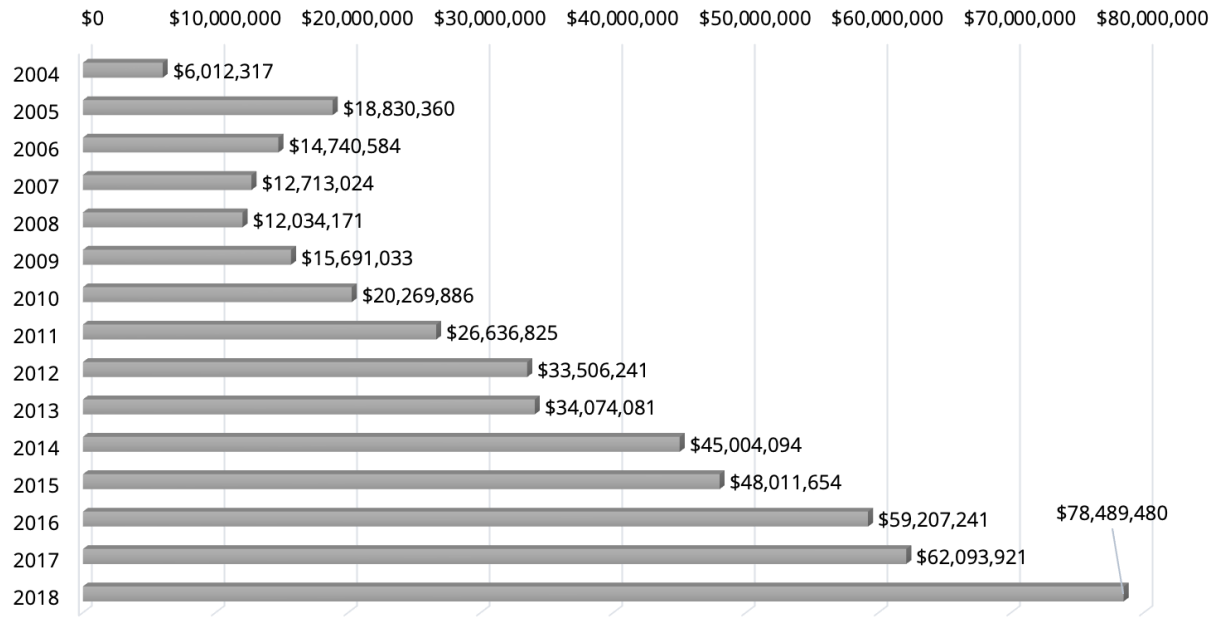


Figure S4. 1. The incremental net monetary benefit (NMB) comparing scenarios with vs. without older adult-specific trial results over time (2004-2018)

