

IMPROVING THE VALIDITY OF NONEXPERIMENTAL COMPARATIVE  
EFFECTIVENESS RESEARCH:

THE IMPACT OF CALENDAR TIME ON PRESCRIBING OF NOVEL  
CHEMOTHERAPEUTIC THERAPIES FOR STAGE III COLON CANCER

Christina DeFilippo Mack, M.S.P.H.

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Epidemiology.

Chapel Hill  
2013

Approved by:

Til Stürmer, MD, MPH, PhD

Alan Brookhart, PhD

Bill Carpenter, MPH, PhD

Anne-Marie Meyer, MSPH,  
PhD

Robert Sandler, MD, MPH

©2013  
Christina DeFilippo Mack, M.S.P.H.  
ALL RIGHTS RESERVED

## **ABSTRACT**

CHRISTINA DEFILIPPO MACK: Improving the validity of nonexperimental comparative effectiveness research: The impact of calendar time on prescribing of novel chemotherapeutic therapies for stage III colon cancer  
(Under the direction of Dr. Til Stürmer)

Oxaliplatin was rapidly adopted for stage III colon cancer treatment after FDA approval in 2004. Uncertainty remains regarding oxaliplatin's superiority to the former chemotherapeutic standard in older patients, the most affected population. The relationship between calendar time and treatment receipt during oxaliplatin's dissemination presents a challenging yet rich methodological research opportunity for comparative effectiveness research (CER).

Stage III colon cancer patients aged 65+ initiating chemotherapy from 2003-2008 were studied using U.S. population-based cancer registry data linked with Medicare claims. We examine changes in treatment receipt using a novel calendar time-specific (CTS) propensity score (PS), which allows covariate predictive values to change over time. We compare this method and a calendar time instrumental variable (IV) with traditional adjustment to enhance understanding of oxaliplatin effectiveness for reducing cancer mortality, a strong driver of all-cause mortality among stage III patients.

PSs for treatment receipt were constructed using logistic models with key components of demographics, tumor substage, grade, and comorbidities. The CTS PS was used to match oxaliplatin-treated and untreated patients within 1-year intervals. The two-level calendar time instrument was anchored at oxaliplatin's approval and based on IV strength and plausibility of assumptions. PS-matched hazard ratios (HR) were estimated using Cox models. Risk differences (RD) were derived from Kaplan-Meier

survival curves. CTS PS and IV results were compared with conventional PS-matched estimates.

Oxaliplatin use increased considerably during the study timeframe, with 8% receipt in the first time period vs. 52% in the last (N=2800). Channeling by comorbidities, income, and age appeared to change over time. The CTS PS improved covariate balance within calendar time strata and yielded an attenuated estimated benefit of oxaliplatin (HR=0.75) compared with the conventional PS (HR=0.69).

The calendar time instrument resulted in 54% compliance (N=2881). The 3-year IV RD (95% confidence interval) was -0.09 (-0.15,-0.03) favoring oxaliplatin; PS-adjusted RD was -0.04 (-0.08,-0.01).

All analyses indicated better survival among oxaliplatin-treated patients. These consistent results based on differing assumptions lend plausibility to the conclusion that oxaliplatin retains effectiveness among older stage III patients. In nonexperimental CER of emerging therapies, calendar time's role as a confounder or instrument should be carefully considered.

## **ACKNOWLEDGEMENTS**

I would like to thank Til Stürmer for the guidance that he has provided me during my time at UNC Chapel Hill. His mentorship in science and methods, as well as his high standards and continuous support, have been invaluable to my education. Alan Brookhart, Bob Sandler, Bill Carpenter and Anne-Marie Meyer, as my dissertation committee, have been incredibly generous with their time and knowledge. They have made many contributions to this work and to my advancement as a researcher. Additionally, I appreciate the inspiration and guidance of Nancy Dreyer, Charlie Poole, Michele Jonsson-Funk, Alice White, and Nancy Colvin.

I am grateful for funding support from NIH grants R01 AG023178, R01 CA124402 and T32KD07634, as well as scholarships from the Merck Foundation and the Harry Guess Memorial Fund.

This dissertation is dedicated to all those who undergo treatment for cancer.

Special love to my brave friends and family: Brionne Clary, Peter DeFilippo, Clara Lane, Vicki Lutzer, Noel Smith, and Ramon Williams.

I give my thanks and love to those who have supported me along the way:

To my husband Peter, who inspired me to pursue this dream and supported me throughout – there is no one I'd rather be on this adventure with.

To our Ethan and Lilliana, who make every moment more fun and rewarding.

To my parents, who made me feel like I could do anything I set my mind to – thank you for the endless support and unconditional love.

To Marylyn, John, Greg and Pat, the most wonderful grandparents a family could ask for – your love and care for our children, coupled with your endless energy, made this possible.

To Clara Lane, whose laughter and perspective I love and miss.

## TABLE OF CONTENTS

LIST OF TABLES.....	x
LIST OF FIGURES.....	xi
LIST OF ABBREVIATIONS.....	xii
CHAPTER 1: STATEMENT OF SPECIFIC AIMS.....	1
CHAPTER 2: BACKGROUND AND REVIEW OF THE LITERATURE.....	6
A. BACKGROUND .....	6
B. DETERRENTS TO OXALIPLATIN RECEIPT .....	10
C. EFFECTIVENESS OF EMERGING THERAPIES .....	13
D. EVALUATION OF CHANGES IN CHANNELING OVER TIME USING PROPENSITY SCORES .....	16
E. INSTRUMENTAL VARIABLES .....	19
F. SUMMARY .....	23
G. TABLES .....	26
CHAPTER 3: METHODS .....	28
A. STUDY DATA .....	28
B. STUDY POPULATION.....	30
C. EXPOSURE, OUTCOME, AND COVARIATES.....	31
a) Treatment exposure.....	32
b) Mortality outcome .....	34
c) Calendar time.....	34
d) Covariates.....	35

D. STATISTICAL ANALYSIS.....	37
E. TABLES & FIGURES .....	48
CHAPTER 4: RESULTS: Calendar Time-Specific Propensity Scores and Comparative Effectiveness Research for Stage III Colon Cancer Chemotherapy .....	56
A. INTRODUCTION.....	56
B. METHODS .....	57
C. RESULTS .....	59
D. DISCUSSION.....	62
E. TABLES & FIGURES .....	68
CHAPTER 5: RESULTS: Comparative Effectiveness of Oxaliplatin vs. 5-fluorouracil in Older Adults: An Instrumental Variable Analysis .....	74
A. INTRODUCTION.....	74
B. METHODS .....	76
C. RESULTS .....	79
D. DISCUSSION.....	81
E. TABLES & FIGURES .....	87
CHAPTER 6: DISCUSSION.....	96
A. SUMMARY OF FINDINGS.....	96
B. PUBLIC HEALTH IMPLICATIONS.....	98
C. STRENGTHS.....	100
D. LIMITATIONS .....	102
E. FUTURE RESEARCH.....	106
F. CONCLUSIONS .....	109
APPENDIX A: Administrative Codes.....	111
APPENDIX B: Covariate Balance .....	112
APPENDIX C: Sensitivity Analyses.....	114



REFERENCES..... 116

## LIST OF TABLES

Table 2.1 Results of Randomized Controlled Trials and Observational Studies Examining Oxaliplatin's Efficacy and Effectiveness .....	26
Table 2.2 NCI's Cancer Intervention and Surveillance Modeling Network chemotherapy model inputs for stage III colorectal cancer* .....	27
Table 3.1 Rationale, ascertainment and data management for covariates that are analyzed in aim 1 and considered potential confounders in aims 2 and 3.....	48
Table 3.2 Crosswalk: Correspondence of Stage III substages with AJCC V6 descriptions .....	50
Table 3.3 Crosswalk: SEER Extent of Disease Version-3 codes corresponding to the American Joint Committee on Cancer version 6 tumor and lymph node descriptions .....	51
Table 4.1 Characteristics and treatment receipt of Stage III Colon Cancer Patients in SEER-Medicare Study Population by Time Period (N=2800).....	69
Table 4.2 Mortality hazard ratios for stage III colon cancer patients treated with oxaliplatin versus 5-FU from May 2003 to April 2006 .....	72
Table 5.1 Characteristics of Stage III Colon Cancer Patients by Treatment Received (N=3660) and Calendar Time Instrument (N=2881) .....	88
Table 5.2 Comparison of Risks and Risk Differences Using Instrumental Variable and Adjusted Analyses .....	91
Table A.1 Administrative codes used to ascertain surgery to qualify patients for this study .....	111
Table B.1 Covariate balance: Percentage and percent difference of patients treated with oxaliplatin of 5-FU only by covariate between full (unmatched) population and matched cohorts generated by the conventional and calendar time-specific PS from April 2003 through May 2006 .....	113
Table C.1 Sensitivity Analyses Results for Instrumental Variable Analysis .....	115

## LIST OF FIGURES

Figure 3.1 Timing for inclusion in study population for aims 1 and 2 .....	52
Figure 3.2 Timing for inclusion in study population for aim 3 .....	52
Figure 3.3 DAG to identify confounders for the association between treatment and mortality .....	53
Figure 3.4 FDA approval history of oxaliplatin for Stage III Colon Cancer .....	54
Figure 3.5 Considerations for Calendar Time as an Instrument Variable .....	55
Figure 4.1 Calendar time periods for stage III colon cancer patients based on first date of 5-FU or oxaliplatin receipt (N=2800) and FDA approval history of oxaliplatin .....	68
Figure 4.2 Changes in channeling by covariate over study time periods (adjusted OR, 95% CI) comparing receipt of oxaliplatin with 5-FU .....	70
Figure 4.3 Comparison of covariate balance between full (unmatched) population and matched cohorts generated by the conventional and CTS PS from April 2003 through May 2006 .....	71
Figure 4.4. Estimated hazard ratios for different PS adjustment methods comparing oxaliplatin with 5-FU for prevention of all-cause mortality, across all study years and within calendar-specific time periods .....	73
Figure 5.1 Dissemination of Oxaliplatin: Receipt of Oxaliplatin vs. 5-fluorouracil for Stage III Colon Cancer by Month and Definition of Calendar Time Instrument Variable (N=3660) .....	93
Figure 5.2 Probability of Overall Survival A) by Calendar Time Instrumental Variable (N=2881) and B) with 5-FU vs. Oxaliplatin in Propensity Score-Matched Analysis (N=2732) .....	94
Figure 5.3 Comparison of Risk Differences Using Instrumental Variable and Adjusted Analyses .....	95

## LIST OF ABBREVIATIONS

5-FU	5-fluorouracil
AIDS	Acquired Immune Deficiency Syndrome
AJCC	American Joint Committee on Cancer
ASCO	American Society of Clinical Oncology
BMI	Body Mass Index
CER	Comparative Effectiveness Research
CHF	Congestive Heart Failure
CI	Confidence Interval
CISNET	Cancer Intervention and Surveillance Modeling Network
COPD	Chronic pulmonary disease
CPT	Current Procedural Terminology
CTS	Calendar Time-Specific
CVD	Cardiovascular disease
DAG	Directed Acyclic Graph
DME	Durable Medical Equipment
EOD	Extent of Disease
FDA	Food and Drug Administration
HAART	Highly Active Antiretroviral Therapy
HCPCS	Health Care Procedure Classification Codes
HMO	Health Maintenance Organization
HR	Hazard Ratio
HRT	Hormone Replacement Therapy
IOM	Institute of Medicine
IV	Instrumental Variable

K-M	Kaplan-Meier
MI	Myocardial Infarction
MOSAIC	Multicenter International Study of Oxaliplatin/5-Fluorouracil/Leucovorin in the Adjuvant Treatment of Colon Cancer
NCCN	National Comprehensive Cancer Network
NDC	National drug codes
OR	Odds Ratio
PEDSF	Patient Entitlement and Diagnosis Summary File
PS	Propensity Score
RCT	Randomized Controlled Trial
RD	Risk Difference
SE	Standard Error
SEER	Surveillance, Epidemiology and End Results
NCI	National Cancer Institute
PCORI	Patient-Centered Outcomes Research Institute
SES	Socioeconomic Status

## **CHAPTER 1**

### **STATEMENT OF SPECIFIC AIMS**

This study aims to help us understand the effectiveness of a new chemotherapeutic, oxaliplatin, compared with the former standard of care, 5-fluorouracil, for prevention of stage III colon cancer mortality in older Americans. Through this important research question, we examine the validity of comparative effectiveness methods in the context of a rapidly disseminating innovation and how predictors of oxaliplatin receipt change over time as this new treatment becomes the standard of care. This study explores the role of calendar time using advanced analytic techniques, with the goal of improving validity of comparative effectiveness research for dynamic therapies. We construct a novel propensity score (PS) which accounts for changes in confounding by indication over time, and compare this with conventional propensity score and regression methods. We then take advantage of oxaliplatin's dissemination patterns by using calendar time as an instrumental variable based on its quick uptake among oncologists. These techniques further understanding of how researchers can study novel treatments in the wake of important policy events and potentially reduce bias in pharmacoepidemiologic studies. In addition, this methods research produces comparative effectiveness estimates of oxaliplatin-containing regimens in reducing cancer mortality, a strong driver of all-cause mortality using a robust suite of analytic techniques, thereby furthering our knowledge of oxaliplatin effectiveness in older stage III colon cancer patients.

The specific aims and hypotheses of this study are:

Aim 1. Evaluate changes in channeling over time for oxaliplatin, a new chemotherapeutic for stage III colon cancer, from off-label use and subsequent FDA approval to wide dissemination into clinical practice by developing a novel calendar-time specific propensity score method.

Rationale (1): Propensity scores assign a predictive value for treatment receipt to key patient-level covariates. PSs are routinely estimated over multiple years, thereby averaging the predictive value of patient characteristics over time, even as treatment paradigms mature. This is a potential gap in PS methodology.

The setting of oxaliplatin as a treatment for stage III colon cancer presents a likely example of changing prescribing patterns. As physicians integrated this innovation into clinical practice and became more familiar with its side effects and clinical effectiveness, patient-level predictors of receipt such as comorbidities, demographics and socioeconomic status may play a changing role in prescribing. Constructing propensity score models within meaningful timeframes during oxaliplatin's rapid adoption by the health care community will assign predictive values to patient characteristics that reflect the dynamic nature of innovative dissemination. Examining these values may provide transparency into these changing prescribing patterns.

Examination of the changes in predictive value of the covariates in a CTS PS may provide insight into predictors of receipt and hint to channeling of treatment over the phases of a drug's lifecycle. This important investigation into changes in confounding by indication will inform future CER studies by developing a novel propensity score.

Hypothesis (1): Channeling of oxaliplatin based on specific patient characteristics will change over time, and the calendar time-specific PS will provide transparency into these changes.

Subaims (1):

- 1a. Identify overall predictors of specific chemotherapy receipt (oxaliplatin vs. 5-FU) in patients receiving chemotherapy.
- 1b. Construct a calendar-time specific (CTS) propensity score within meaningful time periods based on the approval history of oxaliplatin for stage III colon cancer.
- 1c. Evaluate the change in the odds ratio for receipt of oxaliplatin vs. 5-FU for key covariates based on the time of first adjuvant chemotherapy receipt for each patient to identify changes in channeling during oxaliplatin's dissemination period for stage III colon cancer.

Aim 2. Evaluate the ability of the CTS PS to control confounding when estimating oxaliplatin effectiveness for preventing all-cause mortality as compared with the former standard of care, 5-FU without oxaliplatin, in patients with stage III colon cancer.

Rationale (2): In scenarios where a treatment has a dynamic approval, safety or policy history, conventional PS methods that adjust for time and then average estimates of predictive values over an entire study period may leave a gap in control of measured confounding. The CTS PS will reduce this potentially inaccurate averaging of effects, and instead take into account the dissemination patterns of the treatment by allowing patient characteristics to predict treatment differently in separate phases of the drug lifecycle.

Hypothesis (2): By taking changes in channeling over time into account, the CTS PS will provide a more valid HR estimate than a conventionally estimated PS.



Subaims (2):

- 2a. Estimate a hazard ratio (HR) and 95% confidence interval (CI) for all-cause mortality for each calendar time period (aim 1b) and over the full study cohort using both the CTS PS and a conventionally estimated PS.
- 2b. Evaluate the ability of the CTS PS to control confounding by comparing the matched cohorts and HR (95% CI) estimates generated in 2a.

Aim 3. Evaluate potential for using calendar time as an instrument for treatment receipt to produce a valid estimate for all-cause mortality.

Rationale (3): Statistical methods such as multivariate modeling and propensity score adjustment assume no unmeasured confounding (hidden bias)<sup>1,2</sup>. In pharmacoepidemiologic studies using secondary data sources, this is often an unrealistic assumption. For example, although information on patients' comorbidities is often available in administrative data sources, important determinants of treatment receipt such as patient frailty and functional status are not captured. Instrumental variable (IV) analysis is an approach to address the problem of unmeasured confounding.

Oxaliplatin's rapid adoption into practice presents an opportunity to evaluate calendar time as an IV. In this setting, calendar time is clearly and strongly related to overall treatment receipt in the population, but may be assumed to not directly affect patients' outcomes; therefore, one could assume that calendar time is associated with changes in oxaliplatin receipt in a similar manner as random assignment.<sup>3</sup> Because oxaliplatin was generally not used in stage III colon cancer patients prior to its FDA approval in November 2004, patients that were diagnosed and receive their first chemotherapy treatment in 2003-2004 would be "randomized" to be treated using 5-FU without oxaliplatin. Patients diagnosed and treated after oxaliplatin was available would

theoretically be assigned to the oxaliplatin treatment group. In this case, patients who received a different actual treatment than predicted based on calendar time (e.g., oxaliplatin pre-approval or 5-FU alone post-approval) would approximate confounded non-compliance in a randomized controlled trial. We use an ITT analysis strategy, which expects non-compliance.

We will explore the potential for calendar time to serve as an instrument for treatment receipt in dynamic settings, adding to methodological evidence for performing CER for new treatments. Because it is based on different assumptions than prior studies which use conditioning (adjusting) for individual measured covariates, the IV analysis adds to evidence of oxaliplatin's effectiveness in older adults, who bear the greatest burden of colon cancer yet were underrepresented in clinical trials.

Hypothesis (3): Calendar time will serve as an instrument for treatment receipt and yield a valid estimate for oxaliplatin effectiveness which controls for unmeasured as well as measured confounding.

- 3a. Define a 2-level calendar time variable to act as an instrument for treatment receipt.
- 3b. Evaluate the strength of calendar time as an instrumental variable.
- 3c. Estimate risk differences for all-cause mortality using calendar time as an instrument and compare with adjusted regression models.

## **CHAPTER 2**

### **BACKGROUND AND REVIEW OF THE LITERATURE**

#### **A. BACKGROUND**

##### *Impact of colon cancer in the United States*

Colorectal cancer is the third most common cancer in the United States in both men and women, and the second most common when the sexes are combined. Colorectal cancer is a leading cause of cancer mortality, with more than 50,000 people expected to die from it this year. It has an age-adjusted incidence rate of 41.1 cases per 100,000 people per year and a lifetime risk of 5.1%.<sup>139</sup> Colon cancer is primarily a disease of the elderly, with incidence rates strongly increasing with age in both men and women and diagnoses occurring at a median age of 72.<sup>4</sup>

In 2013, over 102,480 new colon cancer cases are expected, approximately a third of which will be diagnosed as stage III.<sup>5</sup> Five-year survival rates for stage III colon cancer are 28%, 46%, and 73% for substages A, B and C respectively.

##### *Chemotherapies for stage III colon cancer*

The standard of care for stage III colon cancer is surgical resection, followed by adjuvant chemotherapy. 5-Fluorouracil (5-FU), an antimetabolite, was the standard adjuvant therapy until it was replaced by FOLFOX, a combination regimen of the newly approved platinum-based treatment oxaliplatin with 5-FU and folinic acid. In 2003, results from the Multicenter International Study of Oxaliplatin/5-Fluorouracil/Leucovorin in the Adjuvant Treatment of Colon Cancer (MOSAIC) stage III clinical trial were

released, showing that FOLFOX improved disease-free survival over 5-FU (rate of 72.2% vs. 65.3%).<sup>6</sup> In 2004, oxaliplatin was approved by the U.S. Food and Drug Administration (FDA) for stage III colon cancer and this innovative treatment began to disseminate throughout U.S. hospitals in increasing amounts. Subsequent clinical trials and nonexperimental studies confirmed oxaliplatin effectiveness (Table 2.1).

#### *Unknown treatment effectiveness among the elderly*

Colon cancer is primarily a disease of older individuals, with 65% of incident cases in the United States occurring in those aged 65 and above.<sup>4</sup> There is uncertainty, however, regarding the effectiveness of oxaliplatin in combination adjuvant chemotherapy for reducing mortality in this population. The seminal randomized controlled trial (RCT) demonstrating oxaliplatin efficacy in patients with stage III colon cancer did not include patients over the age of 75 and had a median patient age of 60,<sup>7</sup> which is twelve years younger than the median age of 72 for U.S. colon cancer patients.<sup>4</sup> Subsequent observational studies and RCTs have produced inconclusive or contradictory results among diverse sub-populations, prompting a need for additional research and robust methodologies for answering this question.

Two 2012 studies concluded that there is unknown effectiveness knowledge of oxaliplatin in elderly subgroups. Tournigand et al studied elderly subgroups from the MOSAIC RCT with the purpose of understanding oxaliplatin use in this subpopulation, due to controversy about its effectiveness.<sup>8</sup> There were only 190 stage III patients aged 70 or older enrolled in MOSAIC, with n=96 exposed to oxaliplatin. In this small subgroup, the investigators found an HR of 0.98 (95% CI, 0.62 to 1.56) and concluded that there was not a statistically significant benefit for overall or disease free survival to the use of oxaliplatin in patients between 70 and 75 years, a statement that can be extrapolated to those over 75 as well. This research, however, was limited due to the small size of the

subgroup and the restriction within MOSAIC to patients younger than 76 years old. Sanoff et al reviewed 5 RCTs and performed studies in 5 separate observational data sources. They found evidence that the addition of oxaliplatin to 5-FU was associated with a survival advantage across the randomized populations as well as the diverse practice settings represented in their data. However, they had to restrict their CER analyses to subjects younger than 75, noting that “RCTs have included too few patients older than 75 years to make robust conclusions about oxaliplatin’s efficacy in older patients” and calling for future research examining oxaliplatin in patients older than 75 years.<sup>9</sup>

Prior studies have also looked at oxaliplatin’s use and effectiveness in the elderly, and have been inconclusive. In a randomized phase III trial of XELOX versus 5-FU, 3-year disease free survival in patients younger than 70 was HR = 0.79, compared with patients older than 70 who had an HR = 0.87. Although both age groups showed an advantage for the oxaliplatin-based regimen over 5-FU only, the authors speculated that the difference may still be clinically significant.<sup>10</sup> It is important to note that these studies were generally looking at relative measures of effect, which may be skewed in older patients because as mortality increases with age, relative measures of effect would move closer to the null.

A 2009 study by the MOSAIC researchers updated RCT results with 6-year survival. These findings highlighted that patients older than 65 did not maintain the survival benefit observed with oxaliplatin in the full population. 463 patients were older than 65, and among these, the risk of death from CRC was not changed by the use of FOLFOX, thus suggesting differences in survival for older vs. younger patients.<sup>11</sup> Similarly, findings from the ACCENT database, 20% of which has patients over the age of 70 (n=2,170), showed improved overall survival, disease-free survival, and time-to-recurrence for oxaliplatin patients younger than 70 years but not in those older than 70.<sup>12</sup>

The PETACC-8 study of adjuvant FOLFOX + cetuximab restricted study entry to patients younger than 70 years of age based on interim results showing higher mortality in the older age group. These authors also noted a potentially higher adverse event rate in older individuals.<sup>13</sup>

#### *Adverse events among oxaliplatin receivers*

Overall, oxaliplatin is considered a relatively safe and effective therapy for colorectal cancer, with manageable side effects.<sup>14</sup> The single contraindication listed on the label of oxaliplatin is a “known allergy to ELOXATIN or other platinum compounds”, although caution is also advised for those on oral anticoagulants. Side effects are well-described and include neuropathy, pulmonary toxicity, neutropenia, gastrointestinal tract toxicity and hepatotoxicity,<sup>15</sup> as well as mild to moderate nausea, vomiting, and diarrhea.<sup>14</sup> Unique among these is neuropathy; oxaliplatin is commonly associated with mild, sensory, and motor axon loss that may not be reversible.<sup>16</sup> This is an adverse event that does not often occur with 5-FU alone and is more intense with oxaliplatin than with other platinum derivatives.<sup>14</sup> Oxaliplatin-induced neurotoxicity typically occurs after several cycles of therapy and is characterized by rapid-onset acute sensory neuropathy and late-onset cumulative sensory neuropathy. The symptoms occur in 40-50% of patients and are reversible in ~75% of patients with a median time to recovery of 13 weeks after treatment discontinuation.<sup>14</sup> A significant number of the patients that do recover, however, retain neuropathy symptoms for longer than 2 years after completing their oxaliplatin regimen.<sup>17</sup>

The recommended dose schedule of oxaliplatin for stage III colon cancer is intravenous administration given every 2 weeks for 12 cycles (6 months total), with dose reductions or discontinuations in patients who experience persistent neurosensory events.

## **B. DETERRENTS TO OXALIPLATIN RECEIPT**

Because the addition of oxaliplatin to 5-fluorouracil-based adjuvant chemotherapy is known to decrease the risk of cancer recurrence or death by an additional 23% for patients with resected colon cancer,<sup>5</sup> most physicians operate under the context that once the decision is made that a patient should receive adjuvant chemotherapy for stage III colon cancer, an oxaliplatin-containing regimen should be standard unless the patient is thought to be unable to tolerate it.<sup>18</sup> Justifiable reasons for not prescribing oxaliplatin may arise from informal clinical risk/benefit evaluations coupled with patient decision-making. The survival benefits of oxaliplatin combination chemotherapy must be weighed against the potential adverse effects of oxaliplatin, and in some cases, patients and/or physicians may decide to forgo the administration of this somewhat harsher treatment.<sup>19</sup> However, inappropriate administration of adjuvant chemotherapy is well described,<sup>20,21,22,23</sup> and patients who should be prescribed oxaliplatin along with 5-FU may not actually receive it. Disparities research, largely focused on receipt of any type of adjuvant chemotherapy rather than oxaliplatin specifically, has shown that certain patient characteristics, such as age, race, marital status, area of residence and SES, may serve as predictors of or deterrents to appropriate chemotherapy receipt.

### *Age*

Young age is a known predictor of adjuvant chemotherapy in general. Observational studies have shown that older patients receive adjuvant therapy at lower rates than their younger peers, with receipt dropping by as much as 78% to 34% with age.<sup>24</sup> The elderly also receive less toxic and shorter chemotherapy regimens.<sup>25</sup> This is

true for oxaliplatin specifically,<sup>26,25</sup> even though increased toxicity of oxaliplatin does not appear to vary by age and the benefits have not conclusively been shown to diminish among older patients.<sup>27,28</sup> In part, this difference in aggressive treatment may be attributed to patient preference; older patients have reported being as willing as younger patients to try chemotherapy but less willing to tolerate severe adverse effects for any given degree of anticancer benefit.<sup>29</sup> There is also prescribing difference; studies have shown that physicians may be unwilling to give aggressive chemotherapeutics to older patients due to lack of RCT evidence in older patients, high comorbidity, and drug toxicity.<sup>30,31</sup>

### *Comorbid conditions*

Comorbidity, which is often higher among the elderly, may be a valid deterrent to oxaliplatin receipt, particularly immediately after arrival to market. Clinical trials do not include patients with high comorbidities, and therefore when a drug first becomes available, physicians do not know how it will affect certain conditions. Oncologists agree that a patient's overall state of health should be considered and that comorbidity and functional status, rather than age, is a better reason to withhold oxaliplatin (one oncologist published the generally held opinion that “frail patients under 70 should perhaps not get oxaliplatin while robust patients over age 70 perhaps should get it”).<sup>18,25</sup>

### *Demographics*

Demographic factors that do not have clinical relevance for chemotherapy have also been shown to predict receipt. Studies have demonstrated that African American patients are less likely to receive adjuvant chemotherapy for resectable colorectal cancer than Caucasian Americans,<sup>32,33,34</sup> suggesting that there may be disparities in the use of the more effective and/or innovative drugs. However, racial differences in the receipt of



oxaliplatin among treated stage III colon cancer patients have not been observed in the Surveillance, Epidemiology and End Results (SEER)-Medicare population.<sup>26</sup> Predictors of oxaliplatin receipt that have been observed in SEER-Medicare include young age, female sex, and married individuals, and overall, patient characteristics, as opposed to physician or hospital factors, have appeared to influence the variation in oxaliplatin use.<sup>35</sup> Socioeconomic status has also been seen to inappropriately influence chemotherapy receipt as well,<sup>23</sup> although this is underexamined for oxaliplatin.

