Walden University
A higher degree. A higher purpose.

Walden University
ScholarWorks

Walden Dissertations and Doctoral Studies

Walden Dissertations and Doctoral Studies Collection

2015

Disparities in Monoclonal Antibody Treatment of Elderly Metastatic Colorectal Cancer Patients

Krista Marie Schroeder
Walden University

Follow this and additional works at: https://scholarworks.waldenu.edu/dissertations

Part of the Epidemiology Commons, and the Public Health Education and Promotion Commons

This Dissertation is brought to you for free and open access by the Walden Dissertations and Doctoral Studies Collection at ScholarWorks. It has been accepted for inclusion in Walden Dissertations and Doctoral Studies by an authorized administrator of ScholarWorks. For more information, please contact ScholarWorks@waldenu.edu.

Walden University

College of Health Sciences

This is to certify that the doctoral dissertation by

Krista M. Schroeder

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Raymond Panas, Committee Chairperson, Public Health Faculty
Dr. Gudeta Fufaa, Committee Member, Public Health Faculty
Dr. Roland Thorpe, University Reviewer, Public Health Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2015

Abstract

Disparities in Monoclonal Antibody Treatment of Elderly Metastatic Colorectal Cancer Patients

by

Krista Marie Schroeder

MHS, Quinnipiac University, 2006

BS, The Ohio State University, 2004

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

August 2015

Abstract

Multiple research studies have demonstrated racial, socioeconomic status (SES), and neighborhood disparities in first-line treatment of colorectal cancer patients, including those with metastatic colorectal cancer. However, disparities in adjunct monoclonal antibody treatment disparities have not been explored. The purpose of this study was to assess racial, SES, and neighborhood disparities in adjunct monoclonal antibody treatment of elderly metastatic colorectal cancer patients. The research was rooted in 3 theories: the fundamental cause theory, the diffusion of innovations theory, and theory of health disparities and medical technology. Data from the SEER-Medicare database and logistic regression were used to assess the relationship between the variables of interest and adjunct monoclonal antibody therapy. In this study, race ($p = 0.070$), SES ($p = 0.881$), and neighborhood characteristics ($p = 0.309$) did not significantly predict who would receive monoclonal antibody therapy. The results demonstrated a potential improvement in historically documented colorectal cancer treatment disparities. Specifically, historical treatment disparities may not be relevant to newer therapies prescribed to patients with severe disease. The difference could be related to improved access to care or a change in treatment paradigm due to the severity of metastatic colorectal cancer. Future studies aimed at understanding the causes of this social change (i.e., reduced treatment disparities) are warranted. Understanding the root cause of the reduced treatment disparities observed in this study could be used to reduce treatment disparities in other cancer populations.

Disparities in Monoclonal Antibody Treatment of Elderly Metastatic Colorectal Cancer Patients

by

Krista Marie Schroeder

MHS, Quinnipiac University, 2006

BS, The Ohio State University, 2004

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

August 2015

Dedication

I dedicate this dissertation project to my Uncle Mel, Grandpa Al, and all colorectal cancer survivors.

I also dedicate this dissertation project to my husband Andrew and our daughter Katie. I hope this project serves as inspiration that she should reach for the stars and follow her dreams. Nothing is unreachable.

Acknowledgments

I would like to thank Dr. Panas (chair; content expert), Dr. Fufaa (committee member; methodology expert) and Dr. Thorpe (URR) for serving on my committee and making this project possible.

This study used the linked SEER-Medicare database. The interpretation and reporting of these data are the sole responsibility of the authors. The authors acknowledge the efforts of the National Cancer Institute; the Office of Research, Development and Information, CMS; Information Management Services (IMS), Inc.; and the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER-Medicare database.

This study employed data from multiple SEER registries including the California Cancer Registry. The California Cancer Registry requires the following statement: The collection of cancer incidence data used in this study was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California, contract HHSN261201000035C awarded to the University of Southern California, and contract HHSN261201000034C awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement # U58DP003862-01 awarded to the California Department of Public Health.

The ideas and opinions expressed herein are those of the author(s) and endorsement by the State of California Department of Public Health, the National Cancer Institute, and the Centers for Disease Control and Prevention or their Contractors and Subcontractors is not intended nor should be inferred. The authors acknowledge the efforts of the National Cancer Institute; the Office of Research, Development and Information, CMS; Information Management Services (IMS), Inc.; and the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER-Medicare database.

Table of Contents

List of Tables vii

List of Figures x

Chapter 1: Introduction to the Study 1

Introduction 1

Background 2

Literature Related to the Study Topic 2

Research Gap 3

Study Importance 4

Problem Statement 5

Purpose of the Study 5

Study Methodology and Intent 5

Study Variables 6

Research Questions and Hypotheses 7

Theoretical Framework for the Study 8

Theoretical Propositions 9

Theoretical Framework Supports the Study Approach 9

Nature of the Study 10

Design Rationale 11

Methodology 11

Definitions 12

Study Variables 12
Study Terms 14
Assumptions 15
Scope and Delimitations 15
Scope 15
Delimitations 16
Limitations and Methods to Address Limitations 16
Design/Methodology Weaknesses 16
Biases 18
Significance 18
Advance Knowledge of the Discipline 19
Potential Practice/Policy Contributions 19
Potential for Positive Social Change 20
Summary 20
Chapter 2: Literature Review 23
Introduction 23
Literature Search Strategy 25
Databases and Search Terms 26
Literature Search Strategy: Summary 32
Theoretical Foundation 32
Fundamental Cause Theory 33

Diffusion of Innovations Theory 36
Theory of Health Disparities and Medical Technology 38
Theoretical Foundation Summary 40
Literature Review 42
Colorectal Cancer Epidemiology 43
Review of Studies Related to Constructs and Methods 43
Evaluation of Strengths and Weaknesses of Previous Approaches 75
Justification of Chosen Variables 78
Review and Synthesis of Studies Related to the Research Questions 78
Summary and Conclusion 80
Chapter 3: Research Method 82
Introduction 82
Research Design and Rationale 82
Research Design 82
Methodology 88
Population 88
Sampling and Sampling Procedures 90
Power Analysis 95
Procedures for Recruitment, Participation, and Data Collection 100
Research Questions and Hypothesizes 105
Data Analysis Plan 106

Hardware 106
Software 107
Data Preparation 107
Statistical Tests 110
Covariates and Confounding Variables 115
Results Interpretation 116
Threats to Validity 118
Threats to External Validity 118
Threats to Internal Validity 119
Threats to Construct or Statistical Conclusion Validity 120
Ethical Procedures 121
Summary 122
Chapter 4: Results 125
Introduction 125
Data Collection 127
Cohort Selection 129
Demographics of the Study Cohort 131
Analysis of Covariates 137
Gender 138
Reporting Registry 139
Medicare Entitlement Characteristics 141

Age at Diagnosis 142
Descriptive Statistics 144
Overall Rate of Adjunct Monoclonal Antibody Treatment 145
Monoclonal Antibody Treatment Rate by Race 146
Monoclonal Antibody Treatment Rate by SES 147
Monoclonal Antibody Treatment Rate by Neighborhood Characteristics 147
Statistical Assumptions 148
Statistical Analysis of Research Questions 149
Monoclonal Antibody Disparities by Race 149
Monoclonal Antibody Disparities by SES 151
Monoclonal Antibody Disparities by Neighborhood Characteristics 153
Final Logistic Regression Model 156
Summary 158
Chapter 5: Discussion, Conclusions, and Recommendations 159
Introduction 159
Interpretation of the Findings 160
Results as They Relate to the Literature 161
Theoretical Framing of the Results 162
Limitations of the Study 165
Recommendations 168
Implications 171

Conclusion 172

References.....175

Appendix A: SEER-Medicare Application Form 195

Appendix B: SEER Medicare Use Agreement 202

Appendix C: Power Analyses 207

Appendix D: Database Generation and Cohort Selection Code (Organized by File Generated).....210

Appendix F: SEER-Medicare Data Approve to Use 238

Appendix G: Logistic Regression Syntax; Monoclonal Antibody Disparities by Race.....240

Appendix H: Logistic Regression Syntax; Monoclonal Antibody Disparities by SES.....241

Appendix I: Logistic Regression Syntax; Monoclonal Antibody Disparities by Neighborhood Characteristics.....242

Appendix J. Logistic Regression Syntax; Model Including All Significant Independent Variables and Covariates.....243

List of Tables

Table 1. Colorectal Cancer Epidemiology Primary Literature From PubMed and Science Direct 28

Table 2. Colorectal Cancer Disparities Primary Literature From PubMed and Science Direct 30

Table 3. Colorectal Cancer Monoclonal Antibody Treatment Disparities PrimaryLiterature From PubMed and Science Direct 31

Table 4. Advanced Age-Related Colorectal Cancer Survival Disparities 46

Table 5. Socioeconomic Colorectal Cancer Survival Disparities 49

Table 6. Racial Colorectal Cancer Survival Disparities 54

Table 7. Neighborhood Characteristics Related to Colorectal Cancer Survival Disparities.....59

Table 8. Access to Care, Screening and Treatment Disparities.....65

Table 9. Monoclonal Antibody Treatment of Elderly Colorectal Cancer Patients.....73

Table 10. Strengths and Weaknesses of Previous Data Collection Methods Related to the Study.....76

Table 11. Logistic Regression Analyses.....114

Table 12. Demographics of the Study Cohort by Year – Race.....132

Table 13. Demographics of the Study Cohort by Year – Gender.....133

Table 14. Demographics of the Study Cohort by Year – Age at Diagnosis.....133

Table 15. Demographics of the Study Cohort by Year – Socioeconomic Status 134

Table 16. Demographics of the Study Cohort by Year – Reporting Registry 135

Table 17. Demographics of the Study Cohort by Year – Neighborhood Characteristics 136

Table 18. Demographics of the Study Cohort by Year – Reason for Original Medicare Entitlement 137

Table 19. Monoclonal Antibody Treatment Rate by Gender 138

Table 20. Differences in Gender based on Monoclonal Antibody Treatment Rate.....139

Table 21. Monoclonal Antibody Treatment Rate by Reporting Registry.....140

Table 22. Differences in Reporting Registry based on Monoclonal Antibody Treatment Rate 140

Table 23. Monoclonal Antibody Treatment by Medicare Entitlement Characteristics 141

Table 24. Differences in Medicare Entitlement Characteristics based on Monoclonal Antibody Treatment Status 142

Table 25. Monoclonal Antibody Treatment by Age at Diagnosis 142

Table 26. Differences in Average Age at Diagnosis based on Monoclonal Antibody Treatment Status 144

Table 27. Monoclonal Antibody Receipt by Diagnosis Year 146

Table 28. Differences in Diagnosis Year Based on Monoclonal Antibody Treatment Status 146

Table 29. Monoclonal Antibody Treatment Rate by Race 147

Table 30. Monoclonal Antibody Treatment Rate by SES Group 147

Table 31. Monoclonal Antibody Treatment Rate by Neighborhood Characteristics 148

Table 32. Logistic Regression Analysis – Race and Monoclonal Antibody Treatment 151

Table 33. Logistic Regression Analysis – SES and Monoclonal Antibody Treatment 153

Table 34. Logistic Regression Analysis – Neighborhood Characteristics and Monoclonal Antibody Treatment.....155

Table 35. Logistic Regression Analysis – Final Model with Diagnosis Age and Reporting Registry 157

List of Figures

Figure 1. Pictorial representation of the software used and files generated during the cohort selection process.....108

Figure 2. Flow chart showing the selection of the cohort based on multiple different inclusion and exclusion criteria. 130

Figure 3. Bar graph showing the distribution of monoclonal antibody treatment by age at diagnosis in the cohort.....143

Chapter 1: Introduction to the Study

Introduction

In the United States, colorectal cancer survival rates, including survival rates of individuals with metastatic colorectal cancer, have been associated with race, socioeconomic status (SES), and neighborhood characteristics (degree of urbanization; Hines, Markossian, Johnson, Dong, & Bayakly, 2014; Lian et al., 2011; Simpson, Pagán, & Chen, 2013). Specifically, Black Americans, individuals of low SES, and individuals residing in rural neighborhoods have been shown to have increased colorectal cancer mortality rates (Centers for Disease Control and Prevention [CDC], 2012; Hines et al., 2014; Lian et al., 2011; Simpson et al., 2013). Differences in colorectal cancer treatment regimen (surgery, radiation, or chemotherapy) have been shown to contribute to these survival disparities (Hao et al., 2011; Le, Ziogas, Lipkin, & Zell, 2008; Rane et al., 2014). In 2004, the first monoclonal antibody therapy for colorectal cancer was approved for treatment of metastatic colorectal cancer when given in adjunct with chemotherapy (Scott, Wolchok, & Old, 2012). A review of Phase II-III clinical trials indicated that monoclonal antibody therapy increases survival in metastatic colorectal cancer patient populations when added to chemotherapy (Tol & Punt, 2010). It is possible that disparities in monoclonal antibody treatment could contribute to disparities in the survival of patients with metastatic colorectal cancer. However, due to the lack of research, it is unknown if there are racial, SES, or neighborhood disparities in monoclonal antibody treatment of metastatic colorectal cancer patients.

This chapter provides an introduction to the research project including a summarization of items that will be described in detail in Chapters 2 and 3. Specifically, I open this chapter with a background section summarizing the literature surrounding the topic, present the research gap, and describe the importance of the study. Following this brief background, the problem statement and study purpose are presented. Afterward an overview of the research methodology and the theoretical basis will be provided. Finally, the scope and limitations of the research project are described.

Background

In this section, I provide a brief summary of literature related to the topic and the research gap, and I describe the importance of the study. Additional details related to the literature search and findings are provided in Chapter 2.

Literature Related to the Study Topic

Colorectal cancer survival disparities based on race, SES, and neighborhood characteristics have been observed in many different studies. Specifically, Black Americans with colorectal cancer, including metastatic colorectal cancer, have reduced survival rates compared to White Americans (Simpson et al., 2013; Sineshaw, Robbins, & Jemal, 2014; Wallace et al., 2013; Wassira, Pinheiro, Symanowski, & Hansen, 2013), low SES populations have reduced colorectal cancer survival rates compared to higher SES populations (Hines et al., 2014; Oliphant et al., 2013b; Wassira et al., 2013), and rural populations have reduced colorectal cancer survival rates compared to urban populations (Henry, Niu, & Boscoe, 2009; Lian et al., 2011).

In addition, multiple researchers have associated treatment regimen differences with these survival disparities, highlighting disparities in both type and aggressiveness of colorectal cancer therapies (Bakogeorgos et al., 2013; Obeidat et al., 2010; Serra-Rexach et al., 2012). Le et al. (2008) found significant differences in first-line treatment (surgery, radiation, and chemotherapy) comparing Black Americans to White Americans. These racial disparity results were supported in subsequent studies by White et al. (2008), Hao et al. (2011), and Hines et al. (2012). In addition, Le et al. found significant differences in first-line treatment (surgery, radiation, and chemotherapy) comparing higher SES to lower SES areas. Another potential contributor to colorectal cancer treatment disparities are the characteristics of the neighborhood patients reside within. Hao et al. (2011) documented urban versus rural disparities in first-line chemotherapy treatment in colorectal cancer. Specifically, populations with urban or suburban zip codes were, respectively, 38% and 52% more likely to receive chemotherapy compared to populations with rural zip codes (Hao et al., 2011).

Research Gap

At highlighted above, multiple studies have provided valuable information on colorectal cancer treatment disparities. However, they were conducted using patient data from the 1990s and early 2000, prior to the approval of monoclonal antibodies (Scott et al., 2012). Currently, there are three monoclonal antibodies (bevacizumab, cetuximab, and panitumumab) approved for the treatment of metastatic colorectal cancer (Scott et al., 2012). Of particular importance to this research study, the literature searches performed

did not uncover any studies addressing racial, SES, or neighborhood (degree of urbanization) disparities in adjunct monoclonal antibody treatment of colorectal cancer patients. Adjunct monoclonal antibody therapy has been shown to reduce metastatic colorectal cancer mortality in multiple populations, including the elderly (Bruera et al., 2013, Cunningham et al., 2013; Naeim et al., 2013). Therefore, it is possible that disparities in monoclonal antibody treatment are contributing to disparities in metastatic colorectal cancer survival. However, given the lack of research, understanding disparities in monoclonal antibody treatment of metastatic colorectal cancer patients is a clear research gap.

Study Importance

Determining if racial, SES, or neighborhood disparities in monoclonal antibody treatment of elderly metastatic colorectal cancer patients exist is a critical first step to eliminating these social inequalities. This study may provide the basis for positive social change in one of two ways. First, if treatment disparities are found, this would illuminate the need for policies that improve access to monoclonal antibodies, thereby helping reduce social inequalities in colorectal cancer survival. Alternatively, if treatment disparities are not found, an opportunity exists to understand the root cause of the reduced treatment disparities, and this knowledge could potentially be used to reduce treatment disparities in other cancer populations.

Problem Statement

Multiple research studies have demonstrated racial, SES, and neighborhood (degree of urbanization) disparities in treatment of colorectal cancer patients with surgery, radiation, or chemotherapy (Aarts, Lemmens, Louwman, Kunst, & Coebergh, 2010; Hao et al., 2011; Le et al., 2008). These treatment disparities have been associated with colorectal cancer survival (Le et al., 2008). However, due to the lack of research, it was unknown if there are racial, SES, or neighborhood disparities in monoclonal antibody treatment of elderly metastatic colorectal cancer patients.

Purpose of the Study

Disparities in chemotherapy, radiation, and/or surgery have been associated with disparities in colorectal cancer survival. However, information regarding disparities in adjunct monoclonal antibody treatment was not found in the literature. Therefore, my aim was to determine whether these historical colorectal cancer treatment disparities have persisted into newer monoclonal antibody therapies and the metastatic colorectal cancer population. The expressed purpose of this study was to determine if there are racial, SES, or neighborhood (degree of urbanization) disparities in first-line adjunct monoclonal antibody treatment of elderly metastatic colorectal cancer patients.

Study Methodology and Intent

In this quantitative research study, I used a retrospective cohort study design and data from the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER)-Medicare linked database, years 2007 to 2012.

Study Variables

The independent, dependent, and control variables used in this research project are described briefly below. Detailed information on each variable, including operational definitions, the location of the data and data codes, can be found in Chapter 3.

Independent variables. The first research question addresses racial/ethnic disparities in first-line monoclonal antibody treatment. Therefore, race was the first independent variable. The second research question addresses SES disparities in first-line monoclonal antibody treatment. The census tract poverty indicator, located in the NCI's Patient Entitlement and Diagnosis Summary File (PEDSF), was used as a surrogate for SES. The final research question addresses neighborhood characteristic (degree of urbanization) differences in first-line monoclonal antibody treatment. The 2003 Rural/Urban Continuum Codes from the Economic Research Service (ERS), which categories counties on an urban/rural scale based on population size, degree of urbanization, and adjacency to a metro or nonmetro area, was used for this purpose (NCI, 2015c).

Dependent variable. The dependent variable for these research questions is receipt of first-line adjunct (in combination with chemotherapy) monoclonal antibody therapy, "yes" or "no."

Covariates. Four covariates were added to the logistic regression analysis to control for confounding. The first covariate used was gender (male or female). The second covariate used was age at diagnosis. The third covariate was reason for original

Medicare entitlement (age or disability). Individuals who qualified for Medicare due to End Stage Renal Disease were excluded from the study. The fourth covariate was the registry that reported the data.

Research Questions and Hypotheses

The three research questions (RQ1-RQ3) addressed by this research project are listed below. In addition, the null and alternative hypotheses have been stated ($H_0$1-$H_0$3 and H_a1-H_a3 respectively).

RQ1: Within the population of elderly (65 years+) individuals diagnosed with metastatic colorectal cancer from January 2007 to December 2011, are there racial disparities in first line adjunct monoclonal antibody treatment?

$H_0$1: There is no significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody treatment based on race.

H_a1: There is a significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody treatment based on race.

RQ2: Within the population of elderly (65 years+) individuals diagnosed with metastatic colorectal cancer from January 2007 to December 2011, are there socioeconomic disparities in first line adjunct monoclonal antibody treatment?

$H_0$2: There is no significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct

monoclonal antibody therapy based on area socioeconomic status (as defined by the PEDSF census tract poverty indicator).

H_a2: There is a significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on area socioeconomic status (as defined by the PEDSF census tract poverty indictor).

RQ3: Within the population of elderly (65 years+) individuals diagnosed with metastatic colorectal cancer from January 2007 to December 2011, are there neighborhood characteristic (degree of urbanization) disparities in first line adjunct monoclonal antibody treatment?

H_03: There is no significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on neighborhood characteristics.

H_a3: There is a significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on neighborhood characteristics.

Theoretical Framework for the Study

The theoretical framework used for this project was a combination of theories similar to the theoretical framework proposed by Chang and Lauderdale (2009). The overarching theory was the fundamental cause theory. However, two additional theories (the diffusion of innovations theory and the theory of health disparities and medical

technology) were used to extend the fundamental cause theory for the purpose of this project. These theories and their relevance to the research topic are described briefly in the following sections.

Theoretical Propositions

The fundamental cause theory was first proposed by Link and Phelan in 1995. The theory argues that social states, such as race and SES, contribute to disease rates/outcomes. Since resources are constantly changing, the authors proposed that it is the beneficial social connections that serve to protect health regardless of the resource mechanism (Phelan et al., 2010). In support of the fundamental cause theory, two other theories were used. First, the diffusion of innovations theory was first proposed by Rogers (1962) but has subsequently been updated with the most recent version of the theory published by Rogers in 2010. The theory presents factors that influence whether or not an individual or population will adopt a new innovation (Rogers, 2010). Second, the theory of health disparities and medical technology was developed by Goldman and Lakdawalla in 2005. The theory surrounds the assumption that richer patients disproportionately use newer therapies.

Theoretical Framework Supports the Study Approach

In line with the fundamental cause theory, individuals of low SES, minority races, or individuals living in rural areas could have systematically less of a given resource due to their social connections. Therefore, it is possible that disparities in monoclonal antibody therapy exist due, in part, to disparities in resource acquisition. Additionally,

according to the diffusion of innovation theory (Rogers, 2010), some populations (high SES, urban neighborhoods) might be more likely to obtain newer treatments (such as monoclonal antibody therapy). Finally, the theory of health disparities and medical technology supports hypothesizes around higher levels of monoclonal antibody use by high SES or nonminority populations.

In summary, these three theories work in concert to provide support for the research questions. The fundamental cause theory attempts to explain disparities based on resource acquisition, whereas the diffusion of innovation theory takes a temporal approach to explain uptake of new technologies or treatments and the theory of health disparities and medical technology proposes that the complexity and quantity of treatment influences who will receive treatment. All three of these theories provide rationale for research into disparities in monoclonal antibody treatment of metastatic colorectal cancer patients, as it is likely that demographics and social status will contribute to disparities in the use of newer, more complex treatments.

Nature of the Study

The specific aim of this quantitative study was to determine if there are racial, SES, or neighborhood characteristic (degree of urbanization) disparities in first-line adjunct monoclonal antibody treatment of elderly metastatic colorectal cancer patients. In this section, I briefly describe the design selected to address this question including the key study variables and an overview of the methodology.

Design Rationale

Quantitative research was the most appropriate research design to answer the research questions. The most complete dataset available to answer these questions was archival data from the NCI's SEER-Medicare database. Therefore, since archival data were used, this quantitative study is strictly observational and no interventions were be performed.

Data for all the variables were obtained from five data sources within the SEER-Medicare database: the PEDSF file (Patient Entitlement and Diagnosis Summary File) and four Medicare claims files, the DME file (Durable Medical Equipment File), the Medicare Part D Event (PDE) file, the Medicare Outpatient Claims file, and the Medicare Carrier Claims file. The PEDSF dataset was available through 2011, while all the claims datasets were available through 2012 with the first full year of PDE data in 2007. Therefore, this project was limited to individuals diagnosed with metastatic colorectal cancer from January 2007 to December 2011. This allowed for assessment of chemotherapy and monoclonal antibody treatment into the 2012 claims datasets for individuals diagnosed in the second half of 2011.

Methodology

Population. The specific population used for this study was elderly (65+) Medicare enrolled individuals diagnosed with metastatic colorectal cancer between January 2007 and December 2011 and treated with first-line chemotherapy within 6 months of diagnosis as in Meyerhardt, Sanoff, Carpenter, and Schrag (2012). This

population was carefully selected for appropriateness in regards to the research question and to control for confounding.

Sampling. This population was sampled using the NCI's SEER-Medicare database. The SEER-Medicare database is a unique research-oriented database resulting from the linkage of the SEER cancer registries database and the Medicare enrollment and claims data files (NCI, 2013a; Warren, Klabunde, Schrag, Bach, & Riley, 2002).

Patient data for the study variables (independent, dependent, covariate/control, and selection variables) from 2007 to 2012 were compiled for all individuals meeting the inclusion and exclusion criteria using MySQL, Python and SPSS.

Analysis. Logistic regression in SPSS was used to model the dichotomous dependent variable based on the independent variables as in Burns and Burns (2008). Specifically, the output of logistic regression predicted which dependent variable group (monoclonal antibody, "yes" or "no") a sample should reside in based on the independent variables (race, SES, neighborhood characteristics; Burns & Burns, 2008). Additional models were planned to control for potential confounding variables.

Definitions

Study Variables

Concise definitions of the independent variables, dependent variable, and control variables/covariates are provided in the following sections.

Independent variables. There are three research questions and three total independent variables included in this research project. Data for all of the independent

variables were available from the PEDSF (NCI, 2015c). The three independent variables included in this research project are defined below:

Neighborhood characteristics (degree of urbanization). Neighborhood was defined using the 2003 Rural/Urban Continuum Codes from the ERS, which categories counties on an urban/rural scale based on population size, degree of urbanization, and adjacency to a metro or nonmetro area (NCI, 2015c). This variable categorizes neighborhoods as Big Metro, Metro, Urban, Less Urban, Rural, or Unknown (NCI, 2015c).

Race. Race used the SEER recode of the patient's self-reported race to allow for analysis of Latino individuals. The SEER recode reports race as one of the 11 selections: Caucasian, Black, American Indian/Alaska Native, Chinese, Japanese, Filipino, Hawaiian, Other Asian or Pacific Islander, Unknown, Caucasian, Spanish origin or surname, Other unspecified (NCI, 2015c). Given the low representation of some of the races in the cohort, American Indian/Alaska Native persons were excluded from the race analysis and Chinese, Japanese, Filipino, Hawaiian and Other Asian or Pacific Islander groups were pooled into a single group entitled Asian or Pacific Islander.

Socioeconomic status (SES). As in an article by Schlichting, Soliman, Schairer, Schottenfeld and Merajver (2012), the Census Tract Poverty Indicator variable was used as a surrogate measure of SES. This variable uses information from the American Community Survey that measures and reports census tract poverty levels (Kentucky Cancer Registry, n.d.).

Dependent variable. The dependent variable for all three research questions was receipt of first-line adjunct (in combination with chemotherapy) monoclonal antibody therapy. As in Meyerhardt et al. (2012), first-line adjunct monoclonal antibody treatment was defined as at least one claim for a monoclonal antibody (bevacizumab, cetuximab, or panitumumab) within 1 month of chemotherapy (which must occur within 6 months of diagnosis). For the purpose of this research project, the dependent variable (first-line monoclonal antibody treatment) was recorded dichotomized as "yes" or "no."

Covariates. Four covariates were added to the logistic regression analysis to control for confounding. These covariates are defined below.

Age at diagnosis. As documented by Medicare.

Gender. Gender was self-reported and defined as male or female.

Reason for original Medicare entitlement. Reason for original Medicare entitlement; age, disability, End Stage Renal Disease (ESRD), or disability and ESRD. Age and disability were the only options for this covariate as individuals with ESRD were excluded from the study.

SEER Registry that reported the data. Which geographically based SEER-Registry reported the data.

Study Terms

Operational definitions for the independent, dependent, and control/covariate variables are documented in the study variable section above. In this section, I provide two additional definitions required for sample selection and study execution.

First-line chemotherapy. Given chemotherapy within 6 months of cancer diagnosis as in Meyerhardt et al. (2012).

Socioeconomic status (SES). SES is a representation of the social and economic state of an individual or population.

Assumptions

There is one large assumption necessary to justify this research project: specifically that monoclonal antibodies improve survival of elderly metastatic colorectal cancer patients. Clinical studies have shown a survival benefit in elderly metastatic colorectal cancer patients treated with adjunct monoclonal antibody therapy (Cunningham et al., 2013). However, real world evidence of improved survival of elderly metastatic colorectal cancer patients associated with monoclonal antibody therapy has not been shown using the SEER-Medicare population.

Scope and Delimitations

Scope

The focus of this study is on identifying disparities in first-line adjunct monoclonal antibody treatment of elderly metastatic colorectal cancer patients. The natural next question is whether or not any observed monoclonal antibody treatment disparities correlate with disparities in survival. A robust analysis of survival would require assessment and controlling for multiple comorbidities both pre- and post-diagnosis. This type of analysis is large, complex, and out of scope for this study.

However, if disparities in adjunct monoclonal antibody therapy are found, a future study could address the impact of these findings on colorectal cancer survival.

Delimitations

This study was limited to U.S. elderly (65+ years) Medicare Part B and Part D enrolled (but with no HMO coverage), SEER registry colorectal cancer patients whose cancer had metastasized at diagnosis and who received first-line chemotherapy. This very specific population was selected to reduce validity threats and bias in the study. However, as a result, it is unclear how these results will generalize to other populations (e.g., elderly population with different health insurance, individuals in a different region of the United States, younger individuals, and individuals who are diagnosed with an earlier stage cancer, and progress to metastatic cancer).

Limitations and Methods to Address Limitations

Design/Methodology Weaknesses

Study design weaknesses. In this study, I employed a retrospective cohort observational study design. Given the sample selection criteria, loss to follow-up (the main concern in cohort studies) should be small. Specifically, assessment of monoclonal antibody therapy (the dependent variable) occurs within 1 month of chemotherapy. Given that receipt of chemotherapy is a selection criterion, the loss to follow-up time was only 1 month long, limiting the impact of this concern.

Data source weaknesses. I used archival data from the SEER-Medicare database to address the research questions. Therefore, the study quality is limited by the validity,

reliability and completeness of the SEER-Medicare database. Given that hospitals, clinicians, and pathologists are responsible for accurate reporting and coding of SEER data, there is the potential for missing or inaccurate data. However, there are published studies that have documented good reliability, validity, and completeness of different subsets of the SEER-Medicare data (Du et al., 2008; Mahnken et al., 2008).

As a result of the database used for this study, concerns regarding extrapolation of the study results do exist. This SEER database includes data from approximately 28% of the U.S. cancer population. However, this sample comes from specific SEER funded cancer registries. These registries are geographically dispersed, located in 13 different states. However, it is unclear whether the data within the SEER database is a true representation of the greater U.S. cancer population. This threat cannot be avoided and will be noted in the limitations section of Chapter 5.

Threats to validity. Study validity could be threatened by confounding variables. Therefore, the logistic regression analysis plans included covariates to test for independent associations between the independent and dependent variables. However, the possibility still exists that other variables, not cited in the literature as associating treatment and thus not included in the models, could confound the analysis.

To increase statistical conclusion validity, great effort was employed to ensure the study was powered appropriately and that the statistical tests did not violate any assumptions. In addition, the statistical analysis plan was explicitly laid out and further,

undocumented analyses were not performed. This effort attempts to limit the type I error rate.

Biases

Given that this is a cohort study, there is the potential for bias in the outcome assessment. For example, it is possible that hospitals that treat primarily black patients are less likely to record treatments compared to hospitals that treat primarily white patients. However, Mahnken et al. (2008) found that the completeness of the SEER-Medicare data did vary nonsignificantly by race and ethnicity. Therefore, if the dataset is relatively complete (all treatments reported regardless of race, SES, or neighborhood characteristics), bias in the outcome assessment should be reduced.

Other potential biases include selection bias. In the case of this study, all SEER-Medicare colorectal cancer patients were included in the cohort if they met the sample selection criteria (there was no random selection from the SEER dataset). This reduced selection bias. Additionally, the SEER sample is large (represents 28% of U.S. cancer cases), also reducing issues associated with selection.

Significance

In this section, I describe the potential significance of the study results. Specifically, how the results could advance knowledge of the discipline, impact practice/policy, and promote positive social change.

Advance Knowledge of the Discipline

Given the survival benefit associated with adjunct monoclonal antibody therapy (Tol & Punt, 2010), it is possible that disparities in adjunct monoclonal antibody therapy are contributing to disparities in metastatic colorectal cancer survival. Increased understanding of colorectal cancer treatment disparities could advance the knowledge of the discipline and provide a new hypothesis to explain why colorectal cancer survival disparities are persisting. If racial, SES, or neighborhood (degree of urbanization) disparities are found, future studies could determine if these disparities contribute to disparities in metastatic colorectal cancer survival.

Additionally, given that the elderly have the highest colorectal cancer burden (risk and survival) of any age group, focusing on this group could potentially increase the impact of the study results. Additionally, limiting this study to elderly colorectal cancer patients will also allow this research to build on treatment disparity research already generated using this population (Le et al., 2008; White et al., 2010).

Potential Practice/Policy Contributions

If racial, SES, or neighborhood (degree of urbanization) disparities are found, research into why these disparities exist could commence. Understanding of the reasoning behind treatment disparities could have a direct influence on practice or policy. For example, it is possible that newer treatment options are less well known to practitioners at rural hospitals compared to urban hospitals, making them less likely to

prescribe adjunct monoclonal antibody therapy. If this was determined, policies could be put in place to educate rural doctors regarding new efficacious treatments.

Potential for Positive Social Change

Determining if racial, SES, or neighborhood disparities in monoclonal antibody treatment of elderly metastatic colorectal cancer patients exist is a critical first step to eliminating these social inequalities. This study may provide the basis for positive social change in one of two ways. First, if treatment disparities are found, this would illuminate the need for policies that improve access to monoclonal antibodies, thereby helping reduce social inequalities in colorectal cancer survival. Alternatively, if treatment disparities are not found, an opportunity exists to understand the root cause of the reduced treatment disparities, and this knowledge could potentially be used to reduce treatment disparities in other cancer populations.

Summary

Colorectal cancer survival disparities based on race, SES, and neighborhood characteristics have been observed in many different studies. Specifically, Black Americans have reduced colorectal cancer survival rates compared to White Americans (Simpson et al., 2013; Sineshaw et al., 2014; Wallace et al., 2013; Wassira et al., 2013), low SES populations have reduced colorectal cancer survival rates compared to higher SES populations (Hines et al., 2014; Oliphant et al., 2013b; Wassira et al., 2013), and rural populations have reduced colorectal cancer survival rates compared to urban populations (Henry et al., 2009; Lian et al., 2011).

In order to address these survival disparities, it is important to understand the potential underlying causes of these disparities. Multiple articles have highlighted disparities in both type and aggressiveness of colorectal cancer therapies (surgery, radiation, and chemotherapy; Bakogeorgos et al., 2013; Obeidat et al., 2010; Serra-Rexach et al., 2012). However, this project is unique as it is unknown if these historical colorectal cancer treatment disparities have extended to the newer monoclonal antibody therapies and individuals with metastatic colorectal cancer. Therefore, the purpose of this study was to determine if there are racial, SES, or neighborhood (degree of urbanization) disparities in first-line adjunct monoclonal antibody treatment of elderly metastatic colorectal cancer patients.

I used archived colorectal cancer patient data available from the NCI's (NCI) SEER-Medicare database. Specifically, a retrospective cohort study was performed, using the SEER-Medicare database, on patients diagnosed with colorectal cancer from January 2007 to December 2011. The data were analyzed using logistic regression with defined covariates.

Determining if racial, SES, or neighborhood disparities in monoclonal antibody treatment of elderly metastatic colorectal cancer patients exist is a critical first step to eliminating these social inequalities. This study may provide the basis for positive social change either through illuminating the need for policies that improve access to monoclonal antibodies (if disparities are found) or providing rationale to further

understand the root cause of the reduced treatment disparities (if disparities are not found).

In this chapter, I provided an introduction to the research project including summarization of items that will be described in detail in Chapters 2 and 3. In the following chapter, I describe, in detail, the literature review performed as a basis for this research study. Within the literature review, the literature gap described briefly in Chapter 1 will be clearly illuminated. Additionally, the theoretical foundation for the research and a rationale for the chosen variables are presented.

Chapter 2: Literature Review

Introduction

In the United States, colorectal cancer accounts for 8.6% of all new cancer cases and 8.8% of all cancer deaths (NCI, n.d.a). Colorectal cancer risk and mortality rate increases with age, with the elderly being most affected by the disease (CDC, 2013). In the United States, elderly individuals (65 years+) account for 60% of all new colorectal cancer cases and 70.9% of all colorectal cancer deaths (NCI, n.d.a). Colorectal cancer survival rates in the United States have been shown to be associated with race, SES, and neighborhood characteristics (degree of urbanization; CDC, 2012; Hao et al., 2011; Naishadham, Lansdorp-Vogelaar, Siegel, Cokkinides, & Jemal, 2011; White, Vernon, Franzini, & Du, 2010). Specifically, Black Americans, individuals of low SES, and individuals residing in rural areas have been shown to have increased colorectal cancer mortality (CDC, 2012; Naishadham et al., 2011; White et al., 2010). White et al. (2010) also showed that racial disparities in colorectal cancer survival persisted when the population was limited to the elderly (65 years+).