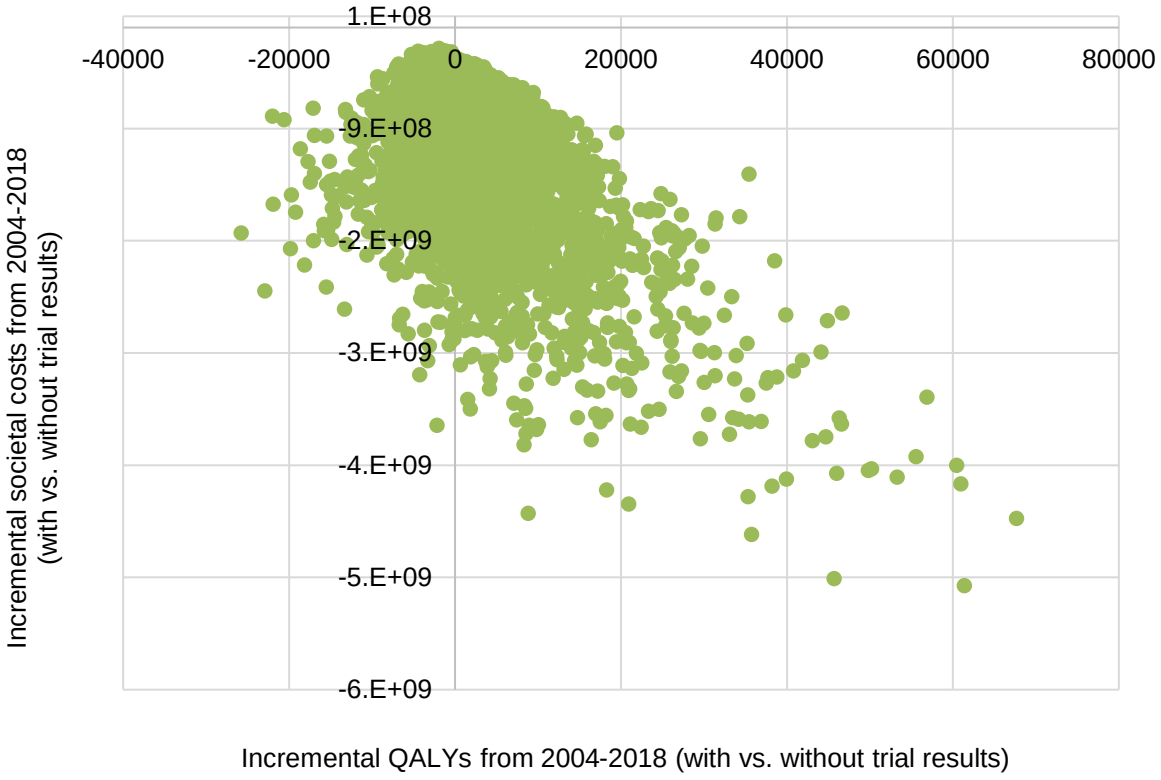


Figure S4. 2. Plane of incremental QALYs and costs in scenarios with vs. without older adult-specific trial results

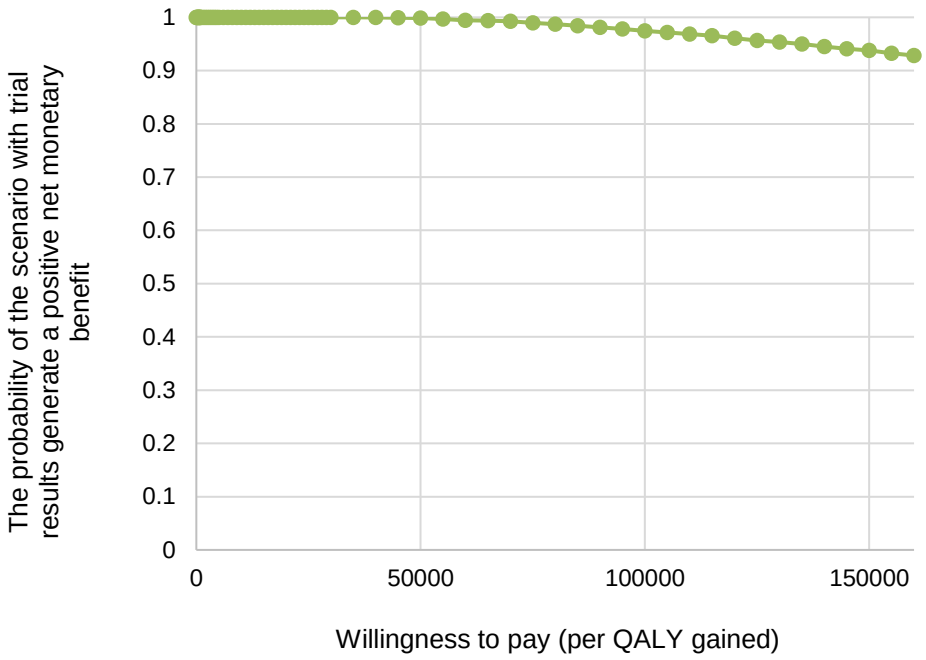


Figure S4. 3. The probability of the scenario with older adult-specific trial results having a positive net monetary benefit across different willingness-to-pay per QALY

# Chapter 5. Conclusions

The contents of this dissertation have worked to examine the challenges and value of conducting older adult-specific clinical trials in clinical areas. To address the potential challenges of investing in older adult-specific clinical trials, we developed a machine learning model that predicts the risk of trial failure using trial-level factors. The prediction model developed in this study can be useful in aiding trial sponsors and researchers in prioritizing and designing clinical trials by providing a quantitative measure for the likelihood of trial success before conducting a study. The study findings also provide a parsimonious set of predictors that are potentially associated with the risk of trial failure. If confirmed by inference studies showing how these predictors are causally associated with trial failure among older adult-specific clinical trials, our results can help researchers modify trial-level factors in a way to reduce the risk of study failure. We believe that adding some other key potential predictors of trial failure (e.g., the clinical importance of the research question and deviation from the standard of care) could help improve the predictive performance of the model.

We also examined the impact of older adult-specific clinical trial results by assessing the causal effect of the trial results on the real-world prescribing patterns, using the phase 3 clinical trials of post-lumpectomy irradiation among older adults with ESBC (i.e., CALGB 9343 and PRIME II) as a case study. We found that the use of post-lumpectomy irradiation has been significantly changed as intended due to the results of CALGB 9343 and PRIME II among older adults over time. The initial 5-year follow-up results of CALGB 9343 were associated with a significant immediate and yearly decrease in the probability of irradiation use and a follow-up disclosure of longer-term CALGB 9343 results further accelerated the rate of decrease. Our study ascertains the causal impact of CALGB 9343 and PRIME II studies on the prescribing

patterns, suggesting that older adult-specific clinical trials, if successfully completed, could significantly change providers' treatment decisions among older patients. Also, the study findings demonstrated the importance of evidence accumulation in changing prescription patterns, meaning that a stepwise approach where evidence is assured by additional findings could be one way to incorporate treatment effect heterogeneity in older people into clinical practice. The changes in the prescribing patterns, however, could be different in other disease or treatment areas, and thus additional case studies are needed to better understand the impact of older adult-specific clinical trials on providers' treatment decisions.