### *Substage*

Substage, an important pathological factor, may also be a predictor of oxaliplatin treatment. Although oxaliplatin is generally recommended for all substages of stage III colon cancer, some physicians feel that the risk of neuropathy may outweigh the survival benefit among patients with substage A, who carry a better chance of survival. For colon cancer, tumor substage within stage III disease strongly influences survival<sup>36</sup> and stage- and substage-specific survival for colon cancer does not increase linearly. A 2004 study by O'Connell et al found worse 5-year stage-specific survival for stage III vs. II (59.5% vs. 82.5%), but when examined on a more granular level, the authors found better survival among stage IIIa patients than stage II: 84.7%, 72.2% for IIa, IIb vs. 83.4% for stage IIIa; stages IIIb and c survival was increasingly worse.<sup>37</sup>

### *Physician factors*

Because the survival benefits of oxaliplatin are relatively well established overall and are integrated in treatment guidelines,<sup>38</sup> the National Cancer Institute's (NCI) Cancer Intervention and Surveillance Modeling Network (CISNET) has generated models to encourage increased uptake in oxaliplatin prescribing. In these models, which highlight racial disparities in oxaliplatin receipt based on the aims of CISNET, NCI projected that

oxaliplatin (as part of FOLFOX) receipt would increase from 0% in 2000 to over 49% in 2010, with an “optimistic but realistic” goal of 81% for all stage III colon cancer patients (Table 2.2). These numbers highlight the gap between appropriate prescribing and actual prescribing, based on both known and described deterrents to receipt. In a study examining practice setting and demographic influences on colon cancer treatment decisions, researchers have found that physician and practice setting characteristics, including organized structures such as organizational affiliation with an NCI Cooperative Group or Community Clinical Oncology Program may facilitate dissemination of new treatment standards.<sup>39,40,41</sup> Another study examining the diffusion of oxaliplatin into community practice demonstrated that exposure to these types of programs was associated with guideline-concordant treatment in general, and oxaliplatin prescribing specifically.<sup>42</sup>

## **C. EFFECTIVENESS OF EMERGING THERAPIES**

### *Nonexperimental Comparative Effectiveness Research*

The healthcare environment in the United States is experiencing supreme change, and comparative effectiveness research (CER) is playing a major role in this movement. The Institute of Medicine (IOM) has defined comparative effectiveness research as “the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels.”<sup>43</sup>

Among other CER-related high expenditure government initiatives, the Patient-Centered Outcomes Research Institute (PCORI) has been funded to establish priorities

and oversee the application of federal expenditures for CER, as well as to guide endeavors based on this research to assist providers, patients and payers to make critical health care decisions.<sup>44</sup> The establishment of PCORI illustrates the high priority placed on CER at this time of national healthcare reform. This type of research is critical for comparing treatments such as 5-FU and oxaliplatin, as the latter is known to be generally more efficacious but can cause harsh side effects. Further, FOLFOX is considerably more expensive than 5-FU alone, at \$29,000 vs. \$6,500 (based on costs for a full regimen in the MOSAIC trial).<sup>45</sup>

Although randomized controlled trials such as MOSAIC are often thought of as the “gold standard” for understanding efficacy and safety, they are not well-suited to answer questions about clinical effectiveness in a real-world, heterogeneous patient population. One major issue is that patients in RCTs are typically younger, healthier and less diverse than the general population, and therefore inferences cannot be drawn about how treatments will perform once they are available to all patients. Diseases such as colon cancer that disproportionately affect the elderly are particularly susceptible to this issue, despite being underrepresented in clinical trials. In this case, nonexperimental CER can fill the gaps left by RCTs for understanding effectiveness and providing evidence for clinical decision-making among all stage III colon cancer patients.

#### *Dissemination of new drugs*

In a report on the aims of healthcare, the IOM asserts that scientific knowledge about best patient care is not applied systematically or expeditiously to clinical practice. According to the Department of Health and Human Services, physician decision making should be evidence-based and patients should receive care based on the best available scientific knowledge, rather than receiving care that varies illogically between clinicians, organizations, or geographies.<sup>46</sup>

On average across disease areas, it takes 17 years for new knowledge generated by RCTs to be incorporated into practice.<sup>46</sup> Few studies have investigated this, and due to the different intensity of symptoms, frequency of clinical care and treatment availability between diseases, dissemination timelines can be expected to vary widely.

Chemotherapeutics are known to disseminate in a unique manner due to the high mortality rates of cancer, and cancer drugs are commonly used off-label and approved for new indications more rapidly than chronic disease drugs. Additionally, oncologists are likely to make practice changes based on conference abstracts or study results rather than full publications regarding new treatments.

Little applicable research has been done to measure the amount of time it takes for new oncology treatments to fully disseminate in the US. McKibbin et al<sup>47</sup> performed a multi-center, retrospective chart review at 11 US oncology clinics which focused on chemotherapy utilization in colorectal cancer patients. They analyzed initial and subsequent chemotherapy use of bevacizumab in the years after FDA approval for advanced colorectal cancer (2003-2006) and found that bevacizumab usage rose to 51% utilization 6 months after FDA approval. Oxaliplatin usage went from 23% to 67% in 2.5 years. The researchers concluded that community oncologists rapidly adopt new chemotherapy regimens into practice.

In 2002, Mariotto et al<sup>48</sup> examined the 1980's dissemination of adjuvant tamoxifen use for post-menopausal breast cancer patients. They found that tamoxifen use reached the 50% level 5 years after publication of Nolvadex Adjuvant Tamoxifen Trial results, and the adoption of adjuvant polychemotherapy took a similar time course. The authors concluded that their results implied a rapid increase in the use of this then-innovation in the 5 years after the 1985 recommendations for tamoxifen use. It is important to note, however, that continuing education and communication among the

physician community has changed drastically since the 1980's with the advent of the internet and other information technology.

In 2005, Buzdar et al<sup>49</sup> used a similar setting to examine the influence of RCT results on hormonal therapy dissemination among post-menopausal women with hormone receptor-positive early breast cancer. Based on efficacy results comparing the new treatment anastrozole with tamoxifen, they found that anastrozole became the dominant adjuvant hormonal therapy for postmenopausal women within 2 years of initial presentation of trial results. They concluded that a major change in clinical practice of adjuvant hormonal replacement therapy occurred and that the dissemination of key clinical data, accompanied by professional commentary and regulatory actions, can rapidly influence the clinical practice of medical oncologists.

#### **D. EVALUATION OF CHANGES IN CHANNELING OVER TIME USING PROPENSITY SCORES**

##### *Channeling*

Channeling is the differential use of a drug by patients with similar profiles and health status. Also known as confounding by indication, channeling occurs when a drug and its comparator are selectively prescribed based on differences in patient profiles.<sup>4</sup> This bias can lead to invalid comparisons of drug effectiveness, particularly for dynamic therapies such as oxaliplatin, which moved relatively quickly from off-label use to FDA approval and subsequent widespread treatment in stage III colon cancer. In cases such as this, differential prescribing by treating physicians based on patient characteristics such as age, substage, socioeconomic status (SES) or comorbidity may change over time as drug information becomes common knowledge or prices change. Drugs that have recently been FDA approved or that have had publicized safety issues, such as

black box warnings added to the label, are often channeled differently toward certain patient groups over the treatment's lifetime.

### *Propensity Scores*

Propensity score theory originated in the 1980's with Rosenbaum and Rubin in an effort to control for selection bias in observational cohort studies. These methods are now commonly used for confounding control in outcome studies. In pharmacoepidemiology, a propensity score is defined as the probability of treatment receipt given observed covariates.<sup>50</sup> A propensity score model estimates the predicted probability of treatment receipt based on the patient's demographics, health status, and other measured characteristics.<sup>51</sup> Under the assumption that all relevant confounders are measured and included in this model (e.g. no unmeasured confounding), a PS can be used to control confounding through matching, stratification, inverse probability of treatment weighting, or adjustment.

This research focuses on PS adjustment through matching, which provides a useful estimate of treatment effectiveness by comparing patients who are potential candidates for the treatment and have similar observed characteristics to those actually treated. Matching aims to adapt the distribution of covariates in the unexposed to those in the exposed by matching unexposed patients with similar probabilities of treatment receipt to the treated patients. This method may lead to the exclusion of exposed patients for whom no unexposed match can be found, thus effectively reducing the size of the treated cohort. Similarly, unexposed patients that do not match are excluded; this, however, is only a minor issue in analyses. Several matching algorithms are available, one of the more common being greedy matching. This creates a match based on the PS and does not reconsider once the match is made.<sup>52</sup> PS-matched analyses produce a

causal estimate that generalizes only to those who received treatment, rather than the full study population.<sup>2</sup>

Propensity scores are often estimated as one value for a study where patients enroll and initiate treatment over multiple calendar years. This may encompass one or more policy-related events for a particular drug. In scenarios where a treatment has a dynamic approval, safety or policy history, estimating a PS over a multi-year study period may leave a gap in control of measured confounding because covariate effects for treatment receipt are averaged across all the years of the study. For the many drugs that experience dynamic changes in use and prescribing, this averaging of effect is inaccurate. New therapeutics in particular often disseminate unevenly throughout the patient population as physicians first become aware of the drug and then learn about its side effects and effectiveness. In instances such as this, where a treatment is channeled differently toward certain patient groups as physicians fine-tune their prescribing of it, propensity score methods should take these changes in prescribing pattern into account. However, pharmacoepidemiologists often use PSs for adjustment in dynamic settings, often without accounting for changes in guidelines and prescription paradigms.

#### *A PS to capture changes in channeling*

Construction of a calendar time-specific (CTS) PS that allows the predictive value of the covariates to change over time may provide a less biased method of estimating effectiveness of new-to-market drugs. Examination of the changes in predictive value of the covariates in a CTS PS may provide insight into predictors of receipt and hint to channeling of treatment over the phases of a drug's lifecycle.

This has been previously implemented by Seeger et al in an examination of statin use and myocardial infarction.<sup>53,54</sup> The authors examined second- to first-line clinical changes in statin use and found that the association between prior non-statin lipid

lowering drug use and statin initiation changed over the course of the study, and that imbalance between compared groups resulting from the change in use over could lead to confounded effect estimates. They accounted for changes in statin use by estimating PSs and matching within half-year time blocks, then stratifying Cox models by these blocks. They found that allowing flexibility in PS estimation showed changes in drug use over time, but did not compare effect estimates with a conventional PS.

In a similar setting of statin use and MI, Rassen et al<sup>55</sup> assessed the performance of an overall PS within subgroups based on patient characteristics and found larger effect differences for small subgroups and few exposed outcomes. They did not, however, look specifically at calendar time. Based on their results, an investigation into treatment-receipt subgroups is warranted.

In research examining statins and MI, Schneeweiss et al<sup>56</sup> noted the particular importance of CER evidence immediately after FDA approval and highlighted the methodological challenge of bias due to confounding by indication of new medications. They describe a sequential cohort study which proposes matching within quarterly or monthly blocks, a recommendation that is in line with creating a CTS PS.<sup>57</sup> The authors do not explicitly compare alternative approaches, such as a conventionally estimated PS, to calculate the differences in estimates.

This research expands upon this work comparing estimates, quantifying bias and defining time periods using policy-related time points.

## **E. INSTRUMENTAL VARIABLES**

Instrumental variable analysis is an econometric method that differs from multivariate or propensity score adjustment in that it does not assume no unmeasured confounding. Instead, a strong instrument may be able to control for unmeasured



confounding through an intense association with the exposure and little to no effect on the outcome other than through its association with the exposure.<sup>68,14,58</sup> This presents a strength in pharmacoepidemiologic studies using secondary data sources, as unmeasured confounding is often an unrealistic assumption. For example, although patient comorbidities are often available to some extent in administrative data sources, important determinants of treatment receipt such as patient frailty and functional status are not captured.

#### *IV assumptions*

IV analyses assume that the instrument 1) strongly affects or is associated with the exposure, 2) is related to the outcome only through its association with the exposure and 3) is unrelated to patient risk factors for the outcome. Instrumental variables generalize to a marginal population, or “compliers” – those whose treatment status depends on the instrument, rather than those who would be treated with the same drug irrespective of the IV (e.g., a patient who is very close to death and would receive the easier-to-tolerate treatment in all levels of the instrument).

#### *History of IVs*

IVs originated in the 1920's and gained research momentum in the 1970's, when Heckman pioneered thinking about sample selection bias and created the 'Heckman two-step procedure', for which he won the Noble Prize.<sup>59</sup> Because IV are difficult to find and many do not feel comfortable with making the necessary assumptions, they have not experienced wide application. McClellan et al<sup>60</sup> used the distance from patients' home addresses to the closest hospital performing coronary interventions as an instrument to estimate the effect of coronary interventions for acute myocardial infarction on mortality. They did this under the presumption that distance to such a hospital

predicts whether a patient with AMI gets a coronary intervention but that this distance would be unrelated to patient factors and not affect mortality otherwise. These researchers found that differential distances to certain types of hospitals are strong independent predictors of treatment intensity, regardless of patient-specific health status. Their finding that catheterization and other revascularization procedures have a modest or possibly no benefit on mortality reduction was particularly important to the IV methods literature due to its difference from previous observational study results, which were likely biased by confounding by indication. Brookhart et al<sup>61</sup> investigated physician-specific prescribing preference as an instrument in the setting of COX-2 inhibitors and gastrointestinal complications and found that use of this instrument substantially reduced unmeasured confounding bias. Stukel et al compared the use of regional catheterization rate as an instrument with multivariable adjustment, PS risk adjustment and PS matching. In their study estimating cardiac catheterization and mortality among elderly Medicare AMI patients, Stukel et al concluded that the instrumental variable analysis appeared to produce less biased effectiveness estimates. In a commentary to Stukel's work, d'Agnostino and d'Agnostino<sup>62</sup> noted that the IV method is useful in certain circumstances and when properly performed. They highlighted the difficulty of finding a suitable instrument and cautioned that it may be difficult to determine which estimate is most believable when comparing adjusted and IV results, and RCTs do not always provide a gold standard comparison. The d'Agnostinos also noted that the instrument should be considered as both a confounder and an instrument in these types of studies.

Pirracchio et al<sup>63</sup> compared PS and regression analyses with IV methods in a study to estimate the benefit of intensive care unit admission on mortality. Using physician specialization as the instrument, the authors found that the IV analysis satisfactorily controlled selection bias in their setting of unmeasured confounding, although the imprecision of these methods presented a disadvantage.

### *Calendar time as an instrument*

Calendar time has been used as an instrument by Cain,<sup>64</sup> Johnston,<sup>65</sup> Shetty,<sup>66</sup> and Zhang<sup>67</sup> et al. These studies illustrate the utility of an IV based on time in settings where trends in medication use create a natural experiment that can be used to strengthen clinical evidence.

Cain et al used calendar time defined by highly active antiretroviral therapy (HAART) availability (pre- and post-1996) as an instrumental variable to study HIV disease progression. The authors found that using calendar time as an IV in the setting of HIV and HAART effectiveness may resolve discrepancies between observational studies and RCTs, specifically in its ability to exposure misclassification common in these types of studies.

Shetty et al used timing surrounding the July 2002 release of Women's Health Initiative data to investigate effectiveness of hormone replacement therapy (HRT) on cardiovascular outcomes. HRT use dropped sharply over a small period of time due to the public release of results showing that HRT increased CVD risk, rather than reducing, per the conclusion of RCT results. The researchers found that decreased HRT use did not reduce acute stroke rate but may be associated with a lower acute myocardial infarction rate. They concluded that calendar time provided a suitable natural experiment in this setting where bias presented an issue for adjustment methods.

Zhang used the timepoint of olanzapine's FDA's approval to look at effectiveness and spending. He concluded that calendar time as an IV could be useful in testing for cost-offsets in other clinical areas and drug classes.

Johnston et al used time period as an instrument to estimate the effectiveness of beta-blocker therapy to prevent all-cause mortality after heart failure hospitalization in data that did not include disease severity, a strong confounder in this setting. They

created a binary variable to indicate patients treated in the hospital pre- and post-1998, a year that a guideline reversal was issued recommending beta blockers for treatment of heart failure. Their results supported the use of calendar time as an IV, as their findings suggested that beta-blocker therapy was associated with decreased mortality and that regression results may have been underestimated due to the unmeasured confounder of disease severity.

Together, these studies illustrate the potential methodological advantage that calendar time presents in dynamic prescribing settings. In our setting, calendar time is clearly and strongly related to overall receipt of oxaliplatin in patients with stage III colon cancer receiving chemotherapy, yet can be assumed to not directly affect patient outcomes. Calendar time, therefore, could theoretically be associated with changes in oxaliplatin receipt in a similar manner as random assignment.<sup>68</sup> Creating an instrument based on calendar time can account for both the unmeasured confounding inherent in administrative database studies and potential present in any data source, as well as the underlying changes in channeling over time based on measured covariates. This might be appropriate when secular trends in medication use are present, often centered around FDA action, publicized safety issues, or major policy, guideline or reimbursement changes. We explore the ability of calendar time to serve as an instrument for treatment receipt and compare the resulting estimate of oxaliplatin effectiveness with propensity score-adjusted estimates.

## **F. SUMMARY**

This research improves the validity of nonexperimental comparative effectiveness research by examining the impact of calendar time on the prescribing of the novel chemotherapeutic oxaliplatin for stage III colon cancer and exploring the optimal method of adjusting for dynamic elements of treatment receipt. We focus on

propensity score and instrumental variable methods for controlling bias in the environment of a new-to-market innovation. Additionally, this research adds evidence to our understanding of chemotherapeutic treatment effectiveness for stage III colon cancer in the real-world, particular in older individuals, and identifies potential deterrents to receiving the gold-standard treatment oxaliplatin at various stages of its diffusion into the standard treatment of stage III colon cancer.

### *Methodological significance*

Estimating propensity scores using a time-specific method to reduce bias has not been examined at this level of detail in previous studies and is a novel concept for observational studies, particularly those that are examining innovative drugs or drugs with safety warnings. If changes in likelihood and determinants of treatment receipt over time continue to be neglected in PS estimation, misspecification of the propensity score model and estimations of biased treatment effects are likely. The use of calendar time as an instrument compared with propensity score adjustment has not been examined in this setting of a rapidly emerging chemotherapeutic, where changes in confounding by indication has been described over time specifically for the target population.

We compare these advanced methods with conventional PSs that both adjust for calendar time as a confounder, do not adjust for time at all, and account for time as a modifier of covariate effects on treatment receipt. In doing so, we illustrate that in settings where calendar time plays an important yet potentially unknown role, constructing multiple models to understand how it performs is important. We underscore that in cases where calendar time is an instrument, a PS that does not adjust for time at all would be preferable over the CTS PS or any PS that accounts for time. Exploring this comprehensive set of results based on instrumental variable methods, the CTS PS and several conventional regression models that diversely account for calendar time,

contributes to a comprehensive picture of how calendar time can be best accounted for and potentially taken advantage of in dynamic settings.

As CER continues to gain momentum as a necessary type of pharmacoepidemiologic research, development of improved methods is essential. Observational research is vulnerable to biases based on the fact that it is based purely on real-world treatment decisions. Channeling bias and the presence of unmeasured confounding are two methodological issues central to issues of validity in nonexperimental CER. Calendar-specific PSs and instrumental variable approaches are advanced methods that have the potential to reduce bias in nonexperimental CER studies, particularly for dynamic drugs, albeit based on different assumptions. These methods will be transferable to any disease where the probability of receipt of a drug changes quickly over time.

#### *Therapeutic significance*

Although oxaliplatin's superiority to 5-FU is well-established in the general population, its effectiveness in older patients, who are most affected by stage III colon cancer, is controversial. This research adds evidence to our understanding of oxaliplatin effectiveness in this important patient population by estimating results based on different assumptions. Additionally, by identifying changes in oxaliplatin's clinical use in this population as it disseminates into clinical practice, this research may provide insight into patient characteristics which merit early receipt of new chemotherapeutics in the U.S.

## G. TABLES

**Table 2.1 Results of Randomized Controlled Trials and Observational Studies Examining Oxaliplatin's Efficacy and Effectiveness**

<b>Trial or Observed population (PI)</b>	<b>HR, 95% CI for overall survival</b>
MOSAIC RCT (André et al <sup>69</sup> )	0.80 (0.65, 0.97)
NSABP C-07 (Kuebler et al <sup>70</sup> )	0.85 (0.72, 1.00)
Pooled RCTs (Sanoff et al <sup>71</sup> )	0.80 (0.70, 0.92)
SEER-Medicare (Sanoff et al)	0.70 (0.60, 0.82)
NYSCR-Medicare (Sanoff et al)	0.58 (0.38, 0.90)

**Table 2.2 NCI's Cancer Intervention and Surveillance Modeling Network chemotherapy model inputs for stage III colorectal cancer\***

Race group	Chemotherapy Input**	2000 level	Level in 2010	
	Stage III Colorectal Cancer Patients		Projected Trends	Optimistic but realistic
White	% not receiving chemotherapy	26%	23%	19%
	% receiving 5-FU	74%	20%	0%
	% receiving FOLFOX	0%	57%	81%
Black	% not receiving chemotherapy	37%	35%	19%
	% receiving 5-FU	63%	16%	0%
	% receiving FOLFOX	0%	49%	81%

\*These model inputs are used as part of NCI's Cancer Intervention and Surveillance Modeling Network (CISNET) efforts to help cancer control planners, program staff and policy makers consider the impact of risk factor reduction, increased early detection, and increased access to optimal treatment on future colorectal cancer mortality rates through simulation modeling of colorectal disease progression in a population with the characteristics of the U.S. population<sup>72</sup>

\*\* Average over all ages



## **CHAPTER 3**

### **METHODS**

This methods research and comparative effectiveness analysis draws qualifying patients from the SEER-Medicare linked database. Separate but similar cohorts were used for aims 1 and 2 vs. aim 3. These differ based on the years of available SEER-Medicare data and the study inclusion / exclusion criteria. The University of North Carolina Office of Human Research Ethics (Study number 12-0139) approved this study.

#### **A. STUDY DATA**

This research is performed in a retrospective cohort using existing nationally-funded registry data linked with insurance claims. The use of retrospectively collected registry data is appropriate for these study questions, as it is a time and cost-efficient way of recruiting a large sample size necessary to address a clinically relevant endpoint such as mortality. In addition, it is the ideal way to look at treatment predictors and dissemination in the real-world.

All aims (1-3) utilize a population-based cohort drawn from SEER-Medicare data. The National Cancer Institute's (NCI) SEER program collects clinical and demographic data on incident cancers and covers approximately 28% of the U.S. population.<sup>73</sup> Data element descriptions for the SEER registry are standardized by NCI and defined in the Patient Entitlement and Diagnosis Summary File (PEDSF) documentation. Healthcare providers give mandatory reports of patient data through region-level registries to SEER, which has a 98% standard for complete case ascertainment. Each registry re-abstracts

medical records yearly for a sample of cases to evaluate the accuracy of each of the data elements collected from the records.<sup>74</sup> Medicare is U.S. public health insurance managed by the Centers for Medicare and Medicaid Services that insures approximately 97% of those aged 65 and older in the U.S.<sup>75</sup> Data are collected through reimbursement claims submitted for inpatient (Part A), outpatient (Part B), and medical supplies (Part B).

The linkage of SEER with Medicare claims Parts A and B provides critical information on patient demographics, cancer characteristics, procedures, inpatient medications, comorbidities, treating facilities, physicians, and all-cause mortality. These data have been used extensively to study cancer patterns of care and outcomes. The two datasets are linked through individual identifiers submitted by participating registries, which are then matched with identifiers contained in Medicare's master enrollment file.<sup>76</sup>

SEER-Medicare contains the large, diverse set of patient participants required by the specific aims outlined in this proposal. Medicare's universal availability to all U.S. citizens older than age 64 provides an appropriate population to study colon cancer, which has a median age of 72 and a risk which increases with age.<sup>139</sup> Because Medicare covers 98% of people aged 65+, the population described above is a scientifically valid population for studying this disease in older Americans. SEER registries are geographically diverse and collect a large amount of data. These registries include patients treated by many types and quality-levels of facilities and physicians, allowing for the study of treatment receipt to be generalizable across the U.S. The patient cohort in this dataset is comprehensive and diverse, and patient selection is relatively unbiased. As these data are available for the relevant years of the oxaliplatin lifecycle, namely from 2002 when its efficacy was tested in stage III colon cancer through its widespread diffusion through the pharmaceutical marketplace, this data is timely for these study questions, with a low cost compared to prospective sources.

The SEER-Medicare dataset for Aims 1 and 2 of this research contains stage III colon cancer diagnoses from 2002-2005 linked to Medicare part A and B claims from 2002-2007. Aim 3 used an extended dataset containing diagnoses from 2002-2007 linked to Medicare part A and B claims from 2002- April 2010.

## **B. STUDY POPULATION**

### *Aims 1 and 2*

The study cohort for aims 1 and 2 sources from U.S. elderly cancer patients living in a SEER registry catchment and includes individuals diagnosed with primary stage III colon cancer between January 1, 2003 and December 31, 2005. Included patients must receive a curative surgery within 90 days of diagnosis and initiate the first course of chemotherapy within 90 days of surgery and 120 days of diagnosis (Figure 3.1). The latter window allows for a max of 150 days from day 1 of the month of diagnosis, in order to conservatively contains treatment timelines recommended by American Society of Clinical Oncology / National Comprehensive Cancer Network quality standards.<sup>38</sup> Qualifying surgeries include resection or excision of colon or large intestine or partial or total colectomy (Appendix A Table 1). Subjects are restricted to those aged 65 and older at the time of diagnosis, because patients enrolled in Medicare that are younger than 65 years of age may be systematically different as a function of different Medicare eligibility requirements. All patients received one of the two treatments of interest: oxaliplatin (used to define the FOLFOX regimen) or 5-FU without oxaliplatin. Exclusion criteria and rationale include: Non-continuous Part A and B Medicare coverage for 12 months pre- and post-diagnosis (or until death) or any health maintenance organization (HMO) coverage during this window, as one or both of these disallow attainment of treatment and procedure claims; diagnosis at autopsy; cancer of

the rectum and rectosigmoid junction, as these patients usually require radiation and evidence to support adjuvant oxaliplatin for these does not exist; receipt of radiation, as this is not standard of care for colon cancer; site of appendix, anus, anal canal, or anorectum, as treatment paradigms for these cancers are vastly different from cancer in the colon; and death within 30 days of surgery, as these patients are unlikely to have received chemotherapy due to fragile health status.

### *Aim 3*

The study cohort for aim 3 included individuals diagnosed with primary stage III colon cancer between January 1, 2003 and December 31, 2007. Included patients must receive a curative surgery within 90 days of diagnosis and initiate the first course of chemotherapy within 110 days of surgery and 120 days of diagnosis, for the reason detailed above (Figure 3.2). Initiation of the first course of chemotherapy was extended for aim 3 because 2.5% (n=242) patients were starting chemotherapy between 90 and 110 days post-surgery. Based on expert opinion, this time period is within the range for typical patients and does not signify patients with unknown issues or disagreement with chemotherapy. Qualifying surgeries included resection or excision of colon or large intestine or partial or total colectomy; the code list is shown in Appendix A Table 1. Similar to the criteria for aims 1 and 2, subjects were restricted to those aged 65 and older at diagnosis and were required to receive either oxaliplatin or 5-FU. Exclusion criteria were the same as described for aims 1 and 2.

## **C. EXPOSURE, OUTCOME, AND COVARIATES**

The validity of this study depends on the correct assignment of chemotherapy treatment (exposure) and mortality (outcome), as well as confounding variables.

Chemotherapy, comorbidities, procedures and provider affiliations were identified through Health Care Procedure Classification Codes (HCPCS), National drug codes (NDC), Current Procedural Terminology (CPT) and ICD-9 codes in the following Medicare claims files: Carrier Claims (formerly, Physician/Supplier Part B or NCH), Outpatient claims (Part B), and the Durable Medical Equipment (DME) files.

#### **a) Treatment exposure**

The treatment groups for the propensity score (aim 1) and the exposed group for the mortality analyses (aims 2 / 3) were defined by the presence of an oxaliplatin claim (J9263 or C9205 over full study and J9999 from 9/1/02-12/31/03). The reference group is categorized as the presence of a 5-FU (J9190, 00013-1036-91, 63323-\*117-10) or capecitabine claim (aim 3 only; J8520 or J8521) and no presence of oxaliplatin within 30 days of 5-FU. General administration codes (J9925, 61517, 96400, 96401, 96402, 96405, 96406, 96408, 96409, 96411, 96412, 96413, 96415, 96416, 96417, 96420, 96422, 96423, 96425, 96440, 96445, 96446, 96450, 96542, 96549, G0355, G0357, G0358, G0359, G0360, G0361, G0362, G0363, Q0083, Q0084, Q0085, S9329, S9330, S9331, 4180F, G8372) were assumed to be 5-FU or capecitabine and listed in the referent group, as these drugs are so inexpensive that the administration rather than the drug is often billed.