Colorectal cancer survival has been shown to be influenced by treatment received. For example, using a population of elderly colorectal cancer patients, Le et al. (2008) found significant differences in first-line treatment (surgery, radiation, and chemotherapy) comparing Black Americans to White Americans and higher SES to lower SES. Hao et al. (2011) also described disparities in first-line chemotherapy treatment in colorectal cancer. Specifically, populations with urban or suburban zip

codes were, respectively, 38% and 52% more likely to receive chemotherapy compared to populations with rural zip codes (Hao et al., 2011). Additionally, urban Black American colorectal cancer patients had 24% reduced rates of chemotherapy compared urban White American colorectal cancer patients (Hao et al., 2011). In a meta-analysis of studies published from 1995 to 2009, Aarts et al. (2010) found that colorectal cancer patients of low SES received less aggressive therapies and less adjunct therapies.

Although these studies have provided valuable information on colorectal cancer treatment disparities, including disparities in elderly populations, they were conducted using patient data from the 1990s and early 2000. The colorectal therapies employed during this period included surgery, radiation and chemotherapy. Beginning in 2004, monoclonal antibody therapies specifically for metastatic colorectal cancer came onto the market (Scott et al., 2012). Currently, there are three monoclonal antibodies (bevacizumab, cetuximab, and panitumumab) approved for the treatment of metastatic colorectal cancer (Scott et al., 2012). A review of clinical trials and observational studies, included in this chapter, indicated that monoclonal antibodies increased progression-free survival and overall survival in elderly metastatic colorectal cancer patients when added to chemotherapy. Therefore, as with chemotherapy, it is possible that disparities in monoclonal antibody treatment could contribute to disparities in colorectal cancer survival.

However, due to the lack of research, it is unknown if there are racial, SES, or neighborhood disparities in monoclonal antibody treatment of elderly metastatic

colorectal cancer patients. Therefore, the purpose of this study was to determine if there are racial, SES, or neighborhood disparities in first-line adjunct monoclonal antibody treatment of U.S. elderly colorectal cancer patients.

This chapter includes three major sections. First, the literature search strategy will be reviewed in detail, including the rationale behind each combination of search terms. This section also includes tables highlighting the relevant articles identified by each combination of search terms. Second, I describe the theoretical foundation supporting the research question will be outlined. Specifically, one overarching theory and two supporting theories provide rationale for potential disparities in colorectal cancer treatment. Third, an in depth review of the literature identified by the literature search will be provided. This section includes multiple subsections reviewing disparities in colorectal cancer survival, underlying causes of the disparities, and the utility of monoclonal antibodies in elderly colorectal cancer patients. Each subsection also contains rationale for inclusion in this document as well as the relationship to the research questions. Additionally, in the literature review section, key variables are outlined and justified. I conclude the chapter with an evaluation of the literature review and a description of how this study will address the gap and extend knowledge of colorectal cancer treatment disparities research.

Literature Search Strategy

In this section of the literature review, I discuss databases and search terms used. Literature searches were conducted to locate articles surrounding colorectal cancer, the

elderly, drug treatments, and disparities in treatments or care. Search terms and articles located are provided in table form in this section, and key articles will be reviewed in Literature Review. As will be described, the reviewed literature informed the design of the study and methodology discussed in Chapter 3.

Databases and Search Terms

Two databases were searched for literature related to colorectal cancer, the elderly, drug treatment, and disparities: PubMed and Science Direct. PubMed is a biomedical literature database developed and maintained by the National Center for Biotechnology Information. Science Direct is a database with content primarily focused on medicine, nursing and allied health. These two databases were selected for their relevance to the research topic. Specifically, they were chosen for their utility in locating journal articles surrounding colorectal cancer epidemiology, drug treatments, and disparities in cancer care. Monoclonal antibody therapies for colorectal cancer, the focal point of the research question, came onto the market in 2004 (Scott et al., 2012). Therefore, the literature searches were restricted from 2004 to 2015. Only peer-reviewed primary literature articles were included in the tables below. However, review articles were read and bibliographies of these sources were also reviewed to ensure the database searches did not miss any important articles.

Colorectal cancer epidemiology searches. PubMed and Science Direct were queried using a combination of the following terms to locate articles regarding colorectal cancer survival disparities in the elderly: *colorectal cancer, survival, epidemiology,*

elderly, and *disparities*. The specific searches are outlined in Table 1. These searches were restricted from 2004 to 2015 and only peer-reviewed primary journal articles were included. However, bibliographies of secondary source articles were reviewed to ensure all important primary articles were captured. Many of these searches produced a significant number of articles. All articles produced in the queries were reviewed; however, only potentially relevant articles are listed in Table 1. Potentially relevant articles met at least one of the following four criteria:

- Provided U.S. colorectal cancer survival data from after 2004.
- Discussed treatment or care of elderly colorectal cancer patients.
- Discussed disparities in colorectal cancer survival, treatment, or care based on race, socioeconomic status, or geography.
- Discussed any disparity in colorectal cancer within the population of elderly U.S. colorectal cancer patients.

Table 1

Colorectal Cancer Epidemiology Primary Literature From PubMed and Science Direct

Search terms	Primary research articles
"Colorectal Cancer"[b], Survival[b], and Epidemiology[a]	Ahmed, Howel & Debrah (2014); Tong et al. (2014); Sineshaw, Robbins & Jemal (2014); Simpson et al. (2013); Allemani et al. (2013); Oliphant et al. (2013a); Mitry et al. (2013); Oliphant et al. (2013b); Wassira, Pinheiro, Symanowski & Hansen (2013); Chien, Schootman & Pruitt (2013); Wallace et al. (2013); Lam, Lu, Kouzminova & Lin (2013); Jafri, Gould, El-Serag, Duan & Davila (2013); Wan, Zhan, Lu & Tiefenbacher (2012); Lansdorp et al. (2012); Renouf et al. (2011); Lian et al. (2011); Cueto, Szeja, Wertheim, Ong & Tsikitis (2011); Tsai et al. (2011); Osterlund et al. (2011); White, Vernon, Franzini & Du (2010); Lejeune et al. (2010); Koroukian et al. (2010); Henry, Niu & Boscoe (2009); Ran et al. (2009); Lang et al. (2009); Kelsall et al. (2009); Le, Ziogas, Lipkin & Zell (2008)
"Colorectal Cancer"[b] and Elderly[b]	Naeim et al. (2013); Bruera et al. (2013); Fu, Tsai, Marshall & Potosky (2013); Bakogeorgos et al. (2013); Cunningham et al. (2013); Serra-Rexach et al. (2012); Singal, Lin, Kuo, Riall & Goodwin (2013); Abdelwahab, Azmy, Abdel-Aziz, Salim & Mahmoud (2012); Jehn, Boning, Kroning, Possinger & Luftner (2012); Price et al. (2012); Wildes et al. (2010); Vrdolijak, Omrcen, Boban & Hrabar (2011); Kozloff et al. (2010); White et al. (2008); Wright, Barlow, Green, Baldwin & Taplin (2007)
"Colorectal Cancer"[b], Survival[b] and Disparities[b]	Sineshaw et al. (2014); Wassira et al. (2013); Wallace et al. (2013); Lansdorp et al. (2012); Wan, Zhan, Lu & Tiefenbacher (2012); Cueto et al. (2011); White et al. (2010); Lejeune et al. (2010); Henry et al. (2009); Le et al. (2008)

Note. [a] Text Word Search; [b] Title Search

Colorectal cancer disparities search. Additional database queries were performed to specifically locate any additional articles regarding colorectal cancer disparities and to exclude any bias searches that could have resulted from including either "survival" or "elderly" in the previous search. For this purpose, PubMed and Science Direct were queried using a combination of the following terms: *colorectal cancer, disparities and one of the following: treatment, race or racial, state, rural, geographical*

or geography, or socioeconomic status. Given the importance of treatment disparities to the research question, the search query that included the terms *colorectal cancer, treatment, and disparities* was queried three times, each time moving one of the three search terms to a text word query instead of a title query. Additionally, given that only one article was identified with the search *colorectal cancer, socioeconomic status, and disparities*, an additional search was performed after eliminating the search term *disparities.* The exact search terms are detailed in Table 2.

Table 2

Colorectal Cancer Disparities Primary Literature From PubMed and Science Direct

Search terms	Primary research articles
"Colorectal Cancer"[b], Treatment[b] and Disparities[b]	Simpson et al. (2013); Wan et al. (2012); Cueto et al. (2011); Haas et al. (2011); White et al. (2010); Lejeune et al. (2010); White et al. (2008); Le et al. (2008)
"Colorectal Cancer"[b], Treatment[a] and Disparities[b]	Zullig et al. (2013); Wassira et al. (2013); Wallace et al. (2013); Cueto et al. (2011); Haas et al. (2011); Crawford, Jones & Richardson (2010); Obeidat et al. (2010); White et al. (2010); Lejeune et al. (2010); Hao et al. (2011); White et al. (2008); Le et al. (2008); McKibbin et al. (2008); Demissie et al. (2004)
"Colorectal Cancer"[b], Treatment[b] and Disparities[a]	Hines & Markossian (2012); Cueto et al. (2011); Haas et al. (2011); White et al. (2010); Lejeune et al. (2010); White et al. (2008); Le et al. (2008); Hassan, Arthurs, Sohn & Steele (2009)
"Colorectal Cancer"[b], "Race or Racial"[b] and Disparities[b]	Sineshaw et al. (2014); Wassira et al. (2013); Wallace et al. (2013); Lansdorp-Vogelaar (2012); Stimpson, Pagan & Chen (2012); Wilkins et al. (2012); Murphy et al. (2011); Laiyemo et al. (2010); Hao et al. (2011); Robbins, Siegel & Jemal (2012); White et al. (2010); White et al. (2008)
"Colorectal Cancer"[b], State[b] and Disparities[b]	Rane et al. (2014); Naishadham, Lansdorp-Vogelaar, Siegel, Cokkinides & Jemal (2011)
"Colorectal Cancer"[b], Rural[b] and Disparities[b]	Rane et al. (2014); Wilkins et al. (2012); Beyer, Comstock, Seagren & Rushton (2011); Cole, Jackson & Doescher (2012)
"Colorectal Cancer"[b],"Geographical or Geography"[b] and Disparities[b]	None
"Colorectal Cancer"[b], "Socioeconomic status"[b] and Disparities[b]	Le et al. (2008)
"Colorectal Cancer"[b] and "Socioeconomic status"[b]	Hines, Markossian, Johnson, Dong & Bayakly (2014); Saldana-Ruiz, Rubin, Colen & Link (2013); Steinbrecher et al. (2012); Kelsall et al. (2009); Le et al. (2008); Gomez, O'Malley, Stroup, Shema & Satariano (2007)

Note. [a] Text Word Search; [b] Title Search

Colorectal cancer monoclonal antibody treatment disparities search. A final database query was performed to specifically locate any articles regarding colorectal cancer disparities and monoclonal antibody therapies. For this purpose, PubMed and Science Direct were queried using a combination of the following terms: *colorectal cancer, disparities and one of the following: antibody, bevacizumab, cetuximab or panitumumab.* Given that very little research has been performed surrounding disparities in monoclonal antibody treatment of colorectal cancer patients, the search terms were relaxed. The term *colorectal cancer* was queried as a title search term; however, the other two search terms in each case were only listed as text search terms. The exact search terms are described in detail in Table 3.

Table 3

Colorectal Cancer Monoclonal Antibody Treatment Disparities Primary Literature From PubMed and Science Direct

Search terms	Primary research articles
"Colorectal Cancer"[b], Antibody[a] and Disparities[a]	McKibbin et al. (2008); Wallace et al. (2013)
"Colorectal Cancer"[b], Bevacizumab[a] and Disparities[a]	McKibbin et al. (2008); Wallace et al. (2013)
"Colorectal Cancer"[b], Cetuximab[a] and Disparities[a]	Wallace et al. (2013)
"Colorectal Cancer"[b], Panitumumab[a] and Disparities[a]	None

Note. [a] Text Word Search; [b] Title Search

It is apparent from this search that very little research has been done surrounding disparities in monoclonal antibody treatment of colorectal cancer patients. Of the two

relevant articles located with this search, neither addressed racial, socioeconomic, or geographical disparities in monoclonal antibody treatment of colorectal cancer patients.

Literature Search Strategy: Summary

The literature search revealed a significant amount of literature surrounding colorectal cancer survival disparities. The search uncovered disparities in survival in many populations including the elderly, blacks, Hispanics, Native Americans, populations with low SES, populations living in rural counties as well as state-to-state geographical disparities. Additionally, the literature search addressed potential underlying causes of these disparities including differences in screening, access to care, and treatment. Of particular relevance to the research topic, the literature search uncovered documented disparities in colorectal cancer treatment. However, the treatment disparities explored in these articles are differences in older therapies (chemotherapy and surgery). The searches did not uncover any articles addressing racial, socioeconomic, or neighborhood disparities in monoclonal antibody treatment of colorectal cancer patients. Given that monoclonal antibody treatment has been shown to improve survival in the elderly, understanding disparities in monoclonal antibody treatment could help explain the persistence of colorectal cancer survival disparities.

Theoretical Foundation

The theoretical framework for this project is a combination of theories similar to the theoretical framework proposed by Chang and Lauderdale (2009). The overarching theory is the fundamental cause theory. However, two additional theories (the diffusion

of innovations theory and the theory of health disparities and medical technology) are used to extend the fundamental cause theory for the purpose of this project. I describe these theories and their relevance to the research topic in detail in the sections below.

Fundamental Cause Theory

The overarching theory for the research questions is the fundamental cause theory. This theory was originally developed to explain why the association between SES and mortality persists across multiple diseases and risk factors (Link & Phelan, 1995). This theory proposes that beneficial social connections protect health and reduce mortality (Link & Phelan, 1995).

Origin and definition of the theory. Link and Phelan (1995) first proposed the fundamental cause theory in 1995. Link and Phelan (1995) argue that social states, such as race and SES, contribute to disease rates for two major reasons. First, social states impact resources available to combat diseases (Link & Phelan, 1995). Second, social state is tied to multiple other disease risk factors (Link & Phelan). An important assumption of this theory is that the relationship between social state and disease persists over time (Phelan et al., 2010). Since resources are constantly changing, it is the beneficial social connections that serve to protect health regardless of the resource mechanism (Phelan et al.).

Utilization of the theory. The fundamental cause theory has been used primarily to address differences in disease and disease mortality rates based on SES. The theory is

especially useful when attempting to explain SES disparities that persist over time and across varying resources.

Of particular importance to the research questions, in 2005, Phelan and Link expanded the theory to help explain why breakthroughs in disease detection or treatment do not often reduce social disparities. The authors argue that social and economic inequalities contribute to resource acquisition (Phelan & Link, 2005). Therefore, as new methods of treating or preventing disease become available, these resources are not distributed equally (Phelan & Link). Instead they are distributed based on "knowledge, money, power, prestige, and beneficial social connections" (Phelan & Link, 2005, p. 227). Therefore, even as the ability to combat disease incidence or mortality improves (as with the introduction of monoclonal antibodies for colorectal cancer) the acquisition of these new resources is not equal and therefore disparities persist.

These arguments were further supported in an article by Chang and Lauderdale (2009). In this work the authors address disparities in cholesterol levels in the era of nongeneric statins (Chang & Lauderdale, 2009). The authors found that in the years before statin therapy, high SES populations had cholesterol levels higher than low SES populations (Chang & Lauderdale). However, looking at cholesterol levels through the statin era, this gradually reversed with the high SES population having lower cholesterol levels compared to the low SES population (Chang & Lauderdale). The authors propose that high SES populations had greater access to statins compared to the lower SES

populations (Chang & Lauderdale). This distribution of resources disparity contributed to the reversal of cholesterol situation comparing the SES groups.

Relationship to the research questions. Although in recent years (2004-2009) new colorectal cancer treatments and technologies appear to be contributing to increasing survival rates in all races, racial disparities in colorectal cancer survival, specifically comparing Black Americans and White Americans, are increasing (Sineshaw et al., 2014). Sineshaw et al. (2014) propose that minority populations have not benefited as greatly from newer treatments or these treatments have not be disseminated into these populations as greatly (Sineshaw et al.). This hypothesis is supported by Simpson et al. (2013). In the Simpson et al. (2013) study, black patients with metastatic colorectal cancer were less likely to receive multimodality therapy compared to White Americans.

In this research project I address disparities in monoclonal antibody treatment of elderly metastatic colorectal cancer patients. Adjunct monoclonal antibody therapy has been shown to be highly efficacious in increasing colorectal cancer survival (Bruera et al., 2013, Cunningham et al., 2013; Naeim et al., 2013). However, monoclonal antibody therapy is new and expensive. Thus, in accordance with the fundamental cause theory, it is possible that minority populations or populations of low SES will be less likely to receive this new resource. These resource acquisition differences could contribute to the persisting disparities in colorectal cancer survival.

Diffusion of Innovations Theory

As an extension of the fundamental cause theory, the diffusion of innovation theory (Rogers, 1962) will be used to help explain why some populations (high SES, certain geographies) might be more likely to obtain newer treatments. This model is much older and more widely used compared to the fundamental cause theory. The model states that widening disparities are influenced by the nature of the new technology and the uptake/diffusion of the technology (Rogers, 1962).

Origin and definition of the theory. The diffusion of innovations theory was first proposed by Rogers in1962. The theory has subsequently been updated with the most recent version of the theory published by Rogers in 2010. The theory proposes five innovation adopter categories based on how long post availability the innovation is expected be adopted (Rogers, 2010). The categories use a normal distribution with innovators (~2.5% of the population) adopting the innovation quickly following availability, followed by early adopters (13.5%), early majority (34%), late majority (34%), and 16% laggards (Rogers, 2010). Of particular importance to the model and the research questions, there are several factors that influence whether or not an individual or population will adopt a new innovation, and thus which adopter category that individual or population will reside within (Rogers, 2010). These factors include relative advantage, compatibility, complexity, triability (ability to "try it out" before committing to adoption) and observability (observed results) (Rogers, 2010). The requirements needed to meet each of these factors can differ from one population to another. Therefore, Rogers (2010)

notes that it is critical to understand the target population and specific social factors that can contribute to innovation adoption (Rogers, 2010).

Utilization of the theory. Relevant to the research questions, the diffusion of innovations theory has been used as a model to explain multiple findings including drug prescribing differences and the uptake of new diagnostic tests. For example, Armstrong, Weiner, Weber, and Asch (2003) used the diffusion of innovations theory to help understand early uptake the BRCA1/2 test for breast cancer susceptibility. The authors found that early adopters of the test were proactive in inquiring about the test and had heard about the test from sources in addition to their doctor or genetic counselor (Armstrong et al., 2003). These adoption characteristics, including access to outside knowledge of the test, could differ based on the social characteristics or demographics of a population.

Additionally, Makowsky, Guirguis, Hughes, Sadowski, and Yuksel (2013) used the diffusion of innovations theory to understand the adoption of pharmacist prescribing behaviors in Alberta, Canada. The authors found that prescribing behaviors varied greatly and was dependent on the innovation, the adopter, readiness of the system, and communication/influence (Makowsky et al., 2013). The authors found that patient focused pharmacists were more likely to prescribe medications compared to disease focused pharmacists (Makowsky et al.). Additionally, pharmacists commented that physician relationships influenced their decision to prescribe (Makowsky et al.). Of importance to this research study, access to patient-centered health care and a

collaborative healthcare system could influence drug prescribing decisions and uptake of new innovations (such as monoclonal antibodies).

Relationship to the research questions. In concert with the fundamental cause theory, the diffusion of innovations theory helps provide support for the research questions. As noted in the section above, racial disparities in colorectal cancer survival, specifically comparing Black Americans and White Americans, are increasing (Sineshaw et al., 2014). Many studies have documented the efficacy of bevacizumab (monoclonal antibody) in metastatic colorectal cancer. However, the uptake of monoclonal antibody therapy has not been immediate or complete. According to Renouf et al. (2011), 5.9% of patients diagnosed with colorectal cancer in 2004 were given bevacizumab compared to 30.6% of colorectal cancer patients diagnosed in 2006. In concert with the diffusion of innovations theory, it is conceivable that different populations could reside in different adopter categories based on social factors such as race and SES. Thus, it is possible that a greater percentage of high SES, nonminority populations fall into innovator or early adopter categories. This could contribute to the widening colorectal cancer survival disparities and supports the research questions aimed at addressing disparities in monoclonal antibody treatment.

Theory of Health Disparities and Medical Technology

As additional support for the fundamental cause theory and the diffusion of innovations theory, the theory of health disparities and medical technology is used (Goldman & Lakdawalla, 2005). This theory states that medical advances are linked to

widening disparities due to disparities in health care utilization (Goldman & Lakdawalla, 2005).

Origin and definition of the theory. The theory of health disparities and medical technology is a relatively new theory developed by Goldman and Lakdawalla in 2005. The theory itself is very similar to the fundamental cause theory, but focuses on the quantity and complexity of newly introduced therapies (Goldman & Lakdawalla, 2005). The theory was originally developed to help explain why better-educated people are healthier (Goldman & Lakdawalla). The authors rooted their theory in basic consumer theory and argue that new medical technologies disproportionately benefit the heaviest health care users; richer patients (Goldman & Lakdawalla). They claim that richer more well-educated patients tend to be the heaviest health care users and use more complex treatment regimens (Goldman & Lakdawalla). Therefore, richer patients disproportionately use newer therapies.

Utilization of the theory. Goldman and Lakdawalla (2005) used the theory of health disparities and medical technology to help explain why the introduction of a complicated anti-retroviral therapy for HIV patients benefited rich well-educated patients disproportionately. The authors found that richer patients were more likely to likely to invest effort into obtaining and adhering to the complex treatment regimen (Goldman & Lakdawalla, 2005).

Additionally, as noted above, the theory relies on the assumption that the heaviest healthcare users benefit most from newer, more complex technologies and treatments

(Goldman & Lakdawalla, 2005). Therefore, Goldman and Lakdawalla (2005) assessed disparities in survival of chronically ill populations and found the largest survival disparity was based on education. Specifically, more educated populations had higher survival rates compared to lower education populations (Goldman & Lakdawalla). This highly educated population, they argue, is the population expected to be the heaviest health care utilizers, supporting their theory.

Relationship to the research questions. Treatment for metastatic colorectal cancer can be multifaceted including surgery (if possible), chemotherapy and monoclonal antibody therapy. Given that the research questions address disparities in adjunct monoclonal antibody therapy, a relatively complex treatment regimen, this theory is appropriate and compliments the other two theories previously presented. Utilizing this theory, it is possible that richer patients use more health care resources compared to poorer individuals and are thus more likely to obtain adjunct monoclonal antibody therapy. This could contribute to disparities in monoclonal antibody treatment based on social status or demographics.

Theoretical Foundation Summary

Utilization of the three theories noted above in drug uptake and disease outcome research was first proposed by Chang and Lauderdale (2009). The overarching theory for this research project is the fundamental cause theory. As commented above, this theory focuses on disparities in resource distribution and can help explain why breakthroughs in disease treatment (such as monoclonal antibody therapy) do not reduce social disparities

in survival. Specifically, as new methods of treating disease become available, these resources are not distributed equally (Phelan & Link, 2005). Instead they are distributed based on "knowledge, money, power, prestige, and beneficial social connections" (Phelan & Link, 2005, p. 227). This theory proposes that beneficial social connections protect health and reduce mortality (Link & Phelan, 1995). In line with the fundamental cause theory, it is possible that disparities in monoclonal antibody therapy exist due to disparities in resource acquisition.

Two additional theories are used to extend the fundamental cause theory for the purpose of this research project. First, the diffusion of innovation theory (Rogers, 2010) is used to help explain why some populations (high SES, certain geographies) might be more likely to obtain newer treatments. This model states that widening disparities are influenced by the nature of the new technology and the uptake/diffusion of the technology (Rogers). For monoclonal antibody therapy, this model is supported by Renouf et al. (2011). The authors found that uptake of monoclonal antibodies for colorectal cancer treatment was not immediate or complete (Renouf et al.). Early adopters of monoclonal antibody therapy could have innovator-like characteristics such as pro-active inquiry into treatment regimens that might contribute to treatment disparities based on social characteristics or demographics. Second, the theory of health disparities and medical technology is used (Goldman & Lakdawalla, 2005). This theory states that medical advances are linked to widening disparities because new medical technologies disproportionately benefit the heaviest health care users; richer patients

(Goldman & Lakdawalla). Given that richer more well-educated patients tend to be the heaviest health care users and use more complex treatment regimens, richer patients disproportionately use newer therapies (Goldman & Lakdawalla). This could contribute to higher monoclonal antibody use by high SES or nonminority populations.

In summary, these three theories work in concert to provide support for the research questions. The fundamental cause theory attempts to explain disparities based on resource acquisition; whereas the diffusion of innovation theory takes a temporal approach to explain uptake of new technologies or treatments; and the theory of health disparities and medical technology proposes that the complexity and quantity of treatment influences who will receive treatment. All three of these theories provide rationale for research into disparities in monoclonal antibody treatment of colorectal cancer patients, as it is likely that demographics and social status will contribute to disparities in utilization of newer, more complex treatments.

Literature Review

In this literature review I have provided an overview of colorectal cancer epidemiology, reviews studies related to disparities in colorectal cancer survival, and examines the potential underlying causes of colorectal cancer survival disparities. I have also addressed studies related to the safety and efficacy of monoclonal antibodies in elderly colorectal cancer patients. In addition, I've described strengths and weaknesses of previous approaches and, on this basis, variables and concepts chosen for the study

were justified. Finally, I provided a review and synthesis of the studies related directly to the research questions.

Colorectal Cancer Epidemiology

In the United States, colorectal cancer accounts for 8.6% of all new cancer cases and 8.8% of all cancer deaths (NCI, n.d.b). Colorectal cancer risk and mortality rate increases with age, with the elderly being most affected by the disease (CDC, 2013). In the U.S., elderly individuals (65 years+) account for 60% of all new colorectal cancer cases and 70.9% of all colorectal cancer deaths (NCI, n.d.a). As identified in the literature search section above, many studies have uncovered disparities in colorectal cancer survival based on demographics such as race, SES, or neighborhood characteristics (degree if urbanization). However, the underlying causes of these disparities, especially in patients with metastatic colorectal cancer, have been only minimally addressed in the literature.

Review of Studies Related to Constructs and Methods

This research study aims to address disparities in adjunct monoclonal therapy of elderly metastatic colorectal cancer patients. This topic is important due to previously observed survival disparities within this population. Therefore, this section will provide an in-depth review of what is known about survival disparities in colorectal cancer and will provide the rationale for further research into treatment disparities.

Disparities in colorectal cancer survival. The research topic for the study addressed disparities in monoclonal antibody therapy based on several demographic

predictors. This topic is important as it could help explain previously observed and, in some cases, widening disparities in colorectal cancer survival. Many studies have addressed disparities in colorectal cancer survival. These studies have uncovered disparities based on age, race, socioeconomic status, and neighborhood characteristics (degree of urbanization). These studies also provide the rationale for the research topic and are described in detail in the sections below.

Disparities in the elderly. Colorectal cancer survival disparities have been documented in the elderly population. Using data from the SEER-Medicare linked database, White et al. (2010) found significant survival disparities comparing elderly black patients to elderly white patients (adjusted Hazard Ratio: 1.24; 95% CI: 1.14-1.35). Similar results were obtained by Gomez et al. (2007), also using the SEER-Medicare linked database. Contrarily, Wallace et al. (2013) observed only modest survival disparities comparing elderly black to elderly white colorectal cancer patients. However, the population in the Wallace et al. (2013) differed from the other two studies. Namely, as opposed to using national-level data, Wallace et al. (2013) limited the population studied to residents of South Carolina and expanded the definition of elderly to include patients over 50 years of age. The White et al. (2010) and Gomez et al. (2007) studies defined elderly as over 65 years of age. The source population differences could explain the discrepancy in results obtained in these studies.

Additionally, Chien et al. (2013) identified spatial disparities in colorectal cancer survival of elderly patients residing in Atlanta and Detroit. Specifically, their research

identified geographical areas with excessive risk of colorectal cancer death (Chien et al., 2013). However, the authors comment that additional research is needed to understand the moderating pathways driving these disparities (Chien et al., 2013).

Finally, disparities in colorectal cancer mortality based on age were identified in a study by Ahmed et al. (2014). Specifically, elderly patients over 80 years of age had higher cancer-specific mortality rates compared to patients 60-80 years of age (Ahmed et al., 2014). Relevant to the research topic, Ahmed et al. (2014) also identified treatment disparities associated with advancing age that were correlated with mortality rates. Similarly, Serra-Rexach et al. (2012) found that a population of colorectal cancer patients younger than 75 years of age had a better cancer-specific survival rate compared to the 75+ years of age group. These results are documented in detail in Table 4.

The survival disparities observed within the population of elderly colorectal cancer patients provide some of the rationale for using this population in this research study. Additionally, many of the key referenced survival disparities studies used data from the SEER-Medicare database (Chien et al., 2013; Lang et al., 2009; White et al., 2010). This is the data source used for this research project. Therefore, this research study builds directly on previous data and methodologies.

Table 4

Advanced Age-Related Colorectal Cancer Survival Disparities

Source	Data source	Sample	Outcome variable(s) assessed	Results
Chien et al. (2013)	SEER Database	Elderly (66+ years) US colorectal cancer patients in two metropolitan areas (Detroit and Atlanta), 1991-2005	Spatial patterns of survival	Identified geographic areas with excess risk of colorectal cancer death. Spatial patterns varied by cancer type and stage.
Wallace et al. (2013)	South Carolina Central Cancer Registry	Pathologically documented colorectal cancer patients diagnosed from 1996-2006.	Median survival time	Elderly patients (50+ years) showed only modest racial disparities in survival (less than 50 years) (Hazard Ratio: 1.16; 95 % CI 1.01-1.32).
White et al. (2010)	SEER Database	Elderly (66+) colorectal cancer patients diagnosed between 1992-2002.	Kaplan-Meier survival methods	Black patients had worse survival compared to white patients (adjusted Hazard Ratio: 1.24; 95% CI: 1.14-1.35). Asian patients had better survival compared to both black and white patients (adjusted Hazard Ratio: 0.80; 95% CI: 0.70-0.92).
Koroukian et al. (2010)	Ohio Cancer Incidence Surveillance System, Vital Records, Medicare Administrative Data, Home Health Care Outcome and Assessment Information Set (OASIS)	Elderly colorectal cancer patients diagnosed between 1999 and 2001 and admitted to home health care 30 days before or after diagnosis	Comorbidities and survival rates	Two or more geriatric syndromes were associated with increased disease-specific mortality (Hazard Ratio: 2.71; 95% CI: 1.80-4.07).

table continues

Lang et al. (2009)	SEER-Medicare Linked Data	Elderly (66+) colorectal cancer patients diagnosed between 1992-2000, follow-up through 2005.	5-year survival	5-year survival improved temporally, with survival improving from 43% to 46.3% in colon cancer and from 39.4% to 42.2% in rectal cancer from 1992-2000.
Serra-Rexach et al. (2012)	Hospital records (administrative database) from a single institution in Madrid, Spain	Colorectal cancer patients diagnosed or treated at the institution	Cancer-specific survival rate	The younger than 75 years of age group had a better cancer-specific survival rate compared to the 75+ years of age group (36.41 months vs. 26.05 months; hazard ratio: 0.66).
Gomez et al. (2007)	SEER-Medicare Linked Data	Colorectal cancer patients (65+ years) diagnosed between 1992-1996 and followed through 1999.	Mortality rate	Unadjusted colorectal cancer mortality rates were higher in Black Americans and Hispanic compared to White Americans and lower in Japanese compared to White Americans. Adjustments for stage eliminated the difference between White Americans and Hispanics and White Americans and Japanese. However, comparing blacks to White Americans, cancer stage and SES only accounted for half of the observed mortality difference.
Ahmed et al. (2014)	Treatment centers in the former Northern Region of England conducted by the Northern Region Colorectal Cancer Audit Group (NORCCAG)	Elderly colorectal cancer patients (60+ years) that had surgery for their cancer between 1998-2003	30 day and 6-month post-operative mortality	30-day and 6-month mortality was the highest in patients 80+ years. Association of age with mortality at 6-months was the highest in patients receiving curative surgery (OR: 3.8, 95% CI: 2.8-5.2) compared to those receiving palliative surgery (OR: 1.5, 95% CI: 1.1-2.1).

Socioeconomic disparities. Socioeconomic disparities in colorectal cancer survival have been documented in the literature. Specifically, populations with lower SES often have higher colorectal cancer mortality rates. Oliphant et al. (2013b) found that both post-operative death and 5-year survival rates were independently predicted by low SES. Similarly, Lian et al. (2011) and Lejeune et al. (2010) found that lower SES areas had higher colorectal cancer-specific mortality rates. Interestingly, in several studies, survival disparities based on race (Le et al., 2008;Wassira et al., 2013; White et al., 2010; Yan et al., 2009) or urban versus rural geography (Henry et al., 2009; Hines et al., 2014) were somewhat attenuated when the authors controlled for SES. Therefore, SES appears to associate with colorectal cancer survival and might even account for some of the observed racial and geographical disparities in survival. Since this research study attempts to uncover treatment differences that could contribute to observed colorectal cancer survival disparities, it is critically important to include SES as a variable. The results from the referenced papers are documented in detail in Table 5.

Table 5

Socioeconomic Colorectal Cancer Survival Disparities

Source	Data source	Sample	Outcome variable(s) assessed	Results
Oliphant et al. (2013b)	Regional audit database from West of Scotland (16 hospitals)	Patients undergoing surgery for colorectal cancer (2001-2004)	Postoperative mortality (30 days post-surgery) and 5-year relative survival rates	Low SES was independently associated with increased postoperative death (Adjusted Odds Ratio: 2.26; 95% CI: 1.45-3.53) and lower 5-year survival rates (Adjusted Relative Excess Risk: 1.25; 95% CI: 1.03-1.51). However, when postoperative deaths were excluded, the relative survival was similar across SES groups.
Wassira et al. (2013)	Nevada Central Cancer Registry	Patients diagnosed with colorectal cancer between 1995-2007.	Survival rates	Black Americans had a 20.1% higher risk of colorectal cancer death compared to White Americans. This increased risk persisted when adjusted for tumor stage, sex, diagnosis period, tumor sub-location, marital status, and SES.
Lam et al. (2013)	Northern California urban county hospital records and national SEER database	Colorectal cancer patients over 11-year time period	5-year mortality	5-year survival rate in this urban underserved population was worse than the national rate (52.9% vs. 64.3%). Colorectal cancer screening was associated with improved survival. Advanced age and later stage cancer was associated with reduced survival. However, insurance status of the patients was not associated with survival.
Lian et al. (2011)	National Institutes of Health AARP Diet and Health Study	Colorectal cancer patients identified from 1995-2003 and followed until 2006.	Survival (Bayesian multilevel survival models)	Lower SES neighborhoods had higher colorectal cancer-specific mortality (Hazard Ratio: 1.2; 95% CI: 1.1-1.5).

table continues

White et al. (2010)	SEER-Medicare Linked Data	Elderly (66+) colorectal cancer patients diagnosed between 1992-2002.	Kaplan-Meier survival methods	Comorbidities and socioeconomic status were associated with a reduction in the survival difference between both blacks and White Americans and Black Americans and Asians.
Lejeune et al. (2010)	Three UK cancer registries linked to area-level socio-economic information	Colorectal cancer patients diagnosed between 1997-2000.	Access to treatment (measured in time from diagnosis to treatment) and 3-year survival	Patients from lower SES areas had poorer overall survival and were also less likely to receive treatment within 6 months of diagnosis. When limited to patients receiving treatment within 1 month of diagnosis, no survival disparity existed based on area SES.
Henry et al. (2009)	New Jersey Cancer Registry	Patients diagnosed with colorectal cancer between 1996-2003.	Survival rates evaluated using spatial scan statistic	There were geographical areas of the state with both higher and lower survival rates. When race/ethnicity and area SES were controlled, the risk of death was attenuated in several areas. However, in some areas, survival disparities persisted.
Yan et al. (2009)	SEER Program	Patients diagnosed with colorectal cancer between 1988-1992.	Survival rate	Black Americans had significantly higher rates of late stage disease and were more likely to reside in lower SES census tracts. After adjusting for age, marital status, sex, SES group, cancer stage, and treatment, race was no longer a significant predictor of overall survival.
Kelsall et al. (2009)	Australia Melbourne Collaborative Cohort Study	Prospective Cohort Study Participants (recruited between 1990-1994; age 40-69) who were subsequently diagnosed with colorectal cancer	Survival rate	Even with universal health care, area-level SES was associated with reduced survival rates (Hazard Ratio: 0.73; 95% CI: 0.53-1.00) from colorectal cancer (after adjustment for age, sex, tumor stage and treatment).

table continues

Le et al. (2008)	California Cancer Registry	Colorectal cancer patients diagnosed between 1994-2003.	Kaplan-Meier survival methods	After adjustment for age, sex, histology, tumor site, and stage, Black Americans had increased cancer-specific death rates compared to White Americans (Hazard Ratio: 1.19; 95% CI: 1.14-1.25). However, when adjusted for area-level SES and treatment, the disparity was reduced (Hazard Ratio: 1.08; 95% CI: 1.03-1.13).
Hines et al. (2014)	Georgia Comprehensive Cancer Registry, Census-tract level geographic residency and SES.	Patients diagnosed with colorectal cancer in Georgia (2000-2007)	Survival rate	Rural residents had a 14% increased risk of death (Hazard Ratio: 1.14, 95% CI: 1.07-1.22). However, when adjusted for SES level, no differences in survival between rural and urban tracts were noted.
Steinbrecher et al. (2012)	California Cancer Registry, U.S. census data	Colorectal cancer patients aged 50+ years	Colorectal cancer cases (1999-2001) and colorectal cancer deaths (1999-2001)	SES and colorectal cancer incidence were positively associated in Hispanics, but negatively associated in Black Americans and White Americans. In White Americans, as SES increased, colorectal cancer mortality declined. However, this was not observed in Hispanics or Black Americans.
Gomez et al. (2007)	SEER-Medicare Linked Data	Colorectal cancer patients (65+ years) diagnosed between 1992-1996 and followed through 1999.	Mortality rate	Unadjusted colorectal cancer mortality rates were higher in Black Americans and Hispanic compared to White Americans and lower in Japanese compared to White Americans. Adjustments for stage eliminated the difference between White Americans and Hispanics and White Americans and Japanese. Comparing Black Americans to White Americans, cancer stage and SES only accounted for half of the observed mortality disparity.