When the downstream clinical and economic benefits of the results from CALGB 9343 and PRIME II phase 3 trials were quantified, we found a significant reduction in the US societal costs of $419 million USD during the 15-year period after the initial publication of the study results. The trial results did not significantly change the clinical outcomes of older adults. Our case study has implications for decision-making among trial sponsors on whether to invest in older adult-specific trials. By showing that older adult-specific trials brought substantial benefits to society, our study may suggest that clinical trials for older adults may be worth the investment. Especially, in situations when older adult-specific trials show negative results, meaning that a drug that works well in younger adults is shown to have no or little benefit among older adults, trial results could be translated into economic benefits without sacrificing clinical outcomes. It should be noted, however, that the size of benefits would vary depending on several factors such as the size of the affected patient population, the level of change in prescribing patterns, and the efficacy and costs of an intervention. Furthermore, more case studies, especially for trials with positive (i.e., significant benefits among older people when there is no benefit among younger people) and neutral results (i.e., effects among older people are not different from those among

younger people), need to be done to have a more comprehensive understanding of how older adult-specific trials generate value.

While there has been a call for increasing efforts to conduct more clinical studies for older adults, conducting clinical trials generally requires a lot of resources. Given limited healthcare resources, trial sponsors and researchers need to consider the potential challenges and value of conducting clinical trials in older adults. The analyses performed in this dissertation allow trial sponsors and researchers to estimate the risk of unsuccessful completion of cancer clinical trials in older adults based on the trial protocol and show the potential value older adult-specific clinical trial results could generate in the real world. We hope that the findings from this study will better inform the investment and prioritization decisions of clinical trials for older adults among trial sponsors and researchers and, ultimately, help improve treatment decisions among older adults.

# Chapter 6. References

1. Vespa J, Armstrong DM, Medina L. Demographic turning points for the United States: Population projections for 2020 to 2060: US Department of Commerce, Economics and Statistics Administration, US …; 2018.
2. Balducci L. Pharmacology of antineoplastic medications in older cancer patients. Oncology. 2009;23(1):78
3. Sedrak MS, Freedman RA, Cohen HJ, Muss HB, Jatoi A, Klepin HD, et al. Older adult participation in cancer clinical trials: A systematic review of barriers and interventions. CA: A Cancer Journal for Clinicians. 2020
4. Scher KS, Hurria A. Under-representation of older adults in cancer registration trials: known problem, little progress. Journal of Clinical Oncology. 2012;30(17):2036-8
5. Palmowski A, Buttgereit T, Palmowski Y, Nielsen SM, Boers M, Christensen R, et al. Applicability of trials in rheumatoid arthritis and osteoarthritis: A systematic review and meta-analysis of trial populations showing adequate proportion of women, but underrepresentation of elderly people. Seminars in arthritis and rheumatism; 2019: Elsevier. p. 983-9.
6. Kanapuru B, Singh H, Kwitkowski V, Blumenthal G, Farrell AT, Pazdur R. Older adults in hematologic malignancy trials: representation, barriers to participation and strategies for addressing underrepresentation. Blood reviews. 2020:100670
7. The Food and Drug Administration, Inclusion of Older Adults in Cancer Clinical Trials, Draft Guidance for Industry. 2020https://www.fda.gov/regulatory-information/search-fda-guidance-documents/inclusion-older-adults-cancer-clinical-trials
8. Hurria A, Levit LA, Dale W, Mohile SG, Muss HB, Fehrenbacher L, et al. Improving the evidence base for treating older adults with cancer: American Society of Clinical Oncology statement. J Clin Oncol. 2015;33(32):3826-33
9. Hurria A, Cohen HJ, Extermann M. Geriatric oncology research in the cooperative groups: A report of a SIOG special meeting. Journal of geriatric oncology. 2010;1(1):40-4
10. Agency EM. Adequacy of guidance on the elderly regarding medicinal products for human use. 2006
11. Raimi-Abraham BT, de Orbe Izquierdo MS, Collignon O, Cerreta F. Regulatory considerations on the enrollment of older adults in oncology clinical trials. people. 2017;4:12
12. De Glas N, Hamaker M, Kiderlen M, De Craen A, Mooijaart S, Van De Velde C, et al. Choosing relevant endpoints for older breast cancer patients in clinical trials: an overview of all current clinical trials on breast cancer treatment. Breast cancer research and treatment. 2014;146(3):591-7
13. Le Saux O, Falandry C, Gan HK, You B, Freyer G, Peron J. Inclusion of elderly patients in oncology clinical trials. Annals of Oncology. 2016;27(9):1799-804
14. Sedrak MS, Freedman RA, Cohen HJ, Muss HB, Jatoi A, Klepin HD, et al. Older adult participation in cancer clinical trials: A systematic review of barriers and interventions. CA: a cancer journal for clinicians. 2021;71(1):78-92
15. Banzi R, Camaioni P, Tettamanti M, Lucca U. Older patients are still under-represented in clinical trials of Alzheimer's disease. Alzheimer's research & therapy. 2016;8(1):1-10
16. The Food and Drug Administration, Inclusion of Older Adults in Cancer Clinical Trials, Draft Guidance for Industry. 2020
17. Williams RJ, Tse T, DiPiazza K, Zarin DA. Terminated trials in the ClinicalTrials. gov results database: evaluation of availability of primary outcome data and reasons for termination. PloS one. 2015;10(5):e0127242