Capecitabine was approved in June 2005 for stage III colon cancer,<sup>77</sup> and therefore was considered equivalent to 5-FU for both the exposed and unexposed groups in Aim 3. Capecitabine is an inactive-form oral tablet that is enzymatically converted to 5-fluorouracil in the tumor. Equivalence for both single agent capecitabine and capecitabine in combination with oxaliplatin (XELOX) has been shown in multiple trials. For the former, the Xeloda in Adjuvant Colon Cancer Therapy trial, which compared 5-FU/LV with oral capecitabine for 6 months in stage III colon cancer, found

disease-free survival in the capecitabine arm was at least equivalent to the control arm (HR 0.87;  $P$  0.001 for noninferiority).<sup>78</sup> The superiority of XELOX over 5-FU/LV alone was shown by Haller et al in 2011, with 3-year disease free survival of 71.0% vs. 67.0% (HR, 95% CI of 0.80, 0.69-0.93) and overall survival of 0.87, 0.72-1.05.<sup>79</sup>

Claims for chemotherapies other than oxaliplatin and 5-FU do not change the exposure category and are not used in the analysis. We also ignored oxaliplatin claims that occurred greater than 30 days after 5-FU/capecitabine receipt, because in this intent-to-treat analysis we are only interested in the first treatment received. However, because late receipt of oxaliplatin may suggest that these individuals had a recurrence or were too sick to initially receive oxaliplatin, a sensitivity analysis excluding these patients ( $n=46$ ) from the analysis was performed for the IV analysis (aim 3).

In addition to being claimed within 90 or 110 days of surgery (aims 1/2 and aim 3, respectively), the defining chemotherapies must also be claimed within 150 days of the first day of the diagnosis month. This ensures that treatment timing is in line with American Society of Clinical Oncology (ASCO) / National Comprehensive Cancer Network (NCCN) quality measures which state that adjuvant chemotherapy should be administered within 120 days of diagnosis for stage III colon cancer.<sup>38</sup> Because SEER does not contain the specific diagnosis date (month/year only) and claims dates may vary slightly from actual date of receipt, the 30 day extension from 120 to 150 days ensures that applicable claims are found and therefore patients treated to this standard timeframe are captured.

Errors in treatment assignment may occur if claims are not submitted through Medicare. To avoid this, individuals who are covered by an HMO or other insurance plan in addition to Medicare are excluded. Misclassification can occur for claims that are filed incorrectly. These errors occur at random and would be nondifferential for mortality.

**b) Mortality outcome**

All-cause mortality is the outcome of interest for aims 2 and 3. Mortality information, including date of death, is derived from Medicare, which is ascertained through the U.S. Social Security (SS) Administration.<sup>80</sup> SEER also collects mortality information, and the SEER mortality was cross-checked with social security records. A status of “deceased” in either database was considered the mortality status of the patient. Patients were assumed to be alive if death was not reported, and therefore there was no loss to follow-up for the mortality outcome. This assumption is valid, as once enrolled in Medicare, patients do not withdraw regardless of whether they attain additional insurance coverage. Medical facilities that report patient data to SEER have 2 years to submit patient data; this delay did not cause an issue because all 2003-April 2010 data was acquired by the time the cohort was finalized in May 2012.

Mortality information was available through December 31, 2007 for aim 2 and April 30, 2010 for aim 3.

**c) Calendar time**

Month of each patient’s first treatment receipt provided the calendar time categories for the propensity score in aim 1 and 2 and dictated the instrumental variable level in aim 3. Patients were assigned to time periods and instrumental variable levels based the date that they received the first qualifying chemotherapy (5-FU only or oxaliplatin), as defined in the “Treatment exposure” section (a) above. For oxaliplatin receivers that received 5-FU prior to oxaliplatin, the date of oxaliplatin receipt was used.

The time periods for aims 1 and 2 are a 3-level categorical variable based on 1-year time periods during the study, from January 1, 2003 through December 31, 2007.

The method for choosing the specific dates around the IV levels is detailed in the Statistical Analyses section, Aim 1.

The calendar time instrumental variable in aim 3 is based on two distinct calendar time periods. These delineate patients who receive treatment before FDA approval of oxaliplatin vs. those who receive treatment when oxaliplatin is a more standard treatment. The definition of the IV takes into account oxaliplatin's adoption curve to maximize the difference in prescribing, the strength of the instrument created by a set of cutpoints and patient exclusions, and the instrument's ability to uphold all IV assumptions. The method for choosing the specific delineations for the calendar time instrument are detailed in the Statistical Analyses section, Aim 3.

#### **d) Covariates**

##### *Surgery*

Date of qualifying surgery is a key covariate that was used to define the study population. To be included in this study, patients needed to undergo a curative surgery within 30 days before or 90 days after diagnosis, the timeframe during which surgery is typically performed. Qualifying surgeries included colectomy or resection of colon or large intestine (Appendix A Table 1). Although this relatively narrow timeframe may exclude a percentage of patients who delay surgery, receipt of surgery after 90 days could be a sign of metastatic disease, and therefore the more restrictive criterion may protect against erroneous categorization of stage (IV instead of III) in SEER. We allow surgery to take place 30 days before the first day of the month of diagnosis because colon cancer diagnosis may occur during a surgery for a separate issue, such as diverticulitis, and claims delay combined with lack of a diagnosis day in SEER necessitates a wide window. Sensitivity analyses from past studies on this dataset have



shown that expanding the cohort to include surgery receipt within 180 days instead of 90 days does not impact results.<sup>26,26</sup>

### *Cancer Stage*

Stage is a key definer of the population of interest. This variable is limited in its susceptibility to stage migration,<sup>81</sup> as well as the potential for misclassification of stage IIb to stage IIIa.<sup>82,83</sup> Although staging guidelines are revised periodically, American Joint Committee on Cancer (AJCC) TNM Staging System adhered to the 6<sup>th</sup> edition guidelines from 2003-2009, which envelopes all diagnoses in the study period of interest for this research and indicates that staging for patients in these analyses is likely consistent.

### *Patient characteristics*

Sex, race, urbanity, age at diagnosis, census median income, number of months of state buy-in, census percentage of non-high school graduation, tumor substage, tumor grade, individual comorbidities of congestive heart failure (CHF), acute myocardial infarction (MI), chronic pulmonary disease (COPD), cardiovascular disease (CVD), diabetes, cerebrovascular disease, peripheral vascular disease, AIDS, dementia, diabetes with complications, mild liver disease, moderate/severe liver disease, past myocardial infarction (old MI), paralysis, moderate/severe renal disease, rheumatologic disease, ulcer disease, cirrhosis, and physician organizational affiliation with a National Cancer Institute (NCI) Cooperative Group containing a colon cancer research portfolio will be the covariates of interest in the propensity scores and regression models for all aims. Table 3.1 shows all covariates and their origin. Comorbidities were identified based on those included in the NCI-Combined Comorbidity Index.<sup>84,85</sup> Adjustment covariates for mortality analyses were identified through literature review, expert opinion and Directed Acyclic Graph (DAG) methodology (Figure 3.3). Variables were analyzed to

ensure that ranges and distributions were as expected. All patient-level covariates were measured within 12 months before diagnosis, and were not updated after the time point at which treatment is assigned. Measurement methods were consistent through the study period for all variables except substage.

Tumor substage was defined differently within the SEER registries for patients diagnosed in 2003 compared to those diagnosed from 2004-2007. Substage calculations for the years prior to 2004 were done using the extent of disease (EOD) and number of node information, as detailed in the crosswalk provided in Tables 3.2 and 3.3. This substage assignment for 2003-diagnosed patients is understood to be consistent with the substage variable that was directly documented in the SEER registry data from 2004-2007.

## **D. STATISTICAL ANALYSIS**

All analyses will be performed using SAS (version 9.2.; SAS Institute, Cary, NC).

### **Aim 1: Evaluate changes in channeling through development of CTS PS**

#### *Predictors of chemotherapy receipt*

Expert knowledge and literature review informed our use of DAG methodology to identify potential confounders between treatment receipt and mortality (Figure 3.3). Age at diagnosis, race, sex, urbanity, tumor grade, tumor substage, socioeconomic status /income, individual comorbidities from the Charlson comorbidity index<sup>85</sup> (congestive heart failure, acute myocardial infarction, chronic obstructive pulmonary disease, cardiovascular disease, diabetes, cerebrovascular disease, peripheral vascular disease, AIDS, dementia, diabetes with complications, mild liver disease, moderate/severe liver disease, past myocardial infarction, paralysis, moderate/severe renal disease, rheumatologic disease, ulcer disease and cirrhosis), functional status, psychological

factors, body mass index (BMI), infection, nutrition, surgical technique, quality of care/physician expertise and time period were identified as important covariates. Among these, psychological factors, BMI, infection, nutrition, and functional status (which encompasses frailty) beyond comorbidities claimed in Medicare are unmeasured and were not possible to directly examine in this thesis. Residual confounding due to this potential unmeasured confounding may remain an issue in adjusted analyses, and this is a limitation of aim 2 that is common to studies performed in claims data. These were further considered in aim 3 when bias due to unmeasured confounding is discussed.

SES/income was represented in the data by census median income (aims 1, 2) or Medicare state buy-in and census percentage of high school graduates (aim 3). Quality of care/physician expertise was represented by physician organizational affiliation with a National Cancer Institute (NCI) Cooperative Group containing a colon cancer research portfolio.<sup>40,41</sup>

Frequency tables and comparisons of these DAG-identified covariates was examined by treatment group to further understand which are functioning as confounders between treatment and mortality. We describe differences in covariate distribution between groups at the time of first treatment for all key covariates. Cox proportional hazards models were used to label a subset of these confounders for mortality as “strong” (defined by  $HR > 1.2$  or  $HR < 0.8$ ), relative to the full group. Variables identified as confounders were included in the propensity score. Comorbidities that affects a very small number (<5) of patients were excluded from the list of adjustment variables.

#### *Definition of time periods for CTS PS*

We defined 3 meaningful time periods during the study cohort of January 1, 2003 and December 31, 2006 based on the history of oxaliplatin dissemination among stage

III colon cancer patients (Figure 3.4). Important milestones included April 2003, the month efficacy trial results were published and November 2004, the month of oxaliplatin FDA approval for stage III colon cancer. We were mindful of the available patient population in each time period but did not aim for equal numbers of patients in each block. We also considered factors such as provision of buffer time surrounding FDA action to allow for dissemination of new information and retaining reasonable numbers of subjects in each of the treatment groups within each time interval to support estimating a multivariable PS in that interval.

#### *Assignment to calendar time*

Patients were categorized into time periods based on the date that they first received either 5-FU (referent) or oxaliplatin (exposed). Date of cancer diagnosis was not used to categorize patients, as we are interested in time-related treatment assignment paradigms.

#### *Propensity Score Models*

Multivariate logistic regression was used to construct all propensity scores. We created 4 PSs: 1) the calendar time-specific PS, 2) the conventionally estimated PS that adjusts for calendar time (primary comparator), 3) a conventionally estimated PS with full interaction terms between calendar time period and each covariate and 4) a conventional PS with no adjustment for calendar time. The final group of covariates included in the propensity scores is as follows: time period (for applicable comparators only), race, age, sex, urbanity, grade, substage, census median income, COPD, diabetes, PVD, CHF, CVD, MI, peptic ulcer disease, mild liver disease, paraplegia/hemiplegia, chronic renal failure, and rheumatologic disease.

The conventionally estimated PSs were estimated across all years. The CTS PS required a separate regression model, inclusive of all key covariates,<sup>86</sup> for each time period to estimate the time-specific propensity of treatment receipt per covariate. Each model included only those patients who received their first qualifying chemotherapy during that time period. This allowed us to estimate a time-specific propensity of treatment receipt per covariate.

#### *Change in relation between patient characteristics and treatment*

To understand if ability to predict treatment receipt changed for individual covariates over time, the change in the odds ratios (OR) and 95% confidence interval (CI) for receipt of oxaliplatin (taken from the logistic PS models) was compared over successive time periods for each covariate in the PS. Qualitative assessment of the change in OR was performed visually using graphs.

### **Aim 2: Compare the CTS PS-adjusted estimate for oxaliplatin effectiveness on mortality with estimates using a conventionally estimated propensity score**

#### *Hazard ratio estimation*

We evaluated the effectiveness of oxaliplatin vs. 5-FU with regards to death in an intent-to-treat approach by constructing Cox models to estimate hazard ratios (HR) and 95% confidence intervals (CI). Proportional hazard assumptions were tested and confirmed using log likelihood tests and graphical methods.

In addition to the overall and year-specific HRs for the CTS PS, seven comparator sets of HRs were generated using the three conventionally-estimated PSs as well as three outcome model comparators. Each set contained one HR for the study overall and a separate HR for each of the 3 study time periods. The outcome model comparators used conventional Cox proportional hazards regression and adjusted for 1)

no confounders (crude), 2) all covariates included in the propensity scores including calendar time, and 3) all covariates included in the propensity scores excluding calendar time

Greedy matching (5-1) was used for PS adjustment. Greedy matching creates an optimal match based on the PS and does not reconsider once the match is made. The greedy match macro<sup>87</sup> was chosen to maximize sample size in the data after matching. For the time period-specific estimates, greedy matching was performed within each time block. To estimate the overall HR for the CTS PS, the resulting matched cohorts will then be combined by stacking the datasets prior to estimating the HR. Conventionally estimated PS were matched across the full cohort; in the year-specific estimates using this method, matching was broken for pairs that received first chemotherapy treatment in different time periods.

#### *Cox model origin*

Cox models used an origin of 90 days after surgery. Although follow-up time is lost, this origin avoids systematic differences in exposure time by treatment group due to timing common to the FOLFOX regimen, where patients start oxaliplatin shortly after they receive 5-FU. Excluding patients for whom death occurs in that 90 day interval may bias results away from the null. These patients (1.6% of population) may differ clinically from the larger cohort, for example, they may have been too frail to survive surgery. However, because this small proportion of patients likely died from surgery rather than lack of effective chemotherapy, the chosen origin is a clinically relevant landmark<sup>88</sup> for looking at treatment effectiveness among colon cancer patients who received surgery.

### *Covariate balance*

To evaluate confounding control, we examined covariate balance between the matched cohorts generated using the CTS vs. the conventional PS. Matching success was compared by describing the percent of exposed patients that were able to be retained in each cohort.<sup>127</sup> The balance of covariates was used as the primary indicator of the performance of the PS. This was evaluated using the absolute difference in percentage by time period for each covariate, with focus on strong confounders. We also report the cumulative balance for each cohort, irrespective of the strength of covariate association with the outcome, and the percent of oxaliplatin-exposed patients retained.<sup>89</sup>

### *Comparison of estimates*

We compared the time period-specific and full cohort HR (95% CI) estimates between the 4 propensity score and 3 standard regression methods. Changes in magnitude were compared using the percent change in HRs ( $(\text{previous time period HR} / \text{current time period HR}) - 1$ ). Qualitative assessment of the change in HRs was performed visually using graphs.

Although the true estimate is unknown<sup>127</sup> and therefore it is not possible to empirically evaluate bias reduction, we also compared both PS estimates' relative closeness to the MOSAIC RCT results, which estimate the HR for disease-free (rather than all-cause) mortality as 0.77 (95% CI: 0.65, 0.91).

### **Aim 3: Evaluate the ability of using calendar time as an instrument for treatment receipt in estimating a valid estimate for mortality**

#### *Definition of calendar time instrument*

We defined a binary measure of calendar time for the instrumental variable. Patients were categorized into levels of the IV based on the month and year of first

qualifying (5-FU or oxaliplatin) treatment receipt. The FDA approval date of oxaliplatin for stage III colon cancer and the observed adoption curve informed our consideration of potential divisions of calendar time for the instrument. We identified the “optimal” IV measure through evaluation of two criteria: 1) the compliance percentage (i.e., the strength of the instrument’s effect on treatment receipt because a strong IV is less affected by violation of assumptions than a weak one)<sup>90</sup> and 2) the shortest overall time-span (to reduce the potential for violating IV assumptions). To achieve the latter, we considered excluding patients treated several years after FDA approval (truncating cohort enrollment while using all follow-up time of included patients) to optimally maintain that calendar time does not directly or indirectly affect the outcome. Additionally, we tested the effect of excluding those treated in the months immediately near FDA approval, when information dissemination and drug access may have been ambiguous. The “optimal” calendar-time intervals for the IV were identified prior to examination of effect estimates.

#### *Evaluation of IV strength*

We calculated the association of the instrument with the exposure using the following formula:  $E[X|Z=1] - E[X|Z=0]$ . The resulting value is the compliance percentage, which we used to measure the strength of the calendar time instrument.

#### *Consideration of validity of IV assumptions*

Instrumental variable analyses assume that the instrument 1) strongly affects or is associated with the exposure, 2) is related to the outcome only through its association with the exposure and 3) is unrelated to patient risk factors for the outcome (Figure 3.5). This research question also assumes monotonicity<sup>68,92,91</sup> or no “defiers”. This essentially requires that there is no patient who would be prescribed oxaliplatin prior to FDA



approval for stage III, yet would receive 5-FU without oxaliplatin post-approval (all characteristics held constant; only timing of first treatment receipt changed). Although these assumptions are statistically untestable because the association between an instrument and the outcome could be mediated through unmeasured paths and the treatment that would be have been received were the treatment date different is impossible to know, we explored these assumptions to the extent possible using measured confounders, subject matter expertise, and time trends in cancer staging and broad oncology treatment paradigms.<sup>92</sup> The assumptions specific to this research question and the method used to investigate and potentially mitigate violation of these conditions are as follows:

Assumption 1: Calendar time is highly correlated with treatment receipt.

Examination methods:

1. Calculation of compliance percentage.
2. Examination of actual treatment concordance with instrument level. We quantify calendar time's relationship to actual treatment receipt by percentage of oxaliplatin receivers in post-approval IV level and percentage of 5-FU receivers in pre-approval IV level.

Assumption 2: Calendar time is related to mortality only through its association with treatment receipt, and therefore it should have no direct effects on the outcome.

Examination methods:

1. Consider possible stage migration by reviewing staging guidelines and talking to tumor board specialist.
2. Investigate possible changes in colon cancer care that may have created an association between time and mortality. Consult literature, guidelines and practicing

oncologists to consider the potential for mortality to change over time during the study time period due to improvement of surgical techniques, new treatment guidelines beyond adjuvant chemotherapy such as addition of radiation or changes in nutrition.

Assumption 3: Calendar time should be independent of and unrelated to confounders.

Examination methods:

1. Compare balance of measured confounders by the instrument with balance across treatment by calculating the prevalence difference for each measured confounder.<sup>68</sup>
2. Unable to examine this assumption for unmeasured confounders beyond the extent to which the prevalence of measured confounders approximate that of measured confounders.

Assumption 4: The study population upholds monotonicity (no “defiers”).<sup>68,92,93</sup>

1. Thought exercise to consider types of patients that may be defiers and the possibility that they are contained in the final study population. This was considered by construction of the following table:

Patient:	“Always treated”	“Compliers”	“Never treated”	“Defiers”
Instrument value:	Healthy patient, high risk cancer	Likely, typical patient	Frail patient, low risk cancer	Monotonicity assumption
Before FDA approval (IV=0)	oxaliplatin	5-FU only	5-FU only	oxaliplatin
After FDA approval (IV=1)	oxaliplatin	oxaliplatin	5-FU only	5-FU

### *Risk difference estimation*

We derived risk of all-cause mortality from Kaplan-Meier (K-M) survival curves and estimated 1-, 2-, and 3-year mortality risk differences (RD) between the 2 instrumental variable levels (main analysis). The RDs were scaled by the compliance percentage (the association of the instrument with the exposure).<sup>90</sup> K-M survival was used to allow use of all available patients and to avoid conditioning on follow-up, which could introduce selection bias. Risk differences were used as opposed to relative measures of effect for multiple reasons. First, these are collapsible and therefore more appropriately compared across methods. Second, as our population of interest is older and therefore is high risk (e.g., these patients are experiencing increasing mortality based on age), relative measures of effect would move closer to the null do to this increasing mortality. The risk difference is not susceptible to this.

We calculated 95% confidence intervals for the IV RD assuming binomial distributions and independent observations and using a standard error (SE) of  $\sqrt{(SE_{unexposed}^2 + SE_{exposed}^2)}$ .<sup>94</sup> We used bootstraps to verify the width of these CIs to 2 significant digits. Balke-Pearl method<sup>95</sup> was used to place bounds around the point estimate for the average treatment effect in the population.

Comparator RD estimates were generated using K-M survival curves and the following methods: 1) unadjusted, 2) a propensity score adjusted for all measured confounders, including calendar time, 3) a PS adjusted for all measured confounders, excluding calendar time, and 4) a PS adjusted using interaction terms between each calendar year and covariate.<sup>96</sup> Directed acyclic graph methodology<sup>97</sup> and expert knowledge were used to identify the following potential confounders: age, sex, race, tumor grade, tumor substage at diagnosis (IIIA-IIIC), urban/rural status, socioeconomic status measured using state buy-in and census percentage of high school graduates, physician organizational affiliation with a National Cancer Institute Cooperative Group,<sup>98</sup> and 13

prevalent comorbidities of COPD, diabetes, CHF, CVD, MI, old MI, PVD, dementia, cirrhosis, peptic ulcer disease, paraplegia/hemiplegia, chronic renal failure, and rheumatologic disease.

We implemented propensity scores using 5-to-1 digit 1:1 matching.<sup>99</sup> All comparator models used the full cohort rather than the restricted IV cohort, and sensitivity analyses were performed in the reduced IV cohort to evaluate differences in selection that may have been induced based on the IV exclusions.

#### *Kaplan-Meier survival curve model origin*

Follow-up began on the date of first treatment receipt for referent (5-FU) patients and 1 day after oxaliplatin receipt for exposed (oxaliplatin) patients. The decision to base the origin on the observed median oxaliplatin start time after 5-FU was made a priori to avoid systematic differences between exposure groups (see rationale for Cox model origin in Aim 2). The median time from the qualifying surgery to the date of chemo start by exposure group was 46 days for oxaliplatin users and 45 days for 5-FU/capecitabine users; we therefore made the origin the date of first treatment for the oxaliplatin group and the day of first treatment + 1 day for the oxaliplatin group. The median origin was thus 46 days for both groups. We had expected a larger difference in start times between the treatment groups, and this decision was carried through despite the small observed difference.

#### *Comparison of estimates*

We graphically compared survival curves between the IV and PS comparator models and RD estimates from each method. Considerations included CI overlap and point estimates relative to the unadjusted estimate. We did not calculate bounds for the comparator estimates, and therefore did not consider the IV bounds in the comparisons.

## E. TABLES & FIGURES

**Table 3.1 Rationale, ascertainment and data management for covariates that are analyzed in aim 1 and considered potential confounders in aims 2 and 3**

<b>Covariate</b>	<b>Rationale</b>	<b>Ascertainment Method</b>
Sex	Possible gender differences exist in aggressiveness of treatment.	Reported through SEER
Race	Racial disparities in treatment receipt have been described, particularly between black and whites.	Reported through SEER
Age at diagnosis	Age is associated with both treatment receipt and mortality.	Reported through SEER
Urbanity	There are possible differences in knowledge of new treatments between rural and urban/metro facilities.	Reported through SEER
Socioeconomic status	Low socioeconomic status may be associated with lack of access to new drugs.	Not available in data; measured through either
Census median income	Income may dictate access to better care.	Reported through SEER
Census percent non-high school graduation	Low education has been shown to be a proxy for socioeconomic status.	Reported through SEER
Number of months with state buy-in coverage	State buy-in programs are an indicator of low-income, as they are based on annual income and are designed to assist Medicare beneficiaries with premiums, deductibles and co-payments.	Reported through SEER
Tumor substage	Advanced tumors may be more/less likely to receive innovative treatments.	Reported through SEER; Undocumented in 2003; calculated using crosswalk (Tables 3.1 and 3.2)

Tumor grade	Serious tumor grade may be more/less likely to receive innovative treatments.	Reported through SEER
Individual comorbidities*	Comorbidities may be deterrent or incentive for innovative treatments	Medicare diagnosis codes within 12 months prior to diagnosis
Physician organizational affiliation with a National Cancer Institute (NCI) Cooperative Group containing a colon cancer research portfolio**	This affiliation may facilitate dissemination of standards of treatment and has been shown to be specifically associated with oxaliplatin prescribing.	Reported through SEER
Time period of treatment	May act as a confounder or instrument.	Based on date of first treatment receipt (5-FU for unexposed or oxaliplatin for exposed) in Medicare claims data

\*Covariates were based on the Charlton score:<sup>85</sup> congestive heart failure (CHF), acute myocardial infarction (MI), chronic pulmonary disease (COPD), cardiovascular disease (CVD), diabetes, cerebrovascular disease, peripheral vascular disease, AIDS, dementia, diabetes with complications, mild liver disease, moderate/severe liver disease, past myocardial infarction, paralysis, moderate/severe renal disease, rheumatologic disease, ulcer disease, cirrhosis

\*\* Qualifying affiliations include American College of Surgeons Oncology Group, Eastern Cooperative Oncology Group, Cancer and Leukemia Group B, Southwest Oncology Group, and the National Surgical Adjuvant Breast and Bowel Project (ACOSOG, ECOG, CALGB, SWOG, NSABP, respectively)

**Table 3.2 Crosswalk: Correspondence of Stage III substages with AJCC V6 descriptions**

<b>CRC Stage III Substage</b>	<b>AJCC Stage Grouping*</b>		
IIIA	T1-T2	N1	M0
IIIB	T3-T4	N1	M0
IIIC	Any T	N2	M0

\* Patients categorized as Stage III N0 with a SEER extent of disease code of T3 or T4<sup>38</sup> were considered IIIA for this study. NX (number of positive nodes unknown) patients were categorized as IIIA if T1 or T2 and IIIB if T3 or T4. If NX with missing tumor information (T), substage was marked as missing.