Racial disparities. Many key studies have reported significant racial disparities in colorectal cancer survival (Hassan et al., 2009; Hines et al., 2012; Le et al., 2008; Wallace et al., 2013; Wassira et al., 2013; White et al., 2010) . For example, a study by Wassira et al. (2013) showed that Black Americans had reduced colorectal cancer survival compared to White Americans. This difference persisted even after adjustment for tumor stage, sex, age, diagnosis period, tumor sub-location, marital status and SES (Wassira et al., 2013). However, the role of race has been debated. A study by Yan et al. (2009) found that after adjusting for age, marital status, sex, SES group, cancer stage, and treatment, race was no longer a significant predictor of overall survival. Therefore, the data is somewhat mixed as to whether race or other factors associated with race predict colorectal cancer survival. However, it is well documented that Black Americans have higher colorectal cancer mortality rates compared to White Americans. Although the data sources are minimal, colorectal cancer survival disparities have also been noted in other minority races including Native Americans (Cuerto et al., 2011) and Hispanics (Jafri et al., 2013; Wan et al., 2012). Authors from several documented studies comment that treatment related data are needed to advance the understanding of racial disparities in colorectal cancer (Wassira et al., 2013; Wallace et al., 2013). This might help explain the difference in results comparing the Wassira et al. (2013) study to the Yan et al. (2009) study; as Yan et al. (2009) controlled for treatment regimen. This point is of particular

importance to this study as disparities in treatment could help explain some of the racial disparities in colorectal cancer survival.

In recent years (2004-2009), new treatments and technologies appear to be contributing to increasing survival rates in all races (Sineshaw et al., 2014). However, these survival increases are greater in White Americans compared to Black Americans (Sineshaw et al., 2014). Therefore, racial disparities in colorectal cancer survival, specifically comparing Black Americans and White Americans, are increasing (Sineshaw et al., 2014). Similar widening disparities were observed in a study by Robbins et al. (2012). Sineshaw et al. (2014) propose that minority populations have not benefited as greatly from newer treatments or these treatments have not be disseminated into these populations as robustly (Sineshaw et al., 2014). This hypothesis is supported by Simpson et al. (2013). In the Simpson et al. (2013) study, black metastatic colorectal cancer patients were less likely to receive multimodality therapy compared to White Americans. When adjusted for differences in treatment, the increased risk of death seen in black versus white patients disappeared (Simpson et al., 2013).

Given that race has been associated with colorectal cancer survival and the presence of studies linking racial disparities to treatment disparities, it is critically important to include race as a variable in this study. It is possible that newer technologies (monoclonal antibody therapies) are driving the recent widening disparities in colorectal cancer based on race. Studies describing racial disparities in colorectal cancer survival are described in detail in Table 6.

Table 6

Racial Colorectal Cancer Survival Disparities

Source	Data Source	Sample	Outcome Variable(s) Assessed	Results
Sineshaw et al. (2014)	SEER Program	US colorectal cancer patients (1992-2009)	5-year cause specific survival rates	Comparing 1992-1997 to 2004-2009, the 5-year survival rate increased in all races except nonHispanic Black Americans (nonHispanic White Americans = 11.4% to 17.7%; nonHispanic Black Americans = 8.6% to 9.8%).
Simpson et al. (2013)	SEER-Medicare Linked Data	Elderly (66+ years) US colorectal cancer patients	Survival (logistic regression)	Black race was associated with a higher mortality rate (Hazard Ratio: 1.15, 95% CI: 1.08-1.22). When the data were adjusted for patient, tumor, and demographic factors, the Hazard Ratio reduced marginally, but was still significant. When the data were adjusted for treatment, there was no longer a significant different in survival comparing white to black race.
Wassira et al. (2013)	Nevada Central Cancer Registry	Patients diagnosed with colorectal cancer between 1995-2007.	Survival rates	Black Americans had a 20.1% higher risk of colorectal cancer death compared to White Americans. This increased risk persisted when adjusted for tumor stage, sex, diagnosis period, tumor sub location, marital status, and SES.
Wallace et al. (2013)	South Carolina Central Cancer Registry	Pathologically documented colorectal cancer patients diagnosed from 1996-2006.	Median survival time	Disparities in survival comparing Black Americans to White Americans increased over the time period. Comparing age groups, younger patients (less than 50 years) showed the greatest racial disparities in survival (Hazard Ratio: 1.34, 95% CI: 1.06-1.71). Elderly male patients (50+ years) showed only modest racial disparities in survival (Hazard Ratio: 1.16, 95% CI: 1.01-1.32). No racial disparity was observed in elderly female patients.

table continues

Jafri et al. (2013)	SEER Program	Hispanic colorectal cancer patients (1993-2007)	1 and 5 year age-specific survival rates	Comparing 1993-1997 to 2003-2007, there were no significant improvements in 5-year survival for Hispanics. There was a modest increase in 1-year survival for young Hispanics.
Lansdorp-Vogelaar et al. (2012)	MISCAN-Colon microsimulation model	Colorectal cancer patients (50 years+) from 1975-2007	Mortality rate	Screening accounted for 19% of the survival disparity between Black Americans and White Americans. 36% of the survival disparity was explained by difference in cancer stage.
Cueto et al. (2011)	SEER database	White and Native American colorectal cancer patients	Survival rates	Survival rates from colon cancer were worse for Native Americans (Hazard Ratio: 1.20; 95% CI: 1.08-1.34), however no difference was found in survival for rectal cancer.
White et al. (2010)	SEER-Medicare Linked Data	Elderly (66+) colorectal cancer patients diagnosed between 1992-2002.	Kaplan-Meier survival methods	Black patients had worse survival compared to white patients (adjusted Hazard Ratio: 1.24; 95% CI: 1.14-1.35). Asian patients had better survival compared to black and white patients (adjusted Hazard Ratio: 0.80; 95% CI: 0.70-0.92).
Henry et al. (2009)	New Jersey Cancer Registry	Patients diagnosed with colorectal cancer between 1996-2003.	Survival rates evaluated using spatial scan statistic	There were geographical areas of the state with both higher and lower survival rates. When race/ethnicity and area SES were controlled, the risk of death was attenuated in several areas. However, in some areas, survival disparities persisted.
Yan et al. (2009)	SEER Program	Patients diagnosed with colorectal cancer between 1988-1992.	Survival rate	After adjusting for age, marital status, sex, SES group, cancer stage, and treatment, race was no longer a significant predictor of overall survival.

table continues

Le et al. (2008)	California Cancer Registry	Colorectal cancer patients diagnosed between 1994-2003.	Kaplan-Meier survival methods	After adjustment for age, sex, histology, tumor site, and stage, Black Americans had increased cancer-specific death rates compared to White Americans (Hazard Ratio: 1.19; 95% CI: 1.14-1.25). However, when adjusted for area-level SES and treatment, the disparity was reduced (Hazard Ratio: 1.08; 95% CI: 1.03-1.13).
Wan et al. (2012)	Texas Cancer Registry	Colorectal cancer patients diagnosed between 1995-2003	Survival and Access to Healthcare	Death rate from Colorectal Cancer was highest in Hispanics and nonHispanic Black Americans and in several geographical regions. Spatial access to oncologists was significantly associated with survival in nonurban areas only. Access to care did not explain all geographic disparities.
Hines et al. (2012)	SEER Cancer Program Data (reporting from Atlanta and Rural Georgia Cancer Registries only)	Colorectal Cancer patients diagnosed between 1992-2007.	Mortality rate	Black Americans had slightly higher death rates from both colon and rectal cancer (Hazard Ratio for colon cancer: 1.11, 95% CI: 1.00-1.24; Hazard Ratio for rectal cancer: 1.24, 95% CI: 1.14-1.35).
Hassan et al. (2009)	Madigan Army Medical Center medical records	Colorectal cancer patients treated at the medical center from 1994-2004.	Stage of diagnosis, Survival rates	NonWhite Americans were diagnosed younger and with a greater proportion of stage III tumors. Age and tumor stage at diagnosis were the only predictors of lower survival.

table continues

Robbins et al. (2012)	SEER cancer program database	Colorectal cancer patients (1985-2008)	Stage-specific mortality rate	Over the time period (1985-2008) mortality rates decreased in both Black Americans and White Americans. However, the decreases were smaller for Black Americans. As a result, black to white mortality ratios increased from 1.17 to 1.41 for localized disease, from 1.03 to 1.30 for regional disease, and from 1.21 to 1.72 for distant-stage disease.
Gomez et al. (2007)	SEER-Medicare Linked Data	Colorectal cancer patients (65+ years) diagnosed between 1992-1996 and followed through 1999.	Mortality rate	Unadjusted colorectal cancer mortality rates were higher in Black Americans and Hispanic compared to White Americans and lower in Japanese compared to White Americans. Adjustments for stage eliminated the difference between White Americans and Hispanics and White Americans and Japanese. However, comparing Black Americans to White Americans, cancer stage and SES only accounted for half of the observed mortality difference.

Neighborhood disparities. The presence of neighborhood level disparities (degree of urbanization) in colorectal cancer survival is controversial. Several studies have shown that controlling for SES attenuated rural versus urban difference in colorectal cancer survival. For example, Hines et al. (2014) reported that rural residents in Georgia had 14% higher risk of colorectal cancer mortality compared to urban residents. However, when the authors adjusted for SES level, the disparity was attenuated (Hines et al., 2014). Additionally, Henry et al. (2009) found geographical areas in New Jersey with higher rates of colorectal cancer death. As in Hines et al. (2014), when the authors controlled for SES, the increased risk of death was attenuated in many areas (Henry et al., 2009).

Lian et al. (2011) took a different approach to the question and directly assessed neighborhood socioeconomic deprivation as a predictor of colorectal cancer mortality rates using national-level data. Contrary to the previously listed studies, in the Lian et al. (2011) article, SES did not explain all the observed geographical variation in mortality. Therefore, although neighborhood SES appears to contribute to colorectal cancer mortality, other neighborhood characteristics might also contribute to colorectal cancer mortality rates. In this study, neighborhood characteristic (degree of urbanization) disparities in monoclonal antibody treatment will be addressed to assess the possibility that rural residents have reduced levels of treatment compared to urban residents. Studies describing neighborhood characteristic disparities in colorectal cancer survival are described in detail in Table 7.

Table 7

Neighborhood Characteristics Related to Colorectal Cancer Survival Disparities

Source	Data Source	Sample	Outcome Variable(s) Assessed	Results
Hines et al. (2014)	Georgia Cancer Registry, Census-tract level geographic residency and SES.	Patients diagnosed with colorectal cancer in Georgia (2000-2007)	Mortality rate	Rural residents had a 14% increased risk of death (Hazard Ratio: 1.14, 95% CI: 1.07-1.22). However, when adjusted for SES level, no difference in survival between rural and urban tracts was noted.
Wan et al. (2012)	Texas Cancer Registry	Colorectal cancer patients diagnosed between 1995-2003	Survival and Access to Healthcare	Death rate from Colorectal Cancer was higher than the state average in several geographical regions. Spatial access to oncologists was significantly associated with survival in nonurban areas only. Access to care did not explain all geographic disparities.
Henry et al. (2009)	New Jersey Cancer Registry	Patients diagnosed with colorectal cancer between 1996-2003.	Survival rates evaluated using spatial scan statistic	There were geographical areas of the state with both higher and lower survival rates. When race/ethnicity and area SES were controlled, the risk of death was attenuated in several areas. However, in some areas, survival disparities persisted.
Chien et al. (2013)	SEER Program-Medicare Linked Data	Elderly US colorectal cancer patients (1991-2005)	Spatial pattern of survival	Identified geographic areas with excess risk of colorectal cancer death. Spatial patterns varied by cancer type and stage.
Lian et al. (2011)	National Institutes of Health AARP Diet and Health Study	Colorectal cancer patients identified from 1995-2003 and followed until 2006.	Survival (Bayesian multilevel survival models)	There was significant geographical variation in colorectal cancer-specific mortality. Lower SES neighborhoods had higher colorectal cancer-specific mortality (Hazard Ratio: 1.2; 95% CI: 1.1-1.5). . However, neighborhood SES did not account for geographical variation in colorectal cancer deaths.

Underlying causes of colorectal cancer survival disparities. As documented in detail in the sections above, there are a multitude of studies describing disparities in colorectal cancer survival based on many different factors including age, race, SES and neighborhood characteristics (degree of urbanization). However, in order to reduce these disparities, it is important to understand the underlying causes of these disparities. This area of research is less well developed, however the three main areas of exploration include observed disparities in access to care, screening and treatment.

Access to care disparities. One potential underlying cause of the colorectal cancer survival disparities are disparities in access to care. Oliphant et al. (2013a) found that access and utilization of a colorectal cancer specialist for surgery as opposed to a nonspecialist was associated with a reduced risk of postoperative death and an increased 5-year survival rate. Likewise, Zullig et al. (2013) found that older age was associated with reduced rate of colorectal cancer specialist referral. Therefore, it is possible that access to a specialist contributes to the colorectal cancer survival disparity seen in the elderly. Additionally, Laiyemo et al. (2010) found racial disparities in follow-up to abnormal colorectal cancer screening results; with Black Americans less likely than White Americans to obtain follow-up within 1 year of abnormal result. Access to care was proposed as a reason for this disparity (Laiyemo et al., 2010) However, if access to care is gauged by healthcare spending, Wright et al. (2007) found that the total cost for colorectal cancer care was slightly higher in Black Americans compared to White

Americans. Therefore, the role that access to care disparities play in overall colorectal cancer survival is still unclear.

Screening disparities. Another potential underlying cause of colorectal cancer survival disparities are disparities in early detection or screening. Multiple papers have highlighted racial differences in colorectal cancer screening. For example, Crawford et al. (2010), Stimpson et al. (2012), and Wilkins et al. (2012) all found that Black Americans had lower colorectal cancer screening rates compared to White Americans. The relevance of these results to survival was addressed by Lansdorp-Vogelaar et al. (2012). In this article, the authors found that 19% of the difference in survival between elderly Black Americans and White Americans with colorectal cancer could be explained by differences in screening (Lansdorp-Vogelaar et al., 2012). Therefore, disparities in early detection via screening do appear to contribute to disparities in colorectal cancer survival. However, according to Lansdorp-Vogelaar et al. (2012) there are additional factors contributing to the differences in survival, including potential disparities in treatment, which need further research. Interestingly, when Stimpson et al. (2012) adjusted their analysis for SES, no racial screening disparities were found.

Cole et al. (2012) found a modest difference in colorectal screening comparing rural to urban zip codes. Specifically, after adjustment for demographics and health characteristics, rural residents had lower screening rates compared to urban residents (48% vs. 54%, respectively) (Cole et al., 2012). These screening rate disparities could be reflecting access to care disparities given the geographical nature of the disparity.

However, this was not address by Cole et al. (2012). Finally, Lang et al. (2009) performed a temporal study to correlate colorectal cancer screening rates and survival rates. Across the time period (1992-2005), earlier detection did not appear to impact survival (Lang et al., 2009). However, the authors did find that technological improvements (new technologies/treatments or better use of current treatments) and demographics were responsible for the largest share of the temporal 5-year survival improvement (Lang et al., 2009).

Treatment disparities. A final potential underlying cause of colorectal cancer survival disparities are disparities in treatment. This area of disparity research has been researched to a greater extent compared to access to care or screening disparities. Multiple papers have found that different demographic variables including age, rural versus urban geography, SES, and race are associated with different treatment regimens. For example, although targeted treatments (such as chemotherapy) have been shown to increase survival rates, Mitry et al. (2013) found that only 10% of elderly received targeted treatments, compared to 40% of younger populations. These results are supported in the literature as age-related disparities in colorectal cancer treatment (surgery, chemotherapy, radiation) have been observed in multiple different studies including Bakogeorgos et al. (2013), Obeidat et al. (2010) and Serra-Rexach et al. (2012). Of importance for this study, this age-related treatment disparity has persisted into newer colorectal cancer therapies. Fu et al. (2013) and Kozloff et al. (2010) observed reduced adjunct bevacizumab treatment with increasing patient age. Therefore, the literature

supports an age-related colorectal cancer treatment disparity, where the elderly are less likely to receive newer treatments.

In addition to age disparities, many articles have addressed racial and SES disparities in colorectal cancer treatments. Le et al. (2008) found significant differences in first-line treatment (surgery, radiation, and chemotherapy) comparing Black Americans to White Americans. These racial disparity results were supported in subsequent studies by Hao et al. (2011), Hines et al. (2012), and White et al. (2008). Interestingly, Zullig et al. (2013) found no racial disparities in treatment when the population was limited to Veteran Affairs (VA) hospitals. In their article, Zullig et al. (2013) proposed that the racial disparities in treatment observed in the aforementioned studies could be related to insurance status or access to care differences.

Compared to higher SES areas, Lejeune et al. (2010) found that patients from lower SES areas were less likely to receive treatment within 6 months of colorectal cancer diagnosis. Additionally, Le et al. (2008) found significant differences in first-line treatment (surgery, radiation, and chemotherapy) comparing higher SES to lower SES areas. Hao et al. (2011) described urban versus rural disparities in first-line chemotherapy treatment in colorectal cancer. Specifically, populations with urban or suburban zip codes were, respectively, 38% and 52% more likely to receive chemotherapy compared to populations with rural zip codes (Hao et al., 2011).

Of particular importance to this study, my literature searches did not uncover any studies addressing racial, SES, or neighborhood disparities in monoclonal antibody

treatment of colorectal cancer. Adjunct monoclonal antibody therapy has been shown to reduce colorectal cancer mortality in multiple populations, including the elderly (as discussed in the next section). It is possible that disparities in monoclonal antibody treatment are contributing to disparities in colorectal cancer survival. However, given the lack of research, disparities in monoclonal antibody treatments are a clear literature gap.

Studies describing access to care, screening and treatment disparities in colorectal cancer patients are described in Table 8.

Table 8

Access to Care, Screening and Treatment Disparities

Access to care disparities

Source	Source	Sample	Outcome Variable(s) Assessed	Results
Singal et al. (2013)	Texas Medicare Database (2000-2009)	Medicare beneficiaries (age 66-75)	Colorectal cancer screening rates	Patients with a primary care physician (PCP) were more likely to be screened for colorectal cancer (OR: 2.05; 95% CI: 2.03-2.07). Ethnic disparities in screening were almost eliminated after accounting for PCP.
Wilkins et al. (2012)	Telephone surveys in two rural Georgia counties	Individuals 50+ years old living in the two highest colorectal cancer mortality counties in Georgia.	Colorectal cancer screening rates	Black Americans were more likely to report colorectal cancer screening barriers. Having a physician recommend screening was associated with the highest screening rates regardless of race.
Zullig et al. (2013)	Veteran Affairs (VA) Health System Database (encompassing 128 centers)	Patients diagnosed with colorectal cancer within the VA system from 2003-2006.	Treatment patterns	Older age at diagnosis was associated with reduced odds of medical oncology referral and surveillance colonoscopy.
Laiyemo et al. (2010)	Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial (10 geographically diverse sites)	Subjects enrolled in the study between 1993-2001 who received trial-sponsored flexible sigmoidoscopy (FSG) and obtained an abnormal result.	Follow-up rates	Black Americans were less likely than White Americans to obtain follow-up colonoscopy within 1 year of abnormal FSG result (62.6% vs. 72.4% respectively).
Wright et al. (2007)	SEER-Medicare Linked Data	Elderly (65+) colorectal cancer patients with stage II-III rectal or stage III colon cancer diagnosed in 1992-1996.	Surgical and post-surgical costs for colorectal cancer care.	The total colorectal cancer cost for Black Americans was higher than White Americans ($44,199 vs. $38,588). However, after adjusting for covariates, this difference was insignificant.

table continues

Oliphant et al. (2013a)	Regional audit database from West of Scotland	Colorectal cancer patients undergoing surgery during two time periods (1991-1994 and 2001-2004)	Postoperative mortality (30 days post surgery) and 5-year relative survival rates.	Both postoperative mortality rates and 5-year survival rates were improved in patients who received surgery from a specialist.

Screening disparities

Source	Data Source	Sample	Outcome Variable(s) Assessed	Results
Lam et al. (2013)	Northern California urban county hospital records and national SEER database	Colorectal cancer patients over 11-year time period	5-year mortality	Colorectal cancer screening was associated with improved survival in this urban underserved population. Insurance status of the patients was not associated with survival.
Lansdorp-Vogelaar et al. (2012)	MISCAN-Colon microsimulati on model	Black colorectal cancer patients (50 years+) from 1975-2007	Mortality rate	Screening accounted for 19% of the survival disparity between Black Americans and White Americans.
Crawford et al. (2010)	Behavioral Risk Factor Surveillance System (BRFSS)	Black, white and Hispanics (age 50+ years) between 2002-2004	Colorectal cancer screening rates	White Americans were more likely than Black Americans or Hispanics to have received either type of colorectal cancer screening (fecal occult blood testing or endoscopy). Persons without insurance or a usual source of care were less likely to obtain screening.
Stimpson et al. (2012)	National Health Interview Study and the linked Area Resource File	Subjects with a personal history of cancer, subjects over 50 years old, or subjects over 40 years old with a family history of cancer.	Colorectal endoscopy rates	White Americans were more likely to report a colorectal endoscopy exam (44%), followed by Black Americans (36%) and Hispanics (28%). The difference in screening rates between White Americans and Black Americans was eliminated when adjusted for area socioeconomic status.
Wilkins et al. (2012)	Telephone surveys in two rural Georgia counties	Individuals 50+ years old living in the two highest colorectal cancer mortality counties in Georgia.	Colorectal cancer screening rates	Black respondents had lower screening rates compared to White Americans (50.4% and 63.4% respectively).

table continues

Cole et al. (2012)	Behavioral Risk Factor Surveillance System (BRFSS)	Individuals (age 50+ years) between 1998-2005	Colorectal cancer screening rates	Although the rates in both rural and urban communities increased over time, after adjustment for demographics and health characteristics, rural residents had lower screening rates compared to urban residents (48% vs. 54% respectively).
Lang et al. (2009)	SEER Database	Elderly (66+) colorectal cancer patients diagnosed between 1992-2000, follow-up through 2005.	5-year survival	Comparing the time periods, earlier detection did not appear to impact survival. Technological improvements (new technologies/treatments or better use of current treatments) and demographics were responsible for the largest share of the temporal 5-year survival improvement.

Treatment disparities

Source	Data Source	Sample	Outcome Variable(s) Assessed	Results
Rane et al. (2014)	West Virginia Cancer registry and the SEER database	Medicare beneficiaries aged 66+ and diagnosed with colorectal cancer from 2003-2006	3-year survival and Receipt of minimally-appropriate colorectal cancer treatment	Compared to the national colorectal cancer sample, individuals diagnosed in West Virginia were diagnosed in early stages, but had poorer survival rates. The West Virginia cohort had lower treatment rates.
Mitry et al. (2013)	French population-based registry	Patients with nonresectable metastatic colorectal cancer diagnosed from 1976-2009	Treatment rates and survival rates	Comparing the 1997-2004 time-period to the earlier time period, patients receiving chemotherapy increased from 19% to 57%. However, from 2005-2009, less than 10% of the elderly received targeted treatments, compared to 40% for younger patients.
Cueto et al. (2011)	SEER database	White and Native American colorectal cancer patients	Survival rates and Treatment Regimens	Native Americans and White Americans had similar surgery recommendation rates. Native Americans were more likely to receive radiation, but less likely to receive sphincter-preserving surgery.

table continues

Lejeune et al. (2010)	Three UK cancer registries linked to area-level socio-economic information	Colorectal cancer patients diagnosed between 1997-2000.	Access to treatment (measured in time from diagnosis to treatment) and 3-year survival	Patients from lower SES areas were less likely to receive treatment within 6 months of diagnosis. When limited to patients receiving treatment within 1 month of diagnosis, no survival disparity existed based on area SES.
Fu et al. (2013)	SEER-Medicare Linked Data	Elderly (65+) colorectal cancer patients, diagnosed from 2005-2009 and given chemotherapy anytime through 2010.	Patient characteristics associated with adjunct bevacizumab use.	57% of newly diagnosed patients received bevacizumab, compared to 44% of patients with progressive disease. Patients aged 80+ were less likely than younger elderly patients (65-69) to receive adjunct bevacizumab (OR: 0.64; CI: 0.57-0.73).
Kozloff et al. (2010)	BRiTE observational cohort study	Colorectal cancer patients	Treatment patterns, safety, progression-free survival, overall survival	Use of bevacizumab as a first-line adjunct therapy decreased with age. Comparing elderly (65+ years) receiving bevacizumab to elderly not receiving bevacizumab, there was a significant increase in median survival with bevacizumab
Wildes et al. (2010)	Barnes-Jewish Hospital Oncology Data Services Registry (1996-2006)	Stage III Colorectal cancer patients (age 65-99)	Chemotherapy treatment rates, death rates	Within this group, advancing age was associated with a reduction chemotherapy treatment. Receiving chemotherapy was associated with lower risk of death for all stratified elderly age groups.
McKibbin et al. (2008)	Medical records from 10 U.S. community practices	Patients with advanced stage colorectal cancer (2003-2006).	Treatment patterns	Elderly individuals (65+ years) were less likely to receive first-line doublet chemotherapy compared to younger (under 65 years) patients (54% vs. 84%). The use of each of the medicines studied (irinotecan, oxaliplatin and bevacizumab) were all lower in elderly patients.
Bakogeorgos et al. (2013)	Single institution comparative study	Colorectal cancer patients (52.8% elderly; 70+)	Treatment, dose, toxicity, efficacy	Elderly patients (70+ years) were more likely to receive only single agent chemotherapy, fewer cycles (6.2 cycles compared to 8.3 for younger patients), and lower doses (42.8% of planned dose compared to 78.4% for younger patients).

table continues

Serra-Rexach et al. (2012)	Hospital records (Administrative database) from a single institution in Madrid, Spain	Colorectal cancer patients diagnosed or treated at the institution	Treatment patterns	The younger than 75 years of age group was more likely to receive surgery, radiation, and chemotherapy and less likely to receive palliative care compared to the older group (75+).
Obeidat et al. (2010)	SEER-Medicare Linked Data	Elderly (66+ years) stage IV colorectal cancer patients diagnosed between 2000-2002 and treated with first-line chemotherapy	Irinotecan treatment (newest chemotherapy option) based on race or age	Black Americans were less likely to be treated with the newer chemotherapy option (Irinotecan) compared to White Americans (OR: 0.641; 95% CI: .453-0.907). Older elderly patients (71+ years) were less likely to receive Irinotecan compared to younger elderly patients (66-70 years).
White et al. (2008)	SEER-Medicare Linked Data	White and Black elderly (65+ years) colorectal cancer patients diagnosed between 1991-2002.	Treatment patterns	Receipt of standard of care (stage specific and defined by the Physician Data Query guidelines) increased in both race groups over time. However, Black Americans were overall 16% less likely to receive standard of care compared to White Americans (OR: 0.85; 95% CI: 0.78-0.90).
Haas et al. (2011)	SEER-Medicare Linked Data	Elderly (65+ years) colorectal cancer patients diagnosed between 1992-2005.	Treatment patterns and mortality	Black Americans and Hispanics were less likely than White Americans to undergo surgery (OD: 0.57, 95% CI: 0.52-0.63 and OR: 0.82, 95% CI: 0.70-0.95 respectively). Similar disparities were seen for receipt of adjunct chemotherapy. Adjustment for area sociodemographics, surgeon capacity, and medical oncologist capacity reduced the observed disparities.
Hao et al. (2011)	Georgia Comprehensive Cancer Registry	Black and white persons diagnosed with stage III colon or stage II/III rectal cancer between 2000-2004	Treatment patterns	Urban and suburban patients were more likely than rural patients to receive chemotherapy (38% and 53% more likely respectively). Urban black patients were 24% less likely to receive chemotherapy compared to urban white patients. There were no racial differences within suburban and rural areas of residence.

table continues

Hines et al. (2012)	SEER Cancer Program Data (reporting from Atlanta and Rural Georgia Cancer Registries only)	Colorectal Cancer patients diagnosed between 1992-2007.	Surgery rates	Black Americans had lower rates of surgery (50% decreased odds for colon cancer; 67% decreased odds for rectal cancer).
Hassan et al. (2009)	Madigan Army Medical Center medical records	Colorectal cancer patients treated at the medical center from 1994-2004.	Stage of diagnosis, treatment patterns	NonWhite Americans were diagnosed younger and with a greater proportion of stage III tumors. No disparities in treatment (surgery, adjunct therapy, disease recurrence) were found comparing nonWhite Americans to White Americans.
Hines et al. (2014)	Georgia Comprehensive Cancer Registry, Census-tract level geographic residency and SES.	Patients diagnosed with colorectal cancer in Georgia (2000-2007)	Treatment patterns	Lower SES census tracts had lower colon cancer surgery rates. Rural and lower-SES census tracts also had reduced rates of chemotherapy.
Demissie et al. (2004)	SEER database	Black and white persons diagnosed with colon or rectal cancer between 1988-1997.	Treatment patterns	Black Americans were less likely than White Americans to receive standard of care. Black Americans were also more likely to refuse recommended treatment.
Zullig et al. (2013)	Veteran Affairs (VA) Health System Database (encompassing 128 centers)	Patients diagnosed with colorectal cancer within the VA system from 2003-2006.	Treatment patterns	Black Americans and White Americans were equally likely to receive National Comprehensive Cancer Network guideline-concordant colorectal cancer care through the VA system.
Koroukian et al. (2010)	Ohio Cancer Surveillance System, Vital Records, Medicare Administrative Data, Home Health Care Outcome and Assessment Information Set (OASIS)	Elderly colorectal cancer patients diagnosed between 1999 and 2001 and admitted to home health care 30 days before or after diagnosis	Treatment patterns	Functional limitation and geriatric syndromes were associated with lower likelihood of treatment (surgery-only or surgery + chemotherapy).

Safety and efficacy of adjunct monoclonal antibody treatment in elderly colorectal cancer patients. In this study, I aim to understand a potential underlying disparity in colorectal cancer survival, namely disparities in adjunct monoclonal antibody treatment. Therefore, for the validity of the hypothesis, it's critically important that monoclonal antibodies extend survival and are safe to use in the study population.

Multiple studies determined that adjunct monoclonal antibody treatment with either bevacizumab or cetuximab was safe in elderly patients with either manageable side effects or side effects equivalent to the younger populations of patients (Abdelwahab et al., 2012; Bruera et al., 2013, Cunningham et al., 2013; Naeim et al., 2013). Additionally, multiple clinical studies showed improved survival in elderly colorectal cancer patients treated with adjunct monoclonal antibody therapy. For example, in a large Phase III study, Cunningham et al. (2013) found that elderly patients that received bevacizumab plus chemotherapy had a median progression-free survival time of 9.1 months compared to 5.1 months in patients treated with chemotherapy only. Progression-free survival improvement was also seen with adjunct bevacizumab in a smaller Phase II study by Vrdoljak et al. (2011). Similar results were obtained with adjunct cetuximab therapy. Specifically, elderly colorectal cancer patients treated with cetuximab plus chemotherapy had a progression-free survival time of 8.8 months, compared to 5.8 months for chemotherapy only (Price et al., 2012).

Several recent population-based observational studies have supported these clinical findings. Kozloff et al. (2010) and Renouf et al. (2011) both found that adjunct bevacizumab therapy increased median survival time in elderly colorectal cancer patients. Additionally, Jehn et al. (2012) found that the efficacy and side effect profile of adjunct cetuximab therapy was similar comparing elderly patients to younger patients.

Table 9 includes both clinical and observational studies citing the efficacy and/or safety of monoclonal antibody treatment in the elderly.

Table 9

Monoclonal Antibody Treatment of Elderly Colorectal Cancer Patients

Clinical Studies with Adjunct Monoclonal Antibody Therapy				
Source	Data Source	Sample	Outcome Variable(s) Assessed	Results
Naeim et al. (2013)	Clinical study (capecitabine plus bevacizumab)	Phase II study in elderly colorectal cancer patients	Survival and Side-Effects	Median overall survival was 12.7 months with manageable side-effects and no new safety signals for capecitabine (chemotherapy) and bevacizumab (anti-VEGF monoclonal antibody) combination.
Bruera et al. (2013)	Clinical study (triple chemotherapy plus bevacizumab)	Phase II study fit elderly colorectal cancer patients	Survival and Side-Effects	Drug activity, efficacy, and safety of triple chemotherapy + bevacizumab were equivalent to the overall colorectal cancer population.
Cunningham et al. (2013)	Clinical Trial (capecitabine +/- bevacizumab)	Phase III Study in elderly (70+) previously untreated colorectal cancer patients	Progression -free survival	Elderly patients had improved progression-free survival with the addition of adjunct bevacizumab to their chemotherapy (median survival 9.1 months vs. 5.1 months; hazard ratio: 0.53, 95% CI: 0.41-0.69). Adverse events were slightly higher in the combination group, but overall the combination was tolerated well.
Vrdoljak et al. (2011)	Clinical Trial (first line capecitabine +/- bevacizumab)	Phase II trial in elderly (70+ years) colorectal cancer patients	Progression -free survival, overall survival	Median progression-free was 11.5 months and overall survival was 21.2 months. Side-effects were similar to those reported in earlier studies with younger age groups.

table continues

Price et al. (2012)	MAX trial data (international randomized controlled trial)	Colorectal cancer patients given capecitabine or capecitabine + bevacizumab	Progression -free survival and toxicity	The addition of bevacizumab to capecitabine to the elderly (75+ years) population increased progression free survival (5.8 months vs. 8.8 months; Hazard ratio: 0.65). Compared to the younger age groups, there were no major toxicity differences.
Abdelwahab et al. (2012)	Clinical Trial (irinotecen plus cetuximab)	Elderly colorectal cancer patients (65 years +) who progressed after at least one previous treatment	Survival and Side-Effects	Median survival time was 7 months and median progression-free survival was 4 months for irinotecen (chemotherapy) + cetuximab (anti-EGFR monoclonal antibody). Side effects were tolerable.
Observational Studies with Adjunct Monoclonal Antibody Therapy				
Source	Data Source	Sample	Outcome Variable(s) Assessed	Results
Jehn et al. (2012)	Multicenter study in Germany using electronic databases (2005-2007)	Colorectal cancer patients (18 years +) treated with cetuximab + irinotecan	Progression -free survival and side effect assessment	Side-effects and efficacy (progression-free survival) of cetuximab + irinotecan were similar comparing the greater than 65 age group to the less than 65 age group.
Kozloff et al. (2010)	BRiTE observational cohort study	Colorectal cancer patients given bevacizumab as a first line adjunct therapy	Treatment patterns, safety, progression -free survival, overall survival	Comparing elderly (65+ years) receiving bevacizumab to elderly not receiving bevacizumab, there was a significant increase in median survival with bevacizumab treatment.
Renouf et al. (2011)	British Columbia cancer agency database	Patients diagnosed with colorectal cancer in 2003/2004 or 2006	Survival pre and post bevacizumab approval (2006)	In the 2003/2004 cohort, 5.9% received bevacizumab; whereas in the 2006 cohort 30.6% received bevacizumab. Overall survival was improved in the 2006 cohort compared to the 2003/2004 cohort (13.8 months to 17.3 months).

Evaluation of Strengths and Weaknesses of Previous Approaches

The vast majority of literature cited in the sections above utilized secondary data sources (such as the SEER-Medicare database) to address their research questions. In a broad sense, this has multiple benefits. First, the data is already collected, organized, and available for use saving both time and cost. Second, using secondary data sources results in assess a larger or more geographically broad sample. Third, since the data is already collected, it is possible to retrospectively use several years of collected data; something that might not have been possible if primary data collection were performed. Alternatively, data for some of the literature cited above was collected or compiled specifically for the study in question. These included the utilization of hospital records, telephone surveys, or clinical trials. The benefit of these approaches is the researchers were able to address very specific questions; which may not have been answered using published database sources. However, these approaches can be time consuming and expensive. Therefore, researchers using these approaches often had smaller sample sizes or addressed a limited sample (either temporally or geographically). Table 10 describes the strengths and weakness of multiple data acquisition approaches.