18. Hauck CL, Kelechi TJ, Cartmell KB, Mueller M. Trial-level factors affecting accrual and completion of oncology clinical trials: A systematic review. Contemporary Clinical Trials Communications. 2021:100843
19. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. Contemporary clinical trials communications. 2018;11:156-64
20. Ruther NR, Jumonville A, Mathiason MA, Emmel AE, Wee SK, Go RS. Speed of accrual into published phase III oncology trials: A comparison across geographic locations. American Society of Clinical Oncology; 2012.
21. Khunger M, Rakshit S, Hernandez AV, Pasupuleti V, Glass K, Galsky MD, et al. Premature clinical trial discontinuation in the era of immune checkpoint inhibitors. The Oncologist. 2018;23(12):1494-9
22. Paul K, Sathianathen N, Dahm P, Le C, Konety BR. Variation in accrual and race/ethnicity reporting in urological and nonurological related cancer trials. The Journal of Urology. 2019;202(2):385-91
23. Nguyen TK, Nguyen EK, Warner A, Louie AV, Palma DA. Failed randomized clinical trials in radiation oncology: what can we learn? International Journal of Radiation Oncology* Biology* Physics. 2018;101(5):1018-24
24. Cheng SK, Dietrich MS, Dilts DM. A sense of urgency: evaluating the link between clinical trial development time and the accrual performance of cancer therapy evaluation program (NCI-CTEP) sponsored studies. Clinical Cancer Research. 2010;16(22):5557-63
25. Stensland KD, McBride RB, Latif A, Wisnivesky J, Hendricks R, Roper N, et al. Adult cancer clinical trials that fail to complete: an epidemic? JNCI: Journal of the National Cancer Institute. 2014;106(9)
26. Korn EL, Freidlin B, Mooney M, Abrams JS. Accrual experience of National Cancer Institute Cooperative Group phase III trials activated from 2000 to 2007. Journal of clinical oncology. 2010;28(35):5197
27. Bennette CS, Ramsey SD, McDermott CL, Carlson JJ, Basu A, Veenstra DL. Predicting low accrual in the National Cancer Institute's Cooperative Group clinical trials. JNCI: Journal of the National Cancer Institute. 2016;108(2)
28. Lyss AP, Lilenbaum RC. Accrual to National Cancer Institute—Sponsored Non–Small-Cell Lung Cancer Trials: Insights and Contributions From the CCOP Program. Clinical Lung Cancer. 2009;10(6):410-3
29. Massett HA, Mishkin G, Rubinstein L, Ivy SP, Denicoff A, Godwin E, et al. Challenges facing early phase trials sponsored by the National Cancer Institute: an analysis of corrective action plans to improve accrual. Clinical Cancer Research. 2016;22(22):5408-16
30. Kim ES, Bernstein D, Hilsenbeck SG, Chung CH, Dicker AP, Ersek JL, et al. Modernizing eligibility criteria for molecularly driven trials. Journal of Clinical Oncology. 2015;33(25):2815-20
31. Duma N, Kothadia SM, Azam TU, Yadav S, Paludo J, Vera Aguilera J, et al. Characterization of comorbidities limiting the recruitment of patients in early phase clinical trials. The Oncologist. 2019;24(1):96-102
32. Gerber DE, Laccetti AL, Xuan L, Halm EA, Pruitt SL. Impact of prior cancer on eligibility for lung cancer clinical trials. Journal of the National Cancer Institute. 2014;106(11):dju302