**Table 3.3 Crosswalk: SEER Extent of Disease Version-3 codes corresponding to the American Joint Committee on Cancer version 6 tumor and lymph node descriptions<sup>100</sup>**

AJCC V6 Description		Corresponding SEER EOD Codes (1988-2003)		
		e10ex1 - e10ex10	e10nd1 - e10nd10	e10pn1 - e10pn10*
<b>T1</b>	Tumor invades muscularis mucosae, polyp (NOS), submucosa, Localized, NOS/confined to colon, NOS	10-16, 30	.	
<b>T2</b>	Tumor invades muscularis propria	20	.	
<b>T3</b>	Tumor invades through muscularis propria into subserosa or into nonperitonealized pericolic tissues	40, 42, 45, 46	.	
<b>T4</b>	Tumor directly invades other organs or structures and/or perforates the visceral peritoneum	50, 55, 57, 60, 65, 66, 70, 75, 80, 85	.	
<b>T0</b>	No evidence of primary tumor	95		
<b>TX</b>	Unknown extension, not documented	99		
<b>N0</b>	No regional lymph node metastasis		0	'00'
<b>N1</b>	Metastasis in 1 to 3 regional lymph nodes		1, 2, 3	'01','02','03'
<b>N2</b>	Metastasis in 4 or more regional lymph nodes		1, 2, 3	'04' to '90'
<b>NX</b>	Positive nodes are documented, but number is unspecified			97
	No nodes were examined			98
	Unknown whether nodes are positive; not			99



Figure 3.1 Timing for inclusion in study population for aims 1 and 2

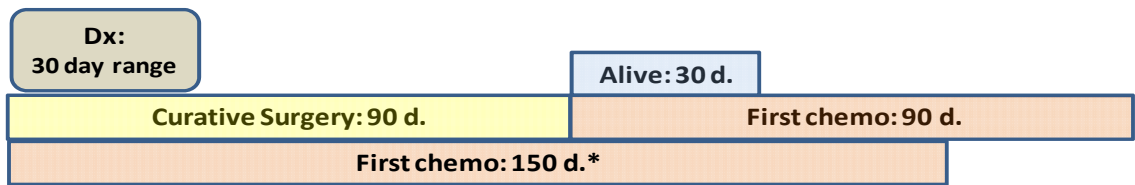
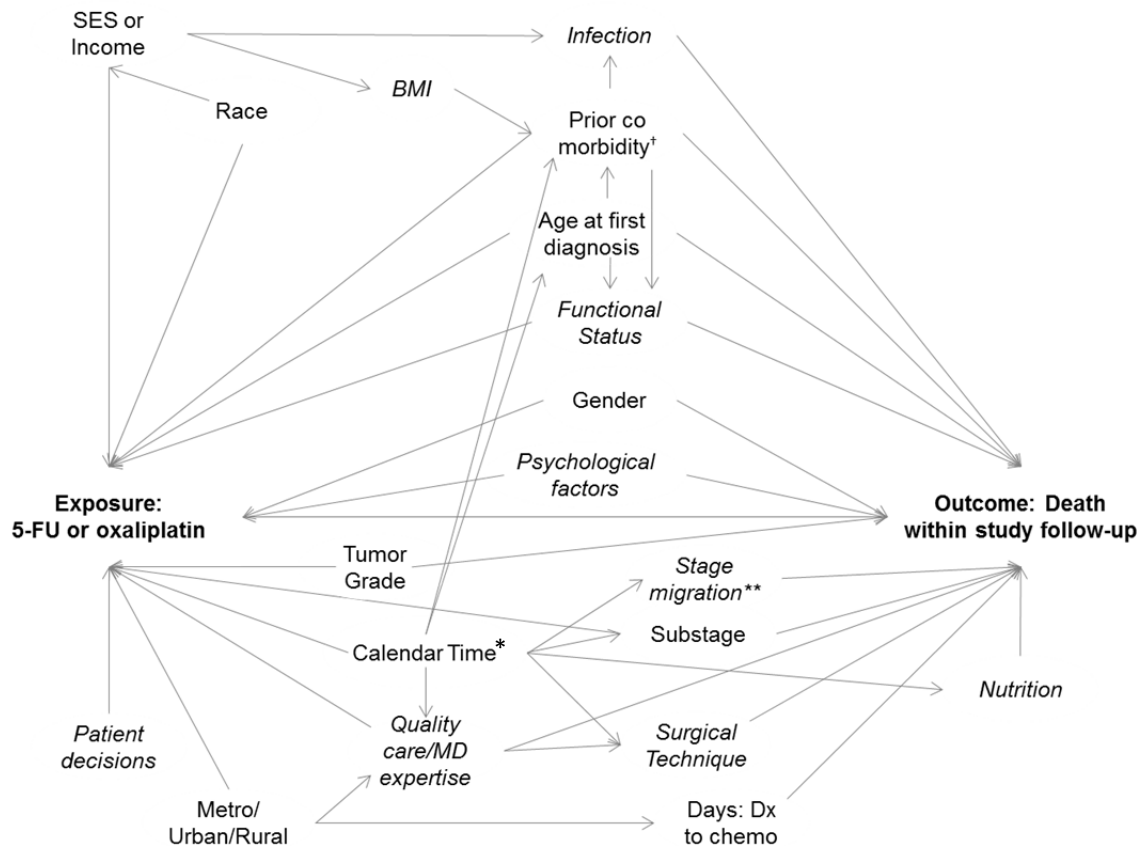


Figure 3.2 Timing for inclusion in study population for aim 3



**Figure 3.3 DAG to identify confounders for the association between treatment and mortality**

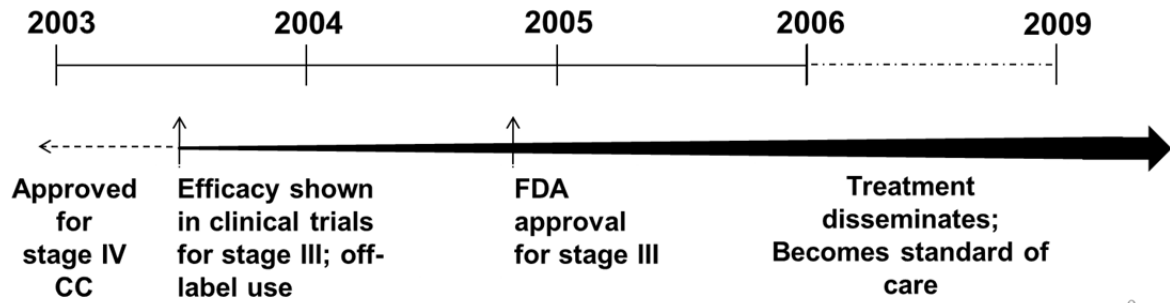


\* This diagram reflects a priori assumptions regarding relationships among treatment, mortality, and covariates that may not be verifiable in this population. Calendar time's potential role as an interaction term is not represented, as interaction terms are not formally recognized as part of DAG methodology.

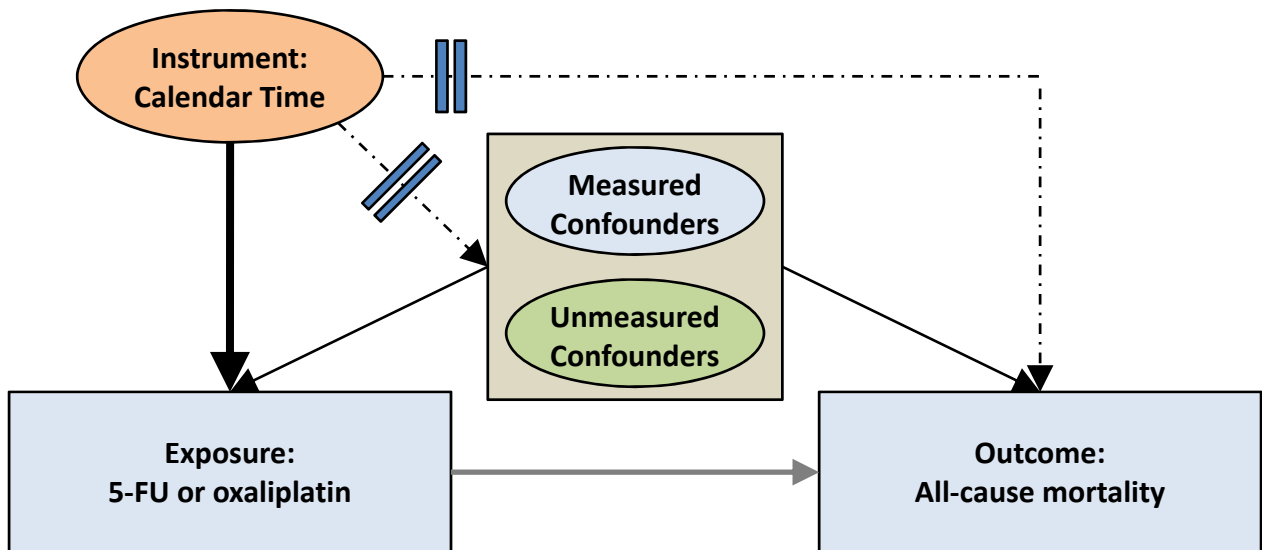
\*\*All included patients are stage III as defined by tumor staging guidelines and medical paradigms at the time of diagnosis; these are subject to stage migration and could change with calendar time.

† Comorbidity prevalence may change due to modifications in diagnostic practices over calendar time. This has been shown for diabetes in this population.

**Figure 3.4 FDA approval history of oxaliplatin for Stage III Colon Cancer**



**Figure 3.5 Considerations for Calendar Time as an Instrument Variable**



## CHAPTER 4

### **RESULTS: Calendar Time-Specific Propensity Scores and Comparative Effectiveness Research for Stage III Colon Cancer Chemotherapy<sup>1</sup>**

#### **A. INTRODUCTION**

Propensity scores (PS) are widely used to control confounding in comparative studies of medical products. A PS is an estimate of the probability that a patient receives one treatment over another, given characteristics of the patient and his/her condition at the time the treatment decision is made.<sup>101,102</sup> PSs are routinely estimated as averages of the effect of patient characteristics on treatment choice over multiple study years. However, many drugs have a dynamic lifecycle, experiencing changes in prescribing based on events and dissemination. A patient characteristic that was once associated with treatment selection may become less relevant over time, or vice versa.

The key assumptions underlying PS methods are that all confounders are accurately measured and the model of treatment receipt given confounders is correct. If calendar time (as a proxy for other changes) is a confounder and prescribing patterns are dynamic, calendar time and its relations to other confounders must be correctly modeled. To our knowledge, few studies consider specific lifecycle events for the drug of interest and incorporate potentially heterogeneous effects of time into PS analyses.<sup>103</sup> This may violate the assumption that the score correctly reflects the underlying

---

<sup>1</sup> This chapter was published by PDS. Citation is: Mack CD, Glynn RJ, Brookhart MA, Carpenter WR, Meyer AM, Sandler RS, Stürmer T. Calendar time-specific propensity scores and comparative effectiveness research for stage III colon cancer chemotherapy. *Pharmacoepidemiol Drug Saf.* 2013 Jan 7 [Epub ahead of print]

propensity for treatment given the confounders.<sup>101</sup> A direct comparison of PS approaches for handling calendar time in dynamic settings has not been performed.

Oxaliplatin, an innovative chemotherapeutic, is a drug that saw striking uptake among stage III colon cancer patients over a short time period, from off-label to widespread use. In this setting, we construct and examine a calendar time-specific (CTS) PS within policy-based time periods of a study cohort to understand possible validity benefits of accounting for changes in confounding by indication over calendar time based on specific patient characteristics. The CTS PS allows the effect of each covariate on the propensity for treatment receipt to be non-uniform over time, taking into account changes in channeling (used here to denote any degree of confounding by indication) relevant to a specific multi-year cohort. Examination of the CTS PS provides insight into prescribing variations and barriers to treatment receipt across calendar years.

## **B. METHODS**

We examined the CTS PS in the context of a CER study of oxaliplatin versus 5-fluorouracil (5-FU) and all-cause mortality in patients with stage III colon cancer, focusing on the early years of oxaliplatin adoption. Based on efficacy results from the MOSAIC trial<sup>104,105</sup> and subsequent FDA approval in November 2004,<sup>106</sup> FOLFOX, defined as the addition of oxaliplatin to 5-FU and folinic acid, replaced 5-FU monotherapy as the standard of care.

Patients were drawn from the Surveillance, Epidemiology and End Results (SEER)-Medicare linked data (described elsewhere)<sup>107,108</sup> and included those diagnosed with stage III colon cancer between 2003 and 2005, with follow-up through 31 December 2006. All patients were traditional Medicare subscribers aged 65+ who received curative

surgery and initiated either oxaliplatin or 5-FU without oxaliplatin within 90 days of surgical resection.

We defined three study time periods, each one year in duration beginning in May 2003, the month MOSAIC trial results were released (Figure 4.1). The second time period encompasses FDA approval and spans the six months pre-approval as well as immediate post-approval dissemination. Patients were categorized into time periods based on their receipt date of 5-FU (referent) or oxaliplatin (exposed). Directed Acyclic Graph methodology<sup>109</sup> and expert knowledge were used to identify potential confounders of age, sex, race, tumor grade, tumor substage at diagnosis (IIIA-IIIC), urban/rural status, income and 11 individual comorbidities.<sup>110</sup>

Multivariable logistic regression was used to estimate PSs. A conventional PS, the primary comparator, was estimated across all years, adjusting for time period. The CTS PS required a separate model for each period to estimate the time-specific propensity of treatment receipt per covariate. To understand if relations between patient characteristics and treatment preference changed for individual covariates over time using the CTS PS, changes in odds ratios (OR) and 95% confidence intervals (CI) for receipt of oxaliplatin were compared graphically over each successive time period (from the CTS PS models) and for the full cohort (from the conventional PS model).

Greedy 5-to-1 digit matching<sup>111</sup> was used for covariate adjustment. CTS PS matching was performed within each time period; matched pairs were pooled to create a full study cohort. For the conventional PS, patients were matched across all three years and the matched cohort was used for both overall and year-specific estimates. For the latter, matching was ignored (broken). To evaluate confounding control, we examined covariate balance between matched cohorts using the absolute difference in percentage by time period for each covariate, with focus on strong confounders. We also report the

cumulative balance for each cohort, irrespective of the strength of covariate association with the outcome, and the percent of oxaliplatin-exposed patients retained.<sup>112</sup>

We compared effectiveness of oxaliplatin vs. 5-FU for prevention of all-cause mortality by constructing Cox models to estimate hazard ratios (HR) and 95% confidence intervals (CI), in an intent-to-treat approach. Cox models used an origin of 90 days after surgery to landmark the analysis.<sup>113</sup> This origin avoids immortal time bias and systematic differences in exposure time by treatment group, and excludes a small proportion of patients that likely died due to surgical complications. An HR for the full study and separate HRs per time period were generated using 4 PSs to adjust for confounding using matching: 1) the calendar time-specific PS, 2) the conventionally estimated PS that adjusts for calendar time (primary comparator), 3) a conventionally estimated PS with full interaction terms between calendar time period and each covariate and 4) a conventional PS with no adjustment for calendar time. For comparison, we also fitted unadjusted and adjusted Cox proportional hazards outcome models. We compared HRs graphically and with percent and absolute differences. The UNC Office of Human Research Ethics (Study number 12-0139) approved this study.

## **C. RESULTS**

Oxaliplatin treatment increased significantly, from 8% (n=86) in 2003-2004 to 52% (n=386) in 2005-2006. Overall, 71% of patients received 5-FU and 29% received oxaliplatin. Exposure group characteristics were similar over the 3 years of the study with the exception of diabetes, which increased in prevalence (Table 4.1).



### *Channeling*

Comorbidities, race, income, urbanity, and age appeared to experience changes in channeling over time. Results for selected covariates are shown to illustrate patterns (Figure 4.2). The adjusted relative odds of oxaliplatin receipt increased over the later two time periods for patients with congestive heart failure (CHF) and decreased across all time periods for those with diabetes. Residence in a high income census area appeared to increase patients' odds of receiving oxaliplatin, particularly prior to FDA approval. Older age was consistently associated with decreased odds of oxaliplatin receipt, and those above 79 became slightly less likely over time to receive it. The effects of tumor grade, COPD and sex on channeling were relatively constant (adjusted OR = 1.0, 0.8/0.9, and 1.0/1.1).

### *Cohort balance*

The CTS PS retained 77% of oxaliplatin-exposed patients (100%, 91%, 59% for each time period, ascending) and the conventional PS retained 79%. Patients were excluded if a suitable match could not be found. In the first time period, the CTS PS was able to include all oxaliplatin-exposed patients because there were relatively few patients receiving oxaliplatin. In later years, the percentage of patients receiving oxaliplatin increased, and therefore oxaliplatin-exposed patients were excluded due to lack of available unexposed (5-FU) matches.

Variables were generally more balanced when using the CTS PS. For example, the balance between 5-FU and oxaliplatin for census income > \$60,000 in 2003-2004 was 30% vs. 37% (balance=6.9) for the conventional PS cohort, compared with 36% vs. 37% for the CTS PS (balance=1.2) (Appendix B Table 1). Because imbalance of the strongest risk factors for mortality leads to more problematic confounding, we focused on the balance of tumor substage (HR=1.8 and 3.6 for substage IIIB and IIIC vs. IIIA), older

age (HR=1.2 for ages 70-74, 1.3 for ages 75-79, and 1.8 for 80+, compared with ages 65-69), undifferentiated/unknown tumor grade (HR=1.3), lower income (HR=1.2), CHF (HR=1.5), COPD (HR=1.4) and diabetes (HR=1.2). Age, COPD, CHF, and income were more balanced across the study years for the CTS PS cohort, although substage and diabetes were less balanced. The balance statistics showing the distribution of the imbalance for each covariate (Figure 4.3) show lower means in each time period for the CTS PS. The entire distribution of balance (quartiles, means, medians) was lower for the CTS PS, showing less imbalance in this cohort, with the exception of 2005-2006. In this time period, the CTS PS cohort had a slightly higher median and maximum than the conventional PS (1.8 vs. 1.5; 8.3 vs. 6.5, respectively); however, it was lower in all other statistics (mean: 1.2 vs. 1.6; 25<sup>th</sup>/75<sup>th</sup> percentiles: 0.3/1.6 vs. 0.4/2.5). Cumulative imbalances of 125 vs. 158 for the CTS compared with the conventional PS further suggest improved balance in the calendar time-specific cohort.

#### *Comparison of hazard ratio estimates*

For the CER analysis, patients receiving oxaliplatin (n=810) were compared with those on a non-oxaliplatin 5-FU regimen (n=1990). Over a median follow-up of 2.65 years, 860 patients (31%) died. The crude mortality rate was 83/1,000 person-years in patients receiving oxaliplatin and 129/1,000 person-years in patients receiving 5-FU. There was a 22% change in HR (HR=0.69 vs. 0.75) between the conventional and CTS PS-adjusted estimates (Table 4.2). Precision between the two methods was similar. The full interaction PS (HR=0.73) generated results similar to the CTS PS in both magnitude and precision. When comparing within time periods, CTS PS-estimated HRs differed more than conventional PS estimates and were closer to the null, with the exception of 2004-2005, in which both estimates moved farther from the null (HR=0.64 and 0.65, respectively) (Figure 4.4).

## D. DISCUSSION

Set in the context of a CER examination of a new chemotherapeutic agent, this study examined a novel approach to propensity score estimation which addresses changes in intervention adoption over time. During oxaliplatin's first three years of rapid adoption for stage III colon cancer, the CTS PS method proved to more adequately address subtle changes in factors associated with treatment selection, under the assumption that calendar time is a confounder or a proxy for confounders, than the commonly applied PS model that assumes uniform effects of patient factors over multiple study years. The CTS PS characterized changes in treatment choices and likely resulted in enhanced confounding control.

Our research expands upon work by Seeger et al<sup>114,115</sup> and Rassen et al<sup>116</sup> by comparing estimates, quantifying bias and defining time periods using policy-related timepoints.<sup>117</sup> In research examining statins and MI, Seeger et al accounted for changes in statin use by estimating PSs and matching within half-year time blocks, then stratifying Cox models by these blocks. They found that allowing flexibility in PS estimation showed changes in drug use over time, but did not compare effect estimates with a conventional PS. Rassen et al assessed the performance of an overall PS within subgroups based on patient characteristics and found larger effect differences for small subgroups and few exposed outcomes. Schneeweiss et al<sup>118</sup> noted the particular importance of CER evidence immediately after FDA approval and highlighted the methodological challenge of bias due to confounding by indication of new medications. They describe a sequential cohort study which proposes matching within quarterly<sup>119</sup> or monthly blocks, but do not explicitly compare alternative approaches.

Our results suggest oncologists may initially have been reluctant to give oxaliplatin to patients with comorbidities such as CHF as they learned of this new drug's effect in patients with characteristics that may have been excluded from clinical trials. A decline in oxaliplatin use was observed in patients with diabetes, suggesting that physicians may have observed neurotoxicity<sup>120,121</sup> that shaped their decision-making in subsequent chemotherapy decisions for these already susceptible patients.<sup>122,123</sup> Similarly, consistent channeling away from older patients may suggest that age-correlated unmeasured variables such as frailty or age discrimination<sup>124</sup> were found to be increasingly relevant over time. Although all patients were covered exclusively by Medicare, higher income areas had increased access to the innovation. This difference dissipated after FDA approval but did not disappear.

The CTS PS produced results closer to the MOSAIC Randomized Controlled Trial (RCT) (HR=0.80, 0.65-0.97),<sup>125</sup> than did the conventional PS and recent observational study findings.<sup>126</sup> Because the true effect among older individuals is unknown, it is not possible to empirically evaluate bias reduction.<sup>89,127</sup> Increased validity of CTS PS estimates can be inferred, however, as there was evidence of changes in confounding by indication over time by individual patient characteristics. Because the CTS PS led to better overall balance of observed covariates within calendar periods and we assume here that calendar time is a confounder, the increased balance reduces confounding bias, at minimum within calendar time period and possibly for the overall estimate. The closeness of the full interaction and CTS PS estimates was expected, as both account for changes in channeling; however, the CTS PS provided the benefit of easier-to-interpret evidence of calendar time's impact on treatment choice.

The differences observed between the year-specific estimates of the CTS PS could be attributed to several factors. In this population, there is evidence of modification by race,<sup>26</sup> and the proportion of African Americans on oxaliplatin ranges

from 6.5% to 8.1% per year. Changes in population mix over time may lead to varied treatment effect estimates in the presence of effect modification. Additionally, unmeasured confounding may change over calendar years and estimating CTS PSs within time periods may allow better identification and management of observations treated contrary to prediction.<sup>128</sup>

Overall matching of exposed patients was similar for both the CTS and conventional PS groups, suggesting practical feasibility of this approach. In this examination, the CTS PS-matched cohort demonstrated greater balance within years than did the conventional PS, as measured by the statistical distribution of individual covariate imbalance, by cumulative absolute difference, and for most, but not all, of the strong confounders. An alternative approach would be to match on the conventional PS within calendar year, which could affect balance comparisons even in the presence of a misspecified model. We matched on the conventional PS across the full cohort after adjusting for calendar year under the assumption that this implementation is most common.

As new drugs are continually entering the market, comparative effectiveness questions of new versus old treatments will continue to arise. PS use has increased exponentially in the last 2 decades<sup>89</sup> and although variables are often liberally included in PS models, vigilance is required in selection.<sup>129,130</sup> Even in settings of dynamic prescribing, calendar time is often not considered in CER specific to its possible role as an instrument, confounder, or modifier of covariate effects on treatment choice. Our assumption is that most researchers would see calendar year as a confounder and thus include it in the PS model; for example, the high-dimensional propensity score algorithm documentation lists year of treatment as the common example of a predefined variable.<sup>131</sup> This is a reasonable assumption in many cases, as time can serve as a proxy for changes in tumor staging, improvements in surgical technique, increases in

provider experience and the use of additional effective treatments that affect common CER outcomes such as mortality and disease recurrence. These factors are unmeasured in these data, as in many claims databases, and controlling for calendar time will limit their potential to confound treatment effects. CTS PSs should be considered in dynamic settings, when calendar time acts as a confounder between the exposure and the outcome and is also a potential modifier due to non-homogenous prescribing or treatment determinants. However, if time is not a confounder but instead an instrument for treatment receipt,<sup>132,133,134</sup> it should not be included in the propensity score model regardless of changes in channeling of the treatment over time. Doing so would result in inflation of the variance and bias if residual confounding is present.<sup>135</sup> As in other settings, the important distinction between a variable (here: calendar time) acting as a confounder or as an instrument cannot be confirmed based on observed data.

Thoughtful consideration of time periods is warranted and ideal choice of calendar time periods is not tested. In this specific example, it was most appropriate to anchor the time periods around efficacy results, when off-label use commenced (period 1), the months surrounding FDA approval (period 2), and the post-approval year when wide dissemination had likely occurred (period 3). In general, drug lifecycle milestones or policy events (e.g., safety warnings) are good candidates for choosing calendar time periods. Providing buffer time around events of interest is needed to allow for dissemination of new information. There is also a need to have reasonable numbers of subjects in each of the treatment groups within time intervals to support estimating a multivariable PS in that interval. In some settings, including pharmacovigilance, allowing for similar numbers of patients per time period may be preferred because such a strategy may be optimal to compare the effect of the treatment over time periods. In any

setting, the ability to divide the full cohort into granular time periods depends on the number of events in each cohort and time period.

While RCTs remain the gold standard for assessing an intervention's effectiveness, they are not always feasible, and their findings often have limited generalizability to the broader population. Comparative effectiveness research using non-experimental data addresses many of the limitations of RCTs, and method development to strengthen CER is critical. Oxaliplatin provides a good practical example for investigating a CTS PS in a non-experimental CER setting. The nature of chemotherapeutic use among oncologists is particularly dynamic; due to rapid disease progression and high mortality, chemotherapies are commonly used off-label, quickly approved for new indications, and rapidly disseminated. These drugs are then used widely, despite unknown effects in populations not included in RCTs such as the elderly and patients with high comorbidity.<sup>136</sup> Age has been associated with receipt of both chemotherapy in general as well as oxaliplatin specifically.<sup>137,138</sup> However, although the median age at colon cancer diagnosis is 72 years, the key RCT establishing oxaliplatin's efficacy had a median age of 60<sup>139,140</sup> and these results cannot be generalized to the older population, especially those over 75. The CTS PS method allows us to not only examine age and other specific characteristics of the general non-RCT population and their association with treatment decisions, but also to see how these things may have changed as the health care community adopted this novel drug and became more familiar with its side effects and clinical use over time.

Limitations of claims data such as lack of information on frailty,<sup>74,141</sup> census-level socioeconomic data, and inexact dates of diagnosis and service apply to these effectiveness results. Medicare is estimated to have 75% sensitivity for picking up 5-FU,<sup>142</sup> and therefore a proportion of the referent group may have been missed.

Comorbidity assessed through claims may also be underestimated for this population, as older age is associated with less aggressive treatment for a number of diseases.<sup>143</sup>

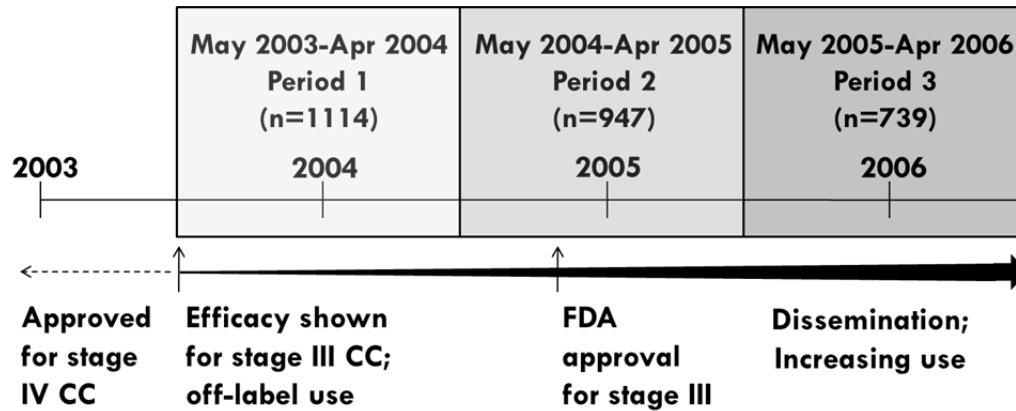
This examination was performed in a single setting and results could be due to chance. The CTS PS should be examined in other settings and over more calendar years. If few potential matches for treated observations exist, the CTS PS may decrease efficiency by diminishing match options. Summary balance measures for PS matching are limited, as they may upweight multi-level variables and ignore individual covariate effects on the outcome, a prerequisite for confounding.

The construct of the calendar time-specific propensity score in the first years of a new drug or after a policy event is likely beneficial to confounding control and validity of estimates in non-experimental CER. The CTS PS allows transparent examination of changes in channeling over time for many covariates at once and is thus useful for understanding determinants of treatment receipt over a drug lifecycle. Creating a CTS PS also prompts researchers to start on the drug life year, which is sensitive to changes in drug prescription patterns, rather than the standard calendar year. Wider implementation of the CTS PS and comparison of estimates with conventional methods is needed in order to further understand the effects of accounting for time in studies of dynamic therapies.



## E. TABLES & FIGURES

**Figure 4.1 Calendar time periods for stage III colon cancer patients based on first date of 5-FU or oxaliplatin receipt (N=2800) and FDA approval history of oxaliplatin**



5-FU=5-Fluorouracil; FDA=Food and Drug Administration; CC=colon cancer

Efficacy results based on the MOSAIC clinical trial, presented in May 2003; FDA approval for stage III colon cancer granted in November 2004.

Time period 1 (May 2003-April 2004) was used as referent time period in the conventional PS model that adjusts for calendar time as a confounder.