Table 10

Strengths and Weaknesses of Previous Data Collection Methods Related to the Study

Previous Data Collection Method	Strengths	Weaknesses
Telephone Surveys	Ability to address a very specific question in a very specific geographical region. Wilkins et al. (2012) were able to address the impact of physician colorectal cancer screening recommendation on screening rates and also identify specific screening barriers.	Telephone surveys are time consuming and limited the scope in Wilkins et al. (2012). This type of approach is not feasible for national-level assessments or large sample sizes. Research is often limited to a cross-sectional approach.
Data Collection from Hospital/Medical Records	If present in medical records, this approach would allow the researchers to address a specific question. Hassan et al. (2009) had specific questions about diagnosis and survival at a single VA center, therefore, this approach was appropriate. This approach also allows the researcher to address differences over time.	This approach also limits the study scope. Hospital record research is likely only feasible if performed at a few hospitals. This would not be feasible on a large scale and could be time consuming.
Clinical Trials	Clinical trials allow the researcher to prospectively address questions such as drug efficacy (Naeim et al., 2013; Bruera et al., 2013; Cunningham et al., 2013). These studies are very well controlled.	This approach is very expensive and time consuming. It is also not feasible for some types of questions due to ethical limitations.
Database (SEER-Medicare Database, Veteran Affairs Database, State-Level Cancer Surveillance Systems, BRFSS)	Researchers used these databases to ask a multitude of research questions including the impact of demographics on colorectal cancer survival and treatment without performing any primary data collection (Naishadham et al., 2011; White et al., 2010; Simpson et al., 2013). This method can save time in data collection and allows for large, broad data sets.	Using secondary databases can limit the questions that can be asked. The data in these sources can be difficult to access and use if not coded uniformly. The multitude of data often requires software for extraction.

Given the research questions address disparities in monoclonal antibody treatment based on demographic variables at a national level, some of these approaches are more feasible than others. For example, telephone surveys are not feasible for this study. It would be difficult to randomly call enough metastatic colorectal cancer patients to adequately assess the use of monoclonal antibody therapy. Additionally, clinical trials would not be feasible due to cost, time, and ethics of the questions asked. Therefore, since the SEER-Medicare linked database contains all the variables needed to address the research question in the study, the secondary data collection method using published databases is appropriate.

The SEER-Medicare database is a large population-based database that contains matched medical records and demographic information for Medicare beneficiaries who have been diagnosed with cancer (NCI, 2015a). For this research project, several SEER-Medicare datasets containing data from colorectal cancer patients 65+ years in age were requested and obtained. The first file obtained was the SEER PEDSF; a file containing demographics and information related to the patient's cancer diagnosis (i.e. information on cancer stage, diagnosis date, previous primary tumors) (NCI, 2015b). In addition to the SEER file, four matched Medicare claims files were obtained. These included the Part D Event (PDE) dataset, the Outpatient Claims dataset, the Durable Medical Equipment (DME) dataset and the Carrier Claims dataset from 2007 to 2012. The PDE dataset includes all drugs prescribed under Medicare Part D; data available only 2007 to 2012 (NCI, 2015b). The remaining datasets contain all Part B claims including diagnosis

and procedure codes (NCI, 2015b). These five patient matched SEER-Medicare datasets contain information on all the variables required to answer the research questions.

Justification of Chosen Variables

The independent variables (race, SES, neighborhood characteristics) selected for the research study were chosen due to their relationship with colorectal cancer survival (either overall survival or survival specific to the metastatic colorectal cancer population). This is critical as this study aims to understand disparities in an underlying factor that could contribute to differences in colorectal cancer survival. The dependent variable, adjunct monoclonal antibody therapy is a new therapy shown to increase survival in metastatic colorectal cancer patients. Disparities in treatment of colorectal cancer patients have been documented, but the findings have not been extended to monoclonal antibody therapy. If racial, SES, or neighborhood disparities in adjunct monoclonal antibody treatment are found, future studies could determine if these disparities contribute to the previously observed disparities in survival.

Review and Synthesis of Studies Related to the Research Questions

The study evaluates racial, SES, and neighborhood (degree of urbanization) disparities in adjunct monoclonal antibody treatment of elderly colorectal cancer patients. As outlined in the literature review sections above, there are documented disparities in colorectal cancer survival based on race, SES, and neighborhood characteristics. The underlying cause(s) of these disparities is not well researched; although many papers have suggested treatment disparities as a needed follow-up study. For example,

Sineshaw et al. (2014) propose that minority populations have not benefited as greatly from newer treatments or these treatments have not be disseminated into these populations as greatly (Sineshaw et al., 2014). However, research has not been performed to determine if disparities in newer colorectal cancer therapies (such as monoclonal antibodies) exist. Given the documented efficacy of monoclonal antibody therapy, this study evaluated adjunct monoclonal antibody treatment to determine if any of the groups exhibiting survival disparities (minority race, low SES, rural neighborhood) also have disparities in monoclonal antibody treatment.

The dependent variable in the study is rate of adjunct monoclonal antibody treatment. A review of clinical trials and post-marketing observational studies indicate that all three available antibodies increase progression-free survival or overall survival when added to chemotherapy (Bruera et al., 2013, Cunningham et al., 2013; Naeim et al., 2013). Therefore, as with chemotherapy, it is possible that disparities in monoclonal antibody treatment could contribute to disparities in colorectal cancer survival.

There are three independent variables in this study. Each independent variable was selected due to its association with colorectal cancer survival. The first independent variable is race. There are many key studies that have associated race with increased risk of colorectal cancer death (Simpson et al., 2013; Sineshaw et al., 2014; Wallace et al., 2013; Wassira et al., 2013). The second independent variable is SES. As with race, many studies have associated low SES with increased risk of colorectal cancer death (Hines et al., 2014; Oliphant et al., 2013b; Wassira et al., 2013). The third independent

variable is neighborhood characteristics. Although less well researched, two studies have identified rural neighborhood as a risk factor for reduced colorectal cancer survival (Henry et al., 2009; Lian et al., 2011).

Summary and Conclusion

As outlined in the literature review above, colorectal cancer survival disparities based on race, SES, and neighborhood characteristics have been observed in many different studies. Specifically, Black Americans have reduced colorectal cancer survival rates compared to White Americans (Simpson et al., 2013; Sineshaw et al., 2014; Wallace et al., 2013; Wassira et al., 2013), low SES populations have reduced colorectal cancer survival rates compared to higher SES populations (Hines et al., 2014; Oliphant et al., 2013b; Wassira et al., 2013), and rural populations have reduced colorectal cancer survival rates compared to urban populations (Henry et al., 2009; Lian et al., 2011).

In order to address these colorectal cancer survival disparities, it is important to address the potential underlying causes of these disparities. Historical research into the underlying causes of colorectal cancer survival disparities can be lumped into three main categories; access to care disparities, screening disparities, and treatment disparities. This study aims to fill a gap and expand on potential treatment disparities in colorectal cancer. Multiple articles have highlighted disparities in both type and aggressiveness of colorectal cancer therapies (surgery, radiation, and chemotherapy) (Bakogeorgos et al., 2013; Obeidat et al., 2010; Serra-Rexach et al., 2012). Although these studies have provided valuable information on colorectal cancer treatment disparities, including

disparities in elderly populations, they were conducted using patient data from the 1990s and early 2000. The colorectal therapies employed during this period included surgery, radiation and chemotherapy. Beginning in 2004, monoclonal antibody therapies for colorectal cancer came onto the market (Scott et al., 2012). Prior to this research project, it was unknown if these historical treatment disparities had extended to the newer monoclonal antibody therapies.

Determining if disparities in monoclonal antibody treatment of elderly metastatic colorectal cancer patients exist is a critical first step to eliminating these social inequalities. If racial, SES, or neighborhood disparities are found, future studies could determine if these disparities contribute to any observed disparities in survival. Additionally, if treatment disparities are found, this would illuminate the need for policies that improve access to monoclonal antibodies, thereby helping reduce social inequalities in colorectal cancer survival. Alternatively, if treatment disparities are not found, an opportunity exists to understand the root cause of the reduced treatment disparities, and this knowledge could potentially be used to reduce treatment disparities in other cancer populations.

The next chapter outlines the specific study design aimed at addressing the literature gap described above. The study methodology and statistical analysis plans will also be described in detail.

Chapter 3: Research Method

Introduction

In this chapter I describe, in detail, the methodology for this research project. This includes information regarding the design and the rationale behind the chosen design. Additionally, the archived dataset (SEER-Medicare) that was used is described and justified. All variables (independent, dependent, and covariate/control) are identified and operationalized. Detailed information about the coding and location of these variables is also provided. I also describe the study population and the sample selection protocol that was used (including inclusion and exclusion criteria and power analysis). Additionally, within this chapter, the analysis procedure used is outlined. This includes justification of the software analysis chosen (SPSS) and details on the descriptive and inferential statistical methods. Finally, information regarding internal, external, construct, and statistical validity, in addition to ethical practices, is presented.

Research Design and Rationale

In this section, I outline, in detail, the study research design and rationale. This includes the selected independent, dependent, and control variables and the time/resource constraints.

Research Design

Study variables. The independent, dependent, and control variables used in this research project are described in detail below. For each variable, the location of the data and data codes are listed.

Independent variables. There were three research questions and three total independent variables included in this research project. All of the independent variables are available from the PEDSF (NCI, 2015c). The PEDSF is linked by a 10 digit patient ID (column #1 in the PEDSF document; patient_id) to the four Medicare claims files used in this research project.

The first research question addressed racial/ethnic disparities in first-line monoclonal antibody treatment. Therefore, race was the first independent variable. For this research project, race was identified using the SEER race recode as listed in column #101 of the PEDSF (rac_recb); 1 = Caucasian, 2 = Black, 3 = American Indian/Alaska Native, 4 = Chinese, 5 = Japanese, 6 = Filipino, 7 = Hawaiian, 8 = Other Asian or Pac. Islander, 9 = Unknown, 11 = Caucasian, Spanish origin or surname, 12 = Other unspecified (NCI, 2015c). Given the low representation of some of the races in the cohort, American Indian/Alaska Native persons were excluded from the race analysis, and Chinese, Japanese, Filipino, Hawaiian and Other Asian or Pacific Islander groups were pooled into a single group entitled Asian or Pacific Islander.

The second research question addressed SES disparities in first-line monoclonal antibody treatment. As in an article by Schlichting et al. (2012), the census tract poverty level was used to estimate of SES. This variable is located in column #146 (census_pov_ind) in the PEDSF document and lists census track poverty as 1 = 0% to < 5% poverty, 2 = 5% to < 10% poverty, 3 = 10% to < 20% poverty, 4 = 20% to 100% poverty, 9 = unknown (NCI, 2015c).

The final research question addressed neighborhood (degree of urbanization) differences in first-line monoclonal antibody treatment. The PEDSF document includes the 2003 Rural/Urban Continuum Codes from the ERS, which categories counties on an urban/rural scale based on population size, degree of urbanization, and adjacency to a metro or nonmetro area (NCI, 2015c). The PEDSF recode of these values was used for this research project. This is column #97 (urbrur) in the PEDSF document and identifies county of residence as 1 = Big Metro, 2 = Metro, 3 = Urban, 4 = Less Urban, 5 = Rural, and 9 = Unknown (NCI, 2015c).

Dependent variable. There were several research questions built into this research project. However, each research question addressed the independent variable's association with the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first-line adjunct monoclonal antibody therapy. Therefore, the dependent variable for these research questions was receipt of first-line adjunct (in combination with chemotherapy) monoclonal antibody therapy, "yes" or "no."

There are three monoclonal antibodies (bevacizumab, cetuximab, and panitumumab) currently approved for first-line treatment of patients with metastatic colorectal cancer when given in combination with chemotherapy (Amgen, 2014; Bristol-Myers Squibb & Eli Lilly and Company, 2013; Genentech, 2013). These monoclonal antibodies can be covered by Medicare Part B (if administered at an outpatient clinic or a physician's office) or Medicare Part D (if administered as a prescription drug from a specialty pharmacy). Therefore, this information was obtained from the Medicare PDE

dataset, the Medicare Outpatient Claims dataset, and the Medicare Carrier Claims dataset. The DME dataset was not used for this variable as data for these three drugs is not present within this dataset.

In the PDE dataset, the brand name for each monoclonal antibody (Avastin, Erbitux, or Vectibix) is listed in column #90 (BN). In addition the following NDC codes for the administered drug will be present in column #35 (PROD_SRVC_ID) of the PDE dataset (NCI, 2015b):

- Bevacizumab (Avastin): 50242006001 or 50242006101 (293 patients from 2007-2012)
- Cetuximab (Erbitux): 66733094823 or 66733095823 (59 patients from 2007-2012)
- Panitumumab (Vectibix): 55513095401 or 55513095601 (23 patients from 2007-2012)

In the Carrier Claims and Outpatient Claims dataset, receipt of one of the monoclonal antibodies is listed as a "J" HCPCS code (columns #93 in the Carrier Claims data set or columns #241 in the Outpatient Claims data set; NCI, 2015b):

- Bevacizumab: J9035 (25,825 patients from 2007-2012)
- Cetuximab: J9055 (7,004 patients from 2007-2012)
- Panitumumab: J9303 (2,026 patients from 2007-2012)

First-line adjunct monoclonal antibody treatment was defined as at least one claim for a monoclonal antibody (bevacizumab, cetuximab, or panitumumab) in one of the

three files above within 1 month of first postdiagnosis chemotherapy claim (which must have occurred within 6 months of diagnosis) as in Meyerhardt et al. (2012). Given the need to compare claim dates to determine if monoclonal antibody was received within 1 month of chemotherapy, the drug dispense date for each monoclonal antibody claim was also used. This is listed an MMDDYYYY in PDE dataset column #27 (srvc_mon, srvc_day, srvc_yr) and in column #32 (from_dtm, from_dtd, from_dty) of the Outpatient Claims and Carrier Claims datasets. For the purpose of this research project, the dependent variable (first-line monoclonal antibody treatment) was recorded dichotomized as "yes" or "no."

Control variables. Multiple covariates were assessed to control for confounding. The first covariate assessed was gender. This is column #41 (m_sex) in the PEDSF file: 1 = male, 2 = female. The second covariate was age at diagnosis. This is column #1881 in the PEDSF file. The third covariate was reason for original Medicare entitlement. This is column #43 in the PEDSF file: 0 = Age, 1 = Disability (individuals with 2 = End Stage Renal Disease or 3 = Disability/End Stage Renal Disease were excluded from the study). The final covariate was which registry reported the data. This is indicated by the first two digits in the patient ID (PEDSF column #1).

Research design. The specific aim of this study was to determine if there are racial, SES, or neighborhood (degree of urbanization) disparities in first-line adjunct monoclonal antibody treatment of elderly, Medicare-enrolled, metastatic colorectal cancer patients. As monoclonal antibody therapy has been shown clinically to improve

survival, it would be unethical to perform an intervention study to address these questions. Therefore, the method used was strictly observational, using a retrospective cohort quantitative research design. Specifically, individuals meeting the selection criteria were selected from the SEER-Medicare colorectal cancer population as outlined in the sample selection section of this document. Disparities in monoclonal antibody treatment were then assessed across different populations (different SES levels, races, and neighborhood characteristics) using logistic regression.

Rationale. All the research questions can be addressed using numerical values and/or discrete categories for all variables. Therefore, quantitative research was the most appropriate research design. Given the utility of the achieved SEER-Medicare database to answer the research questions, I used only secondary data.

Time and resource constraints. I used data from the NCI's SEER-Medicare database. The SEER-Medicare database is a large population-based database that contains matched medical claim records and demographic information for Medicare beneficiaries who have been diagnosed with cancer (NCI, 2015a). Data from five files within this database for years 2007 to 2012 (PEDSF 2007-2011 only) was accessed for a reasonable fee ($1600). Data access required completion of a dissertation proposal, International Review Board (IRB) approval, approval of an application by SEER-Medicare oversight committee, and dissertation chair approval/completion of a data use agreement (NCI, 2015e; E. Yanisko, personal communication, February 5, 2013).

Use of a retrospective research design and archival data reduced the time constraints associated with data collection. However, time constraints still existed. Upon completion of a final data application (which required several iterations), approval by the SEER-Medicare oversight committee took 4 weeks, and data acquisition took an additional 3 weeks (E. Yanisko, personal communication, February 5, 2013). In all, it took 4 months from submission of the first draft application to receipt of the final data set. All the data obtained from SEER-Medicare were received uniformly coded and there were training courses and support staff available to help with the extraction/compilation process (E. Yanisko, personal communication, February 5, 2013).

Methodology

In this section, I describe the population used for this study and the procedures used to sample that population. Since archival data were used and no new data were collected on study subjects, this section does not contain procedures for new data collection.

Population

The research questions addressed disparities in monoclonal antibody therapy of elderly metastatic colorectal cancer patients. The specific population used for this study was elderly (65+) individuals diagnosed with metastatic colorectal cancer between January 2007 and December 2011, treated with first-line chemotherapy within 6 months of diagnosis. Individuals with a previous cancer history and individuals with end stage renal disease were excluded. Since monoclonal antibodies can be prescribed under

Medicare Part B or D, to limit the impact of insurance coverage differences, the population was limited to individuals with evidence of enrollment in Part B and D Medicare. This population is about 65% of the Medicare population (MedPac, 2013). Additionally, given that claims billed to Medicare HMOs (Medicare Part C; Medicare Advantage) would not be included in the SEER-Medicare database, the study cohort was limited to patient without HMO coverage. Given that metastatic colorectal cancer patients have high mortality rates and the time from cohort selection (which required a chemotherapy claim) to assessment of the dependent variable was only 1 month, to limit unnecessary patient exclusion, continuous Medicare Part B and D enrollment was not required.

Tumor resection (if possible) followed by chemotherapy (5-Fluorouracil, Oxaliplatin, or Irinotecan) is the standard of care for metastatic colorectal cancer (NCI, 2015f). Monoclonal antibody therapy is indicated for first-line metastatic colorectal cancer as an adjunct to chemotherapy. Given this, only individuals who received first-line chemotherapy, as indicated by a chemotherapy claim within 6 months of cancer diagnosis, as in Meyerhardt et al. (2012), were included in the study population. Finally, as previously stated, given that adjunct monoclonal antibody therapy is only approved for metastatic colorectal cancer, the population of interest included only patients with metastatic colorectal cancer.

In the United States, colorectal cancer accounts for 8.6% of all new cancer cases and 8.8% of all cancer deaths (NCI, n.d.a). In 2010, 131,607 people in the United States

were diagnosed with colorectal cancer (CDC, 2013). Colorectal cancer risk and mortality rate increases with age, with the elderly being most affected by the disease (CDC, 2013). In the United States, elderly individuals (65 years+) account for 60% of all new colorectal cancer cases and 70.9% of all colorectal cancer deaths (NCI, n.d.a). In 2010, approximately 79 thousand elderly (65+) individuals were diagnosed with colorectal cancer (CDC, 2013). At diagnosis, the tumor in approximately 20% of patients has already metastasized (Roswell Park Cancer Institute, n.d.a). Therefore, the population of elderly persons diagnosed per year with metastatic colorectal cancer is approximately 15.8 thousand.

In terms of prevalence, there are approximately 1.16 million people living with colorectal cancer in the United States (NCI, n.d.a). Of all individuals diagnosed and being treated for colorectal cancer, it is estimated that 55% of these individuals have metastasized tumors (Roswell Park Cancer Institute, n.d.). Therefore, the population of individuals living with metastasized colorectal cancer is approximately 638,000, with the majority of these individuals over the age of 65.

Sampling and Sampling Procedures

The sampling strategy was carefully designed to eliminate as many confounding variables as possible. The detailed sampling procedures are described in the following section.

Sampling procedures. SEER-Medicare datasets for colon and rectal cancer patients from 2007 to 2011 (PEDSF) and 2007 to 2012 (Outpatient Claims file, DME file,

Carrier Claims File and PDE file) were received from the NCI. These datasets included clinical and demographic data from all SEER district reported colorectal cancer cases (NCI, 2015c). In order to be included in the study, an individual must have met the following inclusion and exclusion criteria:

- First colorectal cancer diagnosis at or after 65 years of age (PEDSF column #1881; Age at Diagnosis).
- First diagnosis of colorectal cancer occurred between January 2007 and December 2011 (PEDSF column #1888; Year of Diagnosis).
- Cancer sequence number (number of primary tumors the individual had been diagnosed with up to and including their colorectal cancer diagnosis) of 00 or 01 indicating that the colorectal cancer tumor was their first primary tumor (PEDSF column #1884).
- Covered by Medicare Part B for at least 1 month during the year of diagnosis (PEDSF columns #548). Value >/= 1 for diagnosis year. Presence of a chemotherapy claim in one of the Part B files (DME, Carrier Claims or Outpatient Claims) was also required for inclusion and demonstrates coverage.
- Covered by Medicare Part D for at least 1 month during the year of diagnosis (PEDSF column #221). Value >/= 1 for diagnosis year.
    - Monoclonal antibody therapy can be covered under Medicare Part B or Medicare Part D. Therefore, this selection criterion is included to help

ensure complete medical records for all study participants and limit the impact of insurance coverage on the disparity measures.

- Not Enrolled in a Medicare managed care plan during the year of diagnosis (Medicare Part C) (PEDSF column #550; HMO Months). Any number other than 00 for both the year of diagnosis and the year of treatment will result in exclusion.
    - Managed care plans enrollees have their claims processed by the managed care entities, not Medicare. Thus, these claims will not be in the SEER-Medicare data.
- Colorectal cancer had undergone metastasis at diagnosis (PEDSF column #1953; CS Mets at Dx). Specifically, individuals with codes 00 (none/no distant metastasis) or 99 (unknown/distant metastasis not stated) will be excluded.
    - Monoclonal antibodies are only approved for use in metastatic colorectal cancer patients.
- Patients that qualified for Medicare due to End Stage Renal Disease were excluded (PEDSF column #43; rsncd1). Individuals with codes of 2 or 3 were excluded.
- Given chemotherapy up to 31 days before or 6 months after of diagnosis as determined by comparing diagnosis date (PEDSF column #1886 = Month of Diagnosis; PEDSF column #1888; Year of Diagnosis) to the first

chemotherapy claim date. Given that no day of the month is listed for diagnosis, the diagnosis day was set to the last day of the diagnosis month to ensure all within 6 months of diagnosis chemotherapy claims were captured. Chemotherapy receipt was defined as receiving any of the following as single agents or in combination. The "J" HCPCS codes are listed in columns #93 in the Carrier Claims data set, column #93 in the DME dataset, or columns #241 in the Outpatient Claims data set (NCI, 2015b). The NDC codes are located in column #409 in the DME dataset (NCI, 2015b). Date of receipt was defined using the Claim From Date located in column #32 of all three datasets (NCI, 2015b). Only the earliest chemotherapy claim and claim date for each individual was recorded.

- 5-fluorouracil (5-FU):
  - HCPCS = J9190
  - NDC Codes = 00703301513, 00703301812, 00703301912, 10139006301, 10139006310, 10139006311, 10139006312, 10139006320, 10139006350, 63323011710, 63323011720, 63323011751, 63323011761, 66758004401, 66758004403
- capecitabine:
  - HCPCS = J8520, J8521
  - NDC Codes = 00004110020, 54868414300, 54868414301, 54868414302, 00004110150, 00004110175, 54868526000,

54868526001, 54868526002, 54868526003, 54868526004, 54868526005, 54868526006, 54868526007, 54868526008, 54868526009

- leucovorin:
  - HCPCS = J0640
  - NDC Codes = 00703279301, 00703279701, 00703514001, 00703514501, 00703514591, 00904231560, 25021081310, 25021081430, 25021081530, 25021081567, 25021081630, 25021081667, 55390000901, 55390005110, 55390005210, 55390005301, 55390005401, 55390081810, 55390082401, 55390082501, 55390082601, 62701090030, 62701090099, 62701090125, 63323071050, 63323071100
- floxuridine:
  - HCPCS = J9200
  - NDC Codes = 55390013501, 63323014507
- oxaliplatin:
  - HCPCS = J9263
  - NDC Codes = 00024059010, 00024059120, 00024059240, 00069006701, 00069007001, 00069007401, 00069101001, 00703398501, 00703398601, 25021021120, 25021021250, 41616017640, 41616017840, 47335017640, 47335017840,

61703036318, 61703036322, 63323017530, 63323017650, 63323065010, 63323065017, 63323065020, 63323065027

- irinotecan
  - HCPCS = J9206
  - NDC Codes = 00009111101, 00009111102, 00009752901, 00009752902, 00009752903, 00009752904, 00009752905, 00143970101, 00143970201, 00703443211, 00703443411, 10019093401, 10019093402, 10019093417, 10019093479, 10518010310, 18111000202, 18111000203, 23155017931, 23155017932, 25021020002, 25021020005, 25021021402, 25021021405, 55390029501, 55390029601, 59762752902, 61703034909, 61703034916, 61703034936, 63323019302, 63323019305, 66758004801, 66758004802

Power Analysis

This section describes the G*Power analysis used to estimate sample size required for each of the three research questions. This section also provides justification for the effect sizes, alpha level, and power level chosen.

Alpha level. I used an alpha level 0.05 for this study. In social sciences, an alpha level of 0.05 is often employed (Web Center for Social Research Methods, 2006). This is because an alpha level of 0.05 represents a compromise between Type I and Type II error. This alpha level allows for a 5% chance that the null hypothesis will be rejected

when it is true (i.e. there is no difference between groups; Type I error) (Web Center for Social Research Methods, 2006). Additionally, being conservative and not setting the alpha level too low reduces the risk of Type II error or failing to reject the null hypothesis when it is false (Web Center for Social Research Methods, 2006). Finally, given that the research questions ask if there is a difference between groups and do not specify directionality, a two tailed analysis was used.

Effect size and power level. In G*Power version 3.1.9.2 (Faul, Erdfelder, Buchner, & Lang, 2009) the effect size is represented as an odds ratio, the average probability of "yes" to the dependent variable (in this case monoclonal antibody treatment), and a specified power level. Although it's never been researched, it is reasonable to hypothesize that the racial, SES, and neighborhood (degree of urbanization) differences in monoclonal antibody treatment will mimic historical differences in other colorectal cancer treatments. This hypothesis is in line with the Fundamental Cause Theory which states that even with changing resources, it is the beneficial social connections that serve to protect health regardless of the resource mechanism (Phelan et al., 2010). In other words, groups that showed chemotherapy, radiation, or surgery treatment disparities might show similar monoclonal antibody treatment disparities.

For SES odds ratio estimation, Lejeune et al. (2010) cites the difference in adjusted treatment rates (surgery, radiation, or chemotherapy) between high and low SES districts as 1.15. However, in Hines et al. (2014), the difference in chemotherapy rates comparing high to low SES districts was greater (OR = 0.83). Given that the degree of

disparity differs between the two studies, to be conservative, 1.15 was used as the odds ratio estimate for SES differences. In terms of racial disparities, Obeidat et al. (2010) determined that Black Americans were less likely than White Americans to receive newer chemotherapy options (OR = 0.641). However, White (2008) found more modest differences between Black Americans and White Americans in terms of chemotherapy receipt (OR = 0.85). Therefore, for this power analysis, a conservative OR of 0.85 was used. For urban/rural disparity estimation, the Hines et al. (2014) article found that chemotherapy rates differed comparing urban to rural districts (OR = 0.84). Therefore, an odds ratio of 0.84 was used for this final power analysis.

In addition to the odds ratios, G*Power also requires the average probability of receiving treatment. In an observational study conducted using 2006 treatment data, Renouf et al. (2011) found that 30.6% of metastatic colorectal cancer patients received adjunct bevacizumab. Therefore, given that this study is assessing 2007 to 2012 treatment data, a conservative probability of 0.306 was used.

Finally, adjunct monoclonal antibody treatment of colorectal patients is hypothesized to vary by race, SES and neighborhood (degree of urbanization). However, it is unlikely that any of these variables fully explain differences in treatment. When the magnitude of the expected outcome is unknown, Cohen (1988) suggests selecting a medium power level of 0.5. This medium power level (0.5) means that 6% of the variance in treatment is explained by the independent variable (Cohen, 1988). This power level was used for the research questions.

Independent variable association with control variables. For all three research questions, there are four control variables planned into the logistic model. An R^2 for the relationship between the primary independent variable and the control variables must be defined for the power analysis. A literature search did not return data on the correlation between the independent variables (race, SES, neighborhood characteristics) and the control variables (gender, age at diagnosis, reason for original Medicare entitlement, and reporting registry) in this elderly metastatic colorectal cancer population. Therefore, for the purpose of this analysis, a value representing (by convention) a medium correlation ($R = 0.3$) was selected. Thus, the R^2 value for the relationship between the independent variable in each research question and the control variables was set at $(0.3)^2 = 0.09$.

Power analysis. This research study has three different research questions and three different independent variables, all tested as a predictors of a binary outcome variable (monoclonal antibody treatment; yes or no) using logistic regression. Each independent variable has a different expected odds ratio. Therefore, three different power analyses were performed.

Utilizing the information above regarding available alpha level, effect size, power level and association with control variables, the software program G*Power was used to estimate required sample size for this logistic regression analysis (Faul, et al., 2009). The inputs and outputs of this analysis are represented in Appendix C.

Estimated available sample size. According to the NCI (2015d), the number of newly diagnosed colorectal cancer patients in the SEER dataset was 12,062, 11,489, and

10,564 in 2007 to 2009 respectively. This is an average of 11,372 new cases per year. In order to be included in this count the patients had to be enrolled in both Medicare Part A and Medicare Part B (NCI, 2015d). Since this research study is limited to patients enrolled in Medicare Part B and D (approximately 65% of the Medicare population), the number of patients meeting the study selection criteria is approximately 7,392 patients per year. This study cohort included individuals diagnosed over five years (2007-2011). Therefore, within the SEER dataset (2007-2011), an estimated 36,960 will be colorectal cancer patients enrolled in Medicare Part B and D. Approximately 20% of these patients were expected to be excluded due to enrollment in Medicare managed care/HMO, leaving approximately 29,568 patients. This sample is further reduced as this study only addresses treatment in patients diagnosed with metastatic colorectal cancer. If an estimated 20% of these patients were diagnosed with metastatic colorectal cancer, the expected available sample size for this study over the years 2007 to 2011 is 5,914. The sample size will be limited once more by the number of patients receiving first-line chemotherapy. If an estimated 50% receive first-line chemotherapy, this would result in a sample size of 2,957.

The actual final sample size was 2,241, slightly lower than estimated. However, the sample size is still significantly above the size required for all three research questions (1029, 764, and 666 respectively) with a power of 0.5.

Procedures for Recruitment, Participation, and Data Collection

As stated above, since archival data were used and no new data were collected on study subjects, this section does not contain procedures new data collection. However, this section provides detailed information on the SEER-Medicare database and how the data were accessed.

Data access. This study used data from the NCI's SEER-Medicare database. The SEER-Medicare database is a large population-based database that contains matched medical claim records and demographic information for Medicare beneficiaries who have been diagnosed with cancer (NCI, 2015a). The process for requesting the data is outlined in detail on the SEER Medicare website (NCI, 2015g). Specifically, for this project, an application form (Appendix A), a SEER-Medicare data use agreement (Appendix B), and proof of Institutional Review Board (IRB) approval were submitted (NCI, 2015g). Upon application approval, the data needed for these research questions for years 2007 to 2012 was accessed for a reasonable fee ($1600). The five data files obtained were the Part D Event (PDE) dataset, the Durable Medical Equipment (DME) dataset, the Carrier Claims dataset, the Patient Entitlement and Diagnosis Summary File (PEDSF) and the Outpatient Claims dataset. All colorectal cancer patient data within these five files from 2007 to 2012 (PEDSF 2007-2011 only) was obtained.

SEER-Medicare sample information. The SEER-Medicare database is a unique research-oriented database resulting from the linkage of the Surveillance, Epidemiology, and End Results (SEER) cancer registries database and the Medicare enrollment and

claims data files (NCI, 2015a; Warren et al., 2002). The cancer diagnosis and demographic information within the SEER-Medicare database comes from the SEER program. The SEER program includes NCI contracted registries in Connecticut, Iowa, New Mexico, Utah, Hawaii, Georgia, Louisiana, New Jersey, Puerto Rico, Kentucky and California (NCI, n.d.a). These registries represent approximately 28% of the U.S. cancer population (NCI, 2013). When the SEER database is linked to the Medicare claims databases, such as in the SEER-Medicare dataset, researchers have access to data that includes incidence, cancer site, stage, histology, demographics, medical procedures, initial treatments, and vital status (NCI, n.d.a). Starting in 2007, SEER began linking the Medicare Part D (PDE) file which allows researchers to perform prescription drug research (NCI, 2013). In total, there are 1.6 million persons with cancer included in the dataset and over 900 publications using the dataset (NCI, 2013).

Instrumentation. Given that archived data were used, no instrumentation was used for the current study. All the data within the SEER-Medicare linked database comes from either cancer registry documents or Medicare claim forms. In the case of the SEER program (cancer registry documents), data collection is both passive and active. After hospitals, clinicians and pathologists collect the data, SEER registry personnel actively perform follow-ups to collect data or passively collect data through state databases (NCI, 2011). Once data is compiled by the registries, they incorporate mortality data from the National Center of Health Statistics, Medicare claims records, and information from the U.S. Census Bureau (NCI, 2011). Subsequently quality checks are implemented to

ensure all the data has been linked appropriately and the data is then de-identified (NCI, 2011).

Given that hospitals, clinicians, and pathologists are responsible for accurate reporting and coding of SEER data, there is the potential for missing or inaccurate data. There are multiple published studies that have assessed the reliability and the validity of different subsets of the SEER-Medicare data. Du et al. (2008) assessed the validity and reliability of the SEER report on breast cancer chemotherapy. In this assessment, the authors reviewed medical charts from 1228 women diagnosed with breast cancer and compared these results to the SEER-Medicare reports (Du et al., 2008). For patients who did not receive chemotherapy, the SEER-Medicare data agreed with the chart review over 99% of the time (Du et al., 2008). For the women listed as receiving chemotherapy in the SEER-Medicare dataset, the authors were able to find strong evidence for chemotherapy in the chart review for 97% of the cases, indicating strong validity (Du et al., 2008). In a separate analysis, Du et al. (2008) list the overall reliability (kappa) of the SEER-Medicare breast cancer chemotherapy data as 0.69 (95% confidence interval = 0.63-0.76).

Mahnken et al. (2008) assessed the completeness of the SEER-Medicare database for Oral and Pharyngeal Cancer. When comparing to the SEER limited-use dataset, the authors found that 6.4% of the incident cases were missing from the larger SEER-Medicare linked database (Mahnken et al., 2008). The completeness of the data did vary

nonsignificantly by race and ethnicity and could be partly explained by differences in Medicare coverage (Mahnken et al., 2008).

Operationalization of constructs. All of the variables used in this research project are available within the PEDSF (Patient Entitlement and Diagnosis Summary File), the Outpatient Claims dataset, the DME (Durable Medical Equipment) dataset, the Carrier Claims dataset or the Part D Event (PDE) dataset (NCI, 2015c). The independent and dependent variables are operationalized below. The control/covariate and selection variables have been described in detail in the previous sections.

- Race: Identified as listed in column #101 of the PEDSF (rac_recb); 1=Caucasian, 2=Black, 3=American Indian/Alaska Native, 4=Chinese, 5=Japanese, 6=Filipino, 7=Hawaiian, 8=Other Asian or Pac. Islander, 9=Unknown, 11=Caucasian, Spanish origin or surname, 12=Other unspecified (NCI, 2015c).
- Socioeconomic Status: The PEDSF Census Tract Poverty Indicator will be used as an estimation of SES. This is column #146 (census_pov_ind) in the PEDSF document and uses data from the 2010 Census (NCI, 2015c).
- Neighborhood characteristics (degree of urbanization): This will be defined using the 2003 Rural/Urban Continuum Codes from the Economic Research Service (ERS), which categories counties on an urban/rural scale based on population size, degree of urbanization and adjacency to a metro or nonmetro area (NCI, 2015c). The PEDSF re-code of these values will be used for this

research project. This is column #91 (urbrur) and identifies county of residence as 1=Big Metro, 2=Metro, 3=Urban, 4=Less Urban, 5=Rural, 9=Unknown (NCI, 2015c).

- Monoclonal antibody therapy (yes/no): Receipt of bevacizumab, cetuximab or panitumumab (as identified by a claim in the PDE, Carrier Claims, or Outpatient dataset) within 1 month of first post-diagnosis chemotherapy claim. Location and codes for each file are described below.
    - PDE dataset: The brand names for the three monoclonal antibodies (Avastin, Erbitux, Vectibix) will be listed in column #90 (BN). In addition, the following NDC codes for the administered drug will be present in column #35 (PROD_SRVC_ID): bevacizumab (50242006001 or 50242006101), cetuximab (66733094823 or 66733095823), panitumumab (55513095401 or 55513095601).
    - Outpatient Claims dataset: Receipt of one of the monoclonal antibodies will be listed as a "J" HCPCS code in column #241 (hcpcs): bevacizumab (J9035), cetuximab (J9055), panitumumab (J9303).
    - Carrier Claims dataset: Receipt of one of the monoclonal antibodies will be listed as an HCPCS code (above) in column #93.