33. Hernandez-Torres C, Cheung WY, Kong S, O'Callaghan CJ, Hsu T. Accrual of older adults to cancer clinical trials led by the Canadian cancer trials group–Is trial design a barrier? Journal of geriatric oncology. 2020;11(3):455-62
34. Clinical Trials Transformation Initiative, What is AACT? https://aact.ctti-clinicaltrials.org/
35. The National Institute of Health, U.S. National Library of Medicine, FDAAA 801 and the Final Rule. https://clinicaltrials.gov/ct2/manage-recs/fdaaa
36. US National Library of Medicine, ClinicalTrials.gov, Trends Charts, and Maps. https://clinicaltrials.gov/ct2/resources/trends
37. Carlisle B, Kimmelman J, Ramsay T, MacKinnon N. Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. Clinical Trials. 2015;12(1):77-83
38. Stensland K, Kaffenberger S, Canes D, Galsky M, Skolarus T, Moinzadeh A. Assessing genitourinary cancer clinical trial accrual sufficiency using archived trial data. JCO Clinical Cancer Informatics. 2020;4:614-22
39. Schroen AT, Petroni GR, Wang H, Thielen MJ, Gray R, Benedetti J, et al. Achieving sufficient accrual to address the primary endpoint in phase III clinical trials from US Cooperative Oncology Groups. Clinical Cancer Research. 2012;18(1):256-62
40. Rhys H. Machine Learning with R, the tidyverse, and mlr: Simon and Schuster; 2020.
41. Stone M. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. Journal of the Royal Statistical Society: Series B (Methodological). 1977;39(1):44-7
42. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Annals of internal medicine. 2015;162(1):W1-W73
43. Stark N, Peacock J. Clinical Studies: Europe or the United States? Medical Device and Diagnostic Industry. 2004;26:134-42
44. Bellomo R, Warrillow SJ, Reade MC. Why we should be wary of single-center trials. Critical care medicine. 2009;37(12):3114-9
45. Jaki T, Wason J. Multi-arm multi-stage trials can improve the efficiency of finding effective treatments for stroke: a case study. BMC Cardiovascular Disorders. 2018;18(1):1-8
46. Parmar MK, Carpenter J, Sydes MR. More multiarm randomised trials of superiority are needed. The Lancet. 2014;384(9940):283-4
47. Halpern SD, Karlawish JH, Casarett D, Berlin JA, Townsend RR, Asch DA. Hypertensive patients' willingness to participate in placebo-controlled trials: implications for recruitment efficiency. American heart journal. 2003;146(6):985-92
48. National Academies of Sciences E, Medicine. Developing multimodal therapies for brain disorders: proceedings of a workshop. 2017
49. Gross CP, Herrin J, Wong N, Krumholz HM. Enrolling older persons in cancer trials: the effect of sociodemographic, protocol, and recruitment center characteristics. Journal of clinical oncology. 2005;23(21):4755-63
50. Hamaker ME, Stauder R, van Munster BC. Exclusion of older patients from ongoing clinical trials for hematological malignancies: an evaluation of the National Institutes of Health Clinical Trial Registry. The oncologist. 2014;19(10):1069-75
51. Soto-Perez-De-Celis E, Lichtman SM. Considerations for clinical trial design in older adults with cancer. Expert opinion on investigational drugs. 2017;26(10):1099-102

52. Abbasi J. Older patients (still) left out of cancer clinical trials. JAMA. 2019;322(18):1751-3
53. Steinman MA, Boyd CM, Schmader KE. Expanding evidence for clinical care of older adults: beyond clinical trial traditions and finding new approaches. JAMA. 2021;326(6):475-6
54. Fisher B, Anderson S, Bryant J, Margolese RG, Deutsch M, Fisher ER, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. New England Journal of Medicine. 2002;347(16):1233-41
55. Hughes KS, Schnaper LA, Berry D, Cirrincione C, McCormick B, Shank B, et al. Lumpectomy plus tamoxifen with or without irradiation in women 70 years of age or older with early breast cancer. New England Journal of Medicine. 2004;351(10):971-7
56. Hughes K, Schnaper L, Cirrincione C, Berry D, McCormick B, Muss H, et al. Lumpectomy plus tamoxifen with or without irradiation in women age 70 or older with early breast cancer. Journal of Clinical Oncology. 2010;28(15_suppl):507-
57. Hughes KS, Schnaper LA, Bellon JR, Cirrincione CT, Berry DA, McCormick B, et al. Lumpectomy plus tamoxifen with or without irradiation in women age 70 years or older with early breast cancer: long-term follow-up of CALGB 9343. Journal of clinical oncology. 2013;31(19):2382
58. Kunkler I, Williams L, Jack W, Canney P, Prescott R, Dixon M. Abstract S2-01: The PRIME II trial: Wide local excision and adjuvant hormonal therapy±postoperative whole breast irradiation in women≥ 65 years with early breast cancer managed by breast conservation. AACR; 2013.
59. Kunkler IH, Williams LJ, Jack WJ, Cameron DA, Dixon JM. Breast-conserving surgery with or without irradiation in women aged 65 years or older with early breast cancer (PRIME II): a randomised controlled trial. The Lancet Oncology. 2015;16(3):266-73
60. Soulos PR, James BY, Roberts KB, Raldow AC, Herrin J, Long JB, et al. Assessing the impact of a cooperative group trial on breast cancer care in the medicare population. Journal of clinical oncology. 2012;30(14):1601
61. McCormick B, Ottesen RA, Hughes ME, Javid SH, Khan SA, Mortimer J, et al. Impact of guideline changes on use or omission of radiation in the elderly with early breast cancer: practice patterns at National Comprehensive Cancer Network institutions. Journal of the American College of Surgeons. 2014;219(4):796-802
62. Palta M, Palta P, Bhavsar NA, Horton JK, Blitzblau RC. The use of adjuvant radiotherapy in elderly patients with early-stage breast cancer: Changes in practice patterns after publication of Cancer and Leukemia Group B 9343. Cancer. 2015;121(2):188-93
63. Squeo G, Malpass JK, Meneveau M, Balkrishnan R, Desai RP, Lattimore C, et al. Long-term impact of CALGB 9343 on radiation utilization. Journal of Surgical Research. 2020;256:577-83
64. Taylor LJ, Steiman JS, Anderson B, Schumacher JR, Wilke LG, Greenberg CC, et al. Does persistent use of radiation in women> 70 years of age with early-stage breast cancer reflect tailored patient-centered care? Breast cancer research and treatment. 2020;180(3):801-7
65. Keim-Malpass J, Anderson RT, Balkrishnan R, Desai RP, Showalter SL. Evaluating the long-term impact of a cooperative group trial on radiation use and adjuvant endocrine therapy adherence among older women. Annals of Surgical Oncology. 2020;27(9):3458-65