**Table 4.1 Characteristics and treatment receipt of Stage III Colon Cancer Patients in SEER-Medicare Study Population by Time Period (N=2800)**

		May 2003-Apr	May 2004-Apr	May 2005-Apr
Characteristic (N=2800)		2004 N (%)	2005 N (%)	2006 N (%)
Treatment:*	5-FU without oxaliplatin	1028 (92.3)	609 (64.3)	353 (47.8)
	Oxaliplatin	86 (7.7)	338 (35.7)	386 (52.2)
Race:	Caucasian American	975 (87.7)	832 (87.9)	639 (86.6)
	African American	83 (7.5)	60 (6.3)	53 (7.2)
	Other	54 (4.9)	54 (5.7)	46 (6.2)
Age:	Mean (sd)	75.1 (5.5)	74.5 (5.6)	74.8 (5.6)
Sex:	Female	602 (54.0)	517 (54.6)	419 (56.7)
	Male	512 (46.0)	430 (45.4)	320 (43.3)
Urbanity:	Metro	931 (83.6)	781 (82.5)	616 (83.4)
	Urban	166 (14.9)	146 (15.4)	108 (14.6)
	Rural	17 (1.5)	20 (2.1)	15 (2.0)
Substage:	A	99 (8.9)	115 (12.1)	80 (10.8)
	B	662 (59.8)	542 (57.2)	412 (55.8)
	C	346 (31.3)	290 (30.6)	247 (33.4)
Grade:	Differentiated	749 (67.2)	635 (67.1)	490 (66.3)
	Undifferentiated/Unk	365 (32.8)	312 (32.9)	249 (33.7)
Census Median Income:**	Mean (sd)	49.8 (22.3)	50.5 (24.2)	50.6 (23.6)
	Congestive heart failure (CHF):	37 (3.3)	42 (4.4)	30 (4.1)
	Myocardial Infarction (MI):	32 (2.9)	21 (2.2)	15 (2.0)
	Chronic obstructive pulmonary disease:	66 (5.9)	60 (6.3)	43 (5.8)
	Cerebrovascular Disease (CVD):	31 (2.8)	17 (1.8)	16 (2.2)
	Diabetes:	96 (8.6)	98 (10.3)	115 (15.6)
	Peripheral Vascular Disease (PVD) :	29 (2.6)	24 (2.5)	17 (2.3)

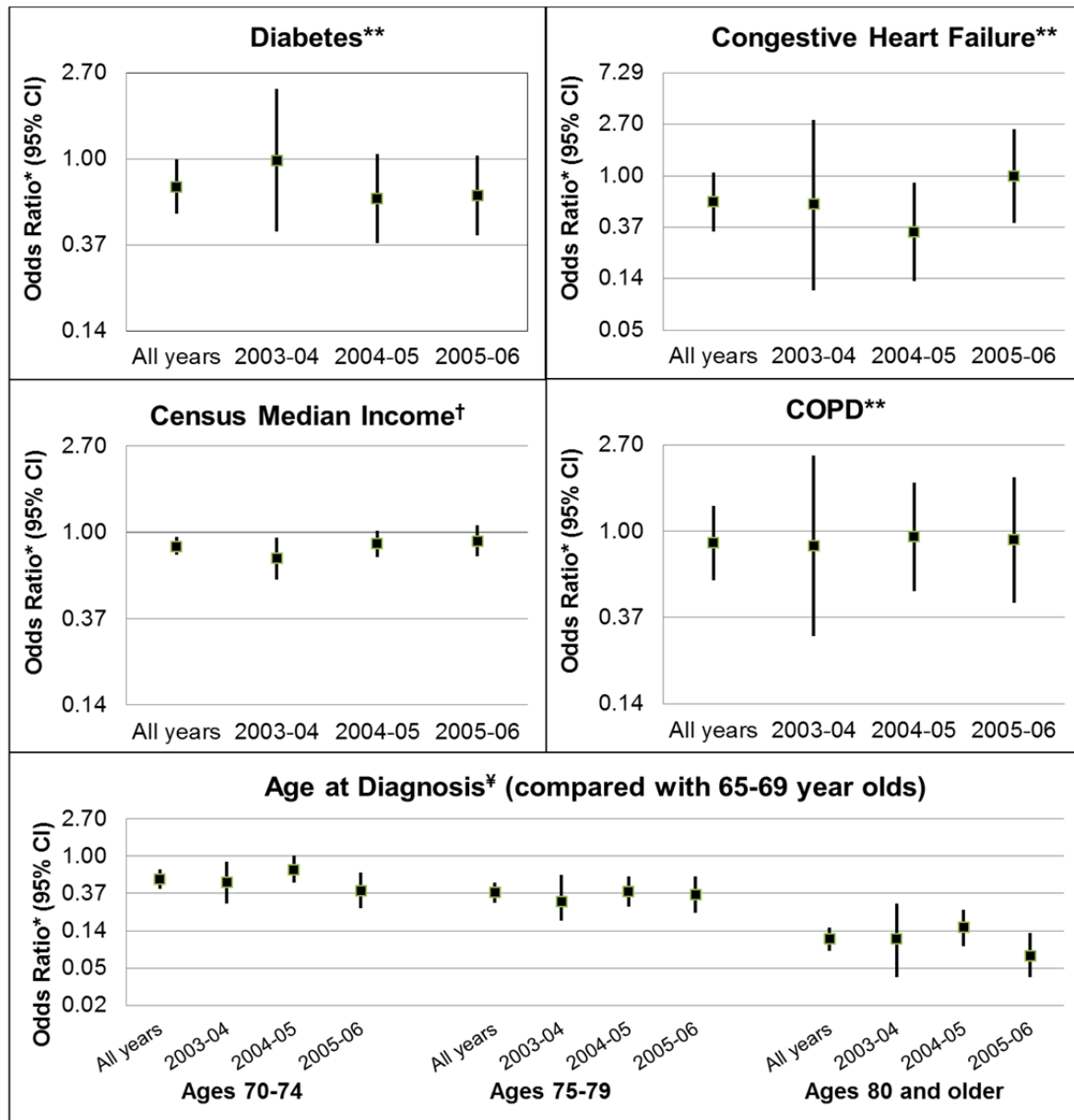
Abbreviations SD, standard deviation; SEER, Surveillance, Epidemiology, and End Results.

Peptic ulcer disease, mild liver disease, paraplegia/hemiplegia, chronic renal failure, and rheumatologic disease are not shown but were also included in propensity score.

\* First chemotherapy treatment received by newly diagnosed patients (intent-to-treat)

\*\*USD, thousands

**Figure 4.2 Changes in channeling by covariate over study time periods (adjusted OR, 95% CI) comparing receipt of oxaliplatin with 5-FU**



Time periods are May through April of the years noted; time period specific estimates are from the CTS PS. Estimates for all years encompass May 2003-April 2006 and are from the conventional PS, adjusted for calendar time. OR scale for CHF and age at diagnosis is expanded due to wide confidence intervals or extreme values.

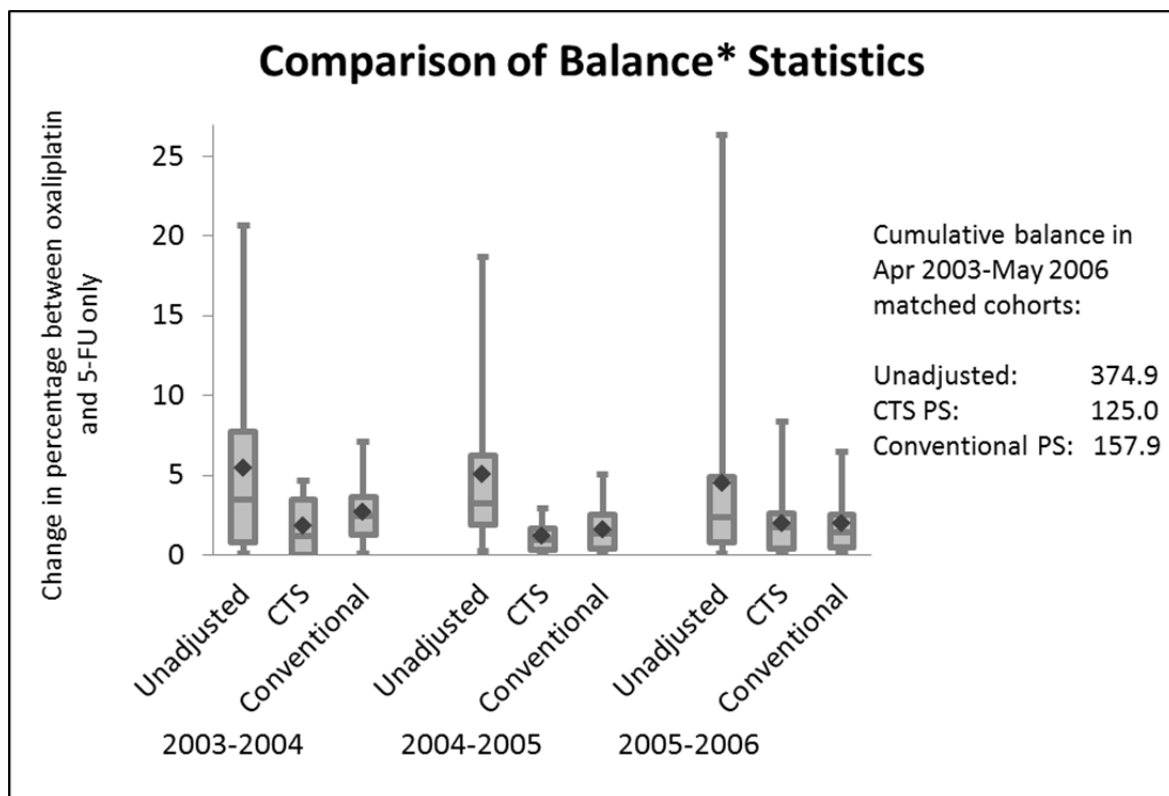
\* Odds ratios are adjusted for all variables included in the propensity score

\*\* Referent group for diabetes, CHF and COPD are those without the condition

† Census median income modeled in quartiles as a continuous covariate, with highest income level as referent

‡ Age modeled as categorical variable, with age group 65-69 as referent

**Figure 4.3 Comparison of covariate balance between full (unmatched) population and matched cohorts generated by the conventional and CTS PS from April 2003 through May 2006**



CTS=Calendar Time-Specific; Time periods are May through April of the years noted.

◆=Mean; Center line=median; Bottom/Top of box=25<sup>th</sup> / 75<sup>th</sup> percentile; Bottom/Top lines=Minimum/Maximum;

Time periods are May through April of the years noted.

\*Balance measured by absolute difference in percentage between exposed and unexposed within covariate level for each time period.

**Table 4.2 Mortality hazard ratios for stage III colon cancer patients treated with oxaliplatin versus 5-FU from May 2003 to April 2006**

<b>Outcome Models</b>	<b>HR (95% CI)*</b>	<b>Percent difference** (absolute change) from CTS PS</b>
Outcome model, unadjusted	0.65 (0.54, 0.77)	-35% (0.15)
Outcome model, adjusted for time <sup>†</sup>	0.71 (0.59, 0.87)	-16% (0.05)
Outcome model, not adjusted for time <sup>†</sup>	0.67 (0.55, 0.80)	-30% (0.12)
<b>Propensity Score Models</b>		
Conventional PS, adjusted for time	0.69 (0.54, 0.90)	-22% (0.08)
Calendar Time-Specific (CTS) PS	0.75 (0.58, 0.98)	--
Full interaction PS	0.73 (0.56, 0.95)	-10% (0.03)
Conventional PS, not adjusted for time	0.67 (0.53, 0.84)	-30% (0.12)

HR=Hazard Ratio; CI=Confidence Interval; CTS=Calendar Time-Specific;

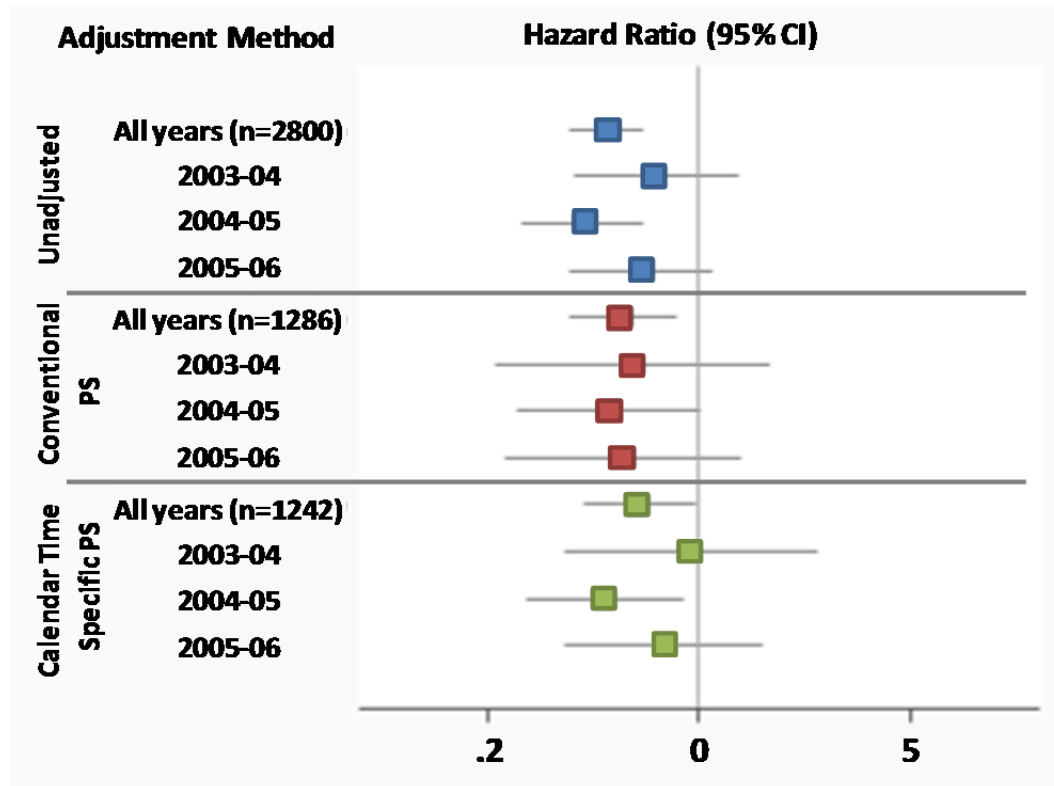
PS=Propensity Score;

\*Although a true estimate is unknown, results can be indirectly compared with MOSAIC RCT results of HR (95% CI)=0.80 (0.65, 0.97).

\*\*Percent difference calculated by:  $\frac{\ln(\text{CTS PS HR}) - \ln(\text{comparison HR})}{\ln(\text{comparison HR})}$

<sup>†</sup> Outcome model comparators used conventional Cox proportional hazards regression for three estimates: unadjusted, adjusted for all covariates included in the propensity scores including calendar time, and adjusted for all covariates included in the propensity scores excluding calendar time

**Figure 4.4. Estimated hazard ratios for different PS adjustment methods comparing oxaliplatin with 5-FU for prevention of all-cause mortality, across all study years and within calendar-specific time periods**



PS=Propensity Score; HR=Hazard Ratio; CL=Confidence Limit; LCL=Lower Confidence Limit; UCL=Upper Confidence Limit; CTS=Calendar Time-Specific.

Time periods are May through April of the years noted.

\*For conventional PS year-specific estimates, the matched cohort was used but matching was broken for pairs that received first chemotherapy treatment in different time periods.

## CHAPTER 5

### **RESULTS: Comparative Effectiveness of Oxaliplatin vs. 5-fluorouracil in Older Adults: An Instrumental Variable Analysis<sup>2</sup>**

#### **A. INTRODUCTION**

Colon cancer (CC) is primarily a disease of older individuals, with 65% of incident cases in the United States occurring in those aged 65 and above.<sup>4</sup> There is uncertainty, however, regarding the effectiveness of oxaliplatin in combination adjuvant chemotherapy for reducing mortality in this population. The seminal randomized controlled trial (RCT) demonstrating oxaliplatin efficacy in patients with stage III CC did not include patients over the age of 75 and had a median patient age of 60, 7 twelve years younger than the median age of 72 for U.S. CC patients.<sup>4</sup> Subsequent observational studies and RCTs have produced inconclusive or contradictory results among diverse sub-populations, prompting a need for additional research and robust methodologies for answering this question.<sup>8,9,12,13,144</sup>

Nonexperimental comparative effectiveness research (CER) quantitatively evaluates the risks and benefits of comparable treatments, thus facilitating informed decision-making among clinicians, patients, purchasers, and policy makers, with the goal of producing real-world estimates of effectiveness. Administrative databases and linked data sources are increasingly used for CER,<sup>145</sup> as they provide large, diverse study populations and long-term follow-up. However, despite the many advantages of repurposing databases for CER, there are limitations to using data that were not collected

---

<sup>2</sup> This chapter was submitted to the journal *Epidemiology*. Authors are Mack CD, Brookhart MA, Carpenter WR, Meyer AM, Glynn RJ, Sandler RS, Stürmer T.

for research.<sup>146</sup> Among them, important patient characteristics such as functional status, frailty, preferences and disease severity all may influence treatment receipt and affect common research outcomes, yet are often missing or measured and recorded with differential fastidiousness across patients in administrative data sources. Because nonexperimental CER studies generally use adjusted regression and propensity score methods that control only for measured differences between treatment groups, the absence of these confounders has the potential to bias effect estimates.<sup>50,147</sup>

In instrumental variable (IV) analyses, an observed variable is used to pseudo-randomize a population based on a context or circumstance that directly affects treatment receipt but is unrelated to the outcome, thus avoiding the assumption of no unmeasured confounding.<sup>3</sup> Some of the most pertinent CER studies examine new-to-market treatments, as these are subject to gaps in knowledge of effectiveness in the wake of clinical trials. Because innovative treatments often experience rapid dissemination upon arrival to the market, treatment decisions may be driven by external factors rather than patient-centric characteristics. In these cases, calendar time proximal to FDA approval or other drug lifecycle events may serve as an instrument for treatment receipt.

Oxaliplatin experienced rapid uptake as part of a multi-agent adjuvant chemotherapy regimen among stage III colon cancer patients over a short time period. Based on 2003 efficacy results from the MOSAIC trial<sup>7,148</sup> and subsequent FDA approval in November 2004, oxaliplatin, in a combination regimen with 5-fluorouracil (5-FU) and folinic acid, rapidly disseminated among stage III CC patients to replace 5-FU monotherapy as the standard of care, despite real-world uncertainty of its effectiveness in older adults underrepresented in the trial. We used calendar time as an instrumental variable and compared results with multivariable propensity score adjustment in



estimating the comparative effectiveness of oxaliplatin vs. 5-FU for reducing all-cause mortality in older adults.

## **B. METHODS**

### *Objective*

We examined the comparative effectiveness of oxaliplatin compared to 5-FU for reduction of all-cause mortality in patients with stage III colon cancer using calendar time as an instrumental variable. We focused on the period before and during oxaliplatin's dissemination to standard of care, with attention on FDA approval for stage III CC as a pivotal timepoint.

### *Data source*

Patients were drawn from Surveillance, Epidemiology and End Results (SEER)-Medicare linked data, which has been described extensively elsewhere.<sup>107,108</sup> The cohort included individuals aged 65+ from 12 US states who were diagnosed with primary stage III colon cancer between 2003 and 2007, with follow-up through April 2010. Included patients received surgical resection within 90 days of diagnosis, survived longer than 30 days, and initiated either oxaliplatin or 5-FU/capecitabine without oxaliplatin within 110 days of surgery and 120 days of diagnosis. Patients who received radiation, were diagnosed at autopsy, or had HMO coverage or incomplete Medicare claims during the 12 months pre- and post-diagnosis (or until death) were excluded.

### *Instrumental variable definition*

We defined a binary measure of calendar time based on the month and year of first treatment receipt. The FDA approval date of oxaliplatin for stage III CC and the

observed adoption curve informed our consideration of potential divisions of calendar time for the instrument. We identified the “optimal” IV measure through evaluation of two criteria: 1) the compliance percentage (i.e., the strength of the instrument’s effect on treatment receipt because a strong IV is less affected by violation of assumptions than a weak one)<sup>90</sup> and 2) the shortest overall time-span (to reduce the potential for violating IV assumptions). To achieve the latter, we considered excluding patients treated several years after FDA approval (truncating cohort enrollment while using all follow-up time of included patients) to optimally maintain that calendar time does not directly or indirectly affect the outcome. Additionally, we tested the effect of excluding those treated in the months immediately near FDA approval, when information dissemination and drug access may have been ambiguous. We examined the instrument in relation to IV method assumptions to the extent possible<sup>92</sup> using measured confounders, expert knowledge, and time trends. These time trends were examined relative to the inception of oxaliplatin-based treatment options as well as other possible changes in colon cancer care that may have created an association between time and mortality. The “optimal” calendar-time intervals for the IV were identified prior to examination of effect estimates.

### *Exposure and outcome*

First treatment receipt was defined as the date of first 5-FU/capecitabine claim with no oxaliplatin claim within 30 days (unexposed) or the date of first oxaliplatin claim, with or without the presence of 5-FU or capecitabine (exposed). We ignored oxaliplatin claims that occurred greater than 30 days after 5-FU/capecitabine receipt, because in this intent-to-treat analysis we are only interested in the first treatment received. However, because late receipt of oxaliplatin may suggest that these individuals had a recurrence or were too sick to initially receive oxaliplatin, a sensitivity analysis excluding these patients from the IV analysis (n=46) was performed. Mortality information was

based on date of death according to Medicare via the U.S. Social Security Administration<sup>19</sup> without loss to follow-up.

### *Analysis*

We derived risk of all-cause mortality from Kaplan-Meier (K-M) survival curves and estimated 1-, 2-, and 3-year mortality risk differences (RD) between the 2 instrumental variable levels (main analysis) or the treatment groups (unadjusted and PS-adjusted). The RDs were scaled by the compliance percentage to estimate the average treatment effect among “compliers”.<sup>90</sup> K-M survival was used to allow use of all available patients and to avoid conditioning on follow-up, which could introduce selection bias. Follow-up began on the date of first treatment receipt for 5-FU and 1 day after for oxaliplatin (based on observed median oxaliplatin start time after 5-FU to avoid systematic differences between exposure groups). RD 95% confidence intervals (CI) were calculated assuming binomial distributions and independent observations, using a standard error of  $\sqrt{(SE_{unexposed}^2 + SE_{exposed}^2)}$ .<sup>94</sup> We used bootstraps to verify the width of these CIs to 2 significant digits. Balke-Pearl method was used to place bounds around the point estimate for the average treatment effect in the population.<sup>95</sup>

Comparator RD estimates were generated using K-M survival curves and the following methods: 1) unadjusted, 2) a propensity score (PS) adjusted for all measured confounders, including calendar time, 3) a PS adjusted for all measured confounders, excluding calendar time, and 4) a PS adjusted using interaction terms between each calendar year and covariate.<sup>96</sup> Directed acyclic graph methodology<sup>97</sup> and expert knowledge were used to identify the following potential confounders: age, sex, race, tumor grade, tumor substage at diagnosis (IIIA-IIIC), urban/rural status, socioeconomic status measured using number of months of state buy-in and census percentage of high school graduates, physician organizational affiliation with a National Cancer Institute

(NCI) Cooperative Group,<sup>23</sup> and 13 prevalent comorbidities from the Charlson comorbidity index (listed in Table 1).<sup>85</sup> We implemented propensity scores using 5-to-1 digit 1:1 matching.<sup>87</sup> Comparator models used the full cohort rather than the restricted IV cohort, and sensitivity analyses were performed in the reduced IV cohort to evaluate differences in selection that may have been induced based on the IV exclusions. All analyses were conducted using SAS version 9.2. (SAS Institute, Cary, NC). The UNC Office of Human Research Ethics (Study number 12-0139) approved this study.

## **C. RESULTS**

Calendar time greatly affected treatment receipt (Figure 5.1). The “optimal” 2-level calendar time instrument grouped patients treated from January 2003 through September 2004 (unexposed n=1449), compared with those treated from March 2005 through May 2007 (exposed n=1432). This excluded patients treated during two separate time periods in the oxaliplatin lifecycle: 1) an interim period and 2) after a post-dissemination truncation date. This IV definition produced oxaliplatin treatment rates of 11% and 65% in the early vs. late arms of the instrument and thus yielded 54% compliance. This indicates a strong association between the instrument (calendar time) and actual treatment, which is an important factor when considering the use of IV methods.

The interim exclusion period removed the immediate months surrounding FDA action (October 2004-February 2005), when differential information access among clinicians in conjunction with possible constraints on drug availability (e.g. medical center formulary updates) may have contributed to ambiguity in clinical use. This excluded patients receiving treatment when oxaliplatin use in this cohort first exceeded 40%, an indication that 5-FU was no longer the standard of care, until oxaliplatin use was (with

one exception) consistently above 50%. The truncated period excluded patients treated 2.5 years after FDA approval (post-May 2007), when the market for innovation was likely to be functionally saturated, with dissemination complete and calendar time less likely to dictate treatment choice. At this time, over 60% of patients in the study population had received oxaliplatin each month for a full year, indicating that it had become the standard of care. This truncation also mitigates the possibility of calendar time affecting mortality through changes in care or diagnostic paradigms beyond the treatments of focus, which limits the potential for the IV analysis to be biased due to violation of the assumption that the instrument only affects mortality through treatment.

Measured patient characteristics were well balanced between instrument levels (Table 1), thereby supporting for the assumption that the IV is unrelated to patient risk factors for the outcome. Prevalence differences (PD) for covariates stratified by the IV compared with treatment assignment were greatly attenuated, which validates the strength of the calendar time instrument and is an indication that the instrument may be independent of unmeasured covariates.<sup>90</sup> Balance for age, substage and cooperative group was particularly improved, with change in PD of as much as 13.6 for treatment to 3.7 for the IV (in the 65-69 year old age category).

The unadjusted 1-, 2- and 3-year mortality risk differences were -0.03, -0.05, -0.06 based on 358, 702, and 956 mortality events, respectively. The IV estimate of the three-year RD for all-cause mortality was -0.09 (-0.15, -0.03), which suggests that for every 100 patients treated with oxaliplatin, 9 additional patients survived to 3 years compared with those treated with 5-FU or capecitabine alone. Fifteen patients would have needed to be treated with oxaliplatin rather than 5-FU to reduce mortality by 1 patient over 3-years. One- and two-year IV RDs were -0.05 (-0.09, -0.01) and -0.07 (-0.12, -0.01)) (Table 2). 1-, 2-, and 3-year IV RD bounds for these estimates were (-0.16, 0.30), (-0.20, 0.26), and (-0.23, 0.24).

Survival curves illustrate decreased risk of mortality both for patients treated later in oxaliplatin's lifecycle (the IV=1 level) and those treated with oxaliplatin (Figure 5.2). IV results were consistent with PS comparators, as all suggested a protective effect and there was substantial CI overlap (Figure 5.3). Point estimates differed in that IV RDs were farther from the null than all other estimates and the unadjusted estimate was in-between the IV and propensity score estimates.

### *Sensitivity analyses*

We 1) ran comparators in reduced IV population rather than full study population, 2) used an IV cutpoint rather than interim exclusion of patients around FDA approval, and 3) removed 46 unexposed patients treated with oxaliplatin after 30 days of chemotherapy inception in sensitivity analyses. None of these analyses substantially differed from main IV and comparator analyses. Absolute changes in risk differences were  $\leq 0.02$  and all CIs encompassed the corresponding RD estimates (Appendix C Table 1).

## **D. DISCUSSION**

In a large study of older stage III CC patients, we found that oxaliplatin reduces all-cause mortality compared with 5-FU alone. These findings were consistent across alternate analytic approaches. Findings were also generally consistent with MOSAIC RCT results for oxaliplatin efficacy in younger groups of patients (2 and 3-year RD derived from MOSAIC K-M survival: -0.02 (-0.05, 0.02); -0.03 (-0.07, 0.01), respectively) (A. de Gramont, written communication, December 2012). The consistency of these effectiveness results in the presence of differing assumptions provides important

information about oxaliplatin effectiveness in older adults, which could aid in decision-making among patients, providers, and policy-makers.

The IV RD suggested a larger protective effect than unadjusted and PS-matched estimates, although considerable overlap among confidence intervals underscores uniformity. The magnitude of oxaliplatin effectiveness cannot be confirmed due to the inability to empirically test assumptions required by IV and PS-adjusted analyses. Taken together, however, these diverse methods provide a useful range of values demonstrating oxaliplatin effectiveness and confidence in its protective effect. The presence of some CI non-overlap requires contemplation of explanatory factors for potential differences in RD estimates.

First, IV and PS-matched estimates generalize to different populations, and therefore these two approaches are producing estimates that apply to potentially dissimilar subgroups.<sup>68</sup> The IV estimates the local average treatment effect in the “compliers”, or those whose initial treatment depends on the calendar time in which their treatment was received (in this case, 54% of the IV cohort; note that this does not imply compliance with the initial treatment during follow-up, as all analyses are intention-to-treat). PS matching estimates the treatment effect in the oxaliplatin treated patients who were successfully matched to 5-FU patients. These populations could meaningfully differ if, for example, patients who received oxaliplatin prior to FDA approval were either healthier-than-average and more likely to succeed on a more aggressive treatment, or had a comparably worse prognosis and were more willing to risk using an off-label treatment. Although the width of the bounds for the IV suggest that our estimate is not informative about the average effect of treatment in the overall population, this is a well-defined clinical population and we think it is unlikely that the average effect of treatment in the compliers would be very different than the average effect in the population. Such bounds were not computed for the PS estimates.

Second, violation of necessary assumptions could introduce bias. Validity of PS estimates could be compromised through unmeasured confounding, which likely exists through lack of data on patient frailty, functional status, and decision-making in these administrative data. Although this would be less pronounced due to comparison with an active treatment, the increased toxicity and cost<sup>149</sup> associated with oxaliplatin may contribute to some unmeasured confounding, particularly in this older population. If time is an instrument rather than a confounder, PS comparators which adjust for calendar year may be even more biased in the presence of unmeasured confounding.<sup>86</sup> We therefore present PSs treating time as both an instrument and a confounder in adjusted analyses.<sup>62</sup>

Validity of IV assumptions could also be compromised. Finding a variable that meets the definition of an IV is challenging. IV analyses assume that the instrument 1) strongly affects or is associated with the exposure, 2) is related to the outcome only through its association with the exposure and 3) is unrelated to patient risk factors for the outcome. While we were able to verify that calendar time's relation to oxaliplatin receipt was strong during the study years, time could affect mortality in ways other than through treatment and, as in all IV analyses, the latter condition is not empirically verifiable. We mitigated this possibility by carefully examining the means through which this is possible and truncating the cohort accordingly. Stage migration or improvements in surgical techniques and other non-chemotherapeutic treatments could create an association between time and mortality. However, AJCC tumor staging guidelines<sup>150</sup> and oncologist interviews suggested that this was unlikely between January 2003 and May 2007. An increase in physician affiliation with NCI cooperative groups, which would improve quality of care for affiliated institutions over time, was not seen in these data.<sup>40,41</sup> The percentage of stage III patients who did not receive any chemotherapy did not change from 2003 to 2007 for most patients, although those 80+ became slightly less likely to



receive adjuvant chemotherapy in later study years. It is possible that in the oldest age groups, sicker patients that may have been included in the study in early years may not qualify in later years, thereby indirectly associating calendar time with mortality. The balance of measured confounders by IV level shown in table 1 supports that assumption 3, also not verifiable, may be upheld.