Research Questions and Hypothesizes

RQ1: Within the population of elderly (65 years+) individuals diagnosed with metastatic colorectal cancer from January 2007 to December 2011, are there racial disparities in first line adjunct monoclonal antibody treatment?

H_{01}: There is no significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody treatment based on race.

H_{a1}: There is a significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody treatment based on race.

RQ2: Within the population of elderly (65 years+) individuals diagnosed with metastatic colorectal cancer from January 2007 to December 2011, are there socioeconomic disparities in first line adjunct monoclonal antibody treatment?

H_{02}: There is no significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on area socioeconomic status (as defined by the PEDSF census tract poverty indicator).

H_{a2}: There is a significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on area socioeconomic status (as defined by the PEDSF census tract poverty indictor).

RQ3: Within the population of elderly (65 years+) individuals diagnosed with metastatic colorectal cancer from January 2007 to December 2011, are there neighborhood characteristic (degree of urbanization) disparities in first line adjunct monoclonal antibody treatment?

H_{03}: There is no significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on neighborhood characteristics.

H_{a3}: There is a significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on neighborhood characteristics.

Data Analysis Plan

In this section, I describe, in detail, the analysis plan used for this research project. This includes the software packages that were used for data compilation, management and analysis, how sample selection occurred, and the statistical methods that were employed.

Hardware

The SEER-Medicare datasets and all raw analysis files were stored on a server leased from Hivelocity and accessed remotely over a secure firewall using VPN. The Hivelocity server had a 2 x 1000GB Hard Drive with 2 x 2.26GHz Quad-Core L5520 Nehalem Xeon Processor and 72GB DDR3 Memory. The operating system was CentOS7.

Software

Python 2.7.5, MySQL 5.6.24 Community Server, MySQL Workbench 6.3.3.0 and BASH (BourneAgain SHell) were employed to create a database to allow access to the data, compile needed variables into tables and select the study cohort based on the listed inclusion and exclusion variables. Data for the selected cohort was then imported into IBM SPSS Statistics 21 and SPSS was used for subsequent descriptive and inferential statistical analyses (IBM, n.d.). For this research project, it was necessary to compile, sort and select data from ~100,000 total cases. Therefore, a high capacity, effective data management program such as MySQL and Python with the ability to query, select and split datasets was critical. Additionally, SPSS was capable of performing logistic regression analyses (IBM, n.d.), as was required by this project.

Data Preparation

Prior to analysis, the data needed to be organized and the cohort needed to be selected. A pictorial representation of the methods followed during the cohort selection process can be found in Figure 1 and the step by step details can be found in Appendix D. Additionally, the exact code used for each step is documented in Appendix D.

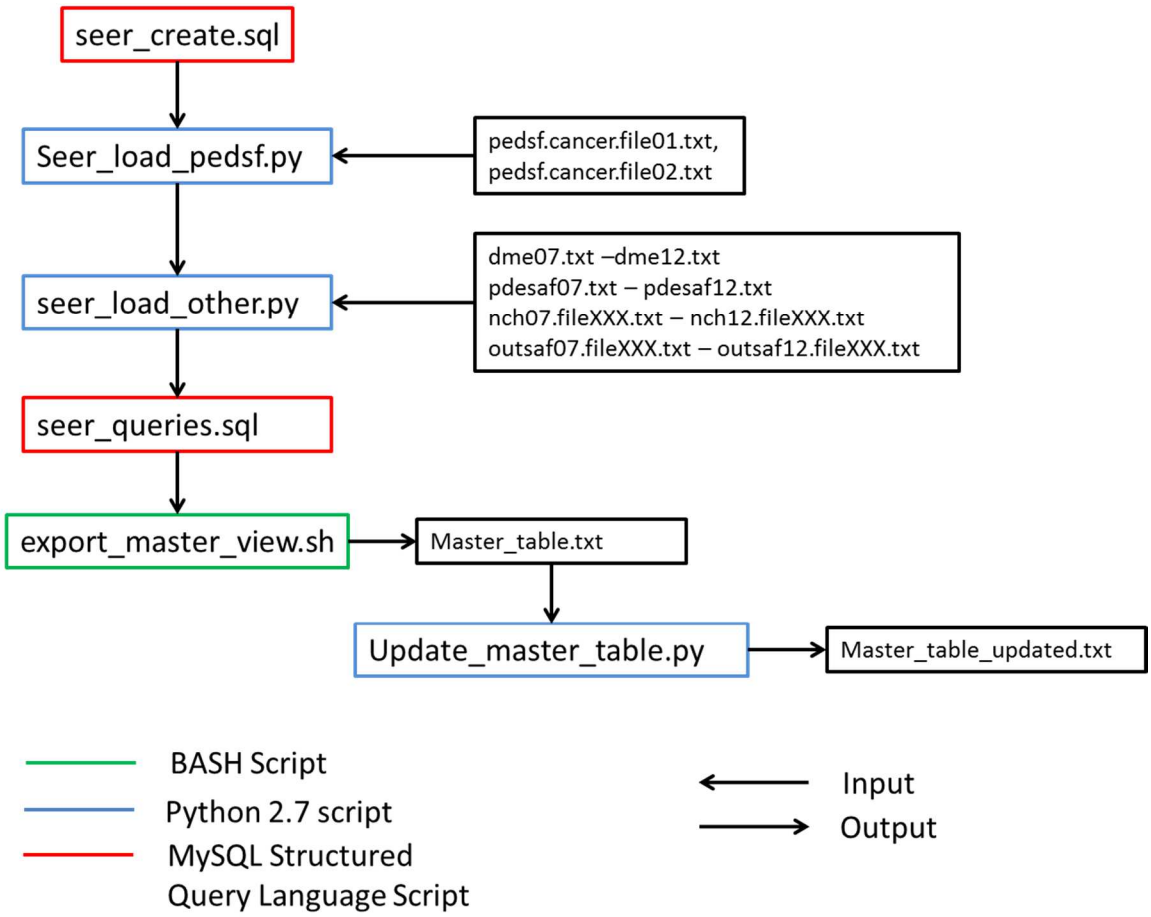

Figure 1. Pictorial representation of the software used and files generated during the cohort selection process.

After the data were compiled and the preliminary cohort was selected using the steps outlined above, the all variables needed for the selected cohort will be imported into SPSS. The variables transferred to SPSS, including several computed variables are documented in Appendix E.

At this point, the cohort in SPSS had already been selected on the following variables:

1. Cases with diagnosis dates from January 2007 to December 2011.
2. Cases with an age at diagnosis of 65 or over.
3. Cases with a sequence number of 00 or 01 (first primary tumor).
4. Cases with cancer metastasis at diagnosis.
5. Evidence (at least 1 month in the diagnosis year) of Medicare Part B and Medicare Part D enrollment.
6. No HMO coverage in the year of diagnosis.
7. Chemotherapy received within 6 months of diagnosis.

Once the data were in SPSS, individuals with Medicare entitlement due to End Stage Renal disease (Reason for original Medicare entitlement codes 2 or 3) were excluded. Additionally, any patient with a chemotherapy claim more than 31 days before diagnosis were excluded. Using the computed column “days” (6a in Appendix D – indicates days from first post diagnosis chemotherapy claim to first monoclonal antibody claim), individuals with a monoclonal antibody within 1 month (31 days) of their

chemotherapy claim date were coded as “yes” in the “Monoclonal antibody” column, all others were coded as “no”.

As a final step in data preparation, codes for each variable were defined (e.g., 1 = male, 2 = female) in the variable view section of SPSS. In addition variable type (categorical or continuous) was defined in the variable view. Once all variables were defined, data analysis (descriptive and inferential) began.

Statistical Tests

This section will describe, in detail, the statistical tests that employed to answer the research questions. Statistical assumptions and interpretation of outputs will also be discussed.

Descriptive statistics. Descriptive statistics in SPSS were used to understand sample distribution. Specifically tables and/or figures were generated to document distribution of the sample in terms of race, gender, age at diagnosis, SES, neighborhood characteristics (degree of urbanization), reporting registry, and reason for original Medicare entitlement over the study period (2007-2011) and by diagnosis year. Additionally, also using SPSS, frequency tables were generated to relate the independent/control variables to the dependent variable as outlined below:

- Frequency of monoclonal antibody therapy by gender.
- Frequency of monoclonal antibody therapy stratified by reporting registry.
- Frequency of monoclonal antibody stratified by reason for original Medicare entitlement (age or disability).

- Frequency of monoclonal antibody therapy stratified by age at diagnosis.
- Frequency of monoclonal antibody therapy by year of diagnosis.
- Frequency of monoclonal antibody therapy by race.
- Frequency of monoclonal antibody therapy across SES ranges (as defined using the Census Tract Poverty Indicator)
- Frequency of monoclonal antibody therapy stratified on neighborhood characteristics (degree of urbanization as defined using the 2003 Rural/Urban Continuum code)

Inferential statistics. Logistic regression is used to model a dichotomous outcome variable (in this case receipt of monoclonal antibody, "yes" or "no") based on other predictor variable(s) (Burns & Burns, 2008). The output of logistic regression predicts which dependent variable group a sample should reside in based on the independent variable(s) (Burns & Burns, 2008). For example, in this study, using a simple ANOVA, it is possible that race would predict (to a certain extent) the chance of monoclonal antibody receipt. However, in this example, in order to determine if the independent variable (race) is independently associated with the dependent variable (monoclonal antibody receipt), the model will need to control for differences in these covariates. In other words, is the association between the independent and dependent variable still significant after adding the covariates to the model? Logistic regression methods will allow this to be accomplished.

There are two main assumptions that must be met in order use logistic regression. First, the dependent variable must be dichotomous (Burns & Burns, 2008). The dependent variable (monoclonal antibody therapy) is dichotomous, meeting this assumption. Second, the categories must be mutually exclusive and exhaustive, because a sample must only be in one group and every sample must be assigned to a group (Burns & Burns, 2008). This assumption is also met with this data set. For all independent, dependent and control variables every sample will fall within one (and only one) of the categories. Burns and Burns (2008) also note that larger samples are often needed for logistic regression and recommend at least 50 cases per predictor. For the analyses described in Table 12, the maximum number of predictor variables will be seven. As noted in the power section above, the sample size (2,241) exceeds this minimum number. Logistic regression does not assume linear relationships between dependent and independent variables, nor does it assume equal variance within groups or normal distribution (Burns & Burns, 2008). Therefore, these conditions do not need to be tested prior to analysis.

There are multiple logistic regression analyses performed on the study data given the different research questions. All logistic regression analyses performed described in Table 11. Covariate analyses to test for independent associations (analyses 4, 5 and 6 in Table 11) were planned to be performed only if a significant association was found between the independent variable and the dependent variable for that research question. For example, if race is found to not associate with monoclonal antibody treatment in

analysis 1, analyses 4 will not be performed as there will be no need to test for variables confounding the relationship (as a relationship does not exist). For analysis 7, a final logistic regression models was generated with any variable (independent or covariate) that was determined to be significantly associated with monoclonal antibody treatment.

Table 11

Logistic Regression Analyses

Analysis #	Dependent Variable	Independent Variable(s)	Covariates	Sample exclusion	Rationale
Analysis 1	Monoclonal Antibody Therapy (yes/no)	Race (9 categories)	None	Patients with Race Unknown or Other/Unspecified will be excluded	Test association between race and monoclonal antibody receipt
Analysis 2	Monoclonal Antibody Therapy (yes/no)	Neighborhood Characteristics (5 categories)	None	Patients with neighborhood characteristics Unknown will be excluded	Test association between neighborhood and monoclonal antibody receipt
Analysis 3	Monoclonal Antibody Therapy (yes/no)	SES; Census Tract Poverty Indicator (4 categories)	None	Patients with Census Tract Poverty Indictor Unknown will be excluded	Test association between area SES and monoclonal antibody receipt
Analysis 4	Monoclonal Antibody Therapy (yes/no)	Race (9 categories)	Age at diagnosis, Gender, Original reason for Medicare entitlement, Reporting registry, Gender	Patients with Race Unknown or Other/Unspecified will be excluded	Test for an independent association between race and monoclonal antibody receipt
Analysis 5	Monoclonal Antibody Therapy (yes/no)	Neighborhood: Rural/ Urban (5 categories)	Age at diagnosis, Gender, Original reason for Medicare entitlement, Reporting registry, Gender	Patients with neighborhood characteristics Unknown will be excluded	Test for an independent association between neighborhood and monoclonal antibody receipt

table continues

Analysis 6	Monoclonal Antibody Therapy (yes/no)	SES; Census Tract Poverty Indicator (4 categories)	Age at diagnosis, Gender, Original reason for Medicare entitlement, Reporting registry, Gender	Patients with Census Tract Poverty Indictor Unknown will be excluded	Test for an independent association between area SES and monoclonal antibody receipt
Analysis 7	Monoclonal Antibody Therapy (yes/no)	Race, Neighborhood Characteristics, SES	Age at diagnosis, Gender, Original reason for Medicare entitlement, Reporting registry, Gender	Samples with Unknown/Unspecified independent values will be excluded	Generate a model of association using all significant variables (independent and covariate).

Covariates and Confounding Variables

Any control variable found to significantly associate with monoclonal antibody treatment during covariate analysis in Chapter 4 was added to the logistic regression analyses to control for confounding. There are four control variables/covariates being controlled for in this research project. The first control variable was gender. Oliver et al. (2013) cites weak correlations between receipt of chemotherapy or surgery and gender. Specifically, women were slightly more likely than men to receive surgery (95.5% vs. 92.2%) and men were slightly more likely than women to receive chemotherapy (38.6% vs. 45.2%) (Oliver et al., 2013). These treatment differences might extend to monoclonal antibody therapy and it is therefore important to control for gender in the statistical model.

The second control variable was age at diagnosis. This study was limited to individuals over 65 years of age. However, even in the over 65 age group, increasing age has been correlated with treatment disparities. Fu et al. (2013) and Kozloff et al. (2010) observed reduced adjunct bevacizumab treatment with increasing patient age. Therefore, it was important to control for age at diagnosis in the model.

The third control variable was reason for original Medicare entitlement (age or disability). Individuals with End Stage Renal Disease (ESRD) or disability + End Stage Renal Disease were excluded. However, underlying disability has the potential to impact treatment prescribed/received and needed to be included in the model.

The final control variable was which registry reported the data. Naishadham et al. (2011) found state to state differences in progress toward reducing colorectal cancer mortality from 1990-2007. Specifically, the southern states showed little to no improvement, whereas the north-eastern states showed improvement often in excess of a 33% reduction in mortality (Naishadham et al., 2011). Therefore, it is possible that geographic region (as indicated by reporting district) could impact treatment received.

Results Interpretation

Descriptive statistics. The planned descriptive statistical analyses were presented in the statistical tests section of this chapter and are reported in frequency tables in Chapter 4. These frequency tables stratify the cohort by each independent/control variable and the dependent variable (monoclonal antibody treatment).

Inferential statistics. The outputs of the logistic regression analyses described in the statistical tests section of this chapter are log odds (b coefficients – slope values), odds ratio (Exp(B)) and p-values. These values are located in the "Variables in the Equation" table in the logistic regression SPSS output (Burns & Burns, 2008). The log odds values ("b") ranges from 0 to infinity and estimates the likelihood of the membership in the target group (monoclonal antibody receipt = yes) versus the other group (monoclonal antibody receipt = no) (Burns & Burns, 2008). For example, if "b" was 0.04 for the independent variable SES (% below poverty level), then for every one unit increase (percentage point) increase in SES, the increase in log odds of receipt of monoclonal antibody receipt increases by 0.04. Negative log odds values mean the relationship is negative (reduces likelihood of monoclonal antibody receipt) instead of positive. Additionally, the Exp(b) has been reported. These are the odds ratios for the predictors (independent variables). The odds ratios indicate the increased (or reduced) chance of monoclonal antibody receipt based on the model of independent variables/covariates. 95% confidence intervals for the odds ratios have also be generated in SPSS. Odds ratios for each research question (race, SES, neighborhood characteristics) were compared with and without covariate inclusion (if necessary) to determine if the independent variable is independently associated with the dependent variable (monoclonal antibody receipt).

A *p*-value (Sig.) for each logistic regression model has also been reported (Burns & Burns, 2008). This value informs whether or not the model was significant for the

independent variable and dependent variable in question. A *p*-value of less than 0.05 has been considered significant.

Finally, two pseudo R Square values (Cox & Snell and Nagelkerke) have been reported for each model. These variables are measures of model fit (values close to 0.0 are a poor fit; values close to 1.0 are a good fit).

Threats to Validity

Threats to External Validity

The external validity of this study could be threatened by population validity or the ability to extrapolate the data to the larger U.S. elderly population. The SEER database includes data from approximately 28% of the U.S. cancer population. However, this sample comes from specific SEER funded cancer registries. These registries are geographically dispersed, located in 13 different states. However, it is possible that the ability to extrapolate the data obtained will be limited. It is unclear whether the data within the SEER database is a true representation of the greater U.S. cancer population. This threat cannot be avoided and will be noted in the limitations section of Chapter 5.

Given that this research is observation (nonexperimental) and used nonsubjective medical record data, many of the other threats to external validity are avoided. For example, ecological validity issues, interaction effect of testing issues, interaction effects of selection biases and the experiment treatment, and reactive effects of experimental arrangements are avoided due to the study design.

Threats to Internal Validity

Internal validity could be threatened by confounding variables. To address this potential issue, research into potential alternative causes of treatment disparities was performed. Historical treatment associations with gender, age at diagnosis, and geographical region (i.e. reporting registry) were found in the literature review. Therefore, the logistic regression analysis included these three variables as covariates to test for independent associations between the independent and dependent variables. In addition, other health issues such as serious disability could influence treatment. Therefore, in addition, the original reason for Medicare entitlement (age or disability) was included as a covariate in the logistic regression model. Individuals who qualified for Medicare due to end stage renal disease (ESRD) and individuals with a previous cancer history were excluded from the study.

The internal validity of this study could also be threatened by maturation. Maturation occurs when changes in the dependent variable occur over time. This study used data over five years of cancer diagnoses and treatments. Therefore, it is possible that the year of diagnosis could impact prescribed treatment. In other words, if fewer individuals diagnosed in 2007 were prescribed a monoclonal antibody compared to diagnosed in 2011, there could be higher levels of disparity in 2007 compared to 2011 (in line with the Diffusions of Innovations Theory). To address this, monoclonal antibody treatment rates have been analyzed by diagnosis year in Chapter 4. If a significant difference exists, Chapter 5 will document a potential for maturation issues in the cohort.

An additional threat to internal validity is sample selection. In the case of this study, all SEER-Medicare colorectal cancer patients were included if they met the sample selection criteria (there was no random selection from the SEER dataset). This reduces selection bias threats. Additionally, the SEER sample is large (represents 28% of U.S. cancer cases) also reducing issues associated with selection.

Experimental mortality is another potential threat to internal validity. This is reduced through sample selection. Since treatment information is reported through the Medicare claims datasets, only samples with evidence of Medicare Part B and Part D enrollment in the year of diagnosis (as indicated by at least 1 month of coverage) were included in the sample. In addition, individuals with any evidence of HMO plan enrollment in the year of diagnosis were excluded as these individuals could have claims not reported in the Medicare data files. Finally, presence of a chemotherapy claim was also required for cohort selection, demonstrating coverage. This should reduce the incidence of missing treatment data after diagnosis. Additionally, the timeframe from cohort selection to measurement of the dependent variable (monoclonal antibody treatment) was only 1 month. This short timeframe should also reduce experimental mortality concerns.

Threats to Construct or Statistical Conclusion Validity

All of the variables used in this research study were explicitly operationalized to reduce threats to construct validity. Many of the other threats to construct validity are reduced due to the observational nature of this study. For example, there should be no

interaction of different treatments, interaction of testing and treatment, evaluation apprehension, or experimenter expectancy issues.

To increase statistical conclusion validity, great effort has been employed to ensure the study was powered appropriately and that the statistical tests have not violated any assumptions. In addition, the statistical analysis plan was explicitly laid out and further analyses were not performed. This will ensure that the type I error rate is not unnecessarily inflated.

Ethical Procedures

Given that this research utilized anonymized medical records and demographic data, human and data ethnic concerns are minimized. However, the efforts to further minimize human and data ethic concerns are described in the following sections.

Ethical treatment of humans. The ethical concerns surrounding this research study are substantially minimized due to the use of anonymized archival data. No new data collections were performed. Therefore, there are no ethical concerns surrounding recruitment materials, data collection, or intervention activities. In addition, zip codes were not requested, minimizing the chance that any data could be accidently unblinded. All of the SEER-Medicare patient files received from the NCI were anonymized and no attempt was made to de-anonymize the data. Finally, an approved IRB application was completed prior to requesting the data.

Ethical treatment of data. All data received from the SEER-Medicare database was pre-anonymized. The raw data were not shared with anyone not associated with this

project (chair, co-chair, statistic support, etc.) and all researchers ensured that attempts to de-anonymize data did not occur. The data were stored securely on a leased Hivelocity server and accessed through a VPN firewall. All the data reported within this dissertation document were aggregate data only; no individual or raw data has been shared publically. There are no conflicts of interest associated with this study.

Agreements to gain access to data. In order to gain access to the SEER-Medicare database, a proposal consisting of a data application, a data use agreement (submitted documents in Appendix A and B) and proof of IRB approval was submitted to the SEER-Medicare liaisons. The Walden University IRB approval (approval #12-11-14-0086341) was utilized to satisfy the IRB requirement. The data use agreement (DUA) was required to be signed by the PhD scientist overseeing the research; in this case, Dr. Raymond Panas (chair). This data application package was approved by the SEER-Medicare (NCI) committee (Appendix F).

Summary

As outlined in the methodology plan above, this study was observational and used archived colorectal cancer patient data available from the NCI's (NCI) SEER-Medicare database. This data within this database contains all information needed to answer the three research questions regarding disparities in monoclonal antibody treatment.

A retrospective cohort study was performed using individuals diagnosed with metastatic colorectal cancer between January 2007 and December 2011. The data were compiled using MySql, Python and SPSS. The outlined sampling procedure limited the

sample to elderly patients eligible for first-line adjunct monoclonal antibody therapy (individuals with metastatic colorectal cancer given first line chemotherapy). Additionally, the sampling procedure aimed to reduce the impact of insurance coverage on claim availability in the dataset by excluding patient covered by managed care organizations/HMOs (as their claims would not be billed through Medicare and would not be represented in the SEER-Medicare dataset).

I present the results in Chapter 4 using descriptive statistics, frequency tables, and logistic regression. Logistic regression is the appropriate inferential statistic due to the dichotomous nature of the dependent variable (monoclonal antibody therapy, "yes" or "no"). Covariate analyses were employed when necessary to determine whether or not the independent variables are independently associated with the dependent variable.

Finally, validity and ethical concerns have been explicitly addressed. Specifically, threats to internal, external, construct, and statistical validity have been addressed and efforts have been taken to minimize threats and maximize the impact of the data. Additionally, human and data ethical concerns have been discussed. The use of archived, anonymized data reduces ethical concerns. However, steps were still taken to protect the data and ensure the anonymized data remains anonymized and secure.

In the following chapter, I describe selection of the study cohort including attrition at each inclusion/exclusion variable. Additionally, I present demographics of the cohort, and descriptive statistics of the independent, control and dependent variables.

Finally, logistic regression modeling is used to test the three research questions surrounding disparities in adjunct monoclonal antibody treatment.

Chapter 4: Results

Introduction

The purpose of this research study is to assess disparities (based on race, SES, and neighborhood characteristics) in adjunct monoclonal antibody treatment of elderly metastatic colorectal cancer patients. In this chapter, I will first outline the selection of the study cohort, present the demographics of the study cohort, and use descriptive statistics to relate the independent variables and the covariates to the dependent variable (monoclonal antibody treatment). Subsequently, using the selected cohort and logistic regression, the three research questions will be tested. The research questions along with null and alternative hypotheses are outlined below:

RQ1: Within the population of elderly (65 years+) individuals diagnosed with metastatic colorectal cancer from January 2007 to December 2011, are there racial disparities in first line adjunct monoclonal antibody treatment?

H_{01}: There is no significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody treatment based on race.

H_{a1}: There is a significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody treatment based on race.

RQ2: Within the population of elderly (65 years+) individuals diagnosed with metastatic colorectal cancer from January 2007 to December 2011, are there socioeconomic disparities in first line adjunct monoclonal antibody treatment?

H_{02}: There is no significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on area socioeconomic status (as defined by the PEDSF census tract poverty indicator).

H_{a2}: There is a significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on area socioeconomic status (as defined by the PEDSF census tract poverty indictor).

RQ3: Within the population of elderly (65 years+) individuals diagnosed with metastatic colorectal cancer from January 2007 to December 2011, are there neighborhood characteristic (degree of urbanization) disparities in first line adjunct monoclonal antibody treatment?

H_{03}: There is no significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on neighborhood characteristics.

H_{a3}: There is a significant association between the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on neighborhood characteristics.

Data Collection

In this research, I used secondary data from the NCI's SEER-Medicare database and thus was strictly retrospective and observational. Therefore, no treatments or interventions were administered and no adverse events were identified. Additionally, all of the data remained anonymized and no attempt was made to identify patients. As requested by the NCI, to conserve anonymity, all data cells with less than 11 counts have been masked.

The SEER-Medicare database is a large population-based database that contains matched medical records and demographic information for Medicare beneficiaries who have been diagnosed with cancer (NCI, 2015a). For this research project, a request was made for five linked SEER-Medicare datasets containing data from colorectal cancer patients for the years 2007 to 2012 (with the exception of the PEDSF with years 2007-2011 only). The first dataset, the PEDSF, is the primary SEER file and contains demographics and information related to the patient's cancer diagnosis (i.e., information on cancer stage, diagnosis date, previous primary tumors; NCI, 2015b). The other four datasets are all Medicare claims datasets. This included the PDE dataset, the Outpatient Claims dataset, the DME dataset, and the Carrier Claims dataset. The PDE dataset includes all drugs prescribed under Medicare Part D and the other three datasets (Outpatient Claims dataset, the DME and the Carrier Claims dataset) contain all Part B claims including diagnosis and procedure codes (NCI, 2015b). These five patients

matched SEER-Medicare datasets contain information on all the variables required to answer the research questions described above.

The data request/application process began in December, 2014 and required several revisions and reviews. The specific concern from the NCI was in regard to data security and storage. The original proposed plan (storage on a home server) and the secondary plan (storage on the Walden server with access via sFTP) were unacceptable to the protocol reviewers. The final approved application included plans to lease a secure server at Hivelocity and access the data via VPN. The raw data files and the analysis files were stored securely on a Hivelocity server throughout the data compilation/analysis process. Only the final summary tables enclosed in this chapter were saved and shared outside the secure server. The entire application process took 4.5 months and data were received and uploaded to the server in April, 2015. Copies of the application and data use agreement can be found in Appendix A and B.

Given that secondary data were used, there were no discrepancies in data collection. However, during the application process, several study design recommendations were received from the NCI. Specifically, the NCI recommended the addition of NDC codes to the DME dataset query, the exclusion of patients with HMO coverage, the exclusion of individuals with a history of other primary tumors, and the exclusion of individuals with end stage renal disease. These changes were all made to the study protocol and Chapter 3 was modified prior to cohort selection/data analysis.

Cohort Selection

The study cohort was selected from a pool of 119,712 patients in the SEER-Medicare database diagnosed with colorectal cancer from January 2007 to December 2012 using MySQL, Python, and SPSS. The methods for this process can be found in Appendix D. A pictorial representation of cohort selection, including the number of eliminated patients at each step in the selection process, is described in Figure 2. All subsequent descriptive and inferential analyses are performed using the 2241 patients in the cohort.

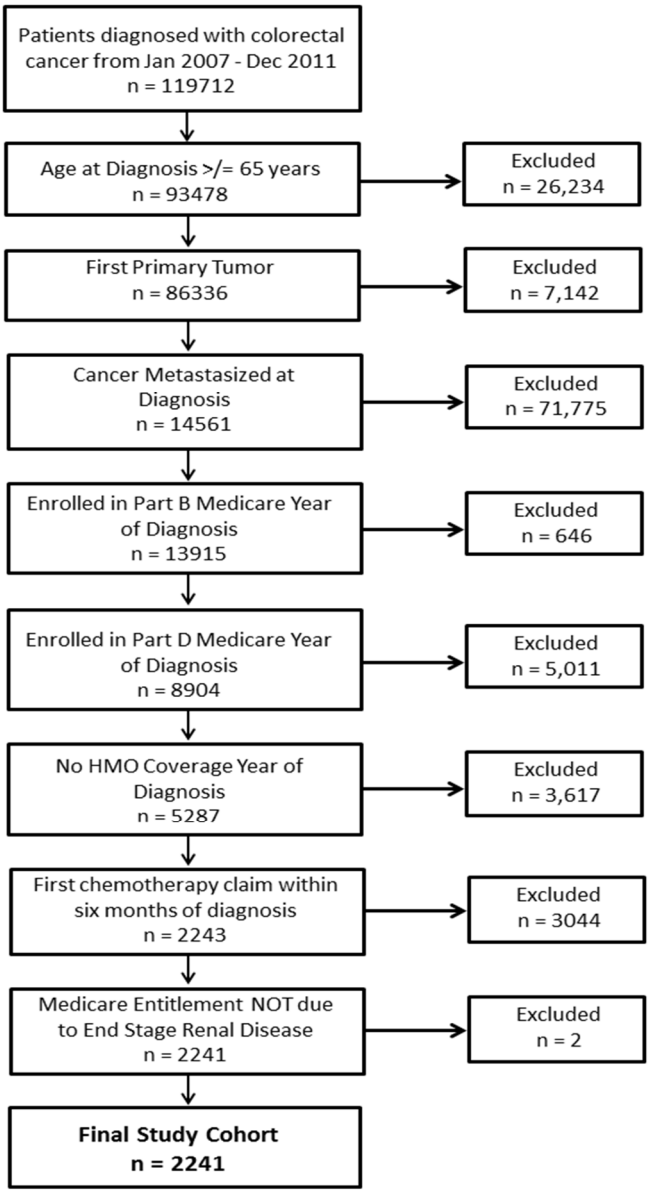

Figure 2. Flow chart showing the selection of the study cohort based on multiple different inclusion and exclusion criteria. The number of individuals excluded at each step in the process is listed.

Demographics of the Study Cohort

The study cohort, as outlined in the cohort selection figure above, includes elderly (65+) patients diagnosed with metastatic colorectal cancer between January 2007 and December 2011 and given chemotherapy within 6 months of diagnosis (as evident by a Medicare claim). Individuals included in the cohort had to have evidence of Part B Medicare coverage (at least 1 month in the year of diagnosis), evidence of Part D Medicare coverage (at least 1 month in the year of diagnosis), but no HMO coverage in diagnosis year. Additionally, they could have no history of a previous primary tumor and could not have been entitled to receive Medicare due to ESRD. The demographics of this cohort are described in detail in Tables 14 through 20.

As described in Table 12, the cohort is 72% white with an additional 8.7% white with Spanish origin or surname. The next most prevalent race is black, which makes up 10.7% of the cohort. For the purpose of this research question, due to low patient numbers in some groups, all Asian and Pacific Islander races (Chinese, Japanese, Hawaiian, Filipino and Other Pacific Islander) were combined into the group called Asian or Pacific Islander. Three other SEER reported race categories, American Indian/Alaska Native, Unknown, and Other Unclassified were excluded from the table and subsequent analyses due to having a total of 11 patients or less over the years 2007 to 2011. Looking at the cohort over diagnosis years 2007 to 2011, the percentage of White Americans and White Americans with Spanish origin or surname remains fairly consistent; however, the percentage of Black Americans in the cohort almost doubles

(from 7.4% to 13.8%) over the timeframe. The incidence of colorectal cancer in Black Americans is known to be slightly higher than in White Americans and Black Americans make up 13.5% of the United States population (NCI, n.d.b). Therefore, the percent of Black Americans in the cohort could be low. However, given that the chemotherapy rates in Black Americans in the general population over the years 2007 to 2011 is unknown and chemotherapy is a selection variable for the cohort, it is difficult to estimate if this racial breakdown is representative of the larger population.

Table 12

Demographics of the Study Cohort by Year - Race

	2007	2008	2009	2010	2011	All Years (2007-2011)
Study cohort (*n*)	407	450	439	488	457	2241
Race (*n*/%)						
White (nonHispanic)	300 (73.7%)	339 (75.3%)	310 (70.6%)	349 (71.5%)	315 (68.9%)	1613 (72%)
Black	30 (7.4%)	38 (8.4%)	51 (11.6%)	57 (11.7%)	63 (13.8%)	239 (10.7%)
Asian or Pacific Islander	33 (8.1%)	38 (8.4%)	40 (9.1%)	31 (6.4%)	38 (8.3%)	180 (8.0%)
White (Spanish origin or surname)	42 (10.3%)	31 (6.9%)	36 (8.2%)	47 (9.6%)	39 (8.5%)	195 (8.7%)

The cohort is distributed equally by gender (48.3% male and 51.7% female) as described in Table 13. Additionally, the gender distribution is fairly constant over the 2007 to 2011 timeframe (with slightly fewer males than females in 2009 and 2010).

Table 13

Demographics of the Study Cohort by Year - Gender

	2007	2008	2009	2010	2011	All Years (2007-2011)
Study cohort (*n*)	407	450	439	488	457	2241
Gender (*n*/%)						
Male	201 (49.4%)	232 (51.6%)	195 (44.4%)	227 (46.5%)	228 (49.9%)	1083 (48.3%)
Female	206 (50.6%)	218 (48.4%)	244 (55.6%)	261 (53.5%)	229 (50.1%)	1158 (51.7%)

As described in Table 14, the average age of diagnosis of the cohort is 74.09 (+/- 6.51) years. This is consistent over the study time frame with the lowest age at diagnosis of 73.53 (+/- 6.07 years) in 2007 and the highest age at diagnosis of 74.48 (+/- 6.61) in 2009. This average age at diagnosis in this cohort is consistent with the average diagnosis age in the general colorectal cancer population given that this study only includes individuals 65 years and older (NCI, n.d.b).

Table 14

Demographics of the Study Cohort by Year – Age at Diagnosis

	2007	2008	2009	2010	2011	All Years (2007-2011)
Age at diagnosis (Average +/- *SD*)	73.53 (+/- 6.07)	74.23 (+/- 6.47)	74.48 (+/- 6.61)	73.82 (+/- 6.77)	74.35 (+/- 6.51)	74.09 (+/- 6.51)

The study cohort is also fairly evenly distributed across the four SES levels as described in Table 15. The highest percentage (28.7%) of individuals in the cohort resided in a census tract with 10% to less than 20% poverty, while the lowest percentage

(21.3%) of individuals in the cohort resided in a census tract with 0% to less than 5% poverty. The distribution is fairly consistent over the 2007 to 2011 timeframe with a slight reduction (25.8% in 2007 to 21.4% in 2011) in the number of individuals residing in the lowest SES areas (20% to 100% poverty).

Table 15

Demographics of the Study Cohort by Year – Socioeconomic Status

	2007	2008	2009	2010	2011	All Years (2007-2011)
Study cohort (*n*)	407	450	439	488	457	2241
Census tract poverty indicator (SES) (*n*/%)						
0% - <5% poverty (highest SES)	85 (20.9%)	104 (23.1%)	103 (23.5%)	99 (20.3%)	87 (19.0%)	478 (21.3%)
5% - <10% poverty	96 (23.6%)	104 (23.1%)	119 (27.1%)	128 (26.2%)	93 (20.4%)	540 (24.1%)
10% - <20% poverty	117 (28.7%)	132 (29.3%)	114 (26.0%)	145 (29.7%)	135 (29.5%)	643 (28.7%)
20% - 100% poverty (lowest SES)	105 (25.8%)	104 (23.1%)	100 (22.8%)	111 (22.7%)	98 (21.4%)	518 (23.1%)
Unknown	*	*	*	*	44 (9.6%)	62 (2.8%)

Note. *Less than 11 patients in group. Value masked for privacy purposes.

The breakdown of reporting registry for the patients in the cohort is described in Table 16. The highest percentage of individuals in the cohort (34.3%) had data reported from California. This is followed by 13.5% in New Jersey and 11.1% in Georgia with the lowest percentage being reported from Utah (1.5%). Frequency of patients within the other reporting registries is described in detail in Table 16. Distribution of the cohort across reporting registry is fairly consistent across the timeframe 2007 to 2011. Given that the SEER-Medicare database relies on data from 13 registries in district geographies,

this geographical breakdown is not representative of the larger U.S. metastatic colorectal cancer population.