66. Christian N, Heelan Gladden A, Friedman C, Gleisner-Patton A, Murphy C, Kounalakis N, et al. Increasing omission of radiation therapy and sentinel node biopsy in elderly patients with early stage, hormone-positive breast cancer. The breast journal. 2020;26(2):133-8
67. Paulsson AK, Fowble B, Lazar AA, Park C, Sherertz T. Radiotherapy utilization for patients over age 60 with early stage breast cancer. Clinical breast cancer. 2020;20(2):168-73
68. Blohmer J-U, Tanko J, Kueper J, Groß J, Völker R, Machleidt A. MarginProbe© reduces the rate of re-excision following breast conserving surgery for breast cancer. Archives of gynecology and obstetrics. 2016;294(2):361-7
69. Kasem I, Mokbel K. Savi Scout® radar localisation of non-palpable breast lesions: systematic review and pooled analysis of 842 cases. Anticancer Research. 2020;40(7):3633-43
70. Haussmann J, Corradini S, Nestle-Kraemling C, Bölke E, Njanang FJD, Tamaskovics B, et al. Recent advances in radiotherapy of breast cancer. Radiation Oncology. 2020;15(1):1-10
71. Angrist JD, Pischke J-S. Mostly harmless econometrics: Princeton university press; 2008.
72. Doll KM, Rademaker A, Sosa JA. Practical guide to surgical data sets: surveillance, epidemiology, and end results (SEER) database. JAMA surgery. 2018;153(6):588-9
73. The National Cancer Institute, Surveillance, Epidemiology and End Results Program, SEER Incidence Database. https://seer.cancer.gov/data-software/
74. Raifman J, Moscoe E, Austin SB, Hatzenbuehler ML, Galea S. Association of state laws permitting denial of services to same-sex couples with mental distress in sexual minority adults: A difference-in-difference-in-differences analysis. JAMA psychiatry. 2018;75(7):671-7
75. Higgerson J, Halliday E, Ortiz-Nunez A, Brown R, Barr B. Impact of free access to leisure facilities and community outreach on inequalities in physical activity: a quasi-experimental study. J Epidemiol Community Health. 2018;72(3):252-8
76. Angrist JD, Pischke J-S. Mastering'metrics: The path from cause to effect: Princeton university press; 2014.
77. Jagsi R, Abrahamse P, Hawley ST, Graff JJ, Hamilton AS, Katz SJ. Underascertainment of radiotherapy receipt in Surveillance, Epidemiology, and End Results registry data. Cancer. 2012;118(2):333-41
78. Noone A-M, Lund JL, Mariotto A, Cronin K, McNeel T, Deapen D, et al. Comparison of SEER treatment data with Medicare claims. Medical care. 2016;54(9):e55
79. Bennett C. Timelines Matter in the Treatment of Breast Cancer. Cancer Network. 2019https://www.cancernetwork.com/view/timelines-matter-treatment-breast-cancer
80. MAYO CLINIC, Radiation therapy for breast cancer. https://www.mayoclinic.org/tests-procedures/radiation-therapy-for-breast-cancer/about/pac-20384940 [Accessed on 2/8/2022]
81. Jobsen J, Van der Palen J, Baum M, Brinkhuis M, Struikmans H. Timing of radiotherapy in breast-conserving therapy: a large prospective cohort study of node-negative breast cancer patients without adjuvant systemic therapy. British journal of cancer. 2013;108(4):820-5
82. Salerno KE. NCCN guidelines update: evolving radiation therapy recommendations for breast cancer. Journal of the National Comprehensive Cancer Network. 2017;15(5S):682-4
83. Quoix E, Zalcman G, Oster J-P, Westeel V, Pichon E, Lavolé A, et al. Carboplatin and weekly paclitaxel doublet chemotherapy compared with monotherapy in elderly patients with advanced non-small-cell lung cancer: IFCT-0501 randomised, phase 3 trial. The Lancet. 2011;378(9796):1079-88
84. Cunningham D, Lang I, Marcuello E, Lorusso V, Ocvirk J, Shin DB, et al. Bevacizumab plus capecitabine versus capecitabine alone in elderly patients with previously untreated