Our analysis has also made the assumption of monotonicity, which assumes there are no patients who would have received oxaliplatin prior to FDA approval yet would not have received it post-approval, all patient characteristics held constant.<sup>3,92,93</sup> Monotonicity is reasonable in this clinical scenario, as it is improbable that a patient would receive oxaliplatin off-label, yet (holding all other considerations constant) for the same patient to receive 5-FU alone after FDA approval. Adverse event reports that would preclude an early oxaliplatin patient from receiving oxaliplatin in a later calendar month were unlikely to be an issue over this time period. It is possible that physician-observed neuropathy may eventually have deterred an oncologist from prescribing oxaliplatin to a diabetic patient in later years, but this is likely to affect no more than a small proportion of patients (if any).<sup>15</sup>

Fourth, the populations are different in size; the full cohort was reduced by PS matching and the IV cohort was reduced through exclusion of interim and truncated treatment dates. These could cause differences either by chance or by the types of patients being removed. Sensitivity analyses showed that PS estimates in the reduced population were relatively consistent with RDs in main analysis, as were IV estimates that used a cutpoint rather than interim exclusion.

Regardless of the reasons for the variation in estimates, the consistency of results between the two methods suggests oxaliplatin effectiveness among older adults, a finding which is robust to the absence of either measured or unmeasured confounding. Given this, the specific pattern of the IV and PS point estimates being on different sides

of the unadjusted estimate may independently provide insight. IV RDs show a more protective effect than PS-adjusted RDs, which fall closer to MOSAIC RCT results and at times closely approached or crossed the null value. The similarity to RCT results provides an important, albeit potentially imprecise benchmark, as these results are not directly comparable due to the cohorts' dissimilarity. The division of the IV and PS results by the crude may be due to differing abilities to control for measured confounding (PS) versus unmeasured confounding (IV). Controlling for an unmeasured confounder that is not addressed by the adjusted estimates but which is accounted for by the IV – e.g., if younger, healthier patients opt away from oxaliplatin due to its association with potential irreversible neuropathy<sup>15</sup> – would be correctly reflected in the more-protective IV estimates, while PS estimates that are unable to control for unmeasured confounding would show oxaliplatin as less effective due to disproportionate numbers of these healthier patients in the 5-FU-exposed group. This suggests that unmeasured confounding by frailty is likely minor given the active comparator and covered to some extent through adjustment for comorbidities. The latter would have the opposite effect on the placement of the point estimates relative to the unadjusted.

Enhanced adjustment for measured confounding may also cause the PS and IV to estimate RDs on opposing sides of the unadjusted estimate. Table 1 illustrates the IV's ability to greatly reduce imbalance between covariates compared to the covariate distribution by treatment. However, small residual differences remain, which could yield more protective IV estimates compared with PS-matched estimates that statistically adjust for measured differences in treatment receipt. For example, although age imbalance was greatly reduced after stratification by calendar time, there were still more young patients (65-69) in the post-FDA approval (exposed IV group and older patients (70+) in the referent. The magnitude of these differences was slight, but may explain the more protective RDs for the IV compared with the matched-PS estimates. Similarly,

there were slightly more patients being treated by cooperative group-affiliated physicians in the exposed IV group, which could translate into this group receiving better care overall and therefore overstating oxaliplatin protectiveness. The predisposition for IV estimates to be exaggerated may also be responsible.<sup>58</sup>

Limitations of claims data, in general and specific to SEER-Medicare, applies to any CER using these data.<sup>142,141</sup> Medicare has an estimated 75% sensitivity for 5-FU,<sup>142</sup> and therefore a proportion of the referent group may have been missed. Comorbidity assessed through claims may be underestimated in this population, as older age is associated with less aggressive treatment for a number of diseases.<sup>143</sup> As in all research, results may be due to chance.

In the presence of emerging therapies, consideration should be given to treatment variability by calendar time and the contribution of dissemination patterns to treatment assignment. Calendar time is routinely conditioned on or used as a confounder,<sup>131</sup> and although this may correctly specify the PS, there is often not much thought given to time's role as a possible instrument. The utility of calendar time as an IV has been shown by Cain,<sup>133</sup> Johnston,<sup>134</sup> and Shetty,<sup>66</sup> et al when, similar to this setting, trends in medication use create a natural experiment that can be used to strengthen clinical evidence.

The importance and utility of nonexperimental CER studies in establishing treatment effectiveness necessitates careful attention to potential biases in estimates. Particularly in this population of older patients, who were underrepresented in oxaliplatin clinical trials yet are the most affected by colon cancer in the real-world, it is critical to employ robust methods to further our understanding of treatment effectiveness. Because we cannot quantify the effect of unmeasured confounding in adjusted analyses or the exact relationship of a natural instrument with exposures and outcomes, the presentation

of a consistent set of results based on these different assumptions build needed confidence in oxaliplatin's protective effect in older adults.

## **E. TABLES & FIGURES**

See following pages.

**Table 5.1 Characteristics of Stage III Colon Cancer Patients by Treatment Received (N=3660) and Calendar Time Instrument (N=2881)**

Patient Characteristic		Treatment				Calendar Time* (Instrument)			
		5-FU		Oxaliplatin		Pre-FDA		Post-FDA	
		(n=2095)		(n=1565)		approval		approval	
		N	(%)	N	(%)	N	(%)	N	(%)
Treatment:	5-FU					1291	(90)	503	(35)
	Oxaliplatin					158	(11)	929	(65)
Year Treated:**	2003-2004	1397	(67)	228	(15)	1449	(100)	0	(0)
	2005	312	(15)	391	(25)	0	(0)	573	(40)
	2006	192	(9)	400	(26)	0	(0)	592	(41)
	2007-2008	194	(9)	546	(35)	0	(0)	267	(19)
Race:	Caucasian American	1822	(87)	1363	(87)	1266	(87)	1249	(87)
	African American	137	(7)	107	(7)	99	(7)	97	(7)
	Other	135	(6)	95	(6)	83	(6)	86	(6)
Age:	65-69	348	(17)	473	(30)	292	(20)	341	(24)
	70-74	564	(27)	508	(33)	434	(30)	410	(29)
	75-79	607	(29)	428	(28)	408	(28)	405	(28)
	80-84	435	(21)	143	(9)	241	(17)	226	(16)
	85+	141	(7)	13	(1)	74	(5)	50	(4)
Sex:	Female	1155	(55)	849	(54)	783	(54)	789	(55)
	Male	940	(45)	716	(46)	666	(46)	643	(45)
Urbanity: Metro		1741	(83)	1338	(86)	1220	(84)	1196	(84)
	Urban	298	(14)	206	(13)	199	(14)	204	(14)
	Rural	56	(3)	21	(1)	30	(2)	32	(2)
Census Median Income:†	Mean (SD)	49.9	(22)	54.1	(25)	50.7	(23)	52.2	(24)
Socioeconomic status:	State buy-in (yes)	241	(0)	173	(1)	168	(0)	160	(1)
Census percent non-high school graduation:	Mean (SD)	18.7	(13)	17.3	(13)	18.4	(13)	18.0	(12)
Cooperative group:*	Yes	926	(44)	743	(48)	651	(45)	668	(47)

Patient Characteristic	Treatment				Calendar Time* (Instrument)			
	5-FU (n=2095)		Oxaliplatin (n=1565)		Pre-FDA approval (n=1449)		Post-FDA approval (n=1432)	
	N	(%)	N	(%)	N	(%)	N	(%)
No	1162	(56)	808	(52)	791	(55)	757	(53)
Substage: A	255	(12)	154	(10)	159	(11)	153	(11)
B	1233	(59)	810	(52)	832	(57)	780	(55)
C	607	(29)	600	(38)	458	(32)	498	(35)
Grade: Well differentiated; differentiated, NOS	101	(5)	82	(5)	80	(6)	64	(5)
Moderately/Intermediately differentiated	1339	(64)	959	(61)	925	(64)	886	(62)
Poorly or undifferentiated; anaplastic	613	(29)	485	(31)	411	(28)	451	(32)
Cell type not determined, stated or applicable	42	(2)	39	(3)	33	(2)	31	(2)
Congestive heart failure (CHF):	160	(8)	72	(5)	99	(7)	82	(6)
Myocardial infarction (MI):	31	(2)	15	(1)	24	(2)	15	(1)
Old myocardial infarction:	47	(2)	31	(2)	32	(2)	31	(2)
Chronic obstructive pulmonary disease (COPD):	212	(10)	166	(11)	130	(9)	157	(11)
Cerebrovascular disease (CVD):	112	(5)	59	(4)	76	(5)	60	(4)
Diabetes (with or without sequelae):	479	(23)	339	(22)	311	(22)	321	(22)
Peripheral Vascular Disease (PVD) :	93	(4)	50	(3)	59	(4)	51	(4)
Ulcer	30	(1)	15	(1)	21	(1)	12	(1)
Chronic renal failure	45	(2)	19	(1)	23	(2)	21	(2)
Rheumatoid arthritis (RA)	241	(11)	173	(10)	168	(11)	160	(10)

Due to SEER-Medicare confidentiality requirements, treatment years are combined and presence of paralysis, dementia, and cirrhosis are not shown.

\* Instrumental variable definition is focused around FDA action in November 2004. Patients treated from October 2004 through February 2005 and after June 2007 were excluded (n=779, or 21% of N=3660, excluded)

\*\* Year treated based on date qualifying treatment (5-FU without oxaliplatin within 30 days or oxaliplatin) was received.

† Median household income in 1,000 US Dollars, based on 2000 data

¥ Physician organizational affiliation with a National Cancer Institute (NCI) Cooperative Group containing a colon cancer research portfolio. Includes the American College of Surgeons Oncology Group (ACOSOG), Eastern Cooperative Oncology Group (ECOG), Cancer and Leukemia Group B (CALGB), Southwest Oncology Group (SWOG), and the National Surgical Adjuvant Breast and Bowel Project (NSABP)

**Table 5.2 Comparison of Risks and Risk Differences Using Instrumental Variable and Adjusted Analyses**

Model	Unexposed*				Exposed*				Risk Diff.**	95% CI	NNT <sup>†</sup>
	# at risk	Events	Risk**	95% CI	# at risk	Events	Risk**	95% CI			
1-year risk											
Unadjusted	1859	236	0.11	0.10, 0.13	1442	123	0.08	0.07, 0.09	-0.03	-0.05, -0.02	29.4
Wald IV estimator <sup>‡</sup>	1284	165	0.11	0.10, 0.13	1307	125	0.09	0.07, 0.10	-0.05	-0.09, -0.01	20.3
Unadjusted for time	1234	132	0.10	0.08, 0.11	1260	106	0.08	0.06, 0.09	-0.02	-0.04, 0	52.6
Adjusted for time	792	86	0.10	0.08, 0.12	806	72	0.08	0.07, 0.10	-0.02	-0.04, 0.01	62.7
Interactions w/ time	675	75	0.10	0.08, 0.12	689	61	0.08	0.06, 0.10	-0.02	-0.05, 0.01	53.6
2-year risk											
Unadjusted	1648	447	0.21	0.20, 0.23	1310	255	0.16	0.15, 0.18	-0.05	-0.08, -0.03	19.8
Wald IV estimator <sup>‡</sup>	1136	313	0.22	0.20, 0.24	1173	259	0.18	0.16, 0.20	-0.07	-0.12, -0.01	15.4
Unadjusted for time	1102	264	0.19	0.17, 0.22	1148	218	0.16	0.14, 0.18	-0.03	-0.06, -0.01	29.7
Adjusted for time	705	173	0.20	0.17, 0.22	736	142	0.16	0.14, 0.19	-0.04	-0.07, 0	28.3
Interactions w/ time	600	150	0.20	0.17, 0.23	622	128	0.17	0.15, 0.2	-0.03	-0.07, 0.01	34.1
3-year risk											
Unadjusted	1388	602	0.29	0.27, 0.31	916	354	0.23	0.21, 0.26	-0.06	-0.09, -0.03	17.8
Wald IV estimator <sup>‡</sup>	1021	428	0.30	0.27, 0.32	1047	353	0.25	0.23, 0.27	-0.09	-0.15, -0.03	11.1
Unadjusted for time	933	367	0.27	0.25, 0.30	799	302	0.23	0.21, 0.25	-0.04	-0.08, -0.01	23.6
Adjusted for time	543	231	0.27	0.24, 0.30	561	205	0.24	0.21, 0.27	-0.03	-0.07, 0.01	33.9
Interactions w/ time	450	205	0.28	0.25, 0.31	488	182	0.25	0.22, 0.28	-0.03	-0.08, 0.01	31.0

\* Unexposed patients received 5-FU or, for the instrumental variable estimator, were treated prior to FDA approval of oxaliplatin (Jan 2003-Sept 2004). Exposed patients received oxaliplatin or were treated after FDA approval for oxaliplatin (Mar 2004-May 2007)

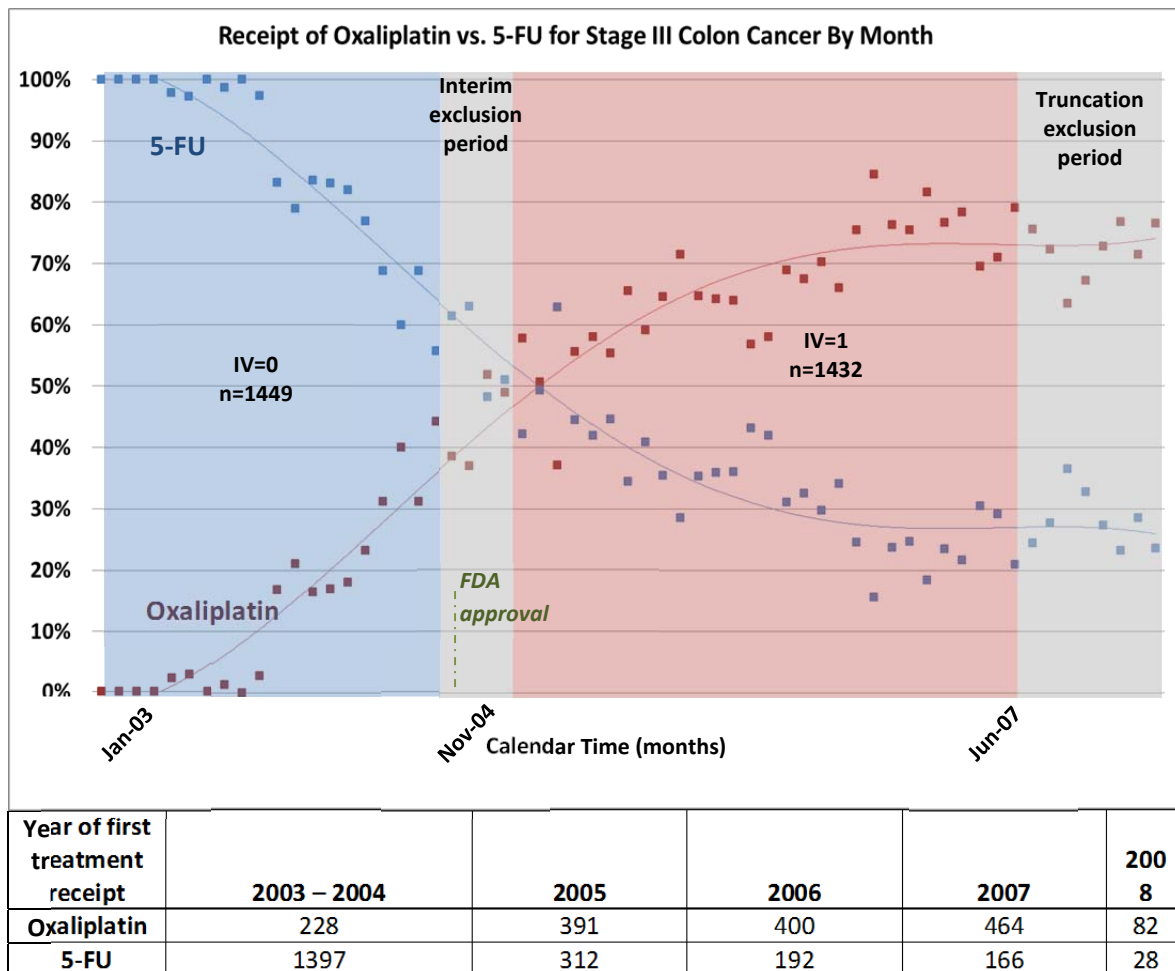
\*\* Estimates of risk and risk difference are accompanied by 95% confidence intervals. Risks taken from Kaplan-Meier survival curve; events / # of patients as shown in table do not take censoring into account and therefore do not calculate risk.

† Number needed to treat calculated by 1/RD.



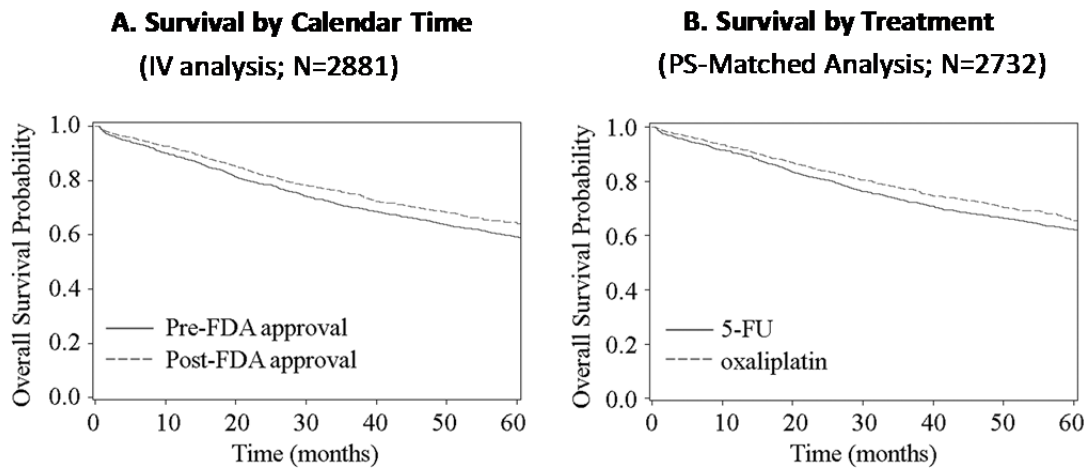
¥ Wald IV estimator is scaled by a compliance percentage of 54%

**Figure 5.1 Dissemination of Oxaliplatin: Receipt of Oxaliplatin vs. 5-fluorouracil for Stage III Colon Cancer by Month and Definition of Calendar Time Instrument Variable (N=3660)**



Points indicate the percentage of patients in each month receiving oxaliplatin or 5-FU. Grey shading indicates excluded patients due to interim period (October 2004–February 2005) and the truncation period of June 2007 and later. For illustrative purposes, diffusion patterns for each treatment are fitted with fourth-order polynomial trendline. The intersection point of lines is not statistically meaningful in terms of dissemination activity. Due to SEER-Medicare confidentiality requirements, treatment years 2003 and 2004 are combined.

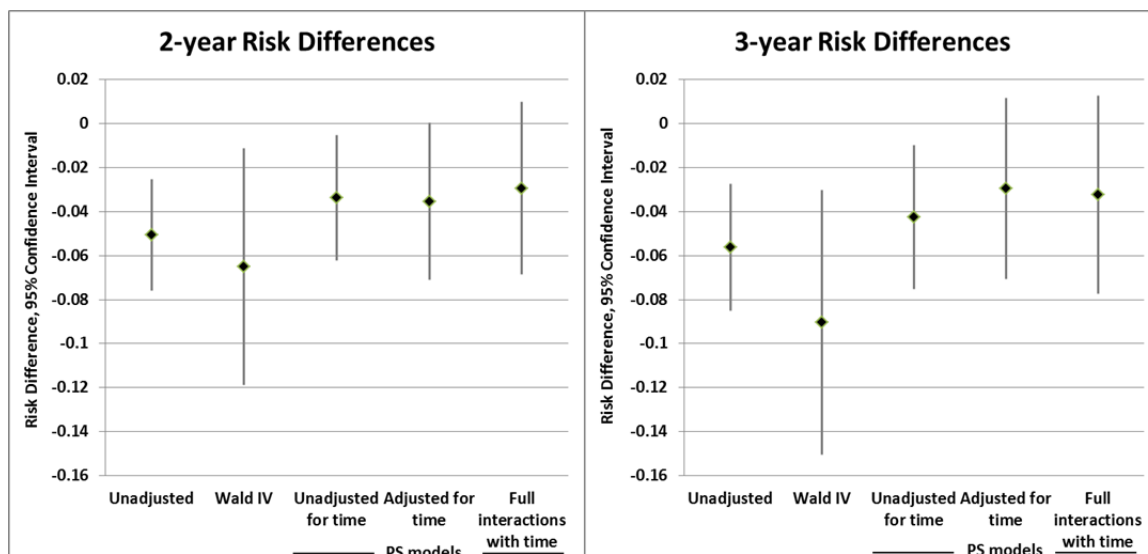
**Figure 5.2 Probability of Overall Survival A) by Calendar Time Instrumental Variable (N=2881) and B) with 5-FU vs. Oxaliplatin in Propensity Score-Matched Analysis (N=2732)**



A) Patient assignment to instrumental variable category is based on month treatment was first received. January 2003-September 2004 (pre-FDA approval, referent) is compared with March 2004-May 2007 (post-FDA approval).

B) Matched propensity score analysis adjusts for age, sex, race, tumor grade, tumor substage at diagnosis (IIIA-IIIC), urban/rural status, socioeconomic status measured using number of months of state buy-in and census percentage of high school graduates, physician organizational affiliation with a National Cancer Institute (NCI) Cooperative Group, and comorbidities.

**Figure 5.3 Comparison of Risk Differences Using Instrumental Variable and Adjusted Analyses**



Estimates of risk difference (RD) are based on risks taken from Kaplan-Meier survival curves. The Wald instrumental variable (IV) estimator is scaled by a compliance percentage of 54%. Comparator estimates are adjusted for the variables presented in Table 1, using propensity scores implemented by matching.

## **CHAPTER 6**

### **DISCUSSION**

#### **A. SUMMARY OF FINDINGS**

Set in the context of a comparative effectiveness examination of a new chemotherapeutic agent, oxaliplatin, this thesis examined novel approaches to propensity score estimation and instrumental variable analyses, which address and take advantage of changes in intervention adoption over time.

The calendar time-specific PS method examined in aim 1 allowed us to examine how age and other specific characteristics of the U.S. colon cancer population are associated with real-world treatment decisions. The CTS PS provided transparency as to how the predictive value of these characteristics changed over time, as the health care community adopted oxaliplatin and became more familiar with its side effects and clinical use. Of note, our results suggested that the effect of CHF, diabetes, age, and income on channeling changed over calendar time as oxaliplatin became the standard of care.

In aim 2, we used this novel PS method to examine the effectiveness of oxaliplatin vs. 5-FU in preventing all-cause mortality for stage III colon cancer among older Americans. We found that during oxaliplatin's first three years of adoption, the CTS PS method more adequately addressed changes in factors associated with treatment selection than the commonly applied PS model that assumes uniform effects of patient factors over multiple study years. Although the true effect among older individuals is unknown and it is therefore impossible to empirically evaluate bias reduction,<sup>89,151</sup>

increased validity of CTS PS estimates was inferred from the fact that by accounting for changes in confounding by indication over time (which we found in aim 1) led to better covariate balance.

Aim 3 found that in the first years of oxaliplatin's lifecycle as a treatment for stage III colon cancer, calendar time greatly affected oxaliplatin receipt. We defined a 2-level calendar time instrument which resulted in 54% compliance, which is a strong IV for a non-randomized, observational setting. This IV analysis, which accounted for unmeasured confounding that was likely present in the aim 2 adjusted analyses, showed a reduction in all-cause mortality among oxaliplatin users compared with patients on 5-FU alone. This finding was consistent across results of all alternate analytic approaches in this aim, as well as those from aim 2 and the seminal RCT (The MOSAIC Trial) results for oxaliplatin efficacy. Aim 3 contributes to an evolving body of instrumental variable literature, successfully showing proof of concept for the use of calendar time as an instrument in dynamic settings.

The consistency of the effectiveness results in the presence of differing assumptions strengthens evidence of oxaliplatin's effectiveness in older adults, who bear the greatest burden of colon cancer yet were underrepresented in clinical trials. This could aid in decision-making among patients, providers, and policy-makers. Taken together, the results from aims 2 and 3 underscore the importance of carefully considering calendar time's role in research of new therapies or any treatment that has experienced a significant event such as a change in indication, safety warning, or major policy change. Even in these settings of dynamic prescribing, calendar time is often not carefully considered in CER specific to its possible role as an instrument, confounder, or modifier of covariate effects on treatment choice. Often, researchers may assume calendar year is a confounder and thus include it in the PS model; for example, the high-dimensional propensity score algorithm documentation lists year of treatment as the

common example of a predefined variable.<sup>152</sup> This is a reasonable assumption in many cases, as time can serve as a proxy for changes in tumor staging, improvements in surgical technique, increases in provider experience and the use of additional effective treatments that affect common CER outcomes such as mortality and disease recurrence. If time is an instrument for treatment receipt instead of a confounder,<sup>153,154,155</sup> it should not be included in the propensity score model regardless of changes in channeling of the treatment over time. Doing so would result in bias and inflation of the variance if residual confounding is present.<sup>156</sup> In this case, calendar time can be used as an instrumental variable and provide valuable knowledge.

## **B. PUBLIC HEALTH IMPLICATIONS**

This research enhances understanding of chemotherapeutic treatment for older stage III colon cancer patients. This is an important disease area, as colon cancer claims the second-highest rate of mortality in the United States.<sup>4,157</sup> Although oxaliplatin's superiority to 5-FU is well-established, there is uncertainty regarding the effectiveness of oxaliplatin in combination adjuvant chemotherapy for reducing mortality in older individuals. This is a critical unanswered question, as CC is primarily a disease of older individuals, with 65% of incident cases in the United States occurring in those aged 65 and above.<sup>4</sup>

This research identified deterrents to receiving the gold-standard treatment oxaliplatin, with a focus on individual patient demographics and comorbidities. Disparities in receipt of this treatment by patient demographics and comorbidity are suspected but have been under-examined. By examining the covariates within the calendar-specific PS across years, we gained insight into patient or practice characteristics which merit early receipt of new chemotherapeutics in the U.S. This may challenge clinicians to examine their prescribing patterns and challenge leadership at

treatment facilities to investigate appropriate prescribing. It may also inform marketing or academic detailing strategies, as it identifies patient groups that are not receiving the best treatment.

Additionally, this work may improve the validity of nonexperimental comparative effectiveness research in the environment of new-to-market innovation by developing and closely examining methods for estimating clinical effectiveness. We showed the strong impact that calendar time can have on the prescribing of a new therapy, and we investigated the optimal method of dealing with dynamic elements of treatment receipt, with focus on propensity score and instrumental variable methods of controlling bias. Knowledge about propensity score and instrumental variable methods is important, as these are becoming increasingly common in outcomes and effectiveness research and it is critical that they provide valid rather than erroneous estimates.

Estimating propensity scores using a calendar-specific method to reduce bias has not been examined this closely in previous studies<sup>10</sup> and is a novel concept for observational studies, particularly those that are examining innovative drugs or drugs with safety warnings that arose after FDA approval. If changes in likelihood and determinants of treatment receipt over time continue to be neglected in PS estimation, misspecification of the propensity score model and estimations of biased treatment effects are likely. The use of calendar time as an instrument, instead of adjusting with propensity scores, has not been examined in a setting where changes in confounding have been described over time specifically for the target population. Exploring this alternative method in conjunction with the insight into changes in channeling provided by the year-specific CTS PS contributes to a comprehensive picture of CER methods to control confounding in nonexperimental settings, and claims data in particular. IV estimates are particularly useful for health services research because they generalize to the population in the sample, rather than only those that were treated.



## C. STRENGTHS

SEER-Medicare data contains high quality patient, treatment and provider information for a large, national, diverse study population. It captures real-world treatment paradigms in an elderly population, which is the appropriate population of focus for research on a disease with a median age of 72.<sup>139</sup> The study population provided by this resource was large enough to study the research questions due to the expansive reach of the SEER registries and the comprehensive coverage of the elderly by Medicare insurance in the U.S. Together, the SEER-Medicare linkage provided a rich data source that contains data on cancer treatments, tumor staging, and many important demographic, health and health services-related covariates which are necessary to control confounding in observational research. Because all data are collected prospectively, there exists no potential for recall bias regarding treatment receipt and procedure dates. The data are population based and there is no loss to follow-up, as Medicare provides coverage until end of life and death information is reliably reported through the U.S. social security administration.