Table 16

Demographics of the Study Cohort by Year – Reporting Registry

Reporting registry	2007 (*n*/%)	2008 (*n*/%)	2009 (*n*/%)	2010 (*n*/%)	2011 (*n*/%)	All Years (2007-2011) (*n*/%)
Study cohort (*n*)	407	450	439	488	457	2241
Reporting registry (*n*/%)						
Hawaii	*	*	*	*	*	35 (1.6%)
Iowa	33 (8.1%)	39 (8.7%)	29 (6.6%)	39 (8.0%)	30 (6.6%)	170 (7.6%)
New Mexico	*	*	12 (2.7%)	*	11 (2.4%)	49 (2.2%)
Seattle	20 (4.9%)	27 (6.0%)	15 (3.4%)	23 (4.7%)	21 (4.6%)	106 (4.7%)
Utah	*	*	*	*	*	34 (1.5%)
Kentucky	31 (7.6%)	37 (8.2%)	38 (8.7%)	28 (5.7%)	27 (5.9%)	161 (7.2%)
Louisiana	18 (4.4%)	22 (4.9%)	28 (6.4%)	42 (8.6%)	43 (9.4%)	153 (6.8%)
New Jersey	44 (10.8%)	61 (13.6%)	63 (14.4%)	71 (14.5%)	63 (13.8%)	302 (13.5%)
Georgia	48 (11.8%)	50 (11.1%)	51 (11.6%)	47 (9.6%)	52 (11.4%)	248 (11.1%)
California	150 (36.9%)	147 (32.7%)	149 (33.9%)	174 (35.7%)	149 (32.6%)	769 (34.3%)

Note. *Less than 11 patients in group. Value masked for privacy purposes.

Neighborhood characteristics (degree of urbanization) for patients within the study cohort are described in Table 17. The cohort is primarily from Big Metro areas (52.7%) with the least number of patients from Rural areas (2.5%). This is fairly consistent across the 2007 to 2011 timeframe.

Table 17

Demographics of the Study Cohort by Year – Neighborhood Characteristics

	2007	2008	2009	2010	2011	All Years (2007-2011)
Study Cohort (*n*)	407	450	439	487	457	2240[a]
Neighborhood Characteristics (*n*/%)						
Big Metro	206 (50.6%)	234 (52.0%)	243 (55.4%)	260 (53.4%)	238 (52.1%)	1181 (52.7%)
Metro	124 (30.5%)	121 (26.9%)	117 (26.7%)	139 (28.4%)	131 (28.7%)	632 (28.2%)
Urban	26 (6.4%)	30 (6.7%)	21 (4.8%)	29 (5.9%)	35 (7.7%)	141 (6.3%)
Less Urban	42 (10.3%)	49 (10.9%)	47 (10.7%)	52 (10.7%)	39 (8.5%)	229 (10.2%)
Rural	*	16 (3.6%)	11 (2.5%)	*	14 (3.1%)	57 (2.5%)

Note. *Less than 11 patients in group. Value masked for privacy purposes.
[a]One missing value

Table 18 describes the cohort based on the reason for original Medicare entitlement (why they first gained access to Medicare). As described in the cohort selection section, individuals who gained Medicare entitlement due to ESRD were excluded from this study. Therefore, individuals in the cohort were only entitled due to Age or Disability. Individuals in the cohort were primarily entitled due to age (90.6%). This is fairly consistent over the time frame with slightly fewer (7.1%) begin entitled due to disability in 2009 and slightly more being entitled due to disability (12.1%) in 2010.

Table 18

Demographics of the Study Cohort by Year – Reason for Original Medicare Entitlement

	2007	2008	2009	2010	2011	All Years (2007-2011)
Study Cohort (*n*)	407	450	439	488	457	2241
Reason for Original Medicare Entitlement[a]						
Age	372 (91.4%)	408 (90.7%)	408 (92.9%)	429 (87.9%)	413 (90.4%)	2030 (90.6%)
Disability	35 (8.6%)	42 (9.3%)	31 (7.1%)	59 (12.1%)	44 (9.6%)	211 (9.4%)

[a]Entitlement due to End Stage Renal Disease Excluded from Cohort

Analysis of Covariates

Four covariates were assessed to determine if they need to be added to the logistic regression model to control for confounding. The first covariate assessed was gender (male or female). The second covariate was age at diagnosis. The third covariate was reason for original Medicare entitlement (Age or Disability only; Individuals that qualified for Medicare due to End Stage Renal Disease were excluded from the study cohort). The fourth covariate was reporting registry.

Distribution of these covariates is described in the following tables and figures. In addition, univariate analysis describing the association of each of the covariates with the dependent variable (monoclonal antibody treatment) is presented. Covariates with significant association with the dependent variable will be carried into the logistic regression model to test for independent association of the dependent and independent variables (if an association was found).

Gender

The first covariate to be analyzed for association with monoclonal antibody treatment rate is gender. Table 19 describes the frequency of the adjunct monoclonal antibody receipt based on gender in the cohort. The frequency is fairly equivalent with slightly fewer women (percentage-wise) receiving adjunct monoclonal antibody therapy.

Table 19

Monoclonal Antibody Treatment Rate by Gender

Gender	Patients (*n*)	Received Monoclonal Antibody Treatment (*n* (%))
Male	1083	460 (42.5%)
Female	1158	470 (40.6%)
Total	2241	930 (41.5%)

Table 20 describes the Chi-Square statistic comparing monoclonal antibody treatment rate to gender. Gender is not associated with monoclonal antibody treatment (Pearson's Chi-Square *p*-value = .365). Therefore, gender will not be included in the logistic regression models.

Table 20

Differences in Gender-Based on Monoclonal Antibody Treatment Rate

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.821[a]	1	.365		
Continuity Correction[b]	.745	1	.388		
Likelihood Ratio	.821	1	.365		
Fisher's Exact Test				.368	.194
Linear-by-Linear Association	.821	1	.365		
N of Valid Cases	2241				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 449.44.
b. Computed only for a 2x2 table

Reporting Registry

The second covariate to be analyzed for association with monoclonal antibody treatment is reporting registry. Table 21 describes the frequency of adjunct monoclonal antibody receipt based on reporting registry within the cohort. There are some registries with higher than average rates and some with lower than average rates. For example, Utah, Hawaii, and Seattle have relatively high treatment rates (61.8%, 57.1% and 53.8% respectively) while Connecticut, California, and New Jersey have lower treatment rates (33.1%, 37.7% and 38.4%).

Table 21

Monoclonal Antibody Treatment Rate by Reporting Registry

Reporting Registry	Patients (*n*)	Received Monoclonal Antibody Treatment (*n* (%))
Connecticut	121	40 (33.1%)
Detroit	93	48 (51.6%)
Hawaii	35	20 (57.1%)
Iowa	170	70 (41.2%)
New Mexico	49	23 (46.9%)
Seattle	106	57 (53.8%)
Utah	34	21 (61.8%)
Kentucky	161	69 (42.9%)
Louisiana	153	61 (39.9%)
New Jersey	302	116 (38.4%)
Georgia	248	115 (46.4%)
California	769	290 (37.7%)
Total	2241	930 (41.5%)

Table 22 describes the Chi-Square statistic comparing monoclonal antibody treatment rate to reporting registry. Reporting registry is associated with monoclonal antibody treatment (Pearson's Chi-Square *p*-value = .001). Therefore, reporting registry will be included in the logistic regression models to control for confounding.

Table 22

Differences in Reporting Registry Based on Monoclonal Antibody Treatment Rate

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	32.378[a]	11	.001
Likelihood Ratio	32.138	11	.001
Linear-by-Linear Association	2.470	1	.116
N of Valid Cases	2241		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 14.11.

Medicare Entitlement Characteristics

The third covariate to be analyzed for association with monoclonal antibody treatment is original Medicare entitlement characteristics (age or disability). Table 23 describes the frequency of adjunct monoclonal antibody receipt based on entitlement characteristics. Within the cohort, there are slightly higher rates of monoclonal antibody receipt in patients entitled due to age compared to patients entitled due to disability.

Table 23

Monoclonal Antibody Treatment by Medicare Entitlement Characteristics

Entitlement Characteristics[a]	Patients (*n*)	Received Monoclonal Antibody Treatment (*n* (%))
Age	2030	854 (42.1%)
Disability	211	76 (36.0%)
Total	2141	930 (41.5%)

[a]Patients entitled due to End Stage Renal Disease were excluded from both Cohort

Table 24 describes the Chi-Square statistic comparing monoclonal antibody treatment rate to original Medicare entitlement characteristics. Medicare Entitlement Characteristics are not associated with monoclonal antibody treatment (Pearson's Chi-Square *p*-value = .090). Therefore, entitlement characteristics will not be included in the logistic regression models.

Table 24

Differences in Medicare Entitlement Characteristics Based on Monoclonal Antibody Treatment Status

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.882[a]	1	.090
Likelihood Ratio	2.921	1	.087
Linear-by-Linear Association	2.880	1	.090
N of Valid Cases	2241		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 87.56.

Age at Diagnosis

The final covariate to be analyzed for association with monoclonal antibody treatment rate is age at diagnosis. Table 25 describes the mean age of diagnosis for individuals in the cohort who received and did not receive adjunct monoclonal antibody treatment. The mean age of diagnosis is higher for individuals who did not receive adjunct monoclonal antibody therapy.

Table 25

Monoclonal Antibody Treatment by Age at Diagnosis

Monoclonal Antibody Treatment	N	Mean Age at Diagnosis	Std. Deviation	Std. Error Mean
Yes	930	73.03	6.002	.197
No	1311	74.84	6.742	.186

Figure 3 graphically represents the distribution of age at diagnosis for those who received and did not receive monoclonal antibody therapy. The percentage of individuals who received monoclonal antibody declines with increasing diagnosis age.

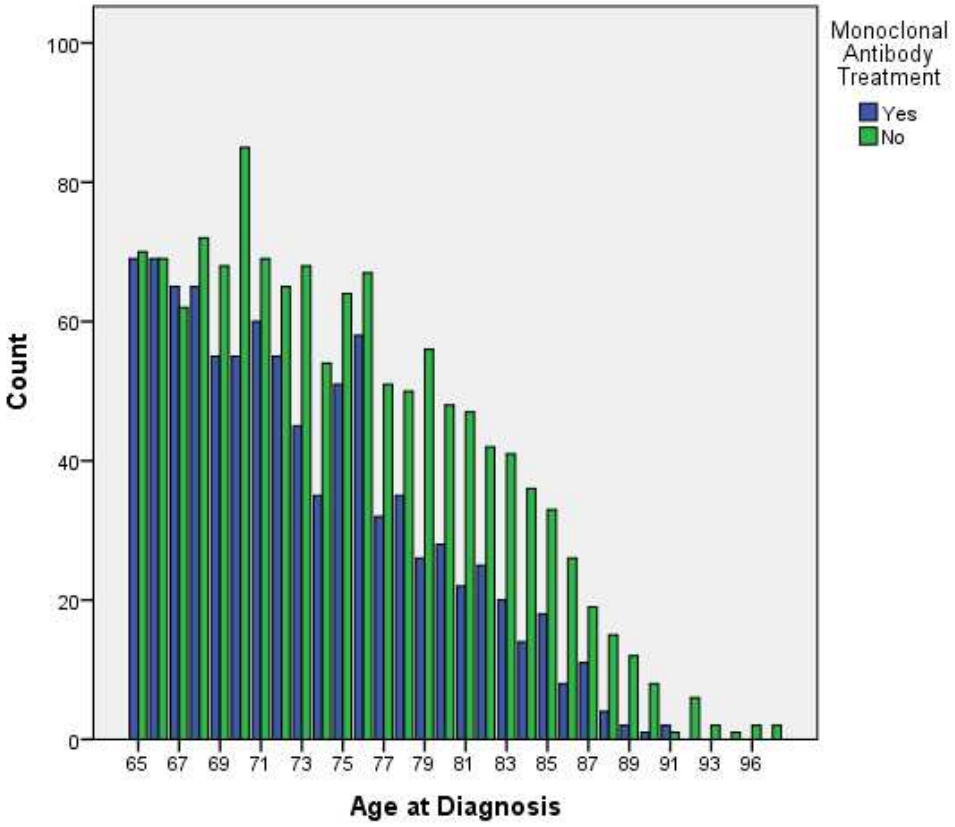

Figure 3. Bar graph showing the distribution of monoclonal antibody treatment by age at diagnosis.

Table 26 presents the results from an independent samples *t*-test comparing monoclonal antibody treatment to age at diagnosis for the cohort. As shown, age at diagnosis is significantly associated with monoclonal antibody treatment (*p*-value = .000). Therefore, age at diagnosis will be included in the logistic regression models to control for confounding.

Table 26

Differences in Average Age at Diagnosis Based on Monoclonal Antibody Treatment Status

		Levene's Test for Equality of Variances		*t*-test for Equality of Means					
		F	Sig.	*t*	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
Age at Diagnosis	Equal variances assumed	18.087	**.000**	-6.525	2239	**.000**	-1.803	.276	(-2.345, -1.261)
	Equal variances not assumed			-6.655	2127.67	**.000**	-1.803	.271	(-2.334, -1.272)

As presented in the sections above, two of the covariates are significantly associated with monoclonal antibody receipt. Therefore, to control for confounding, reporting registry and age at diagnosis will be included in the logistic regression models. Specifically, if a significant association is found between any of the independent variables (race, SES, neighborhood characteristics) and monoclonal antibody treatment, an additional model will be run to include these three covariates as potential confounders. Including the covariates will test for an independent association between the independent and dependent variables.

Descriptive Statistics

The results for the three research questions presented in this paper are outlined

and presented here. This section first uses descriptive statistics to present the distribution of adjunct monoclonal antibody therapy based on race, SES, and neighborhood characteristics. Specifically, this section describes the overall rate of adjunct monoclonal antibody treatment and the relationship between monoclonal antibody treatment rate and the three independent variables. These relationships were also stratified on diagnosis year to test for maturation issues.

Overall Rate of Adjunct Monoclonal Antibody Treatment

As presented in Table 27, the average adjunct monoclonal antibody treatment rate was 41.5%. The monoclonal antibody treatment rate declines slightly from 2007 to 2011. As shown in Table 28, this decrease in treatment rate from 2007 to 2011 approaches significance (Pearson Chi-Square p-value = .067). Given that the difference in monoclonal antibody treatment is nonsignificant over the study timeframe, concerns described in Chapter 3 regarding maturation are minimized.

Table 27

Monoclonal Antibody Receipt by Diagnosis Year

	2007	2008	2009	2010	2011	Total
Total in Study Cohort (*n*)	407	450	439	488	457	2241
Monoclonal Antibody Receipt (*n* (%))	181 (44.5%)	192 (42.7%)	196 (44.6%)	194 (39.8%)	167 (36.5%)	960 (41.5%)

Table 28

Differences in Diagnosis Year Based on Monoclonal Antibody Treatment Status

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8.762[a]	4	.067
Likelihood Ratio	8.806	4	.066
Linear-by-Linear Association	6.510	1	.011
N of Valid Cases	2241		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 168.90.

Monoclonal Antibody Treatment Rate by Race

The first research question in this project addresses disparities in adjunct monoclonal antibody treatment by race. Table 29 presents the monoclonal antibody treatment rates in the cohort by race. The White (nonHispanic) group had the highest rates of monoclonal antibody treatment (43.1%). The other three race groups (Black Americans, Asian or Pacific Islander, and White (Spanish origin or surname)) had monoclonal antibody treatment rates that were 4.8%-7.2% lower.

Table 29

Monoclonal Antibody Treatment Rate by Race

Race	Total in Study Cohort (*n*)	Monoclonal Antibody Treatment Rate (n (%))
White (nonHispanic)	1613	695 (43.1%)
Black	239	88 (36.8%)
Asian or Pacific Islander	180	69 (38.3%)
White (Spanish origin or surname)	195	70 (35.9%)

Monoclonal Antibody Treatment Rate by SES

The second research question in this project addresses disparities in adjunct monoclonal antibody treatment by SES. Table 30 presents the monoclonal antibody treatment rates in the cohort by SES (using the census tract poverty indicator). The rate of monoclonal antibody treatment is fairly consistent across SES groups.

Table 30

Monoclonal Antibody Treatment Rate by SES Group

Census Tract Poverty Indicator (SES)	Total in Study Cohort (*n*)	Monoclonal Antibody Treatment Rate (*n* (%))
0% - <5% (highest SES)	478	199 (41.6%)
5% - <10%	540	218 (40.4%)
10% - <20%	643	273 (42.5%)
20% - 100% (lowest SES)	518	220 (42.5%)

Monoclonal Antibody Treatment Rate by Neighborhood Characteristics

The third research question in this project addresses disparities in adjunct monoclonal antibody treatment based on neighborhood characteristics (degree of urbanization). Table 31 presents the monoclonal antibody treatment rates in the cohort

by Neighborhood (i.e. Big Metro, Metro, Urban, Less Urban or Rural). The rate of monoclonal antibody is lowest in the rural neighborhood group (31.6%).

Table 31

Monoclonal Antibody Treatment Rate by Neighborhood Characteristics

	Total in Study Cohort (*n*)	Monoclonal Antibody Treatment Rate (*n* (%))
Big Metro	1181	484 (41.0%)
Metro	632	261 (41.3%)
Urban	141	67 (47.5%)
Less Urban	229	99 (43.2%)
Rural	57	18 (31.6%)

Statistical Assumptions

Logistic regression will be used in the subsequent sections to test for associations between the independent and dependent variables. There are two main assumptions that must be met in order use logistic regression. First, the dependent variable must be dichotomous (Burns & Burns, 2008). The dependent variable in this case (monoclonal antibody therapy; "yes" or "no") is dichotomous, meeting this assumption. Second, the categories must be mutually exclusive and exhaustive, meaning that a sample must only be in one group and every sample must be assigned to a group (Burns & Burns, 2008). This assumption is also met with this data set. For all independent and dependent variables every sample falls within one (and only one) of the categories. Burns and Burns (2008) also note that larger samples are often needed for logistic regression and recommend at least 50 cases per predictor. When all independent variables and covariates are included, the maximum number of predictor variables will is six (three

significant covariates and three independent variables). Therefore, for these analyses, Burns and Burns would recommend at least 300 cases. The total sample size exceeds this number since the cohort has 2,241 patients. Logistic regression does not assume linear relationships between dependent and independent variables, nor does it assume equal variance within groups or normal distribution (Burns & Burns, 2008). Therefore, these conditions do not need to be tested prior to analysis.

Statistical Analysis of Research Questions

In previous sections in this chapter, cohort demographics were presented, covariates were assessed and descriptive statistics comparing the independent and dependent variables were introduced. This section uses logistic modeling to test for significant and independent associations between the independent and dependent variables. Specifically, this section will answer the three research questions presented in the project and determine whether or not, in these two cohorts, there are disparities in adjunct monoclonal antibody therapy based on race, SES, or neighborhood characteristics.

Monoclonal Antibody Disparities by Race

Prior to analysis, individuals with race American Indian/Alaska Native, unknown, or other unspecified (there were 11 patients or less in the cohort in each of these groups) were excluded. This eliminated a total of 14 patients. Race was coded as a categorical variable for dummy coding and White (nonHispanic) was listed as the reference variable.

Black American was dummy coded as (1), Asian/Pacific Islander was dummy coded as (2) and White (Spanish origin or surname) was dummy coded as (3).

A binary logistic regression model was then executed using SPSS to determine if race predicts monoclonal antibody treatment. The syntax for this analysis can be found in Appendix G. Table 32 shows the result of the logistic regression analysis including a model summary and statistics for the association between race and monoclonal antibody treatment. The pseudo R Square values (0.003 for the Cox & Snell and 0.004 for Nagelkerke) signify a poor model fit for race and monoclonal antibody treatment.

The odds ratios for race versus monoclonal antibody treatment can be found in the Exp(B) column in the Table 32. The odds ratios are all compared back to the reference group, in this case White (nonHispanic). The odds ratio for White (nonHispanic) vs. Black American is 1.299 (95% CI: 0.981, 1.720), the odds ratio for White (nonHispanic) vs. Asian/Pacific Islander vs. is 1.218 (95% CI: 0.888, 1.671), and the odds ratio for White (nonHispanic) vs. White (Spanish origin or surname) is 1.352 (95% CI: 0.993, 1.841). Given that the overall model significance was 0.070 and all confidence intervals overlap 1.0, race is not a significant predictor of monoclonal antibody treatment. However, Black Amerian and White (Spanish origin or surname) do approach significance with a *p*-values of 0.068 and 0.056 respectively.

Table 32

Logistic Regression Analysis – Race and Monoclonal Antibody Treatment

Model Summary								
-2 Log likelihood		Cox & Snell R Square				Nagelkerke R Square		
3013.928[a]					.003			.004
Variables in the Equation								
Race (dummy code)	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
White (nonHispanic) (reference)			7.071	3	**.070**			
Black (1)	.262	.143	3.337	1	.068	1.299	.981	1.720
Asian/Pacific Islander (2)	.197	.161	1.493	1	.222	1.218	.888	1.671
White (Spanish origin or surname) (3)	.302	.158	3.664	1	.056	1.352	.993	1.841
Constant	.278	.050	30.632	1	.000	1.321		

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

The logistic regression model shows that race is not a significant predictor of adjunct monoclonal antibody treatment in this cohort. Therefore, the null hypothesis should be accepted for the first research question relating race to monoclonal antibody treatment. Additionally, given the lack of association between race and monoclonal antibody therapy, no covariates need to be added to the models to control for confounding.

Monoclonal Antibody Disparities by SES

Prior to analysis, individuals with a census tract poverty indicator value of "unknown" were excluded from the cohort (the census tract poverty indicator variable is

the measure of SES for this research project). This eliminated a total of 44 patients. SES was coded as a categorical variable for dummy coding and the 0% to <5% poverty group was listed as the reference variable. The 5% to <10% poverty group was dummy coded as (1), the 10% to <20% poverty group was dummy coded as (2) and the 20% to 100% poverty group was dummy coded as (3).

A binary logistic regression model was then executed using SPSS to determine if SES predicts monoclonal antibody treatment. The syntax for this analysis can be found in Appendix H. Table 33 shows the result of the logistic regression analysis including a model summary and statistics for the association between SES and monoclonal antibody treatment. The pseudo R Square values (0.000 for the Cox & Snell and 0.000 for Nagelkerke) signify no model fit for SES and monoclonal antibody treatment.

The odds ratios for SES level versus monoclonal antibody treatment can be found in the Exp(B) column in the Table 33. The odds ratios are all compared back to the reference group, in this case 0% to <5% poverty. Therefore, the odds ratio for 0% to <5% poverty vs. 5% to <10% poverty is 1.054 (95% CI: 0.820, 1.353), the odds ratio for 0% to <5% poverty vs. 10% to <20% poverty is 0.967 (95% CI: 0.761, 1.229), and the odds ratio for 0% to <5% poverty vs. 20% to 100% poverty is 0.966 (95% CI: 0.751, 1.243). Given that the overall model significance was $p = 0.881$ and all confidence intervals overlap 1.0, SES (as measured using the census tract poverty indicator variable) is not a significant predictor of monoclonal antibody treatment.

Table 33

Logistic Regression Analysis – SES and Monoclonal Antibody Treatment

Model Summary								
-2 Log likelihood			Cox & Snell R Square			Nagelkerke R Square		
2960.649[a]			.000			.000		
Variables in the Equation								
SES (dummy code)	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
0% to <5% poverty (reference)			.668	3	**.881**			
5% to <10% poverty (1)	.052	.128	.167	1	.683	1.054	.820	1.353
10% to <20% poverty (2)	-.034	.122	.077	1	.782	0.967	.761	1.229
20% to 100% poverty (3)	-.034	.128	.072	1	.789	0.966	.751	1.243
Constant	.338	.093	13.262	1	.000	1.402		

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

The logistic regression model shows that SES is not a significant predictor of adjunct monoclonal antibody treatment. Therefore, the null hypothesis should be accepted for the second research question relating SES to monoclonal antibody treatment. Additionally, given the lack of association between SES and monoclonal antibody therapy, no covariates need to be added to the models to control for confounding.

Monoclonal Antibody Disparities by Neighborhood Characteristics

Prior to analysis, one individual with a missing Rural/Urban Continuum code was eliminated. The 2003 Rural/Urban Continuum code groups were coded as categorical variables for dummy coding and the Big Metro group was listed as the reference variable.

Metro was dummy coded as (1), Urban was dummy coded as (2), Less Urban was dummy coded as (3) and Rural was dummy coded as (4).

A binary logistic regression model was then executed using SPSS to determine if Neighborhood Characteristics (degree of urbanization) predicts monoclonal antibody treatment. The syntax for this analysis can be found in Appendix I. Table 34 shows the result of the logistic regression analysis including a model summary and statistics for the association between neighborhood characteristics and monoclonal antibody treatment. The pseudo R Square values (0.002 for the Cox & Snell and 0.003 for Nagelkerke) signify a poor model fit for neighborhood characteristics and monoclonal antibody treatment.

The odds ratios for neighborhood characteristics versus monoclonal antibody treatment can be found in the Exp(B) column in the Table 34. The odds ratios are all compared back to the reference group, in this case Big Metro. Therefore, the odds ratio for Big Metro vs. Metro is 0.987 (95% CI: 0.811, 1.201), the odds ratio for Big Metro vs. Urban is 0.767 (95% CI: 0.540, 1.089), and the odds ratio for Big Metro vs. Less Urban is 0.912 (95% CI: 0.685, 1.214) and the odds ratio for Big Metro vs. Rural is 1.505 (95% CI: 0.851, 2.662). Given that the overall model significance was $p = 0.309$ and all confidence intervals overlap 1.0, neighborhood characteristics are not a significant predictor of monoclonal antibody treatment.

Table 34

Logistic Regression Analysis – Neighborhood Characteristics and Monoclonal Antibody Treatment

Model Summary		
-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
3034.950[a]	.002	.003

Variables in the Equation								
Neighborhood Characteristics (dummy code)	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
Big Metro (reference)			4.792	4	**.309**			
Metro (1)	-.013	.100	.017	1	.897	0.987	.811	1.201
Urban (2)	-.265	.179	2.204	1	.138	0.767	.540	1.089
Less Urban (3)	-.092	.146	.400	1	.527	0.912	.685	1.214
Rural (4)	.408	.291	1.970	1	.160	1.505	.851	2.662
Constant	.365	.059	37.993	1	.000	1.440		

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

The logistic regression model shows that neighborhood characteristics are not a significant predictor of adjunct monoclonal antibody treatment. Therefore, the null hypothesis should be accepted for the third research question relating neighborhood characteristics (degree of urbanization) to monoclonal antibody treatment. Additionally, given the lack of association between neighborhood characteristics and monoclonal antibody therapy, no covariates need to be added to the models to control for confounding.

Final Logistic Regression Model

The three independent variables tested (race, SES, neighborhood characteristics) did not significantly predict the dependent variable group (monoclonal antibody therapy; yes or no) in the logistic regression models above. However, two of the covariates (diagnosis age and reporting registry) did significantly associate with monoclonal antibody receipt in the covariate analysis section of this chapter. Therefore, as a final analysis, a logistic regression model has been generated using these two significant variables to determine how well these variables fit the model and if they are both independently associated with monoclonal antibody treatment.

Prior to analysis, the categorical variables reporting registry was coded as a categorical variable for dummy coding. Binary logistic regression models were executed using SPSS to determine if the model predicts monoclonal antibody treatment. The Syntax for these queries can be found in Appendix J. Table 35 shows the result of the logistic regression analysis including a model summary and statistics for the association between the two variables (age at diagnosis, reporting registry) and monoclonal antibody treatment. The pseudo R Square values (0.032 for the Cox & Snell and 0.042 for Nagelkerke) signify a weak model fit.

Table 35

Logistic Regression Analysis – Final Model With Diagnosis Age and Reporting Registry

Model Summary		
-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
2969.819a	**.032**	**.042**

Variables in the Equation

Variable (dummy code)	B	S.E.	Wald	df	Sig.	Exp (B)	95% CI for EXP(B)	
							Lower	Upper
Reporting Registry = Connecticut			28.953	11	.002			
Reporting Registry = Detroit (1)	-.721	.286	6.331	1	.012	.486	.277	.853
Reporting Registry = Hawaii (2)	-.870	.396	4.816	1	.028	.419	.193	.911
Reporting Registry = Iowa (3)	-.293	.251	1.368	1	.242	.746	.456	1.219
Reporting Registry = New Mexico (4)	-.497	.348	2.038	1	.153	.608	.307	1.204
Reporting Registry = Seattle (5)	-.807	.277	8.466	1	.004	.446	.259	.769
Reporting Registry = Utah (6)	-1.017	.406	6.279	1	.012	.362	.163	.801
Reporting Registry = Kentucky (7)	-.298	.253	1.389	1	.239	.742	.452	1.219
Reporting Registry = Louisiana (8)	-.160	.257	.387	1	.534	.852	.515	1.411
Reporting Registry = New Jersey (9)	-.186	.229	.662	1	.416	.830	.530	1.300
Reporting Registry = Georgia (10)	-.464	.234	3.932	1	.047	.629	.397	.995
Reporting Registry = California (11)	-.124	.209	.350	1	.554	.883	.586	1.332
Age at Diagnosis	.043	.007	38.437	1	.000	1.044	1.030	1.058
Constant	-2.549	.558	20.892	1	.000	.078		

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

Although the pseudo R Square values indicate that this is a weak model, the model does significantly predict monoclonal antibody therapy (0.002 for reporting registry and 0.000 for age at diagnosis). Additionally, given that both variables maintained significance in the model, reporting registry and age at diagnosis are independently associated with monoclonal antibody treatment.

Summary

In this research project, I explored disparities in adjunct monoclonal antibody treatment of elderly metastatic colorectal cancer patients. The three independent variables tested for association with monoclonal antibody treatment were race, SES, and neighborhood characteristics (degree of urbanization). There were also four covariates assessed for association with monoclonal antibody treatment; gender, age at diagnosis, reporting registry and reason for Medicare entitlement. Two of the covariates (age at diagnosis and reporting registry) were found to significantly associate with monoclonal antibody therapy. Logistic regression modeling of the three independent variables (race, SES and neighborhood characteristics) showed that none of the variables significantly predicted who would receive monoclonal antibody therapy. These data support accepting the null hypotheses for all three research questions (H_{01}, H_{02} and H_{03} listed at the beginning of this chapter). A final logistic regression model was then executed with all variables shown to associate with monoclonal antibody therapy (age at diagnosis and reporting registry). These two variables independently associated with monoclonal antibody therapy in the logistic regression model.

In the next chapter I will interpret the finding presented in this chapter. Namely, I will compare the results to what was previously found in the literature and to the theoretical framework this study was based on. Additionally, limitations of the research, recommendations for future research and implications of the study will be presented.

Chapter 5: Discussion, Conclusions, and Recommendations

Introduction

Colorectal cancer survival disparities based on race, SES, and neighborhood characteristics (degree of urbanization) have been documented in the literature. Disparities in treatment regimen have been associated with these survival disparities. However, assessments of disparities in adjunct monoclonal antibody treatment (a newer treatment specifically for metastatic colorectal cancer) were not found in the literature. Therefore, in this study, my aim was to determine whether these historical colorectal cancer treatment disparities, observed with surgery, radiation, and chemotherapy, extended into the newest class of treatments, monoclonal antibody therapies. The specific aim of this quantitative study was to assess racial, SES, or neighborhood characteristic (degree of urbanization) disparities in first-line adjunct monoclonal antibody treatment of elderly metastatic colorectal cancer patients.

To answer these research questions, a retrospective cohort study was performed using secondary data from the NCI's SEER-Medicare database. From this database, 2,241 patient records that met the inclusion and exclusion criteria were selected for the study. The demographics of this cohort including any differences across diagnosis year were described in Chapter 4 (Tables 12-18). Generally, the cohort was stable over diagnosis years 2007 to 2011 with a few exceptions including a linear increase in the number of Black American patients in the cohort from diagnosis year 2007 to diagnosis year 2011 (7.4% of the cohort to 13.8% of the cohort over the timeframe). Adjunct

monoclonal antibody treatment rate did not significantly change over the 2007 to 2011 timeframe although a steady decline was observed that almost reached significance (p = 0.067). The covariates age at diagnosis and reporting registry were significantly associated with adjunct monoclonal antibody therapy, while the covariates gender and reason for original Medicare entitlement (age or disability) were not. Logistic regression was then used to model the independent variables as predictors of the dependent variable. None of the three independent variables (race, SES, and neighborhood characteristics) predicted monoclonal antibody therapy. As a result, the null hypothesis was accepted for all three research questions. However, of note, the race vs. monoclonal antibody model did approach significance (overall model p = 0.070), with black race (p = 0.068) and White, Spanish origin or surname (p = 0.056) being almost significantly different than the reference group, White American race. As a final analysis, a model was generated with the two variables that did significantly associate with monoclonal antibody therapy (age at diagnosis and reporting registry). This final model was significant and both variables were independently associated with adjunct monoclonal antibody therapy.

Interpretation of the Findings

The results of this study were outlined in length in Chapter 4 and summarized above. In this section, I will interpret the findings as they relate to the literature review and the theoretical foundation of this study.

Results as They Relate to the Literature

Multiple studies outlined in the literature review found disparities in either the type of aggressiveness of colorectal cancer treatment based on race, SES, and neighborhood characteristics (degree of urbanization; Bakogeorgos et al., 2013; Obeidat et al., 2010; Serra-Rexach et al., 2012). This study is unique in that I examined a population of newly diagnosed metastatic colorectal cancer patients for disparities in treatment with adjunct monoclonal antibody treatment (a newer therapy specifically for metastatic colorectal cancer). Unlike the vast majority of the literature, in this research, I did not identify significant treatment disparities based on race, SES, or neighborhood characteristics. Therefore, the historically observed treatment disparities may not be relevant to newer therapies (monoclonal antibodies) prescribed to patients with severe disease (metastatic colorectal cancer). The difference could be related to improved access to care or a change in treatment paradigm due specifically to the severity of metastatic colorectal cancer.

Although the null hypothesis was accepted for all three research questions (monoclonal antibody therapy disparities based on race, SES, or neighborhood characteristics), a few literature reported findings were confirmed. First, age at diagnosis was selected as a covariate due to previous literature documentation of an association between diagnosis age and adjunct monoclonal antibody therapy. Specifically, Fu et al. (2013) and Kozloff et al. (2010) observed reduced adjunct bevacizumab treatment with increasing patient age. These results were confirmed in this study. Age at diagnosis was

independently and significantly associated with adjunct monoclonal antibody therapy ($p = 0.000$). In addition, another covariate, reporting registry was significantly associated with monoclonal antibody treatment. SEER reporting registries are geographically dispersed across regions of the United States. Previous research has shown marked differences in colorectal cancer survival improvements across geographical regions of the United States. Specifically, over the timeframe 1990 to 2007, the southern states showed little to no improvement, whereas the north-eastern states showed improvement often in excess of a 33% reduction in mortality (Naishadham et al., 2011). It is possible that regional treatment disparities influence these mortality rates. This was not addressed as a primary question in this study due to the incomplete coverage of U.S. regions by the SEER registry data, but could be explored in future studies.

Theoretical Framing of the Results

The overarching theory for this research project was the fundamental cause theory. According to this theory, as new methods of treating diseases become available, these resources are not distributed equally (Phelan & Link, 2005). Instead they are distributed based on "knowledge, money, power, prestige, and beneficial social connections" (Phelan & Link, 2005, p. 227). Based on this theory and the literature review documenting disparities in older colorectal cancer therapies, disparities in monoclonal antibody treatment based on race, SES, or neighborhood characteristics (degree of urbanization) would have been expected. However, disparities in race, SES, and neighborhood characteristics were not observed in this study. Therefore, the findings

are seemly inconsistent with this theory. However, most of the previous research into colorectal cancer treatment disparities was performed with the general colorectal cancer population. This study is addressing disparities in patients with advanced disease, specifically colorectal cancer that had metastasized at diagnosis. Therefore, it is possible, given the severity of the disease and the high mortality rate, that some of the social barriers to treatment were eliminated. For example, maybe patients with metastatic colorectal cancer are more likely to see a specialist or have a doctor advocate on their behalf. These are questions that could be explored in future studies. Alternatively, given that monoclonal antibodies were approved in 2004 and I assessed patients diagnosed between 2007 and 2011, it is possible that the fundamental cause theory (which would support disparities in new treatments) is no longer applicable. For example, if by 2007, uptake of monoclonal antibodies was already at a maximum level, monoclonal antibodies would no longer be considered a new method of treatment and their use may not be subject to unequal distribution.

Two additional theories were used to extend the fundamental cause theory for the purpose of this research project: the diffusion of innovations theory and the theory of health disparities and medical technology. The diffusion of innovation theory states that widening disparities are influenced by the nature of the new technology and the uptake/diffusion of the technology (Rogers, 2010). Individuals who were treated with monoclonal antibody shortly after approval would be expected to have innovator-like characteristics such as proactive inquiry into treatment regimens that might contribute to

treatment disparities based on social characteristics or demographics. Renouf et al. (2011) found that 5.9% of patients diagnosed with colorectal cancer in 2004 were given the VEGF monoclonal antibody bevacizumab compared to 30.6% of colorectal cancer patients diagnosed in 2006. In the Renouf et al. study, according to the diffusion of innovations theory, higher disparities in treatment would have been expected in 2004 compared to 2006 given that treatment was in an uptake phase. However, monoclonal antibodies were newly approved in 2004, likely accounting for this sharp increase in treatment rate between 2004 and 2006. In this study, the rate of monoclonal antibody treatment was 44.5%, 42.7%, 44.6%, 39.8%, and 36.5% in diagnosis years 2007, 2008, 2009, 2010, and 2011 respectively. Therefore, in the years examined in this study (2007-2011), the overall rate of monoclonal antibody therapy had reached a plateau. Given this, the diffusion of innovations theory, which is most useful in explaining a disparity during the uptake of a technology, is likely not valid for this research study.