metastatic colorectal cancer (AVEX): an open-label, randomised phase 3 trial. The lancet oncology. 2013;14(11):1077-85
85. Roth JA, Etzioni R, Waters TM, Pettinger M, Rossouw JE, Anderson GL, et al. Economic return from the Women's Health Initiative estrogen plus progestin clinical trial: a modeling study. Annals of internal medicine. 2014;160(9):594-602
86. Mukesh MB, Barnett GC, Wilkinson JS, Moody AM, Wilson C, Dorling L, et al. Randomized controlled trial of intensity-modulated radiotherapy for early breast cancer: 5-year results confirm superior overall cosmesis. Journal of clinical oncology. 2013;31(36):4488-95
87. Haviland JS, Owen JR, Dewar JA, Agrawal RK, Barrett J, Barrett-Lee PJ, et al. The UK Standardisation of Breast Radiotherapy (START) trials of radiotherapy hypofractionation for treatment of early breast cancer: 10-year follow-up results of two randomised controlled trials. The lancet oncology. 2013;14(11):1086-94
88. Whelan TJ, Pignol J-P, Levine MN, Julian JA, MacKenzie R, Parpia S, et al. Long-term results of hypofractionated radiation therapy for breast cancer. New England Journal of Medicine. 2010;362(6):513-20
89. Owen JR, Ashton A, Bliss JM, Homewood J, Harper C, Hanson J, et al. Effect of radiotherapy fraction size on tumour control in patients with early-stage breast cancer after local tumour excision: long-term results of a randomised trial. The lancet oncology. 2006;7(6):467-71
90. Hunter D, Mauldon E, Anderson N. Cost-containment in hypofractionated radiation therapy: a literature review. Journal of medical radiation sciences. 2018;65(2):148-57
91. Pignol J-P, Olivotto I, Rakovitch E, Gardner S, Sixel K, Beckham W, et al. A multicenter randomized trial of breast intensity-modulated radiation therapy to reduce acute radiation dermatitis. Journal of Clinical Oncology. 2008;26(13):2085-92
92. Shaitelman SF, Schlembach PJ, Arzu I, Ballo M, Bloom ES, Buchholz D, et al. Acute and short-term toxic effects of conventionally fractionated vs hypofractionated whole-breast irradiation: a randomized clinical trial. JAMA oncology. 2015;1(7):931-41
93. Bekelman JE, Sylwestrzak G, Barron J, Liu J, Epstein AJ, Freedman G, et al. Uptake and costs of hypofractionated vs conventional whole breast irradiation after breast conserving surgery in the United States, 2008–2013. Jama. 2014;312(23):2542-50
94. Jagsi R, Falchook AD, Hendrix LH, Curry H, Chen RC. Adoption of hypofractionated radiation therapy for breast cancer after publication of randomized trials. International Journal of Radiation Oncology* Biology* Physics. 2014;90(5):1001-9
95. Anderson SJ, Wapnir I, Dignam JJ, Fisher B, Mamounas EP, Jeong J-H, et al. Prognosis after ipsilateral breast tumor recurrence and locoregional recurrences in patients treated by breast-conserving therapy in five National Surgical Adjuvant Breast and Bowel Project protocols of node-negative breast cancer. Journal of Clinical Oncology. 2009;27(15):2466
96. Sen S, Wang S-Y, Soulos PR, Frick KD, Long JB, Roberts KB, et al. Examining the cost-effectiveness of radiation therapy among older women with favorable-risk breast cancer. JNCI: Journal of the National Cancer Institute. 2014;106(3)
97. Social Security Actuarial Life Table 2019. https://www.ssa.gov/oact/STATS/table4c6.html, accessed 5/10/2022
98. United States Cancer Statistics - Mortality: 1999 - 2016 Archive, WONDER Online Database. United States Department of Health and Human Services, Centers for Disease Control and Prevention. http://wonder.cdc.gov/CancerMort-v2016.html, accessed 5/10/22
99. Jagsi R, Griffith KA, Moran JM, Matuszak MM, Marsh R, Grubb M, et al. Comparative effectiveness analysis of 3D-conformal radiation therapy versus intensity modulated radiation