These analyses enhanced understanding of oxaliplatin use and identified early reasons for not receiving this gold-standard treatment. Although it is not the primary aim of this study, the estimate of oxaliplatin effectiveness toward preventing all-cause mortality was generalizable to the U.S. elderly due to the high case ascertainment and heterogeneous population in SEER-Medicare. This confirmed clinical trial results and extended the results of several recent observational studies to an older population.

Comparative effectiveness research using non-experimental data addresses many of the limitations of RCTs; however, in exchange researchers must contend with confounding. Thus, methods development to strengthen CER is critical. Oxaliplatin

provides a good practical example for investigating a CTS PS in a non-experimental CER setting. The nature of chemotherapeutic use among oncologists is particularly dynamic; due to rapid disease progression and high mortality, chemotherapies are commonly used off-label, quickly approved for new indications, and rapidly disseminated. These drugs are then used widely, despite unknown effects in populations not included in RCTs such as the elderly and patients with high comorbidity.<sup>158</sup>

These aims provide a thorough examination and comparison of two advanced statistical adjustment methods: propensity scores and instrumental variables. The ability of these methods to control confounding based on their underlying assumptions, with particular attention to how they identify and handle unmeasured confounding, was thoroughly examined. The lack of functional status data is a common and pervasive issue in research using claims data, and therefore this examination will be applicable to many studies. Investigation into these methodological issues in using re-purposed observational data sources is both timely and significant in light of the current U.S. focus on healthcare quality and comparative effectiveness research.

Calendar time-specific PSs are a novel and potentially less biased method of estimation, particularly for drugs that are new or have safety issues. This method is transferable to any disease where the probability of receipt of a drug changes over time. This concept was shown to be feasible and useful to study drug effectiveness, provide insight into predictors of receipt, and hint toward the presence of unmeasured confounding.

Potential issues in this research were thought through and addressed through sensitivity analyses or secondary “checks” such as bootstrapping for the IV CIs. Cox model origins were mindful of possible biases that occur with real-world uses of these chemotherapeutics. Conservative bounds were estimated for the IV results, to

underscore the potential for bias in IV analyses and the inability for researchers to empirically verify the validity of IV assumptions.

Logistically, this SEER-Medicare population was readily accessible, as the data were already collected and de-identified by NCI. Use of this data is therefore efficient and comparatively inexpensive compared to primary data collection.

## **D. LIMITATIONS**

### *Claims data*

Although they provide a rich resource for epidemiologic research, claims data are administrative in nature and therefore not created for research purposes. The limitations of claims data have been well described.<sup>74,80,159</sup> In Medicare, these limitations include lack of data for HMO enrollees, no documentation of reasons and results for tests and procedures, and variability in coding practices over time. We employed state-of-the art methods to minimize these acknowledged biases, many of which are outlined in this proposal. The nature of these research questions seeks to better understand how to best control for one of the most limiting issues of pharmacoepidemiologic claims data research: unmeasured confounding due to patient decision making, unrecorded frailty, functional status, unknown comorbidity, or other factors that present a deterrent to treatment.

We worked to identify errors and inconsistencies in the SEER-Medicare data in several ways. All variables were analyzed to ensure that ranges and distributions are as expected, and we cross-checked variables that are collected in both SEER and Medicare. Disagreement between SEER surgery variable and Medicare claims was minimal, although it did exist. The SEER data indicated no surgery for 17 out of 8007

patients (0.2% of an interim study population) who had surgery in claims. This indicates relatively good agreement between variables.

In SEER-Medicare, there is an estimated 80% concordance with diagnosis date and timing of claims for colon cancer. However, only month and year of diagnosis are provided and exact index dates must be estimated. We took this into account by building a conservative time buffer into all of our analyses. Specifically, we added 30 days to recommended or respected treatment windows and used the first day of the previous month as an index date for finding surgeries and chemotherapies. Comorbidity assessed through claims may also be underestimated for this population, as older age is associated with less aggressive treatment for a number of diseases.<sup>160</sup>

### *Mortality*

All-cause mortality, the outcome of interest, may be an inferior endpoint to cancer-specific mortality. In SEER however, the cancer-specific mortality variable is of limited reliability, as it is believed that cause of death for a stage III colon cancer patient would be documented as colon cancer regardless of the circumstances of death. Disease-free survival is another possible outcome; however, recurrence is difficult to ascertain in SEER-Medicare data and is generally considered unreliable, as patients with a recurrence may not be treated with surgery or may use oral drugs that wouldn't be contained in the claims.<sup>10</sup>

In focusing only on mortality as a study endpoint, this thesis is limited in how much it can contribute to a complete interpretation of comparative effectiveness between oxaliplatin and 5-FU. We do not perform a full harm/benefit analysis, in that we do not look at side effects or disease recurrence as clinical endpoints, nor did we consider patient reported outcomes in this work. This was out of the scope of the SEER-Medicare

data, as well as our study aims. These effectiveness results can inform other CER studies that wish to understand the comparison of oxaliplatin and 5-FU in its entirety.

#### *Treatment ascertainment and patient classification*

Medicare is estimated to have 75% sensitivity for picking up 5-FU and lower for capecitabine.<sup>74</sup> Due to this, the study cohort may erroneously exclude a proportion of the referent group that will appear to be untreated in the claims data. Although this reduced sample size, it likely excluded referent group patients nondifferentially based on mortality and should therefore not affect estimates. To avoid missing a considerable proportion of FOLFOX users due to these low sensitivities, we used evidence of treatment with oxaliplatin to define the exposure of interest, rather than requiring evidence of all medications in the FOLFOX regimen.

Although capecitabine and 5-FU are considered therapeutically equivalent, it is possible that capecitabine and 5-FU are channeled toward patients with different profiles. Capecitabine can be taken from home, whereas 5-FU requires an in-facility infusion; therefore receivers may be more likely to be rural, elderly, or have low functional status. In this case, there may be differential exclusion of patients with certain characteristics from the referent group and if these characteristics are strong predictors of mortality, this could bias study results. This is unlikely to affect study results, as less than 1% of patients received capecitabine. Exposure misclassification would be possible for those that started 5-FU toward the end of the claims window defined by this study and started oxaliplatin after the window. This misclassification would likely be small and tend to bias results toward the null. There are similar coding issues with oxaliplatin, namely that it is claimed with the generic code J999 at the beginning of its lifecycle, during which time there is also an unknown section of patients taking oxaliplatin through clinical trials. This is mitigated by the fact that final data collection for the MOSAIC trial

occurred in April 2003 and it is therefore unlikely that MOSAIC patients would receive their first treatment during the proposed study period.

Exposure misclassification would be possible for those that started 5-FU toward the end of the 150-day claims window defined by this study and started oxaliplatin after the 150 days window. The oxaliplatin receipt would not be detected by our coding and these patients would be classified into the referent group. This misclassification is likely small and would tend to bias results toward the null. There is also potential misclassification for those patients that received general administration, 5-FU or capecitabine first, and then oxaliplatin more than 30 days later (n=54 received oxaliplatin within 38 to 131 days after first administration of the referent treatment). These patients were classified into the referent treatment group, which upholds the principals of this intent-to-treat analysis. We performed a sensitivity analysis excluding these patients in aim 3, which allows us to consider the possibility that these patients may initially be considered too sick to start the harsher oxaliplatin, but after doing better than expected on 5-FU, oxaliplatin was added. It is also possible that they experienced a recurrence, therefore spurring the addition of oxaliplatin later in treatment.

#### *Changes other than oxaliplatin introduction over study period*

Measurement error may occur due to stage migration over time, or the Will Roger's effect.<sup>81</sup> This error would be associated with changes in diagnostic accuracy and colon cancer staging across the years of the study. A significant change in diagnostic practices or staging would affect all CER analyses; however, 6th edition staging guidelines applied from 2003-2009 and no major improvements in diagnostic technology occurred during this time. Moderate stage migration, however, would cause issues in the IV analysis, where changes in time are assumed to be unassociated with mortality. Here, even small changes in staging over time in either direction would be strongly associated

with mortality, as there are large differences in prognosis between stages II, III and IV. We researched potential changes and found that staging should be consistent at least until 2007. In this year, there could be possible upstaging due to quality improvement initiatives that aimed to increase lymph node count, which could increase staging accuracy and possibly upgrade a patient from a stage II to III (i.e. the population diagnosed in 2007 could contain more people with earlier disease). This issue was attenuated by using a shorter time period in IV analysis through cohort truncation.

### *Generalizability*

Generalizability of these effectiveness results are limited to the elderly and to patients that are exclusively covered by Medicare. Although this is an older, less wealthy population than U.S. patients overall, it is a high risk population for colon cancer mortality and therefore an ideal group to study.

As in all research, these results could be due to chance.

## **E. FUTURE RESEARCH**

This thesis has produced several questions which will be addressed in the following ongoing research projects.

### *(1) Considerations for Creating a Calendar Time Instrumental Variable in Specific Settings of Nonexperimental Comparative Effectiveness Research*

Our results showed that calendar time is potentially a very strong instrumental variable (IV) for new-to-market therapies that have experienced dramatic changes in clinical practice. This is an important option in large database studies where unmeasured confounding is likely. In this dissertation, we performed a brief investigation

on defining the optimal approach for defining a calendar time IV from the continuum of time; however, deeper consideration is warranted.

A follow-up study will evaluate approaches for creating a calendar time IV in this setting of CER comparing 5-FU and oxaliplatin for stage III colon cancer. Fifteen variations of a calendar time IV will be constructed to delineate patients treated prior to vs. after Ox uptake, anchored around Ox FDA approval in November 2004. We will examine the use of cutpoints vs. interim exclusion ranges during transition months surrounding FDA approval; removal of time distant from FDA action that may violate IV assumptions; and categorical vs. continuous IV. We will evaluate IV strength based on percent compliers, prevalence difference ratios, and IV assumption legitimacy and will compare risk difference (RD) estimates between instrument variants using Kaplan-Meier survival curves.

## *(2) Outlining a framework for the use of calendar time in studies of drug effectiveness*

Our manuscript will aim to provide awareness of the important role of calendar time in CER and different perspectives on how calendar time can be utilized. We will outline a framework for the use of calendar time in studies of drug effectiveness in the form of a decision tree for how to use calendar time in CER.

The premise for this is based on the importance of date of treatment receipt in studies of therapies that have experienced dramatic changes in clinical practice. Calendar time's role as a potential confounder, modifier or instrumental variable (IV) is a critical consideration in comparative effectiveness research (CER). The appropriate use of calendar time in CER studies of emerging and dynamic therapies is critical for validity of results. Researchers must explicitly consider and make assumptions regarding time, and in some cases should conduct multiple analyses with time as both an IV and



confounder to produce a range of results under differing assumptions. This will increase confidence in results.

Our recommendation will be that researchers should begin CER studies of dynamic therapies by considering clinical setting and covariate availability in the data source to assess whether assumptions of unmeasured confounding are plausible. If important confounders are missing, consideration regarding whether calendar time would act as a strong instrument based on policy is warranted. In this case, time should not be included in the PS and IV assumptions should be tested to the extent possible. In some cases, multiple analyses should be conducted, using calendar time as a confounder, modifier and instrument, with one method chosen a priori and other(s) presented to create a range of feasible estimates.

*(3) Examining the CTS PS in the context of safety: Calendar Time Specific Propensity Score Estimation to Address Channeling Bias in Comparative Effectiveness Estimates for Second Generation Antipsychotics*

We are implementing the CTS PS in the setting of second-generation antipsychotics, using safety reports and new guidelines as the catalyst for changes in channeling. Our objective is to demonstrate channeling among new users of second generation antipsychotics following a Food and Drug Administration safety advisory and to evaluate the impact of channeling on cardiovascular risk estimates over time. Using Florida Medicaid data from 2001-2006, we will examine adults with schizophrenia, bipolar disorder or psychosis initiating second generation antipsychotics. We will use propensity scores to match olanzapine initiators with other second generation antipsychotic initiators. To evaluate channeling away from olanzapine following an FDA safety advisory, we will estimate calendar time-specific propensity scores and compare

the performance of these calendar time-specific propensity scores with conventionally-estimated propensity scores on estimates of cardiovascular risk.

Our findings have shown increased channeling away from olanzapine for several key cardiovascular risk factors, which corresponded with the timing of the FDA advisory. Hazard ratio estimates varied by propensity score estimation strategy but bias by unmeasured confounding was predominant in this setting.

## **F. CONCLUSIONS**

Date of treatment receipt is an important variable in studies of therapies that experience changes in clinical practice. In these settings, researchers should begin observational CER studies by considering calendar time's role as a potential confounder, modifier or instrument. Additionally, covariate availability in the data source should be assessed to ascertain whether assumptions of unmeasured confounding are plausible. If important confounders are missing, attention regarding whether calendar time would act as a strong instrument based on policy is particularly warranted. In this case, time should not be controlled for (e.g., included in the PS) and IV assumptions should be tested to the extent possible.

The appropriate use of calendar time is important for validity of results and promotes a useful understanding of prescribing paradigms. Researchers must explicitly consider and make assumptions regarding time. In some cases, multiple analyses should be conducted using time as both an IV and confounder, with one method chosen a priori and others presented to create a range of feasible estimates. This range, which would depend on different sets of assumptions, can increase confidence in results.

The construct of the calendar time-specific propensity score in the first years of a new drug or after a policy event is likely beneficial to confounding control and validity of estimates in non-experimental CER. Use of the CTS PS will allow transparent

examination of changes in channeling over time for many covariates at once and is thus useful for understanding determinants of treatment receipt over a drug lifecycle. Creating a CTS PS also prompts researchers to account for time based on the drug life year, which has clinical meaning and is sensitive to changes in drug prescription patterns, rather than the more arbitrary standard calendar year. Wider implementation of the CTS PS and comparison of estimates with conventional methods is needed in order to further understand the effects of accounting for time in studies of dynamic therapies.

The IV and PS analyses presented in this thesis both indicated better survival among patients treated with oxaliplatin, albeit with different point estimates. As these results were based on different assumptions, this body of work adds to evidence of oxaliplatin's effectiveness in older adults, who bear the greatest burden of colon cancer.

## APPENDIX A: Administrative Codes

**Table A.1. Administrative codes used to ascertain surgery to qualify patients for this study**

ICD-9	Description
45.4*	local excision or destruction of lesion or tissue of large intestine
45.52	isolation of segment of large intestine (resection of colon for interposition)
45.7*	partial excision of large intestine
45.8	total intra-abdominal colectomy
45.9*	intestinal anastomosis
46.04, 46.03	resection, exteriorized segment large intestine
48.4*	pull-through resection of rectum
48.5*	abdominoperineal resection of rectum
48.6*	other resection of rectum
HCCS/CPT	Description
44140-44147	colectomy, partial
44150-44156, 44157-44158	colectomy, total
44110	Excision 1 or more lesions of small or large intestine not requiring anastomosis exteriorization or fistulization, single enterotomy
44111	Excision 1 or more lesions of small or large intestine not requiring anastomosis exteriorization or fistulization, multiple enterotomies
44130	enteroenterostomy, anastomosis of intestine w. or w/o cutaneous enterostomy (separate procedure)
44139	mobilization(take down) of splenic flexure performed in conjunction w/partial colectomy
44160	colectomy, partial w/ removal of ileum w/ ileocolostomy
44204-44212	laproscopy, colectomy, partial

\* Refers to all additional numbers, e.g. 45.4\* = 45.41, 45.42, etc.

## **APPENDIX B: Covariate Balance**

**Table B.1. Covariate balance\*: Percentage and percent difference of patients treated with oxaliplatin of 5-FU only by covariate between full (unmatched) population and matched cohorts generated by the conventional and calendar time-specific PS from April 2003 through May 2006**

Time Period*:	Unadjusted (unmatched) cohort**									Calendar Time-Specific PS-matched cohort†									Conventional PS-matched cohort**								
	(N=2800)									(N=1242)									(N=1286)								
	2003-2004			2004-2005			2005-2006			2003-2004			2004-2005			2005-2006			2003-2004			2004-2005			2005-2006		
Covariate	OX	5-FU	Δ in %	OX	5-FU	Δ in %	OX	5-FU	Δ in %	OX	5-FU	Δ in %	OX	5-FU	Δ in %	OX	5-FU	Δ in %	OX	5-FU	Δ in %	OX	5-FU	Δ in %	OX	5-FU	Δ in %
Race: Caucasian American	88.4	87.6	0.8	86.1	89.0	-2.9	87.8	85.3	2.5	88.4	90.7	-2.3	87.9	87.0	1.0	87.2	84.6	2.6	88.4	93.3	-4.9	86.8	89.7	-2.9	88.9	84.4	4.5
African American			0.7	6.5	6.2	0.3	6.2	8.2	-2.0	8.1	9.3	-1.2	6.5	7.2	-0.7	7.5	9.6	-2.2	8.1	4.5	3.6	6.6	7.1	-0.5	6.7	9.0	-2.3
Other			-1.5	7.4	4.8	2.7	6.0	6.5	-0.5	3.5	0.0	3.5	5.5	5.9	-0.3	5.3	5.7	-0.4	3.5	2.2	1.2	6.6	3.2	3.4	4.4	6.6	-2.2
Age: 65-69	38.4	17.7	20.7	34.3	17.6	16.7	30.3	11.3	19.0	38.4	43.0	-4.7	29.0	30.0	-1.0	13.6	15.8	-2.2	38.4	34.8	3.5	29.9	31.9	-2.0	13.8	14.8	-0.9
70-74	32.6	28.4	4.2	34.6	26.4	8.2	29.5	24.4	5.2	32.6	29.1	3.5	36.8	33.9	2.9	34.6	32.9	1.8	32.6	37.1	-4.5	35.9	31.3	4.6	32.8	33.2	-0.4
75-79			-7.6	22.2	28.4	-6.2	31.1	28.9	2.2	22.1	20.9	1.2	24.4	24.1	0.3	36.4	35.1	1.3	22.1	21.3	0.7	24.3	26.1	-1.8	39.5	35.7	3.9
80+			-17.2	8.9	27.6	-18.7	9.1	35.4	-26.3	7.0	7.0	0.0	9.8	12.1	-2.3	15.4	16.2	-0.9	7.0	6.7	0.2	9.9	10.6	-0.8	13.8	16.4	-2.6
Sex: male vs. female	50.0	54.4	-4.4	52.1	56.0	-3.9	54.4	59.2	-4.8	50.0	52.3	-2.3	52.8	52.4	0.3	52.6	60.5	-7.9	50.0	52.8	-2.8	53.3	53.2	0.1	54.5	59.4	-4.9
Urbanity: Metro	90.7	83.0	7.7	85.8	80.6	5.2	85.2	81.3	3.9	90.7	90.7	0.0	85.7	88.6	-2.9	83.8	82.0	1.8	90.7	86.5	4.2	84.9	87.7	-2.9	81.4	82.0	-0.5
Urban			-7.3			-3.3			-2.4	8.1	8.1	0.0	13.4	10.4	2.9	14.9	16.7	-1.8	8.1	13.5	-5.3	14.1	11.3	2.9	16.6	16.8	-0.2
Rural			-0.4			-1.9			-1.5	1.2	1.2	0.0	1.0	1.0	0.0	1.3	1.3	0.0	1.2	0.0	1.2	1.0	1.0	0.0	2.0	1.2	0.7
Substage: A			0.4	9.5	13.6	-4.2	11.1	10.5	0.7	9.3	9.3	0.0	10.4	11.7	-1.3	8.3	9.6	-1.3	9.3	9.0	0.3	10.2	12.3	-2.1	9.1	9.0	0.1
B			-10.6	51.5	60.4	-8.9	51.3	60.6	-9.3	50.0	53.5	-3.5	56.0	55.7	0.3	61.4	58.8	2.6	50.0	47.2	2.8	55.6	55.2	0.4	57.7	57.4	0.3
C	40.7	30.5	10.2	39.1	25.9	13.1	37.6	28.9	8.7	40.7	37.2	3.5	33.6	32.6	1.0	30.3	31.6	-1.3	40.7	43.8	-3.1	34.2	32.6	1.6	33.2	33.6	-0.4
Grade: Differentiated v. Not	66.3	67.3	-1.0	66.0	67.7	-1.7	66.6	66.0	0.6	66.3	68.6	-2.3	65.1	66.4	-1.3	71.9	69.3	2.6	66.3	67.4	-1.1	65.5	65.8	-0.3	68.4	64.3	4.0
Income: > 60,000	37.2	20.4	16.8	26.3	20.9	5.5	24.6	22.1	2.5	37.2	36.0	1.2	23.1	26.1	-2.9	23.7	20.6	3.1	37.2	30.3	6.9	22.7	27.7	-5.0	21.3	23.0	-1.6
> 45,000 - 60,000	23.3	25.8	-2.5	26.9	24.5	2.5	26.7	21.8	4.9	23.3	27.9	-4.7	28.0	27.7	0.3	26.3	24.1	2.2	23.3	30.3	-7.1	27.0	25.8	1.2	25.7	24.2	1.5
> 34,000 - 45,000	24.4	30.7	-6.3	27.5	29.1	-1.5	28.0	33.7	-5.7	24.4	24.4	0.0	28.3	27.4	1.0	26.3	34.6	-8.3	24.4	25.8	-1.4	29.6	27.1	2.5	28.5	34.8	-6.4
<=34,000	15.1	23.1	-7.9	19.2	25.6	-6.4	20.7	22.4	-1.7	15.1	11.6	3.5	20.5	18.9	1.6	23.7	20.6	3.1	15.1	13.5	1.6	20.7	19.4	1.4	24.5	18.0	6.5
Congestive Heart Failure			-1.1			-4.1	3.1	5.1	-2.0	2.3	2.3	0.0	2.0	2.6	-0.7	3.5	3.5	0.0	2.3	2.2	0.1	2.0	2.6	-0.6	4.7	3.3	1.5
COPD			-0.1	5.0	7.1	-2.0	5.4	6.2	-0.8	5.8	4.7	1.2	5.5	5.9	-0.3	3.5	3.9	-0.4	5.8	3.4	2.4	5.3	4.8	0.4	4.7	4.9	-0.2
Cerebrovascular disease			0.8			-1.4			-0.2	3.5	2.3	1.2	1.0	0.3	0.7	1.8	1.3	0.4	3.5	2.2	1.2	1.0	1.3	-0.3	2.0	2.5	-0.5
Diabetes			0.7	8.3	11.5	-3.2	13.2	18.1	-4.9	9.3	7.0	2.3	8.8	6.2	2.6	15.8	14.9	0.9	9.3	7.9	1.4	8.9	9.7	-0.8	16.2	13.9	2.3
Peripheral vascular disease			3.5			-0.7			0.1	5.8	2.3	3.5	2.3	2.0	0.3	2.6	2.2	0.4	5.8	2.2	3.6	1.6	3.2	-1.6	2.0	2.5	-0.5
Myocardial infarction			1.9			0.2			0.6	4.7	3.5	1.2	2.3	2.3	0.0	1.3	1.3	0.0	4.7	2.2	2.4	2.3	2.3	0.0	0.8	2.0	-1.3
<b>Total imbalance</b>			<b>136.4</b>			<b>125.5</b>			<b>113.0</b>			<b>46.5</b>			<b>29.0</b>			<b>49.4</b>			<b>67.9</b>			<b>40.0</b>			<b>50.0</b>
<b>Total imbalance, all years</b>									<b>374.9</b>									<b>125.0</b>									<b>157.9</b>

\*Balance measured by absolute difference in percentage between exposed and unexposed within covariate level for each time period.

\*\*Small percentage values have been removed to mask data per SEER-Medicare Data Use Agreement Requirements.

† Calendar Time-Specific PS was matched within calendar year. Conventional PS was adjusted for time period but matched across full cohort.

\* Time periods are May through April of the years noted

## **APPENDIX C: Sensitivity Analyses**

**Table C.1. Sensitivity Analyses Results for Instrumental Variable Analysis**

Analysis Method	1-year RD	95% CI	$\Delta$ in RD*	2-year RD	95% CI	$\Delta$ in RD*	3-year RD	95% CI	$\Delta$ in RD*
<b>Primary study results</b>									
Wald IV estimator**	-0.05	-0.09, -0.01		-0.07	-0.12, -0.01		-0.09	-0.15, -0.03	
Unadjusted for time	-0.02	-0.04, 0		-0.03	-0.06, -0.01		-0.04	-0.08, -0.01	
Adjusted for time	-0.02	-0.04, 0.01		-0.04	-0.07, 0		-0.03	-0.07, 0.01	
Full interactions with time	-0.02	-0.05, 0.01		-0.03	-0.07, 0.01		-0.03	-0.08, 0.01	
<b>Sensitivity #1: Estimate propensity score comparator RDs in reduced IV population</b>									
Unadjusted for time	-0.03	-0.06, -0.01	0.012	-0.04	-0.08, -0.01	0.009	-0.060	-0.1, -0.03	0.022
Adjusted for time	-0.01	-0.04, 0.02	-0.007	-0.03	-0.07, 0.02	-0.009	-0.040	-0.09, 0.01	0.008
Full interactions with time	-0.02	-0.06, 0.01	0.004	-0.01	-0.06, 0.04	-0.020	-0.02	-0.08, 0.03	-0.007
<b>Sensitivity #2: Define calendar time IV using Dec. 2004/Jan. 2005 cutpoint rather than interim exclusion of patients</b>									
Wald IV estimator**	-0.05	-0.09, -0.01	0.004	-0.07	-0.12, -0.01	0.002	-0.08	-0.14, -0.02	-0.008
<b>Sensitivity #3: Exclude referent patients who received oxaliplatin &gt;30 days after receiving 5-FU</b>									
Adjusted for time	-0.05	-0.09, -0.01	-0.002	-0.06	-0.12, -0.01	-0.002	-0.09	-0.15, -0.03	-0.004

\* Absolute change is calculated by subtracting the RD of equivalent method used in the main analysis from the RD of sensitivity analysis

\*\* The Wald instrumental variable estimator is scaled by a compliance percentage of 54% for the primary study results and 50% for sensitivity analysis #2.



## REFERENCES

1. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41–55.
2. Glynn RJ, Schneeweiss S, Stürmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic and Clinical Pharmacology and Toxicology* 2006;98:253-9
3. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf.* 2010 Jun;19(6):537-54. Review.
4. Altekruse SF, Kosary CL, Krapcho M, et al (eds). SEER Cancer Statistics Review, 1975-2007. Bethesda, MD: National Cancer Institute; 2010. Available at: [http://seer.cancer.gov/csr/1975\\_2007/](http://seer.cancer.gov/csr/1975_2007/), Accessed October 29, 2012.
5. American Cancer Society. Cancer Facts & Figures 2013. Atlanta, Ga: American Cancer Society; 2013.
6. National Cancer Institute. Highlights from ASCO 2003. May 31, 2003. Available at: <http://www.cancer.gov/asco2003/highlights>, Accessed October 30, 2010.
7. André T, Boni C, Mounedji-Boudiaf L, et al. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med.* 2004;350:2343–51.
8. Tournigand C, André T, Bonnetain F, et al. Adjuvant therapy with fluorouracil and oxaliplatin in stage II and elderly patients (between ages 70 and 75 years) with colon cancer: subgroup analyses of the Multicenter International Study of Oxaliplatin, Fluorouracil, and Leucovorin in the Adjuvant Treatment of Colon Cancer trial. *J Clin Oncol.* 2012;30(27):3353-60.
9. Sanoff HK, Carpenter WR, Martin CF, et al. Comparative Effectiveness of Oxaliplatin vs Non-Oxaliplatin-containing Adjuvant Chemotherapy for Stage III Colon Cancer. *J Natl Cancer Inst.* 2012;104(3):211-27.
10. Haller DG, Cassidy J, Tabernero J, et al. Efficacy findings from a randomized phase III trial of capecitabine plus oxaliplatin versus bolus 5-FU/LV for stage III colon cancer (NO16968): Impact of age on disease-free survival (DFS). *J Clin Oncol.* 2010;28(15s). Abstract 3521.
11. André T, Boni C, Navarro M, et al: Improved overall survival with oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment in stage II or III colon cancer in the MOSAIC trial. *J Clin Oncol* 27:3109–3116, 2009.
12. McCleary NAJ, Meyerhardt J, Green E, et al. Impact of older age on the efficacy of newer adjuvant therapies in >12,500 patients with stage II/III colon cancer: findings from the ACCENT database [abstract]. *J Clin Oncol.* 2009;27(15s). 4010.
13. Marshall JL, de Gramont A, Köhne CH, et al. Accomplishments in 2008 in the Adjuvant Treatment of Colon Cancer. *Gastrointest Cancer Res.* 2009;3(suppl):S2-S7.