As additional support for the fundamental cause theory, the theory of health disparities and medical technology was used (Goldman & Lakdawalla, 2005). This theory states that medical advances are linked to widening disparities because new medical technologies disproportionately benefit the heaviest health care users, richer patients (Goldman & Lakdawalla, 2005). Given that richer more well-educated patients tend to be the heaviest health care users and use more complex treatment regimens, richer patients disproportionately use newer therapies (Goldman & Lakdawalla, 2005). This theory was used to hypothesize that there might be higher monoclonal antibody use by

high SES or nonminority populations. This study did not demonstrate a social inequality in receipt of monoclonal antibody as would be expected based on this theory. However, it is possible, given the severity of the disease and the high mortality rate, that some of the social/economic barriers to treatment in poorer populations were eliminated.

In summary, the lack of disparity in treatment of metastatic colorectal cancer patients with monoclonal antibodies is seemly inconsistent with the three theories used to provide rationale for the research questions. Possible reasons for this include the extreme severity of disease studied and plateau in the uptake of monoclonal antibody therapy during the study period.

Limitations of the Study

This study has multiple limitations. First, due to the nature of the data, this study used a very specific population of elderly individuals enrolled in Medicare Part B and D, but with no HMO (Medicare Part C; Medicare Advantage) plan. This population does differ systematically from the general Medicare population and from the population of Medicare enrollees with an HMO plan. For example, the population sampled for this study has a higher average income rate compared to the general Medicare population (America's Health Insurance Plans, 2015). Additionally, lower frequencies of racial populations including Black Americans, Asians, and individuals with Hispanic origin are observed in the population sampled for this study compared to the general Medicare population (America's Health Insurance Plans, 2015). Therefore, the sample used in this study may not be completely representative of the overall Medicare population.

However, given the potential for missing data with HMO plan enrollees, this limitation was unavoidable.

Second, all individuals selected for inclusion in the study had colorectal cancer that had metastasized at diagnosis. This was necessary as the SEER registry data only identifies metastasis state at diagnosis. There is no way using the SEER registry data to identify patients that progress to metastasis. Likewise, the Medicare claims files cannot provide information about metastasis. Claims data are generated for the purpose of payment (billed items) and not for the purpose of research. Cancer stage and progression cannot be determined using claims data. Therefore, the sample selected for this study is entire composed of individuals who were diagnosed with severe metastasized colorectal cancer tumors. As such, it is unclear how well these results would extrapolate back to individuals diagnosed in an earlier stage that subsequently progress to metastasis.

Third, the study sample was limited to individuals with a cancer sequence number of 00 or 01 indicating that the colorectal cancer tumor was their first tumor. Given the potential implications to treatment for individuals with a history of other cancers, this exclusion criterion was necessary. However, it is unclear how the results of this study would extrapolate to individuals with a history of other primary tumors.

Fourth, given the nature of the data used (Medicare claims) in this study, I used only included Medicare enrolled individuals aged 65+ at diagnosis. As such is it unclear if these results can be extrapolated to the population of metastatic colorectal cancer patients under the age of 65 and with different insurance coverages.

The fifth limitation is the ability to assess survival. In order to assess survival relative to monoclonal antibody therapy, it would have been necessary to assess and control for any comorbidities. This comorbidity analysis would have required research into ICD-9 codes present in the claims files during the pre- and post-diagnosis period for each patient. This would have been a significant undertaking and was determined to be out of scope for this research study. Therefore, death dates were not collected, and survival was not assessed in this study.

Sixth, the population sampled came from the 13 geographically dispersed SEER registries. The ability to extrapolate the results to the larger U.S. wide population of metastatic colorectal cancer patients is limited by the representativeness of the SEER database. This SEER database includes data from approximately 28% of the U.S. cancer population. However, it is unclear whether the data within the SEER database are a true representation of the greater U.S. cancer population.

Seventh, I used archival data from the SEER-Medicare database to address the research questions. Therefore, the study quality is limited by the validity, reliability, and completeness of the SEER-Medicare database. Given that hospitals, clinicians, and pathologists are responsible for accurate reporting and coding of SEER data, there is the potential for missing or inaccurate data. However, there are published studies that have documented good reliability, validity, and completeness of different subsets of the SEER-Medicare data (Du et al., 2008; Mahnken et al., 2008).

Finally, adjunct monoclonal antibody treatment in this study was defined as being given any approved monoclonal antibody (Avastin, Erbitux, and Vectibix) in combination with first line chemotherapy. The different antibodies were not separated out during the analysis. Of importance, Erbitux and Vectibix (both EGFR antibodies) are only indicated for adjunct therapy after testing for a specific KRAS genetic mutation. Thus, there is an additional variable influencing whether or not a person is prescribed an EGFR antibody. In this study, of the individuals that received adjunct monoclonal antibody therapy, there were only 72 individuals that received an EGFR antibody; the remaining 858 individuals received Avastin, a VEGF antibody. Given this, it is unclear how the study results would extrapolate to a study looking at only adjunct EGFR antibody therapy.

In summary, as a result of the database used for this study and the necessary inclusion and exclusion criteria, concerns regarding extrapolation of the study results do exist. These limitations will be taken into account in the recommendations and implications sections below.

Recommendations

Multiple previous reports have documented colorectal cancer treatment disparities based on race, SES, and neighborhood. Given that treatment disparities were not observed in this study, these historical treatment disparities may not be relevant to newer therapies (adjunct monoclonal antibodies) prescribed to patients with severe disease (metastatic colorectal cancer). The difference could be related to improved access to care

or a change in treatment paradigm due specifically to the severity of metastatic colorectal cancer. Future studies aimed at understanding the causes of this observed difference is warranted. Understanding the root cause of the reduced treatment disparities observed in this study could potentially be used to reduce treatment disparities in other cancer populations. For example, given the severity of metastatic colorectal cancer, is it possible that patients are more likely to see a specialist or have a doctor advocate on their behalf, thus reducing the disparities in treatment? These are questions that could be explored in future research studies.

As commented previously, race did approach significance in predicting monoclonal antibody treatment (overall model $p = 0.070$), with Black American race ($p = 0.068$) and White, Spanish origin or surname ($p = 0.056$) being almost significantly different than the reference group, White American race. Therefore, future confirmatory studies to confirm (or reject) the lack of racial disparities in monoclonal antibody treatment of metastatic colorectal cancer patients are needed.

Reporting registry was a significant predictor of adjunct monoclonal antibody therapy. Given the reported U.S. regional differences in colorectal cancer survival, future studies could follow-up on regional disparities in adjunct monoclonal antibody treatment and potentially the influence of treatment on survival. A GIS study to overlay treatment regimen and survival rate of metastatic colorectal cancer patients could be especially powerful.

An additional variable that almost reached significance was reason for original Medicare entitlement (age or disability). Rates of monoclonal antibody treatment were lower in individuals entitled to Medicare due to disability compared to those entitled due to age ($p = 0.090$). Given the marginal association, future studies assessing disparities in monoclonal antibody treatment of metastatic colorectal cancer patients based on disability are warranted.

When assessing the demographics of the study cohort, the number of Black Americans in the cohort increased linearly and doubled from 2007 to 2011. Given that the cohort selection criteria required first line chemotherapy, it is possible that this increase is due to increased first line chemotherapy treatment rates in elderly Black Americans with metastatic colorectal cancer. Exploring this was out of scope for this study. However, it is an interesting finding that could be addressed in future studies.

The overall adjunct monoclonal antibody treatment rate in the cohort declined by 8% from diagnosis year 2007 to diagnosis year 2011. This difference in monoclonal antibody treatment rate based on diagnosis year was almost significant ($p = 0.067$). Given the reported efficacy of adjunct monoclonal antibody treatment, this decline is unexpected. One possible explanation for the low and declining treatment rate is the cost of monoclonal antibody therapies and the cost-effectiveness of adding monoclonal antibody treatment to first-line chemotherapy. Although monoclonal antibodies have been shown to improve survival rates in individuals with metastatic colorectal cancer, a low perceived cost-effectiveness of adding monoclonal antibodies to chemotherapy could

be driving down their use (Lange et al., 2014). Prospective studies designed to illicit the rationale for prescribing behavior or qualitative studies to assess provider/patient feelings about adjunct monoclonal antibody therapy and their benefit versus their high cost might be the next logical step.

Implications

The results demonstrate a potential improvement in historically documented colorectal cancer treatment disparities. Specifically, historical treatment disparities may not be relevant to newer therapies (adjunct monoclonal antibodies) prescribed to patients with severe disease (metastatic colorectal cancer). The difference could be related to improved access to care or a change in treatment paradigm due specifically to the severity of metastatic colorectal cancer. Future studies aimed at understanding the causes of this observed social change (i.e. reduced treatment disparities) are warranted. Understanding the root cause of the reduced treatment disparities observed in this study could potentially be used to reduce treatment disparities in other cancer populations.

Although disparities in treatment based on race, SES, and neighborhood were not observed, I did demonstrate, in this study, differences in adjunct monoclonal antibody treatment rate based on reporting registry and age at diagnosis. The age at diagnosis disparity replicates results found in two other studies and highlights the need for interventions (at the policy or practice level) to improve access to monoclonal antibody therapy for those of advanced age. Additionally, the regional differences in monoclonal antibody treatment rate demonstrated by the reporting registry differences highlight the

need to understand treatment patterns at a regional level across the United States. Remediation in the form of policies to improve access could help increase monoclonal antibody treatment rates in low treatment rate regions.

Finally, the study uncovered an overall decline in the percent of elderly metastatic colorectal cancer patients given first line chemotherapy who received adjunct monoclonal antibody therapy (8% decline from 2007 to 2011). Monoclonal antibodies are efficacious and safe. Therefore, this study highlights the need for additional research into adjunct monoclonal antibody treatment patterns and prescribing habits. Understanding this unexpected decline in adjunct monoclonal antibody treatment rates could illicit social change by improving antibody treatment rates and, in turn, colorectal cancer survival.

Conclusion

Previous research highlighted social disparities in colorectal cancer treatment that, in part, explained disparities in colorectal cancer survival. However, all previous research looked at conventional therapies (chemotherapy, radiation, surgery) in the general colorectal cancer population. This research project was unique in that it explored disparities in adjunct monoclonal antibody treatment within a specific population of elderly metastatic colorectal cancer patients. In this study, race, SES, and neighborhood characteristics were not associated with adjunct monoclonal antibody therapy. The results demonstrate a potential improvement in historically documented colorectal cancer treatment disparities. Specifically, historical treatment disparities may not be relevant to newer therapies prescribed to patients with severe disease. The difference could be

related to improved access to care or a change in treatment paradigm due specifically to the severity of metastatic colorectal cancer. Future studies aimed at understanding the causes of this social change (i.e. reduced treatment disparities) are warranted. Understanding the root cause of the reduced treatment disparities observed in this study could potentially be used to reduce treatment disparities in other cancer populations.

I accepted the null hypothesis for all three research questions. However, two covariates (reporting registry and age at diagnosis) did significantly predict adjunct monoclonal antibody treatment rate. The age at diagnosis disparity in monoclonal antibody treatment of metastatic colorectal cancer patients supports previous findings by Fu et al. (2013) and Kozloff et al. (2010). In addition, two other findings were uncovered. First, the number of Black American individuals in the cohort doubled from 2007 to 2011. The most likely cause, assuming insurance enrollment is stable in the population, is an increase in first-line chemotherapy rate. This is encouraging and should be explored further. Additionally, over the years of this study (2007 - 2011) there was an 8% decline in the overall rate of monoclonal antibody treatment. Monoclonal antibodies have been shown to be safe and efficacious when added to chemotherapy. Given that all the individuals in the cohort received chemotherapy (which has multiple side effects), the low treatment rate and specifically the decline in adjunct monoclonal antibody treatment (which has been shown to be safe and effective) was unexpected. Further research into prescribing habits and who receives adjunct monoclonal antibody therapy is warranted. One potential topic for further research is the perceived cost-effectiveness of monoclonal

antibodies for metastatic colorectal cancer. This could be explored from both provider and the patient angles. If an explanation for the low and declining rate of monoclonal antibody treatment is found, policies or support systems could be put into place to increase adjunct monoclonal treatment rates and possibly increase survival rates for individuals with metastatic colorectal cancer.

References

Aarts, M., Lemmens, V., Louwman, M., Kunst, A., & Coebergh, J. (2010). Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome. *European Journal of Cancer (Oxford, England: 1990), 46*(15), 2681-2695. doi:10.1016/j.ejca.2010.04.026

Abdelwahab, S., Azmy, A., Abdel-Aziz, H., Salim, H., & Mahmoud, A. (2012). Anti-EGFR (cetuximab) combined with irinotecan for treatment of elderly patients with metastatic colorectal cancer (mCRC). *Journal of Cancer Research and Clinical Oncology, 138*(9), 1487-1492. doi:10.1007/s00432-012-1229-8

Ahmed, S., Howel, D., & Debrah, S. (2014). The influence of age on the outcome of treatment of elderly patients with colorectal cancer. *Journal of Geriatric Oncology. Advance Online Publication.* doi: 10.1016/j.jgo.2013.12.005

Allemani, C., Rachet, B., Weir, H., Richardson, L., Lepage, C., Faivre, J.,... Coleman, M. (2013). Colorectal cancer survival in the USA and Europe: a CONCORD high-resolution study. *BMJ Open, 3*(9), e003055. doi:10.1136/bmjopen-2013-003055

America's Health Insurance Plans. (2015). *Medicare advantage demographics report*. Retrieved from http://www.medpac.gov/documents/september-2014-meeting-presentation-medicare-advantage-demographics-and-enrollment-patterns.pdf?sfvrsn=0

Amgen. (2014). *Highlights of prescribing information: Vectibix (panitumumab).* Retrieved from http://pi.amgen.com/united_states/vectibix/vectibix_pi.pdf

Armstrong, K., Weiner, J., Weber, B., & Asch, D. (2003). Early adoption of BRCA1/2 testing: who and why. *Genetics in Medicine, 5*(2), 92-98. doi: 10.1002/ajmg.a.20025

Bakogeorgos, M., Mountzios, G., Kotsantis, G., Economopoulou, P., Fytrakis, N., & Kentepozidis, N. (2013). Chemotherapy compliance, tolerance and efficacy in elderly and non-elderly patients with metastatic colorectal cancer: a single institution comparative study. *Journal of the Balkan Union of Oncology, 18*(3), 629-634. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24065475

Beyer, K., Comstock, S., Seagren, R., & Rushton, G. (2011). Explaining place-based colorectal cancer health disparities: Evidence from a rural context. *Social Science & Medicine, 72*(3), 373-382. doi:10.1016/j.socscimed.2010.09.017

Bristol-Myers Squibb, Eli Lilly and Company. (2013). *Highlights of prescribing information: Erbitux (cetuximab).* Retrieved from http://packageinserts.bms.com/pi/pi_erbitux.pdf

Bruera, G., Cannita, K., Giordano, A., Vicentini, R., Ficorella, C., & Ricevuto, E. (2013). Effectiveness and safety of intensive triplet chemotherapy plus bevacizumab, FIr-B/FOx, in young-elderly metastatic colorectal cancer patients. *Biomed Research International*, *2013*, Article ID 143273. doi:10.1155/2013/143273

Burns, R., & Burns, R. (2008). Chapter 24: Logistic regression. In R. Burns, & R. Burns, *Business research methods and statistics using SPSS*. (pp. 568-588). London, UK: SAGE Publications Ltd.

Centers for Disease Control and Prevention. (2012). *Colorectal cancer rates by race and ethnicity*. Retrieved from http://www.cdc.gov/cancer/colorectal/statistics/race.htm

Centers for Disease Control and Prevention. (2013). *Colorectal cancer risk by age.* Retrieved from http://www.cdc.gov/cancer/colorectal/statistics/age.htm

Chang, V., & Lauderdale, D. (2009). Fundamental cause theory, technological innovation, and health disparities: the case of cholesterol in the era of statins. *Journal of Health and Social Behavior, 50*(3), 245-260. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2885132/

Chien, L., Schootman, M., & Pruitt, S. (2013). The modifying effect of patient location on stage-specific survival following colorectal cancer using geosurvival models. *Cancer Causes & Control, 24*(3), 473-484. doi:10.1007/s10552-012-0134-4

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Lawrence Erlbaum.

Cole, A., Jackson, J., & Doescher, M. (2012). Urban-rural disparities in colorectal cancer screening: Cross-sectional analysis of 1998-2005 data from the Centers for Disease Control's Behavioral Risk Factor Surveillance Study. *Cancer Medicine, 1*(3), 350-356. doi:10.1002/cam4.40

Crawford, N., Jones, C., & Richardson, L. (2010). Understanding racial and ethnic disparities in colorectal cancer screening: Behavioral Risk Factor Surveillance System, 2002 and 2004. *Ethnicity & Disease, 20*(4), 359-365. Retrieved from

http://www.researchgate.net/publication/49821711_Understanding_racial_and_ethnic_disparities_in_colorectal_cancer_screening_Behavioral_Risk_Factor_Surveillance_System_2002_and_2004

Cueto, C., Szeja, S., Wertheim, B., Ong, E., & Tsikitis, V. (2011). Disparities in treatment and survival of white and Native American patients with colorectal cancer: a SEER analysis. *Journal of the American College of Surgeons, 213*(4), 469-474. doi:10.1016/j.jamcollsurg.2011.05.026

Cunningham, D., Lang, I., Marcuello, E., Lorusso, V., Ocvirk, J., Shin, D.,…Saunders, M. (2013). Bevacizumab plus capecitabine versus capecitabine alone in elderly patients with previously untreated metastatic colorectal cancer (AVEX): An open-label, randomised phase 3 trial. *Lancet Oncology, 14*(11), 1077-1085. doi:10.1016/S1470-2045(13)70154-2

Demissie, K., Oluwole, O., Balasubramanian, B., Osinubi, O., August, D., & Rhoads, G. (2004). Racial differences in the treatment of colorectal cancer: a comparison of surgical and radiation therapy between whites and blacks. *Annals of Epidemiology, 14*(3), 215-221. doi:10.1016/j.annepidem.2003.08.002

Du, X., Key, C., Dickie, L., Darling, R., Geraci, J., & Zhang, D. (2006). External validation of Medicare claims for breast cancer chemotherapy compared with medical chart reviews. *Medical Care, 44*(2), 124-131. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2567101/

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. Retrieved from http://link.springer.com/article/10.3758%2FBRM.41.4.1149

Fu, A., Tsai, H., Marshall, J., Freedman, A., & Potosky, A. (2013). Utilization of bevacizumab in US elderly patients with colorectal cancer receiving chemotherapy. *Journal of Oncology Pharmacy Practice*. Advance Online Publication. doi: 10.1177/1078155213507010

Genentech. (2013). *Highlights of prescribing information: Avastin (bevacizumab).* Retrieved from http://www.avastin.com/patient

Goldman, P., & Lakdawalla, N. (2005). A theory of health disparities and medical technology. *Contributions to Economic Analysis and Policy, 4*, 1–30.

Gomez, S., O'Malley, C., Stroup, A., Shema, S., & Satariano, W. (2007). Longitudinal, population-based study of racial/ethnic differences in colorectal cancer survival: impact of neighborhood socioeconomic status, treatment and comorbidity. *BMC Cancer,* 7193. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2228311/

Haas, J., Brawarsky, P., Iyer, A., Fitzmaurice, G., Neville, B., & Earle, C. (2011). Association of area sociodemographic characteristics and capacity for treatment with disparities in colorectal cancer care and mortality. *Cancer, 117*(18), 4267-4276. doi:10.1002/cncr.26034

Hao, Y. Y., Landrine, H. H., Jemal, A. A., Ward, K. C., Bayakly, A. R., L., ... Ward, E. M. (2011). Race, neighbourhood characteristics and disparities in chemotherapy for colorectal cancer. *Journal of Epidemiology & Community Health, 65*(3), 211-217. doi:10.1136/jech.2009.096008

Hassan, M., Arthurs, Z., Sohn, V., & Steele, S. (2009). Race does not impact colorectal cancer treatment or outcomes with equal access. *American Journal Of Surgery, 197*(4), 485-490. doi:10.1016/j.amjsurg.2008.01.017

Henry, K., Niu, X., & Boscoe, F. (2009). Geographic disparities in colorectal cancer survival. *International Journal Of Health Geographics, 848*. doi:10.1186/1476-072X-8-48

Hines, R., & Markossian, T. (2012). Differences in late-stage diagnosis, treatment, and colorectal cancer-related death between rural and urban African Americans and whites in Georgia. *The Journal Of Rural Health: Official Journal Of The American Rural Health Association And The National Rural Health Care Association, 28*(3), 296-305. doi:10.1111/j.1748-0361.2011.00390.x

Hines, R., Markossian, T., Johnson, A., Dong, F., & Bayakly, R. (2014). Geographic residency status and census tract socioeconomic status as determinants of colorectal cancer outcomes. *American Journal Of Public Health*. Advance Online Publication. doi:10.2105/AJPH.2013.301572

IBM. (n.d.). *SPSS software*. Retrieved from http://www-01.ibm.com/software/analytics/spss/

Jafri, N., Gould, M., El-Serag, H., Duan, Z., & Davila, J. (2013). Incidence and survival of colorectal cancer among hispanics in the United States: a population-based study. *Digestive Diseases & Sciences, 58*(7), 2052-2060. doi:10.1007/s10620-012-2454-3

Jehn, C., Böning, L., Kröning, H., Possinger, K., & Lüftner, D. (2012). Cetuximab-based therapy in elderly comorbid patients with metastatic colorectal cancer. *British Journal Of Cancer, 106*(2), 274-278. doi:10.1038/bjc.2011.554

Kahi, C., Imperiale, T., Juliar, B., & Rex, D. (2009). Effect of screening colonoscopy on colorectal cancer incidence and mortality. *Clinical Gastroenterology And Hepatology, 7*(7), 770-775. doi:10.1016/j.cgh.2008.12.030

Kelsall, H., Baglietto, L., Muller, D., Haydon, A., English, D., & Giles, G. (2009). The effect of socioeconomic status on survival from colorectal cancer in the Melbourne Collaborative Cohort Study. *Social Science & Medicine, 68*(2), 290-297. doi:10.1016/j.socscimed.2008.09.070

Kentucky Cancer Registry. (n.d.). *Census Poverty Indicator*. Retrieved from http://www.kcr.uky.edu/informatics/public-projects/census-poverty-indicator

Koroukian, S., Xu, F., Bakaki, P., Diaz-Insua, M., Towe, T., & Owusu, C. (2010). Comorbidities, functional limitations, and geriatric syndromes in relation to treatment and survival patterns among elders with colorectal cancer. *Journals Of Gerontology Series A: Biological Sciences & Medical Sciences, 65*(3), 322-329. doi:10.1093/gerona/glp180

Kozloff, M., Berlin, J., Flynn, P., Kabbinavar, F., Ashby, M., Dong, W., ... Grothey, A. (2010). Clinical outcomes in elderly patients with metastatic colorectal cancer receiving bevacizumab and chemotherapy: results from the BRiTE observational cohort study. *Oncology, 78*(5-6), 329-339. doi:10.1159/000320222

Laerd Statistics. (n.d.). *Two-way ANOVA in SPSS*. Retrieved from https://statistics.laerd.com/spss-tutorials/two-way-anova-using-spss-statistics.php

Laiyemo, A., Doubeni, C., Pinsky, P., Doria-Rose, V., Bresalier, R., Lamerato, L., ... Berg, C. (2010). Race and colorectal cancer disparities: health-care utilization vs different cancer susceptibilities. *Journal Of The National Cancer Institute, 102*(8), 538-546. doi:10.1093/jnci/djq068

Lam, V., Lu, A., Kouzminova, N., & Lin, A. (2013). Characteristics of colorectal cancer survival in an urban county hospital. *Journal Of Gastrointestinal Cancer, 44*(1), 68-72. doi:10.1007/s12029-012-9446-6

Lang, K., Korn, J., Lee, D., Lines, L., Earle, C., & Menzin, J. (2009). Factors associated with improved survival among older colorectal cancer patients in the US: a population-based analysis. *BMC Cancer, 9227*. doi:10.1186/1471-2407-9-227

Lange, A., Prezler, A., Frank, M., Kirstein, M., Vogel, A., & von der Schulenburg, J. (2014). A systematic review of cost-effectiveness of monoclonal antibodies for metastatic colorectal cancer. *European Journal of Cancer, 50*(1), 40-49. doi:10.1016/j.ejca.2013.08.008

Lansdorp-Vogelaar, I., Kuntz, K., Knudsen, A., van Ballegooijen, M., Zauber, A., & Jemal, A. (2012). Contribution of screening and survival differences to racial disparities in colorectal cancer rates. *Cancer Epidemiology, Biomarkers & Prevention, 21*(5), 728-736. doi:10.1158/1055-9965.EPI-12-0023

Le, H., Ziogas, A., Lipkin, S., & Zell, J. (2008). Effects of socioeconomic status and treatment disparities in colorectal cancer survival. *Cancer Epidemiology, Biomarkers & Prevention, 17*(8), 1950-1962. doi:10.1158/1055-9965.EPI-07-2774

Lejeune, C., Sassi, F., Ellis, L., Godward, S., Mak, V., Day, M., & Rachet, B. (2010). Socio-economic disparities in access to treatment and their impact on colorectal cancer survival. *International Journal Of Epidemiology, 39*(3), 710-717. doi:10.1093/ije/dyq048

Lian, M., Schootman, M., Doubeni, C. A., Park, Y., Major, J. M., A., & ... Schatzkin, A. (2011). Geographic variation in colorectal cancer survival and the role of small-area socioeconomic deprivation: A multilevel survival analysis of the NIH-AARP diet and health study cohort. *American Journal Of Epidemiology, 174*(7), 828-838. doi:10.1093/aje/kwr162

Link, B., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior, 35*, 80-94. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7560851

Mahnken, J., Keighley, J., Cumming, C., Girod, D., & Mayo, M. (2008). Evaluating the completeness of the SEER-Medicare linked database for oral and pharyngeal cancer. *Journal of Registry Management, 35*(4), 145-148. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2950706/

Makowsky, M., Guirguis, L., Hughes, C., Sadowski, C., & Yuksel, N. (2013). Factors influencing pharmacists' adoption of prescribing: qualitative application of the diffusion of innovations theory. *Implementation Science, 8109*. doi:10.1186/1748-5908-8-109

McKibbin, T., Frei, C., Greene, R., Kwan, P., Simon, J., & Koeller, J. (2008). Disparities in the use of chemotherapy and monoclonal antibody therapy for elderly advanced colorectal cancer patients in the community oncology setting. *Oncologist, 13*(8), 876-885. doi:10.1634/theoncologist.2008-0061

MedPac. (2013). *Report to the congress: Medicare payment policy.* Retrieved from http://www.medpac.gov/chapters/Mar13_Ch15.pdf

Meyerhardt, J., Li, L., Sanoff, H., Carpenter, W., & Schrag, D. (2012). Effectiveness of bevacizumab with first-line combination chemotherapy for Medicare patients with stage IV colorectal cancer. *Journal of Clinical Oncology, 30*(6), 608-615. doi: 10.1200/JCO.2011.38.9650

Mitry, E. E., Rollot, F. F., Jooste, V. V., Guiu, B. B., Lepage, C. C., Ghiringhelli, F. F., & ... Bouvier, A. M. (2013). Improvement in survival of metastatic colorectal

cancer: Are the benefits of clinical trials reproduced in population-based studies?. *European Journal of Cancer, 49*(13), 2919-2925. doi:10.1016/j.ejca.2013.04.001

Murphy, G., Devesa, S., Cross, A., Inskip, P., McGlynn, K., & Cook, M. (2011). Sex disparities in colorectal cancer incidence by anatomic subsite, race and age. *International Journal of Cancer. Journal International Du Cancer, 128*(7), 1668-1675. doi:10.1002/ijc.25481

Naeim, A., Ward, P., Wang, H., Dichmann, R., Liem, A., Chan, D., & ... Hecht, J. (2013). A phase II trial of frontline capecitabine and bevacizumab in poor performance status and/or elderly patients with metastatic colorectal cancer. *Journal Of Geriatric Oncology, 4*(4), 302-309. doi:10.1016/j.jgo.2013.05.001

Naishadham, D., Lansdorp-Vogelaar, I., Siegel, R., Cokkinides, V., & Jemal, A. (2011). State disparities in colorectal cancer mortality patterns in the United States. *Cancer Epidemiology, Biomarkers & Prevention, Cosponsored By The American Society Of Preventive Oncology, 20*(7), 1296-1302. doi:10.1158/1055-9965.EPI-11-0250

National Cancer Institute. (2011). *Data flow in NCI's SEER registries.* Retrieved from http://seer.cancer.gov/about/factsheets/SEER_Data_Flow_.pdf

National Cancer Institute. (2013). *SEER Medicare data*. Retrieved from http://appliedresearch.cancer.gov/seermedicare/overview/seermed_fact_sheet.pd

National Cancer Institute. (2015a). *SEER-Medicare: Brief description of the SEER-Medicare database*. Retrieved from http://appliedresearch.cancer.gov/seermedicare/overview/

National Cancer Institute. (2015b). *SEER-Medicare: Medicare claims files*. Retrieved from http://healthcaredelivery.cancer.gov/seermedicare/medicare/claims.html

National Cancer Institute. (2015c). *SEER-Medicare: SEER program & data*. Retrieved from http://appliedresearch.cancer.gov/seermedicare/aboutdata/program.html

National Cancer Institute. (2015d). *SEER-Medicare: Number of cases for selected cancers appearing in the data.* Retrieved from http://appliedresearch.cancer.gov/seermedicare/aboutdata/cases.html

National Cancer Institute. (2015e). *SEER-Medicare: Cost of acquiring data.* Retrieved from http://appliedresearch.cancer.gov/seermedicare/obtain/cost.html

National Cancer Institute. (2015f). *Treatment option overview for colon cancer.* Retrieved from http://www.cancer.gov/cancertopics/pdq/treatment/colon/HealthProfessional/page4

National Cancer Institute. (2015g). *SEER-Medicare: Required documents & instructions for submitting requests.* Retrieved from http://appliedresearch.cancer.gov/seermedicare/obtain/requests.html

National Cancer Institute. (n.d.a). *About the SEER registries.* Retrieved from http://seer.cancer.gov/registries/

National Cancer Institute. (n.d.b). *SEER stat fact sheets: Colon and rectum cancer*. Retrieved from http://seer.cancer.gov/statfacts/html/colorect.html

Obeidat, N., Pradel, F., Zuckerman, I., Trovato, J., Palumbo, F., DeLisle, S., & Mullins, C. (2010). Racial/ethnic and age disparities in chemotherapy selection for colorectal cancer. *American Journal of Managed Care, 16*(7), 515-522. Retrieved from http://www.ajmc.com/journals/issue/2010/2010-07-vol16-n07/AJMC_10jul_Obeidat_515to522/

Oliphant, R., Nicholson, G., Horgan, P., Molloy, R., McMillan, D., & Morrison, D. (2013a). Contribution of surgical specialization to improved colorectal cancer survival. *The British Journal of Surgery, 100*(10), 1388-1395. doi:10.1002/bjs.9227

Oliphant, R., Nicholson, G., Horgan, P., Molloy, R., McMillan, D., & Morrison, D. (2013b). Deprivation and colorectal cancer surgery: longer-term survival inequalities are due to differential postoperative mortality between socioeconomic groups. *Annals of Surgical Oncology, 20*(7), 2132-2139. doi:10.1245/s10434-013-2959-9

Oliver, J. S., Martin, M. Y., Richardson, L., Kim, Y., & Pisu, M. (2013). Gender differences in colon cancer treatment. *Journal of Women's Health, 22*(4), 344-351. doi:10.1089/jwh.2012.3988

Österlund, P., Soveri, L., Isoniemi, H., Poussa, T., Alanko, T., & Bono, P. (2011). Hypertension and overall survival in metastatic colorectal cancer patients treated

with bevacizumab-containing chemotherapy. *British Journal Of Cancer, 104*(4), 599-604. doi:10.1038/bjc.2011.2

Phelan, J., & Link, B. (2005). Controlling disease and creating disparities: a fundamental cause perspective. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 60*, 227-33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16251587

Phelan, J., Link, B., & Tehranifar, P. (2010). Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. *Journal Of Health And Social Behavior, 51*(Suppl), S28-S40. doi:10.1177/0022146510383498

Price, T., Zannino, D., Wilson, K., Simes, R., Cassidy, J., Van Hazel, G., & ... Tebbutt, N. (2012). Bevacizumab is equally effective and no more toxic in elderly patients with advanced colorectal cancer: a subgroup analysis from the AGITG MAX trial: an international randomised controlled trial of Capecitabine, Bevacizumab and Mitomycin C. *Annals Of Oncology, 23*(6), 1531-1536. doi:10.1093/annonc/mdr488

Rane, P., Madhavan, S., Sambamoorthi, U., Kalidindi, S., Kurian, S. & Pan, X. (2014). Treatment and survival patterns among elderly Medicare beneficiaries with colorectal cancer: A comparative analysis between a rural state cancer registry and national data. *ISPOR 19th Annual International Meeting Reseach Abstracts, 17*(3), A97. doi:10.1016/j.jval.2014.03.563

Renouf, D., Lim, H., Speers, C., Villa, D., Gill, S., Blanke, C., ... Kennecke, H. (2011). Survival of metastatic colorectal cancer in the bevacizumab era: a population-based analysis. *Clinical Colorectal Cancer, 10*(2), 97-101. doi:10.1016/j.clcc.2011.03.004

Robbins, A., Siegel, R., & Jemal, A. (2012). Racial disparities in stage-specific colorectal cancer mortality rates from 1985 to 2008. *Journal of Clinical Oncology, 30*(4), 401-405. Retrieved from jco.ascopubs.org/content/30/4/401.full.pdf

Rogers, E. (1962). *Diffusion of Innovations*. New York, NY: Free Press.

Rogers, E. (2010). *Diffusion of innovations* (4th ed.). New York, NY: The Free Press.