therapy (IMRT) in a prospective multicenter cohort of patients with breast cancer. International Journal of Radiation Oncology* Biology* Physics. 2022;112(3):643-53
100. Shih Y-CT, Dong W, Xu Y, Shen Y. Assessing the cost-effectiveness of updated breast cancer screening guidelines for average-risk women. Value in Health. 2019;22(2):185-93
101. Ward MC, Vicini F, Al-Hilli Z, Chadha M, Pierce L, Recht A, et al. Cost-effectiveness analysis of endocrine therapy alone versus partial-breast irradiation alone versus combined treatment for low-risk hormone-positive early-stage breast cancer in women aged 70 years or older. Breast Cancer Research and Treatment. 2020;182(2):355-65
102. Wheeler SB, Rotter JS, Baggett CD, Zhou X, Zagar T, Reeder-Hayes KE. Cost-effectiveness of endocrine therapy versus radiotherapy versus combined endocrine and radiotherapy for older women with early-stage breast cancer. Journal of Geriatric Oncology. 2021;12(5):741-8
103. Hanmer J, Lawrence WF, Anderson JP, Kaplan RM, Fryback DG. Report of nationally representative values for the noninstitutionalized US adult population for 7 health-related quality-of-life scores. Medical Decision Making. 2006;26(4):391-400
104. Halasz LM, Patel SA, McDougall JA, Fedorenko C, Sun Q, Goulart BH, et al. Intensity modulated radiation therapy following lumpectomy in early-stage breast cancer: patterns of use and cost consequences among Medicare beneficiaries. Plos one. 2019;14(9):e0222904
105. Smith BD, Jiang J, Shih Y-C, Giordano SH, Huo J, Jagsi R, et al. Cost and complications of local therapies for early-stage breast cancer. JNCI: Journal of the National Cancer Institute. 2017;109(1)
106. Shih Y-CT, Dong W, Xu Y, Etzioni R, Shen Y. Incorporating baseline breast density when screening women at average risk for breast cancer: a cost-effectiveness analysis. Annals of internal medicine. 2021;174(5):602-12
107. Grady I, Grady S, Chanisheva N. Long-term cost of breast cancer treatment to the United States Medicare Program by stage at diagnosis. The European Journal of Health Economics. 2021;22(9):1365-70
108. Suh WW, Hillner BE, Pierce LJ, Hayman JA. Cost-effectiveness of radiation therapy following conservative surgery for ductal carcinoma in situ of the breast. International Journal of Radiation Oncology* Biology* Physics. 2005;61(4):1054-61
109. Fuel Cost Calculator. https://www.calculator.net/fuel-cost-calculator.html?tripdistance=1&tripdistanceunit=miles&fuelefficiency=25&fuelefficiencyunit=mpg&gasprice=4&gaspriceunit=gallon&x=42&y=26, Accessed 5/10/22
110. U.S. BUREAU OF LABOR STATISTICS, BLS Reports, Highlights of women's earnings in 2020. 2021https://www.bls.gov/opub/reports/womens-earnings/2020/home.htm
111. Kuo T-M, Mobley LR. How generalizable are the SEER registries to the cancer populations of the USA? Cancer causes & control. 2016;27(9):1117-26
112. Schwartz NR, Flanagan MR, Babigumira JB, Steuten LM, Roth JA. Cost-effectiveness analysis of adjuvant neratinib following trastuzumab in early-stage HER2-positive breast cancer. Journal of Managed Care & Specialty Pharmacy. 2019;25(10):1133-9
113. Greenup RA, Blitzblau RC, Houck KL, Sosa JA, Horton J, Peppercorn JM, et al. Cost implications of an evidence-based approach to radiation treatment after lumpectomy for early-stage breast cancer. Journal of oncology practice. 2017;13(4):e283-e90
114. Martin L, Hutchens M, Hawkins C, Radnov A. How much do clinical trials cost. Nat Rev Drug Discov. 2017;16(6):381-2

115. Sadler GR, Ko CM, Malcarne VL, Banthia R, Gutierrez I, Varni JW. Costs of recruiting couples to a clinical trial. Contemporary clinical trials. 2007;28(4):423-32
116. Forsat ND, Palmowski A, Palmowski Y, Boers M, Buttgereit F. Recruitment and retention of older people in clinical research: a systematic literature review. Journal of the American Geriatrics Society. 2020;68(12):2955-63

# VITA

Woojung Lee earned her BS in Systems Biology from Yonsei University in South Korea. She received her PharmD from Seoul National University and is a licensed pharmacist in South Korea. She also completed her PhD in Health Economics and Outcomes Research (HEOR) from the Comparative Health Outcomes, Policy and Economics (CHOICE) Institute at the University of Washington. During her PhD, she completed her dissertation titled *Investing in Clinical Trials for Older Adults: The Value and Challenges of Older Adult-Specific Clinical Trials.* She has deep formal training in biostatistics, econometrics, pharmaco-economics and epidemiology, and health policy. Her research interests include the assessment of clinical and economic value of pharmaceuticals, real-world evidence for medical technologies, patient medication experience, and geriatric oncology.