14. Cassidy J, Misset JL. Oxaliplatin-related side effects: characteristics and management. *Semin Oncol*. 2002 Oct;29(5 Suppl 15):11-20. Review.
15. ELOXATIN: Highlights of prescribing information. Sanofi-Aventis U.S. LLC. 2009. Available at: [http://www.accessdata.fda.gov/drugsatfda\\_docs/label/2009/021492s011,021759s009lbl.pdf](http://www.accessdata.fda.gov/drugsatfda_docs/label/2009/021492s011,021759s009lbl.pdf), Accessed 20 November 2011.
16. Burakgazi AZ, Messersmith W, Vaidya D, et al. Longitudinal assessment of oxaliplatin-induced neuropathy. *Neurology*. 2011 Sep 6;77(10):980-6. Epub 2011 Aug 24
17. Weickhardt A, Wells K, Messersmith W. Oxaliplatin-induced neuropathy in colorectal cancer. *J Oncol*. 2011;2011:201593. Epub 2011 Dec 12.
18. Barton, M. K. (2012), Oxaliplatin in the adjuvant treatment of colon cancer. *CA: A Cancer Journal for Clinicians*, 62: 3–4.
19. Sanoff HK, Goldberg RM, Pignone MP. Assessing the quality of initial consultations regarding adjuvant colon cancer therapy. *Clin Colorectal Cancer*. 2010 Apr;9(2):113-8.
20. Potosky AL, Harlan LC, Kaplan RS, et al. Age, sex, and racial differences in the use of standard adjuvant therapy for colorectal cancer. *J Clin Oncol*. 2002;20:1192–1202.
21. Haas JS, Brawarsky P, Iyer A, et al. Association of area sociodemographic characteristics and capacity for treatment with disparities in colorectal cancer care and mortality. *Cancer*. 2011 [epub ahead of print].
22. Abrams TA, Brightly R, Mao J, et al. Patterns of adjuvant chemotherapy use in a population-based cohort of patients with resected stage II or III colon cancer. *J Clin Oncol*. 2011 Aug 20;29(24):3255-62. Epub 2011 Jul 18.
23. Hodgson DC, Fuchs CS, Ayanian JZ. Impact of patient and provider characteristics on the treatment and outcomes of colorectal cancer. *J Natl Cancer Inst* 2001 ; 93 : 501 – 15.
24. Schrag D, Cramer LD, Bach PB, et al. Age and adjuvant chemotherapy use after surgery for stage III colon cancer. *J Natl Cancer Inst*. 2001;93:850-857.
25. Kahn KL, Adams JL, Weeks JC, et al. Adjuvant chemotherapy use and adverse events among older patients with stage III colon cancer. *JAMA*. 2010 Mar 17;303(11):1037-45.
26. Mack CD, Carpenter W, Meyer AM, et al. Racial disparities in receipt and comparative effectiveness of oxaliplatin for stage III colon cancer in older adults. *Cancer*. 2011 Nov 9. [Epub ahead of print]
27. Sargent DJ, Goldberg RM, Jacobson SD, et al. A pooled analysis of adjuvant chemotherapy for resected colon cancer in elderly patients. *N Engl J Med*. 2001;345:1091-1097.
28. Goldberg RM, Tabah-Fisch I, Bleiberg H, et al. Pooled analysis of safety and efficacy of oxaliplatin plus fluorouracil/leucovorin administered bimonthly in elderly patients with colorectal cancer. *J Clin Oncol*. 2006 Sep 1;24(25):4085-91.

29. Yellen SB, Cella DF, Leslie WT. Age and clinical decision making in oncology patients. *J Natl Cancer Inst* 1994;86:1766–70.
30. Ayanian JZ, Zaslavsky AM, Fuchs CS, et al. Use of adjuvant chemotherapy and radiation therapy for colorectal cancer in a population-based cohort. *J Clin Oncol*. 2003;21(7):1293-1300.
31. Jessup JM, Stewart A, Greene FL, Minsky BD. Adjuvant chemotherapy for stage III colon cancer. *JAMA*. 2005;294(21):2703-2711.
32. Haas JS, Brawarsky P, Iyer A, et al. Association of area sociodemographic characteristics and capacity for treatment with disparities in colorectal cancer care and mortality. *Cancer*. 2011 [epub ahead of print].
33. Morris AM, Billingsley KG, Hayanga AJ, et al. Residual treatment disparities after oncology referral for rectal cancer. *J Natl Cancer Inst*. 2008;100:738–744.
34. Potosky AL, Harlan LC, Kaplan RS, et al. Age, sex, and racial differences in the use of standard adjuvant therapy for colorectal cancer. *J Clin Oncol*. 2002;20:1192–1202.
35. Lund, JL. (2011). *The validity of administrative data and patterns of chemotherapy use among elderly colorectal cancer patients*. (Doctoral Dissertation). University of North Carolina at Chapel Hill. Retrieval information pending.
36. Kornmann M, Formentini A, Ette C, et al. Prognostic factors influencing the survival of patients with colon cancer receiving adjuvant 5-FU treatment. *Eur J Surg Oncol*. 2008 Dec;34(12):1316-21. Epub 2008 Mar 3.
37. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J Natl Cancer Inst*. 2004 Oct 6;96(19):1420-5.
38. American Society of Clinical Oncology and National Comprehensive Cancer Network, Inc. ASCO/NCCN Quality Measures: Breast and Colorectal Cancers. April 2007 UPDATE. Available at: <http://www.asco.org/ASCO/Downloads/Cancer%20Policy%20and%20Clinical%20Affairs/NCCN/ASCO%20NCCN%20Quality%20Measures%20table%20web%20posting%20with%20CoC%200507.pdf>, Accessed Mar 20, 2010].
39. Minasian L, Carpenter WR, Weiner BJ, et al. (2010). Translating Research into Evidence-Based Practice: The National Cancer Institute's Community Clinical Oncology Program. *Cancer*. 116(19):4440-4449.
40. Carpenter WR, Reeder-Hayes K, Bainbridge J, et al. (2011). The Role of Organizational Relationships and Research Networks in the Diffusion of Breast Cancer Treatment Innovation. *Medical Care*. 49(2):172-9.
41. McFall SL, Warnecke RB, Kaluzny AD, Ford L. Practice setting and physician influences on judgments of colon cancer treatment by community physicians. *Health Serv Res*. 1996 Apr;31(1):5-19.

42. Carpenter B, Meyer AM; Wu, Qaqish, et al. Translating research into practice: the role of provider based research networks in the diffusion of an evidence-based colon cancer treatment innovation. Under review.
43. National Research Council. "What Is Comparative Effectiveness Research?" Initial National Priorities for Comparative Effectiveness Research. Washington, DC: The National Academies Press, 2009.
44. Patient-Centered Outcomes Research Institute. Available at <http://www.pcori.org/>, Accessed December 15, 2012.
45. Aballéa S, Chancellor JV, Raikou M, et al. Cost-effectiveness analysis of oxaliplatin compared with 5-fluorouracil/leucovorin in adjuvant treatment of stage III colon cancer in the US. *Cancer*. 2007 Mar 15;109(6):1082-9.
46. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. National Academy of Sciences. 2000. Available at [http://www.nap.edu/html/quality\\_chasm/reportbrief.pdf](http://www.nap.edu/html/quality_chasm/reportbrief.pdf), Accessed December 15, 2012.
47. McKibbin T, Kwan P., Simon J., Koeller J. Rapid adoption of chemotherapy regimens for advanced colorectal cancer (CRC) in the community oncology setting. *J Clin Oncol* 26: 2008 (May 20 suppl; abstr 6563).
48. Mariotto A, Feuer EJ, Harlan LC, et al. Trends in use of adjuvant multi-agent chemotherapy and tamoxifen for breast cancer in the United States: 1975-1999. *J Natl Cancer Inst* 2002;94:1626–1634.
49. Buzdar A, Macahilig C. How rapidly do oncologists respond to clinical trial data? *Oncologist*. 2005 Jan;10(1):15-21.
50. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41–55.
51. Stürmer T, Joshi M, Glynn RJ, et al. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol* 2006; 59: 437–47.
52. Lori S. Parsons. Reducing Bias in a Propensity Score Matched-Pair Sample Using Greedy Matching Techniques. Ovation Research Group, Seattle, WA. Paper 214-26.
53. Seeger JD, Walker AM, Williams PL, et al. A propensity score-matched cohort study of the effect of statins, mainly fluvastatin, on the occurrence of acute myocardial infarction. *Am J Cardiol* 2003; 92: 1447–1451.
54. Seeger JD, Williams PL, Walker AM. An application of propensity score matching using claims data. *Pharmacoepidemiol Drug Saf* 2005; 14: 465-476.
55. Rassen JA, Glynn RJ, Rothman KJ, et al. Applying propensity scores estimated in a full cohort to adjust for confounding in subgroup analyses. *Pharmacoepidemiol Drug Saf* 2012 Jul;21(7):697-709.

56. Schneeweiss S, Gagne JJ, Glynn RJ, et al. Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. *Clin Pharmacol Ther* 2011 Dec;90(6):777-90. Epub 2011 Nov 2.
57. Gagne JJ, Glynn RJ, Rassen JA, et al. Active safety monitoring of newly marketed medications in a distributed data network: application of a semi-automated monitoring system. *Clin Pharmacol Ther* 2012 Jul;92(1):80-6. Epub 2012 May 16
58. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006;17:360e72.
59. Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica*. 47(1): 153–161.
60. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA*. 1994 Sep 21;272(11):859-66.
61. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology*. 2006 May;17(3):268-75.
62. D'Agostino RB Jr, D'Agostino RB Sr. Estimating treatment effects using observational data. *JAMA*. 2007;297(3):314-6.
63. Pirracchio R, Sprung C, Payen D, Chevret S. Benefits of ICU admission in critically ill patients: whether instrumental variable methods or propensity scores should be used. *BMC Med Res Methodol*. 2011 Sep 21;11:132.
64. Cain LE, Cole SR, Greenland S, et al. Effect of highly active antiretroviral therapy on incident AIDS using calendar period as an instrumental variable. *Am J Epidemiol*. 2009;169(9):1124-32.
65. Johnston KM, Gustafson P, Levy AR, et al. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Stat Med*. 2008;27:1539–1556.
66. Shetty KD, Vogt WB, Bhattacharya J. Hormone replacement therapy and cardiovascular health in the United States. *Med Care*. 2009;47(5):600-6.
67. Zhang Y. Cost-saving effects of olanzapine as long-term treatment for bipolar disorder. *J Ment Health Policy Econ*. 2008 Sep;11(3):135-46.
68. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf*. 2010 Jun;19(6):537-54. Review.
69. André T, Boni C, Navarro M, et al. Improved Overall Survival With Oxaliplatin, Fluorouracil, and Leucovorin As Adjuvant Treatment in Stage II or III Colon Cancer in the MOSAIC Trial. *J Clin Oncol*. 2009 July 1;27(19):3109–3116. 2009.

70. Kuebler JP, Wieand HS, O'Connell MJ, et al. Oxaliplatin combined with weekly bolus fluorouracil and leucovorin as surgical adjuvant chemotherapy for stage II and III colon cancer: results from NSABP C-07. *J Clin Oncol*. 2007 Jun 1;25(16):2198–2204.
71. Sanoff HK, Carpenter WR, Martin CF, et al. Comparative Effectiveness of Oxaliplatin vs Non-Oxaliplatin-containing Adjuvant Chemotherapy for Stage III Colon Cancer. *J Natl Cancer Inst*. 2012 Jan 20. [Epub ahead of print]
72. Colorectal Cancer Mortality Projections, National Cancer Institute, NIH, DHHS, Bethesda, MD, December 2007. Available at: <http://cisnet.cancer.gov/projections/colorectal/>, Accessed Dec 27, 2010.
73. National Cancer Institute. SEER-Medicare: Brief Description of the SEER-Medicare Database. Available at: <http://healthservices.cancer.gov/seermedicare/overview/>, Accessed Sept 22, 2010.
74. Warren JL, Klabunde CN, Schrag D, et al. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care*. 2002 Aug;40(8 Suppl):IV-3-18.
75. Office of Budget and Management. The 2002 Budget Guide. The White House. 2002. Available at: <http://www.whitehouse.gov/omb/rewrite/budget/fy2002/guide02.html>, Accessed: November 2009.
76. National Cancer Institute. SEER-Medicare: Brief Description of the SEER-Medicare Database. Available at: <http://healthservices.cancer.gov/seermedicare/overview/>, Accessed 22 September 2009.
77. National Cancer Institute. FDA Approval for Capecitabine. Jan 10, 2011. Available at: <http://www.cancer.gov/cancertopics/druginfo/fda-capecitabine>, Accessed February 1, 2012.
78. de Gramont A, de Gramont A, Chibaudel B, et al. GERCOR French Oncology Research Group. The evolution of adjuvant therapy in the treatment of early-stage colon cancer. *Clin Colorectal Cancer*. 2011 Dec;10(4):218-26.
79. Haller DG, Tabernero J, Maroun J, et al. Capecitabine plus oxaliplatin compared with fluorouracil and folinic acid as adjuvant therapy for stage III colon cancer. *J Clin Oncol*. 2011 Apr 10;29(11):1465-71. Epub 2011 Mar 7.
80. Bach PB, Guadagnoli E, Schrag D, et al. Patient demographic and socioeconomic characteristics in the SEER-Medicare database applications and limitations. *Med Care*. 2002 Aug;40(8 Suppl):IV-19-25.
81. Feinstein AR, Sosin DM, Wells CK. The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *N Engl J Med*. 1985 Jun 20;312(25):1604-8.
82. Kornmann M, Formentini A, Ette C, et al. Prognostic factors influencing the survival of patients with colon cancer receiving adjuvant 5-FU treatment. *Eur J Surg Oncol*. 2008 Dec;34(12):1316-21. Epub 2008 Mar 3.

83. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J Natl Cancer Inst.* 2004 Oct 6;96(19):1420-5.
84. National Cancer Institute. SEER-Medicare: Calculation of Comorbidity Weights. Dec 17, 2010. Available at: <http://healthservices.cancer.gov/seermedicare/program/comorbidity.html>, Accessed June 26, 2011.
85. Klabunde CN, Legler JM, Warren JL, et al. A refined comorbidity measurement algorithm for claims-based studies of breast, prostate, colorectal, and lung cancer patients. *Ann Epidemiol.* 2007 Aug;17(8):584-90. Epub 2007 May 25.
86. Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection in propensity score models: some insights from a simulation study. *American Journal of Epidemiology* 2006;163:1149-56.
87. Lori S. Parsons. Reducing Bias in a Propensity Score Matched-Pair Sample Using Greedy Matching Techniques. Ovation Research Group, Seattle, WA. Paper 214-26.
88. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *J Clin Oncol*, 1 (1983), 710–719
89. Stürmer T, Joshi M, Glynn RJ, et al. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol* 2006;59:437-47.
90. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results [electronic article]. *Int J Biostat.* 2007;3(1).
91. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc.* 1996;81:444–455.
92. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology.* 2006; 17(4): 360-72.
93. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc.* 1996;81:444–455.
94. Rothman KJ. Analysis of Crude Data. In: Rothman KJ, ed. *Modern epidemiology*. 1st ed. Boston/Toronto: Little, Brown and Company; 1986:172.
95. Balke A, Pearl J. Bounds on Treatment Effects From Studies With Imperfect Compliance. *Journal of the American Statistical Association.* 1997; 92:1171-1176.
96. Mack CD, Glynn RJ, Brookhart MA, et al. Calendar Time-Specific Propensity Scores and Comparative Effectiveness Research for Stage III Colon Cancer Chemotherapy. *Pharmacoepidemiol Drug Saf.* 2013. [Epub ahead of print]

97. Glymour MM, Greenland S. Causal Diagrams. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008:183-194.
98. Carpenter WR, Reeder-Hayes K, Bainbridge J, et al. The Role of Organizational Relationships and Research Networks in the Diffusion of Breast Cancer Treatment Innovation. *Med Care*. 2011;49(2):172-9.
99. Lori S. Parsons. Reducing Bias in a Propensity Score Matched-Pair Sample Using Greedy Matching Techniques. Ovation Research Group, Seattle, WA. Paper 214-26.
100. National Cancer Institute. SEER Extent of Disease – 1998. Codes and coding Instructions. Third edition. Eds. Fritz A, Ries L. January 1998, changes marked through April 2002. Available at: [seer.cancer.gov/manuals/EOD10Dig.pub.pdf](http://seer.cancer.gov/manuals/EOD10Dig.pub.pdf), Accessed 14 January 2012.
101. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 1983; 70: 41–55.
102. Stürmer T, Schneeweiss S, Brookhart MA, et al. Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal antiinflammatory drugs and short-term mortality in the elderly. *Am J Epidemiol*. 2005 May 1;161(9):891-8.
103. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; 127: 757–763
104. National Cancer Institute. Highlights from ASCO 2003. May 31, 2003. Available at: <http://www.cancer.gov/asco2003/highlights>, Accessed October 30, 2010.
105. André T, Boni C, Mounedji-Boudiaf L, et al. Multicenter International Study of Oxaliplatin/5-Fluorouracil/Leucovorin in the Adjuvant Treatment of Colon Cancer (MOSAIC) Investigators. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med* 2004;350:2343–51.
106. National Cancer Institute. FDA Approval for Oxaliplatin. Available at: <http://www.cancer.gov/cancertopics/druginfo/fda-oxaliplatin>, Accessed 27 December 2012.
107. Warren JL, Klabunde CN, Schrag D, et al. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care*. 2002 Aug;40(8 Suppl):IV-3-18.
108. National Cancer Institute. SEER Program Overview. <http://seer.cancer.gov/> (accessed 24 Sept 2010).
109. Glymour MM, Greenland S. Causal Diagrams. In *Modern epidemiology*. Rothman KJ, Greenland S, Lash TL. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2008: 183-194.



110. Sanoff HK, Carpenter WR, Martin CF, et al. Comparative Effectiveness of Oxaliplatin vs Non-Oxaliplatin-containing Adjuvant Chemotherapy for Stage III Colon Cancer. *J Natl Cancer Inst.* 2012 Feb 8;104(3):211-27. Epub 2012 Jan 20.
111. Parsons LS. Reducing bias in a propensity score matched pair sample using greedy matching techniques. Proceedings of the 26th annual SAS users group international conference, Cary, NC. <http://www2.sas.com/proceedings/sugi26/p214-26.pdf>.
112. Stürmer T, Joshi M, Glynn RJ, et al. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol* 2006;59:437-47.
113. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *J Clin Oncol* 1983 Nov;1(11):710-9.
114. Seeger JD, Walker AM, Williams PL, et al. A propensity score-matched cohort study of the effect of statins, mainly fluvastatin, on the occurrence of acute myocardial infarction. *Am J Cardiol* 2003; 92: 1447–1451.
115. Seeger JD, Williams PL, Walker AM. An application of propensity score matching using claims data. *Pharmacoepidemiol Drug Saf* 2005; 14: 465-476.
116. Rassen JA, Glynn RJ, Rothman KJ, et al. Applying propensity scores estimated in a full cohort to adjust for confounding in subgroup analyses. *Pharmacoepidemiol Drug Saf* 2012 Jul;21(7):697-709. Epub 2011 Dec 8.
117. Mack CD, Glynn RJ, Stürmer T. Calendar Time-Specific Propensity Scores and Hazard Ratio Estimation. *Pharmacoepidemiol Drug Saf* 2011; 20(suppl 1):S296; Oral presentation, 27th International Conference on Pharmacoepidemiology and Therapeutic Risk Management, Chicago, IL 2011.
118. Schneeweiss S, Gagne JJ, Glynn RJ, et al. Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. *Clin Pharmacol Ther* 2011 Dec;90(6):777-90. Epub 2011 Nov 2.
119. Gagne JJ, Glynn RJ, Rassen JA, et al. Active safety monitoring of newly marketed medications in a distributed data network: application of a semi-automated monitoring system. *Clin Pharmacol Ther* 2012 Jul;92(1):80-6. 2011.369. Epub 2012 May 16
120. Burakgazi AZ, Messersmith W, Vaidya D, et al. Longitudinal assessment of oxaliplatin-induced neuropathy. *Neurology* 2011 Sep 6;77(10):980-6. Epub 2011 Aug 24
121. Weickhardt A, Wells K, Messersmith W. Oxaliplatin-induced neuropathy in colorectal cancer. *J Oncol* 2011;2011:201593. Epub 2011 Dec 12.
122. Dyck PJ, Kratz KM, Karnes JL, et al. The prevalence by staged severity of various types of diabetic neuropathy, retinopathy, and nephropathy in a population-based cohort: the Rochester Diabetic Neuropathy Study. *Neurology* 1993 Apr;43(4):817-24.

123. Ramanathan RK, André T, Rothenberg ML, et al. Diabetes mellitus (DM) and the incidence and time to onset of oxaliplatin induced peripheral sensory neurotoxicity (PSN) in patients with colorectal cancer. A pooled analysis of 3 randomized studies. American Society of Clinical Oncology. 2008 Gastrointestinal Cancers Symposium. Abstract 347. [http://www.asco.org/ascov2/Meetings/Abstracts?&vmview=abst\\_detail\\_view&confID=53&abstractID=10742](http://www.asco.org/ascov2/Meetings/Abstracts?&vmview=abst_detail_view&confID=53&abstractID=10742) (accessed 9 May 2012).
124. Haas JS, Brawarsky P, Iyer A, et al. Association of area sociodemographic characteristics and capacity for treatment with disparities in colorectal cancer care and mortality. *Cancer* 2011 Sep 15;117(18):4267-76. Epub 2011 Mar 16.
125. André T, Boni C, Navarro M, et al. Improved Overall Survival With Oxaliplatin, Fluorouracil, and Leucovorin As Adjuvant Treatment in Stage II or III Colon Cancer in the MOSAIC Trial. *J Clin Oncol* 2009 July 1;27(19):3109–3116. 2009.
126. Sanoff HK, Carpenter WR, Martin CF, et al. Comparative effectiveness of oxaliplatin vs non-oxaliplatin-containing adjuvant chemotherapy for stage III colon cancer. *J Natl Cancer Inst* 2012 Feb 8;104(3):211-27. Epub 2012 Jan 20.
127. Schneeweiss S, Patrick AR, Stürmer T, et al. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *MedCare* 2007 Oct;45(10 Supl 2):S131-42.
128. Stürmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution--a simulation study. *Am J Epidemiol* 2010 Oct 1;172(7):843-54. Epub 2010 Aug 17.
129. Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. *Am J Epidemiol* 2006 Jun 15;163(12):1149-56. Epub 2006 Apr 19.
130. Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med* 2007 Feb 20;26(4):734-53.
131. Rassen JA, Doherty M, Huang W, Schneeweiss S. Pharmacoepidemiology Toolbox. Boston, MA. <http://www.hdpharmacoepi.org> (accessed 25 July 2012).
132. Mack CD, Brookhart MA, Glynn RJ, T Stürmer. Calendar Time as an Instrumental Variable in Nonexperimental Comparative Effectiveness Research of Dynamic Therapies. *Am J Epidemiol* 2012; 175(suppl 11):181-S; Oral presentation, 45th Annual meeting Society for Epidemiologic Research, Minneapolis, USA. 2012.
133. Cain LE, Cole SR, Greenland S, et al. Effect of highly active antiretroviral therapy on incident AIDS using calendar period as an instrumental variable. *Am J Epidemiol* 2009 May 1;169(9):1124-32. Epub 2009 Mar 24.
134. Johnston KM, Gustafson P, Levy AR, et al. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Stat Med* 2008;27:1539-1556.

135. Brookhart MA, Stürmer T, Glynn RJ, et al. Confounding control in healthcare database research: challenges and potential approaches. *Med Care* 2010 Jun;48(6 Suppl):S114-20.
136. Bugeja G, Kumar A, Banerjee AK. Exclusion of elderly people from clinical research: a descriptive study of published reports. *BMJ* 1997 Oct 25;315(7115):1059.
137. Sanoff HK, Carpenter WR, Freburger J, et al. Comparison of adverse events during 5-fluorouracil versus 5-fluorouracil/oxaliplatin adjuvant chemotherapy for stage III colon cancer: A Population-based analysis. *Cancer* 2012 Jan 31. [Epub ahead of print]
138. Abraham A, Habermann EB, Rothenberger DA, et al. Adjuvant chemotherapy for stage III colon cancer in the oldest old: Results beyond clinical guidelines. *Cancer* 2012 Jul 17. [Epub ahead of print]
139. Altekruse SF, Kosary CL, Krapcho M, et al (eds). SEER Cancer Statistics Review, 1975-2007, National Cancer Institute. Bethesda, MD, [http://seer.cancer.gov/csr/1975\\_2007/](http://seer.cancer.gov/csr/1975_2007/), based on November 2009 SEER data submission, posted to the SEER web site, 2010 (accessed 20 Mar 2010).
140. André T, Boni C, Mounedji-Boudiaf L, et al. Multicenter International Study of Oxaliplatin/5-Fluorouracil/Leucovorin in the Adjuvant Treatment of Colon Cancer (MOSAIC) Investigators. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med* 2004;350:2343–51.
141. Cooper GS, Virnig B, Klabunde CN, et al. Use of SEER-Medicare data for measuring cancer surgery. *Med Care* 2002;40:IV-43-8.
142. Warren JL, Harlan LC, Fahey A, et al. Utility of the SEER-Medicare data to identify chemotherapy use. *Med Care* 2002;40:IV-55-61.
143. Glynn RJ, Knight EL, Levin R, Avorn J. Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology* 2001 Nov;12(6):682-9.
144. McCleary NAJ, Meyerhardt J, Green E, et al. Impact of older age on the efficacy of newer adjuvant therapies in >12,500 patients with stage II/III colon cancer: findings from the ACCENT database [abstract]. *J Clin Oncol*. 2009;27(15s). 4010.
145. Stürmer T, Jonsson Funk M, Poole C, et al. Nonexperimental comparative effectiveness research using linked healthcare databases. *Epidemiology*. 2011;22(3):298-301. Review.
146. Meyer AM, Carpenter WR, Abernethy AP, et al. Data for cancer comparative effectiveness research: past, present, and future potential. *Cancer*. 2012; 118(21):5186-97.
147. Glynn RJ, Schneeweiss S, Stürmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol*. 2006;98:253-9
148. de Gramont A., Banzi M., Navarro M., et al. Oxaliplatin/5-FU/LV in adjuvant colon cancer: Results of the international randomized mosaic trial [abstract]. Presented at

2003 ASCO Annual Meeting, Chicago, IL, US, May 31–June 3, 2003. *Proc Am Soc Clin Oncol* 22: 2003 (abstr 1015).

149. Aballéa S, Chancellor JV, Raikou M, et al. Cost-effectiveness analysis of oxaliplatin compared with 5-fluorouracil/leucovorin in adjuvant treatment of stage III colon cancer in the US. *Cancer*. 2007;109(6):1082-9.

150. American Joint Committee on Cancer. Summary of Changes: Understanding the Changes from the Sixth to the Seventh Edition of the AJCC Cancer Staging Manual. Chicago, IL: American Joint Committee on Cancer; 2011.

151. Schneeweiss S, Patrick AR, Stürmer T, et al. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *MedCare* 2007 Oct;45(10 Supl 2):S131-42.

152. Rassen JA, Doherty M, Huang W, Schneeweiss S. Pharmacoepidemiology Toolbox. Boston, MA. <http://www.hdpharmacoepi.org> (accessed 25 July 2012).

153. Mack CD, Brookhart MA, Glynn RJ, T Stürmer. Calendar Time as an Instrumental Variable in Nonexperimental Comparative Effectiveness Research of Dynamic Therapies. *Am J Epidemiol* 2012; 175(suppl 11):181-S; Oral presentation, 45th Annual meeting Society for Epidemiologic Research, Minneapolis, USA. 2012.

154. Cain LE, Cole SR, Greenland S, et al. Effect of highly active antiretroviral therapy on incident AIDS using calendar period as an instrumental variable. *Am J Epidemiol* 2009 May 1;169(9):1124-32. Epub 2009 Mar 24.

155. Johnston KM, Gustafson P, Levy AR, et al. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Stat Med* 2008;27:1539-1556.

156. Brookhart MA, Stürmer T, Glynn RJ, et al. Confounding control in healthcare database research: challenges and potential approaches. *Med Care* 2010 Jun;48(6 Suppl):S114-20.

157. Kohler BA, Ward E, McCarthy BJ, et al. Annual report to the nation on the status of cancer, 1975-2007, featuring tumors of the brain and other nervous system. *J Natl Cancer Inst*. 2011 May 4;103(9):714-36. Epub 2011 Mar 31.

158. Bugeja G, Kumar A, Banerjee AK. Exclusion of elderly people from clinical research: a descriptive study of published reports. *BMJ* 1997 Oct 25;315(7115):1059.

159. Cooper GS, Virnig B, Klabunde CN, et al. Use of SEER-Medicare data for measuring cancer surgery. *Med Care*. 2002;40:IV-43-8.

160. Glynn RJ, Knight EL, Levin R, Avorn J. Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology* 2001 Nov;12(6):682-9.