Rosenberg, D. (1997). *Trend analysis and interpretation: Key concepts and methods for maternal and child health professionals.* Retrieved from http://mchb.hrsa.gov/publications/pdfs/trendanaylsis.pdf

Roswell Park Cancer Institute. (n.d.). *Metastatic cancer: If your cancer has spread.* Retrieved from https://www.roswellpark.org/cancer/colorectal/treatment/metastatic-cancer

Saldana-Ruiz, N., P., Rubin, M. S., Colen, C. G., & Link, B. G. (2013). Fundamental causes of colorectal cancer mortality in the United States: Understanding the importance of socioeconomic status in creating inequality in mortality. *American Journal of Public Health, 103*(1), 99-104. doi:10.2105/AJPH.2012.300743

Schlichting, J., Soliman, A., Schairer, C., Schottenfeld, D., & Merajver, S. (2012). Inflammatory and non-inflammatory breast cancer survival by socioeconomic

position in the Surveillance, Epidemiology, and End Results database, 1990-2008. *Breast Cancer Research and Treatment, 134*(3), 1257-1268. doi:10.1007/s10549-012-2133-2

Scott, A., Wolchok, J., & Old, L. (2012). Antibody therapy of cancer. *Nature Reviews. Cancer, 12*(4), 278-287. doi:10.1038/nrc3236

Serra-Rexach, J. A., Jimenez, A. B., GarcÃ-a-Alhambra, M. A., Pla, R., VidÃ¡n, M., RodrÃ-guez, P.,... MartÃ-n, M. (2012). Differences in the therapeutic approach to colorectal cancer in young and elderly patients. *Oncologist, 17*(10), 1277-1285. doi:10.1634/theoncologist.2012-0060

Simpson, D., Martínez, M., Gupta, S., Hattangadi-Gluth, J., Mell, L., Heestand, G.,... Murphy, J. (2013). Racial disparity in consultation, treatment, and the impact on survival in metastatic colorectal cancer. *Journal of the National Cancer Institute, 105*(23), 1814-1820. doi:10.1093/jnci/djt318

Sineshaw, H., Robbins, A., & Jemal, A. (2014). Disparities in survival improvement for metastatic colorectal cancer by race/ethnicity and age in the United States. *Cancer Causes & Control. 25*(4), 419-423. doi:10.1007/s10552-014-0344-z

Singal, A., Lin, Y., Kuo, Y., Riall, T., & Goodwin, J. (2013). Primary care physicians and disparities in colorectal cancer screening in the elderly. *Health Services Research, 48*(1), 95-113. doi:10.1111/j.1475-6773.2012.01433.x

Steinbrecher, A., Fish, K., Clarke, C., West, D., Gomez, S., & Cheng, I. (2012). Examining the association between socioeconomic status and invasive colorectal

cancer incidence and mortality in California. *Cancer Epidemiology, Biomarkers & Prevention, 21*(10), 1814-1822. doi:10.1158/1055-9965.EPI-12-0659

Stimpson, J. P., Pagán, J. A., & Chen, L. (2012). Reducing racial and ethnic disparities in colorectal cancer screening is likely to require more than access to care. *Health Affairs, 31*(12), 2747-2754. doi:10.1377/hlthaff.2011.1290

Tan, C., & Du, X. (2012). KRAS mutation testing in metastatic colorectal cancer. *World Journal of Gastroenterology, 18*(37), 5171-5180. doi:10.3748/wjg.v18.i37.5171

Tol, J., & Punt, C. (2010). Monoclonal antibodies in the treatment of metastatic colorectal cancer: a review. *Clinical Therapeutics, 32*(3), 437-453. doi:10.1016/j.clinthera.2010.03.012

Tsai, W., Hsieh, P., Yeh, C., Chiang, J., Tang, R., Chen, J.,... Wang, J. (2011). Long-term survival benefits of adjuvant chemotherapy by decreasing incidence of tumor recurrence without delaying relapse in stage III colorectal cancer. *International Journal of Colorectal Disease, 26*(10), 1329-1338. doi:10.1007/s00384-011-1214-8

Vrdoljak, E., Omrčen, T., Boban, M., & Hrabar, A. (2011). Phase II study of bevacizumab in combination with capecitabine as first-line treatment in elderly patients with metastatic colorectal cancer. *Anti-Cancer Drugs, 22*(2), 191-197. doi:10.1097/CAD.0b013e3283417f3e

Wallace, K., Hill, E., Lewin, D., Williamson, G., Oppenheimer, S., Ford, M.,... Alberg, A. (2013). Racial disparities in advanced-stage colorectal cancer survival. *Cancer Causes & Control, 24*(3), 463-471. doi:10.1007/s10552-012-0133-5

Wan, N., Zhan, F., Lu, Y., & Tiefenbacher, J. (2012). Access to healthcare and disparities in colorectal cancer survival in Texas. *Health & Place, 18*(2), 321-329. Retrieved from http://www.researchgate.net/publication/51835922_Access_to_healthcare_and_disparities_in_colorectal_cancer_survival_in_Texas

Warren, J., Klabunde, C., Schrag, D., Bach, P., Riley, G. (2002). Overview of the SEER-Medicare data: Content, research applications, and generalizability to the United States elderly population. *Med Care, 40*(8 Suppl), 3-18. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12187163

Wassira, L., Pinheiro, P., Symanowski, J., & Hansen, A. (2013). Racial-ethnic colorectal cancer survival disparities in the mountain west region: the case of Blacks compared to Whites. *Ethnicity & Disease, 23*(1), 103-109. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23495630

Web Center for Social Research Methods. (2006). *Statistical power.* Retrieved from http://www.socialresearchmethods.net/kb/power.php

White, A., Liu, C., Xia, R., Burau, K., Cormier, J., Chan, W., & Du, X. (2008). Racial disparities and treatment trends in a large cohort of elderly African Americans and

Caucasians with colorectal cancer, 1991 to 2002. *Cancer, 113*(12), 3400-3409. doi:10.1002/cncr.23924

White, A., Vernon, S., Franzini, L., & Du, X. (2010). Racial disparities in colorectal cancer survival: to what extent are racial disparities explained by differences in treatment, tumor characteristics, or hospital characteristics? *Cancer, 116*(19), 4622-4631. doi:10.1002/cncr.25395

Wildes, T., Kallogjeri, D., Powers, B., Vlahiotis, A., Mutch, M., Spitznagel, E.,...& Piccirillo, J. (2010). The benefit of adjuvant chemotherapy in elderly patients with Stage III colorectal cancer is independent of age and comorbidity. *Journal of Geriatric Oncology, 1*(2), 48-56. doi:10.1016/j.jgo.2010.08.003

Wilkins, T., Gillies, R., Harbuck, S., Garren, J., Looney, S., & Schade, R. (2012). Racial disparities and barriers to colorectal cancer screening in rural areas. *Journal of the American Board of Family Medicine, 25*(3), 308-317. doi: 10.3122/jabfm.2012.03.100307

Wright, G., Barlow, W., Green, P., Baldwin, L., & Taplin, S. (2007). Differences among the elderly in the treatment costs of colorectal cancer: how important is race? *Medical Care, 45*(5), 420-430. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17446828

Yan, B., Noone, A., Yee, C., Banerjee, M., Schwartz, K., & Simon, M. (2009). Racial differences in colorectal cancer survival in the Detroit Metropolitan area. *Cancer (0008543X), 115*(16), 3791-3800. doi:10.1002/cncr.24408

Zullig, L., Carpenter, W., Provenzale, D., Weinberger, M., Reeve, B., & Jackson, G. (2013). Examining potential colorectal cancer care disparities in the Veterans Affairs health care system. *Journal of Clinical Oncology, 31*(28), 3579-3584. doi:10.1200/JCO.2013.50.4753

Appendix A: SEER-Medicare Application Form

APPLICATION FOR SEER-MEDICARE DATA
(Please complete all information in this form)

I. Contact information

Project Title: Disparities in Monoclonal Antibody Treatment of Elderly Metastatic Colorectal Cancer Patients

Principal Investigator: (*students or fellows may NOT be listed as the PI*)

Name:	Dr. Raymond Panas
Institution:	Walden University
Address:	[redacted]
City, State Zip	
Email:	
Phone	

Student / fellow contact:

Name:	Krista Schroeder, M.H.S.
Institution:	Walden University
Address:	[redacted]
City, State Zip	
Email:	
Phone:	

II. Project Description:

A. Title: Disparities in Monoclonal Antibody Treatment of Elderly Metastatic Colorectal Cancer Patients

B. Brief overview of your project (one or two sentences): This project will determine whether or not there are disparities in first line adjunct monoclonal antibody treatment of elderly metastatic colorectal cancer patients.

C. Cancer sites being requested (e.g. Lung): Colorectal Cancer

D. Description of the Project (between 1-5 pages):

Research Questions:

The three research questions (RQ1-RQ3) that will be addressed by this research project are listed below. In addition, the null and alternative hypotheses have been stated (Ho1-Ho3 and Ha1-Ha3 respectively).

RQ1: Within the population of elderly (65 years+) individuals diagnosed with metastatic colorectal cancer from January 2007 to December 2010, are there racial disparities in first line adjunct monoclonal antibody treatment?

Ho1: There are no significant differences in the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody treatment based on race.

Ha1: There are significant differences in the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody treatment based on race.

RQ2: Within the population of elderly (65 years+) individuals diagnosed with metastatic colorectal cancer from January 2007 to December 2010, are there socioeconomic disparities in first line adjunct monoclonal antibody treatment?

Ho2: There are no significant differences in the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on area socioeconomic status (as defined by % census area under the poverty level or % census area with less than a HS education).

Ha2: There are significant differences in the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on area socioeconomic status (as defined by % census area under the poverty level or % census area with less than a HS education).

RQ3: Within the population of elderly (65 years+) individuals diagnosed with metastatic colorectal cancer from January 2007 to December 2010, are there neighborhood characteristic (urban versus rural) disparities in first line adjunct monoclonal antibody treatment?

Ho3: There are no significant differences in the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on neighborhood characteristics.

Ha3: There are significant differences in the percentage of elderly (65 years+) metastatic colorectal cancer patients who receive first line adjunct monoclonal antibody therapy based on neighborhood characteristics.

Description of the Study Subject/Sites/Phases to be Included in the Analysis:

The specific sample utilized for this analysis will be elderly (65+) Medicare enrolled individuals diagnosed with metastatic colorectal cancer between January, 2007 and December 2010, treated with first-line chemotherapy within 6 months of diagnosis and enrolled in Medicare Part B and D.

Rationale: The research questions address disparities in monoclonal antibody therapy of elderly colorectal cancer patients. Monoclonal antibody therapy is indicated only as an adjunct to chemotherapy (not as a stand-alone drug). Therefore, only individuals who received first-line chemotherapy will be included in this analysis. Additionally, given that monoclonal antibody therapy is only approved for metastatic colorectal cancer, only patients with metastatic colorectal cancer will be included in this analysis.

Cohort Selection Criteria:

In order to be included in the study, an individual must meet the following 6 inclusion criteria:

1) First colon or rectal cancer diagnosis at or after 65 years of age (PEDSF column #2457; Age at Diagnosis).
2) First diagnosis of colorectal cancer occurred between January 2007 and December 2010 (PEDSF column #2451; Year of Diagnosis).
3) Covered by Medicare Part B for 12 months during both the year of diagnosis and the year of treatment (PEDSF columns #1268).
4) Covered by Medicare Part D for 12 months during both the year of diagnosis and the year of treatment (PEDSF column #823).
5) Therapy initiated between 2007 and 2011 (PEDSF column #2493; Year Therapy Started).
6) Given chemotherapy within 6 months of diagnosis as determined by comparing diagnosis date (PEDSF column #2449 = Month of Diagnosis; PEDSF column #2451; Year of Diagnosis) to the chemotherapy claim date. Chemotherapy receipt will be defined as receiving any of the following: 5-fluorouracil (5-FU, J9190), capecitabine (J8520, J8521), leucovorin (J0640), floxuridine (J9200), oxaliplatin (J9263), or irinotecan (J9206). These "J" HCPCS codes are listed in columns #93 in the NCH data set; columns #241 in the Outpatient Claims data set; or column #93 in the DME data set (National Cancer Institute, 2013b).

Patients will be excluded from the study if their colorectal cancer had not undergone metastasis at diagnosis; as monoclonal antibodies are only approved for metastatic colorectal cancer. PEDSF column #2553 (CS Mets at Dx) defines whether or not the cancer had metastasized at diagnosis. Individuals with codes 00 (no distant metastasis) or 99 (distant metastasis unknown/not documented) will be excluded from the study.

Independent Variables:

There are three research questions and four total independent variables included in this research project. All of the independent variables are available from the PEDSF (Patient Entitlement and Diagnosis Summary File). The PEDSF is linked by a ten digit patient ID (column #1 in the PEDSF document; patient_id) to all other SEER-Medicare files utilized in this research project.

Race is the first independent variable. Race will be identified as listed in column #663 of the PEDSF (rac_recb); 1=Caucasian, 2=Black, 3=American Indian/Alaska Native, 4=Chinese, 5=Japanese, 6=Filipino, 7=Hawaiian, 8=Other Asian or Pac. Islander, 9=Unknown, 11=Caucasian, Spanish origin or surname, 12=Other unspecified. Socioeconomic status (SES) is the second independent variable and will be defined as in an article by Schlichting, Soliman, Schairer, Schottenfeld and Merajver (2012). Specifically, percent of census tract residents living below the poverty level and percent of census track residents aged 25+ with less than a high school (HS) education will be used as estimates of SES. These variables can be found in column #438 (ctpov00) and column #396 (ctnon00) in the PEDSF document. Neighborhood (rural vs. urban) is the third independent variable. Column #91 (urbrur) in the PEDSF document identifies county of residence as 1=Big Metro, 2=Metro, 3=Urban, 4=Less Urban, 5=Rural, 9=Unknown.

Covariates:

Multiple covariates will be added to the logistic regression analysis to control for confounding. The first covariate utilized will be gender. This is column #41 (m_sex) in the PEDSF file; 1=Male, 2=Female. The second covariate utilized will be year of birth. This is column #37 in the PEDSF file. The third covariate will be reason for original Medicare entitlement. This is column #43 in the PEDSF file; 0=Age, 1=Disability, 2=End Stage Renal Disease, 3=Disability/End Stage Renal Disease. The final covariate will be which registry reported the data. This is indicated by the first 2 digits in the patient ID (PEDSF column #1).

Outcome Variable:

The outcome (dependent) variable for the three research question is receipt of first-line (concurrent chemotherapy and within 6 months of diagnosis) monoclonal antibody therapy. This outcome variable is dichotomized as "yes" or "no".

In the PDE dataset, the brand name for the monoclonal antibody (Avastin, Erbitux, or Vectibix) will be listed in column #90 (BN). In addition the following NDC11 codes for the administered drug will be present in column #35 (PROD_SRVC_ID) of the PDE dataset:

- Avastin: 50242-0060-01
- Erbitux: 66733-0948-23
- Vectibix: 55513-0954-01

In the NCH and Outpatient Claims data sets receipt of one of the monoclonal antibodies will be listed as a "J" HCPCS code (column #93 in the NCH data set and column #241 in the Outpatient Claims data set). The DME file does not include these treatment codes and will not be used for this variable.

- Bevacizumab: J9035
- Cetuximab: J9055
- Panitumumab: J9303

List of requested files and how they will be used:

Colorectal cancer patient data from the PDE, NCH, DME, PEDSF and Outpatient Claims files for the years 2007-2011 are being requested. Information on the variables that will be obtained from each of the four files and how they will be used are outlined below. The variables extracted from the five files will be used to either select the data or as independent, dependent or control (covariate) variables.

1) Part D Event (PDE) dataset (2007-2011):
    a. Column #90 (BN) – presence of the brand name Avastin, Erbitux, or Vectibix (dependent variable)
    b. Column #35 (PROD_SRVC_ID) – presence of treatment codes 50242-0060-01, 66733-0948-23, or 55513-0954-01 (dependent variable)
2) NCH Carrier Claims dataset (2007-2011):
    a. Columns #93 – presence of treatment codes J9035, J9055, or J9303 (dependent variable)
    b. Columns #93 – presence of treatment codes J9190, J8520, J8521, J0640, J9200, J9263, or J9206 (selection variable)
3) DME (Durable Medical Equipment) dataset (2007-2011):
    a. Column #93 – presence of treatment codes J9190, J8520, J8521, J0640, J9200, J9263, J9206 (selection variable)
    b. NOTE: The DME equipment file will not be used for the dependent variable (monoclonal antibody treatment) as the DME summary file does not include patients with these treatment codes.
4) Patient Entitlement and Diagnosis Summary File (2007-2011)
    a. Column #41 (m_sex) – gender identifier (covariate)
    b. Column #37 (year of birth) – age (covariate)
    c. Column #43 – reason for original Medicare entitlement (covariate)
    d. Column #1 (first two digits) – reporting registry (covariate)
    e. Column #663 (rac_recb) – race (independent variable)
    f. Column #438 (ctpov00) – SES measure #1 (independent variable)
    g. Column #396 (ctnon00) – SES measure #2 (independent variable)
    h. Column #91 (urbrur) – neighborhood characteristics (independent variable)
    i. Column #2553 (CS Mets at Dx) – Metastasis at diagnosis (selection variable)
    j. Column #2457 – Age at Diagnosis (selection variable)
    k. Column #2451 – Year of Diagnosis (selection variable)
    l. Columns #1268 – Medicare Part B enrollment (selection variable)
    m. Column #823 – Medicare Part D enrollment (selection variable)
    n. Column #2493 – Year Therapy Started (selection variable)
    o. Column #2449 – Month of Diagnosis (selection variable)
5) Outpatient Claims dataset (2007-2011).
    a. Columns #241 – presence of treatment codes J9035, J9055, or J9303 (dependent variable)
    b. Columns #241 – presence of treatment codes J9190, J8520, J8521, J0640, J9200, J9263, or J9206 (selection variable)

The 5% population (non-cancer and/or other cancer) is NOT being requested.

Description of the Personnel Involved:

This data request supports the dissertation research of Krista Schroeder, M.H.S. Krista has ten years of experience as a pre-clinical research biologist specializing in the

discovery and development of large molecule therapies. Upon completion of this research project, Krista will have a PhD in Public Health-Epidemiology from Walden University.

The chair of Krista Schroeder's dissertation committee and the Principle Investigator (PI) for this research project is Dr. Raymond Panas. Dr. Panas is a contributing member of the faculty as Walden University. Dr. Panas will provide direct oversight and review of the student's work. He will also work with the student to address questions or concerns regarding the dissertation process, data collection, data analysis, and development of the final dissertation manuscript.

Timeline for Completion:

Upon receipt of the datasets, Krista Schroeder under the direct oversight of Dr. Panas will compile and analyze the data as described briefly in this application and in detail in her approved dissertation proposal. The current timeline for completion of this research project and her dissertation is September, 2015. She also plans to publish the results in a journal by year's end 2015.

References:

Schlichting, J., Soliman, A., Schairer, C., Schottenfeld, D., & Merajver, S. (2012). Inflammatory and non-inflammatory breast cancer survival by socioeconomic position in the Surveillance, Epidemiology, and End Results database, 1990-2008. Breast Cancer Research And Treatment, 134(3), 1257-1268. doi:10.1007/s10549-012-2133-2

Column numbers and variable descriptors were located using the following two National Cancer Institute web pages:

National Cancer Institute. (2013a). SEER-Medicare: SEER program & data. Retrieved from http://appliedresearch.cancer.gov/seermedicare/aboutdata/program.html

National Cancer Institute. (2013b). SEER-Medicare: Medicare claims files. Retrieved from http://appliedresearch.cancer.gov/seermedicare/medicare/claims.html#medpar

E. Data Storage and Protection:
Please be aware that Cloud Storage of data does NOT meet privacy rules and will not be approved for storing SEER-Medicare data.

A monthly lease will be established on a Hivelocity server located at their Tampa, FL data center (8010 Woodland Center Boulevard #700, Tampa, FL 33614). Storage of data at this data center is both HIPAA and PCI compliant (2014 audits). They also passed their 2014 SSAE-16 audit. The datacenter is under constant video surveillance. Entrance to both the building and the datacenter floor requires a keycard. In addition, to enter the datacenter, a keypad code is required. Each keycard has its own unique keypad login and each keycard has the name and picture of that particular employee. Hivelocity uses centralized controls for real-time data security and monitoring and would contact Dr. Panas and Krista Schroeder if the server were part of a data breach. Given their real-time security monitoring system, this contact will occur immediately after the data breach is identified (well within 24 hours). Additionally, the server will be set up with KernelCare so that all security kernel updates are applied automatically as soon as the patch is available without the need for a reboot. The server will use a Jinuper SRX-210 firewall with VPN features through the firewall (VPN set

up with encryption). Dr. Raymond Panas and Krista Schroeder will be given access to the server through this VPN and access will be password protected. The SEER-Medicare data DVDs will be sent to Dr. Panas' home and he will upload the data directly to the secure server through the VPN (no data will be downloaded onto his personal computer). He will then store the DVDs in a physical lockbox in his home office. No attempt will be made to identify individual patients, hospitals or physicians and any publications and presentations of the data will not allow identification of patients, hospitals or physicians (no data points with less than 11 patients will be reported). Finally, we agree to contact the NCI and obtain approval if the data needs to be moved for any reason (change host provider, etc.).

F. Funding Source: If your organization is a consulting firm, contractor, or pharmaceutical company, then your application must include a letter from the funder indicating that you are free to work and publish your findings without limitations by the funder. This letter must come from a person in authority on company letterhead.

There is no funding source for this project. Krista Schroeder will pay all fees associated with the data files in support of her dissertation research.

G. Restricted Variables: Selected variables are not released without the permission of the Principal Investigator of each of the SEER Registries. These variables include census tract of the patient, zip code of the patient, physician or hospital, and unencrypted provider numbers. If you are requesting access to any of these variables, you must include the justification in your description of the project and also submit the completed request form for restricted variables. Please see http://healthservices.cancer.gov/seermedicare/privacy/variables.html for information. NCI will provide a researcher with contact information for each of these registries; however it is the responsibility of the researcher to obtain permission from each registry.

Restricted variables are not being requested for this research project.

III. Data Files Requested:

Please list specific SEER-Medicare data files and years of data required. Project description **must** describe how each file will be used.

Medicare claims prior to 1998 are available only for cases diagnosed with cancer before 2003. Cases diagnosed 2003-2005 have claims from 1998+; cases diagnosed 2006 -2007 have claims from 2000+; cases diagnosed 2008-2009 have claims from 2002+; cases diagnosed 2010-2011 have claims 2004+.

Name of file	Years available	Years requested
Patient Entitlement and Diagnosis Summary File (PEDSF) (SEER cases)	1973-2011	2007-2011
Summarized Denominator File (SUMDENOM) 5% non-cancer sample	1991-2011	
5% PEDSF ** 5% cancer sample (Medicare enrollment data only)	1991-2011	
MEDPAR	1991-2012	

NCH - Carrier (physician/supplier)	1991-2012	2007-2011
Outpatient	1991-2012	2007-2011
Home Health (HHA)	1991-2012	
Hospice	1991-2012	
Durable Medical Equipment (DME)	1994-2012	2007-2011
Part D Event (PDE)	2007-2012	2007-2011
*Hospital File	1996, 1998, 2000-2012	

*NCI approval required. Investigator must provide a description of how the hospital data are to be used in this application.

(** To receive the 5%PEDSF, you must also be approved to receive the SUMDENOM file. In this case, your SUMDENOM file will include persons in the 5% sample who never had any cancer (SUMSTAT = 1) and patients in the 5% sample reported to have cancer, but no additional information (SUMSTAT = 2).

Appendix B: SEER Medicare Use Agreement

Investigator: Raymond M. Panas, PhD
Date: February 9, 2015
Project title: Disparities in Monoclonal Antibody Treatment of Elderly Metastatic Colorectal Cancer Patients

SEER-MEDICARE DATA USE AGREEMENT (DUA)
PRINCIPAL INVESTIGATOR

Information pertaining to an individual's health status and medical treatment is sensitive. Therefore, specific laws, including the Privacy Act of 1974 and the Health Insurance Portability and Accountability Act of 1996, have been enacted to ensure the confidentiality of health information. In utilizing health data for research purposes, it is absolutely necessary to ensure, to the extent possible, that uses of such data will be limited to research. Uses for any other reason, particularly those resulting in personal disclosures, will be prosecuted to the full extent of the law. In addition, release of information about providers, i.e., the physicians and hospitals that provide care for cancer patients, may compromise the willingness of these providers to cooperate with the activities of the cancer registries. Therefore, considerations regarding the privacy of providers are also of great importance

In order for the National Cancer Institute to provide the linked SEER-Surveillance, Epidemiology and End Results (SEER)-Medicare data to you, it is necessary that you agree to the following provisions:

1. You agree that the statements and methods made in your attached research proposal are complete and accurate.

2. You will not use the data for purposes other than described in your research proposal.

3. You will not permit others to use the data except for collaborators involved with the research as described in your proposal. Access to the SEER-Medicare data shall be limited to the minimum number of individuals necessary to achieve the purpose stated in your proposal. The number of locations where the data are located shall also be minimized and specific location details must be provided in your proposal's data storage and management plan. If you plan to move the data to a new location at your institute you must contact NCI in writing prior to moving the data for instruction on how to handle the SEER-Medicare data.

4. You will establish and maintain the appropriate administrative, technical, and physical safeguards to protect the confidentiality of the data and to prevent unauthorized use or access to it, as described in your proposal. The safeguards shall provide a level and scope of security that is not less than the level and scope of security requirements established by the Office of Management and Budget (OMB) in OMB Circular No. A-130, Appendix III--Security of Federal Automated Information Systems (http://www.whitehouse.gov/omb/circulars/a130/a130.html) as well as Federal Information Processing Standard 200 entitled "Minimum Security Requirements for Federal Information and Information Systems" http://csrc.nist.gov/publications/fips/fips200/FIPS-200-final-march.pdf); and, Special Publication 800-53 "Recommended Security Controls for Federal Information Systems" (http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf). You agree to allow NCI to conduct on-site inspections to ensure compliance with the data storage / confidentiality /security policies.

5. You agree not to place the SEER-Medicare data on personal computers, portable devices and removable media without permission. Portable devices include any non-fixed equipment that contains an operating system which may be used to create, access or store SEER-Medicare data. This includes but is not limited to laptops, personal digital assistants (PDAs), and smart phones. Removable media include, but are not limited to: CDs, DVDs, MP3 players, removable memory, and USB drives (thumb drives). If approved, all data stored on any of these devices must be password protected AND encrypted. Approved encryption standards must be FIPS-140 compliant and include Advanced Encryption Algorithm (AES) that uses a 128, 192, or 256-bit key size. In the event that the data are lost or stolen, you agree to report the loss to the SEER-Medicare contact within 24-hours/first business day of discovering the loss. Cloud storage does not meet privacy rules and is not acceptable for storing SEER-Medicare data.

6. You may use an institutionally provided VPN to link to a time sharing system for data access. In this case, the remote PC may support the VPN but the SEER-Medicare data must remain on the institution's server.

7. You will store all media on which the SEER-Medicare data are delivered in a secure location, such as a locked file cabinet in a locked office, only accessible by you or appropriate designated staff.

8. You must maintain all datasets containing restricted variables physically separate from any other SEER-Medicare files. Separate access controls with strong user authentication (username/password, digital certifications, etc.) must be established to allow limited access to these files. You should be able to track all access to these files.

9. All SEER-Medicare data must reside at your institution under your purview. If you plan to move to a different institution, you must contact NCI in writing prior to moving for instructions on how to handle the SEER-Medicare data. You may not duplicate any SEER-Medicare files prior to moving nor can you take SEER-Medicare data with you without written permission from NCI. If you chose not to take the data with you, you must destroy the files or designate a new PI prior to moving.

10. You will not attempt to link nor permit others to link the SEER-Medicare data with individually identified records in another database without the written consent from the applicable SEER registries.

11. No one having access to the data will attempt to learn the identity of any persons with cancer in these data and/or their physicians or treating hospitals. In the event that you discover or are able to deduce the identity of a specific patient or provider (individual or institution), you agree that you will not attempt to contact these individuals or institutions.

12. No findings or information derived from the SEER-Medicare data may be released if such findings contain any combination of data elements that might allow the deduction of a patient's or providers' (individual or institution) identity. In tables, cell sizes less than 11 (eleven) must be suppressed. Also, no use of percentages or other mathematical formulas may be used if they result in the display of a cell 10 or less. Mapping of data related to reflect incidence, treatment, or survival at the registry-specific level or at other small areas is not permitted without prior approval from NCI and the involved registries. Although it is permissible to report registry names with registry-specific cancer rates (e.g., incidence, complications, mortality), registry names must

be anonymized when reporting the quality or completeness of registry-specific data (e.g., case or treatment ascertainment). You agree that NCI shall be the sole judge as to whether the anonymization sufficiently precludes one from identifying or deducing the identity of a specific patient, provider (individual or institution) or registry with a reasonable degree of certainty.

13. You agree to provide the SEER-Medicare contact with a copy of all manuscripts to be submitted for publication prior to submission. You further agree not to submit such findings to any third party until receiving NCI's approval from the SEER-Medicare contact to do so. NCI agrees to make a determination about approval and to notify the user within 4 weeks after receipt of any findings. NCI's review of the findings is for the purpose of assuring that data confidentiality is maintained and that individual patients and/or providers (individual and institution) cannot be identified. NCI may withhold approval for publication if NCI reviewers determine that the format in which data are presented may result in identification of individual patients and/or providers (individual and institution). NCI may also review that the focus of the paper is consistent with research questions that were described in the request for SEER-Medicare data. NCI may decline to approve manuscripts that are beyond the scope of what is in the data request.

14. You agree that in the event NCI determines or has a reasonable belief that you have violated any terms of this agreement, NCI may request that you return the data and all derivative files to NCI. You understand that as a result of NCI's determination or reasonable belief that a violation of this agreement has taken place, NCI may refuse to release further SEER-Medicare data to you for a period of time to be determined by NCI.

15. All files received may be retained for a maximum of five years. At the completion of the project or five years from receipt all files including all back-up files and original media must be destroyed and notification of destruction must be sent to NCI. Investigators who need to retain files beyond that period must contact NCI.

Please indicate the SEER-Medicare files you will use:

☒	Patient Entitlement and Diagnosis Summary File) (PEDSF)	Years	2007-2011
☐	Summarized Denominator File (SUMDENOM)	Years	
☐	MEDPAR	Years	
☒	NCH - Carrier (physician/supplier)	Years	2007-2011
☒	Outpatient	Years	2007-2011
☐	Home Health Agency	Years	
☐	Hospice	Years	
☒	Durable medical equipment	Years	2007-2011
☒	Part D Event (PDE)	Years	2007-2011

These files will include:

☒ Cancer cases ☐ Non-cancer cases

Signature of Principal Investigator (In the case of students and fellows, the department chair or advisor from the student's academic institution must sign the data request)

Your signature indicates that you agree to comply with the above stated provisions. Deliberately making a false statement regarding any matter within the jurisdiction of any department or agency of the Federal Government violates 18 USC 1001 and is punishable by a fine up to $10,000 or up to five years in prison.

Raymond M. Panas
Name – (printed or typed)
Walden University
Institution/Organization
[redacted]
Electronic Signature - Raymond M. Panas/02092015/16:29:14
Signature
February 9, 2015
Date

Appendix C: Power Analyses

Logistic regression – A priori: Compute required sample size – given α, power and effect size: SES Input/Output Table

Input Parameters:

Tail(s) = Two
Odds ratio = 1.15
Pr(Y=1 I X = 1) H0 = 0.306
α err prob = 0.05
Power (1-β err prob) = 0.50
R^2 other X = 0.09
X distribution = Normal
X parm υ = 0
X parm σ = 1

Output:

Critical z = 1.96
Total sample size = **1029**
Actual power = 0.500

Logistic regression – A priori: Compute required sample size – given α, power and effect size: Race

Input Parameters:

Tail(s) = Two
Odds ratio = 0.85
Pr(Y=1 I X = 1) H0 = 0.306
α err prob = 0.05
Power (1-β err prob) = 0.50
R^2 other X = 0.09
X distribution = Normal
X parm υ = 0
X parm σ = 1

Output:

Critical z = 1.96
Total sample size = **764**
Actual power = 0.500

Logistic regression – A priori: Compute required sample size – given α, power and effect size: Rural/Urban Neighborhood

Input Parameters:

Tail(s) = Two
Odds ratio = 0.84
Pr(Y=1 I X = 1) H0 = 0.306
α err prob = 0.05
Power (1-β err prob) = 0.50

R^2 other X = 0.09
X distribution = Normal
X parm υ = 0
X parm σ = 1

Output:

Critical z = 1.96
Total sample size = **666**
Actual power = 0.500

Appendix D: Database Generation and Cohort Selection Code (Organized by File Generated)

Steps used to compile the study cohort:

1. Build the database (seer_create.sql; Appendix D): Script that creates the database and tables for use on this project.
2. Import the PEDSF data (seer_load_pedsf.py; Appendix D): Python script that connects to the database 'seer_db' created above and then runs sql insert queries to load the data from the seer text files into the database.
    a. This script loaded the data from the pedsf.cancer.file01.txt and pedsf.cancer.file02.txt files into the PEDSF table.
3. Import the Claims File data (seer_load_other.py; Appendix D): Python script that connects to the database 'seer_db' created above and then runs sql insert queries to load the data from the seer text files into the database.
    a. This script loaded the following datasets (each dataset has multiple files):
        i. Durable Medical Equipment files to the seer_db.DME table (files loaded: dme07.txt, dme08.txt, dme09.txt, dme10.txt, dme11.txt, dme12.txt)
        ii. Part D Event files to the seer_db.PDE table (files loaded: pdesaf07.txt, pdesaf08.txt, pdesaf09.txt, pdesaf10.txt, pdesaf11.txt, pdesaf12.txt)

iii. Carrier Claims files to the seer_db.NCH table (files loaded: nch07.file001.txt, nch07.file002.txt, nch07.file003.txt, nch07.file004.txt, nch07.file005.txt, nch07.file006.txt, nch07.file007.txt, nch08.file001.txt, nch08.file002.txt, nch08.file003.txt, nch08.file004.txt, nch08.file005.txt, nch08.file006.txt, nch08.file007.txt, nch08.file008.txt, nch09.file001.txt, nch09.file002.txt, nch09.file003.txt, nch09.file004.txt, nch09.file005.txt, nch09.file006.txt, nch09.file007.txt, nch09.file008.txt, nch10.file001.txt, nch10.file002.txt, nch10.file003.txt, nch10.file004.txt, nch10.file005.txt, nch10.file006.txt, nch10.file007.txt, nch10.file010.txt, nch11.file001.txt, nch11.file002.txt, nch11.file003.txt, nch11.file004.txt, nch11.file005.txt, nch11.file006.txt, nch11.file007.txt, nch11.file008.txt, nch12.file001.txt, nch12.file002.txt, nch12.file003.txt, nch12.file004.txt, nch12.file005.txt, nch12.file006.txt)

iv. Outpatient claims files to the seer_db.OUTPATIENT table (files loaded: outsaf07.file001.txt, outsaf07.file002.txt, outsaf07.file003.txt, outsaf07.file004.txt, outsaf07.file005.txt, outsaf08.file001.txt, outsaf08.file002.txt, outsaf08.file003.txt, outsaf08.file004.txt, outsaf08.file005.txt, outsaf08.file006.txt,

outsaf09.file001.txt, outsaf09.file002.txt, outsaf09.file003.txt, outsaf09.file004.txt, outsaf09.file005.txt, outsaf09.file006.txt, outsaf10.file001.txt, outsaf10.file002.txt, outsaf10.file003.txt, outsaf10.file004.txt, outsaf10.file005.txt, outsaf10.file006.txt, outsaf10.file007.txt, outsaf11.file001.txt, outsaf11.file002.txt, outsaf11.file003.txt, outsaf11.file004.txt,outsaf11.file005.txt, outsaf11.file006.txt, outsaf11.file007.txt, outsaf12.file001.txt, outsaf12.file002.txt, outsaf12.file003.txt, outsaf12.file004.txt, outsaf12.file005.txt, outsaf12.file006.txt)

4. Execute SQL queries to select data based on inclusion and exclusion criteria (seer_queries.sql; Appendix D): The selection process is described below.
    a. Filters claim data to only those matching Chemotherapy or Monoclonal antibody NDC or HCPCS codes needed for this research study.
    b. Filters claim data to first claim after diagnosis.
    c. Combines filtered claim data and pedsf data into one master table
    d. Adds a 'diagnosis_date' column which is the last day of the month in which the patient was diagnosed. This date is combination of smPEDSF.date_yr and smPEDSF.date_mo.
5. Database export (export_master_view.sh; Appendix D): BASH script - runs a simple SELECT query to export the seer_db.master_view to master_table.txt

6. Pre-SPSS processing (update_master_table.py; Appendix D): Python script – computes and adds five columns to the master_table.txt and saves it as master_table_updated.txt. Columns added are below.
    a. days – The number days between first chemotherapy treatment and first monoclonal antibody treatment.
    b. first_chemo – The date of the first chemotherapy treatment for that patient
    c. chemo_type – The type (ex. hcpcs code) for the first chemotherapy treatment
    d. first_ma - The date of the first monoclonal antibody treatment for that patient
    e. ma_type – The type (ex. hcpcs code) for first monoclonal antibody treatment

Code used to generate the study cohort:

seer_build.py

```
# -*- coding: utf-8 -*-
# seer_build.py
# MySQL Workbench Python script
# Written in MySQL Workbench 6.3.2
# Author: Andrew Schroeder, May 2015
# Description: Connects to MySQL server and executes 'seer_create.sql'
#   SQL script to create the seer_db

import mysql.connector
from mysql.connector import errorcode
import sys
import re

PATH_TO_FILE = "seer_create.sql"

try:
    cnx = mysql.connector.connect(user='root',
        password='', host='localhost')
```

```
except mysql.connector.Error as err:
    if err.errno == errorcode.ER_ACCESS_DENIED_ERROR:
        print("Something is wrong with your user name or password")
    elif err.errno == errorcode.ER_BAD_DB_ERROR:
        print("Database does not exist")
    else:
        print(err)
    sys.exit(1)

cursor = cnx.cursor()
sql = ""
for line in open(PATH_TO_FILE):
    ln = line.rstrip()
    if re.match("^--.*", ln):
        pass
    elif re.match("^\s*$", ln):
        pass
    else:
        sql += " " + ln

for q in sql.split(";"):
    if q == "":
        pass
    else:
        print (q + ";")
        cursor.execute(q + ";")
cnx.close()
```

seer_load_pedsf.py

```
# -*- coding: utf-8 -*-
# seer_load_pedsf.py
# MySQL Workbench Python script
# Written in MySQL Workbench 6.3.2
# Author: Andrew Schroeder  May 2nd 2015

import mysql.connector
from mysql.connector import errorcode
import sys
import re

class COLS:
    patient_id = 0
    registry = 1
    race = 2

PEDSF_1_FILE = "/samba/seer/data/pedsf.cancer.file01.txt"
PEDSF_2_FILE = "/samba/seer/data/pedsf.cancer.file02.txt"
pedsf_tbl_cols = ['patient_id', 'reporting_id', 'rac_recb',
```

```
    'census_pov_ind', 'urbrur', 'm_sex', 'birthyr', 'age_dx',
'date_yr',
    'date_mo', 'rsncd1', 'seq_num', 'ptbcnt_2007', 'ptbcnt_2008',
    'ptbcnt_2009', 'ptbcnt_2010', 'ptbcnt_2011', 'ptbcnt_2012',
    'hmocnt_2007', 'hmocnt_2008', 'hmocnt_2009',
    'hmocnt_2010', 'hmocnt_2011', 'hmocnt_2012',
    'ptd_2007', 'ptd_2008', 'ptd_2009', 'ptd_2010', 'ptd_2011',
'ptd_2012',
    'cs_met_1', 'cs_met_2', 'cs_met_3', 'cs_met_4', 'cs_met_5',
    'cs_met_6', 'cs_met_7', 'cs_met_8', 'cs_met_9', 'cs_met_10']
pedsf_txt_cols =
    [
    [1,10],[1,2],[101,2],[146,1],[97,1],[41,1],
    [37,4],[1881,3],[1888,4],[1886,2],
    #rsncd1
    [43,1],[1884,2],
    #ptbcnt 2007-2012
    [1252,2],[1296,2],[1340,2],[1384,2],[1428,2],[1472,2],
    #hmocnt 2007-2013
    [1254,2],[1298,2],[1342,2],[1386,2],[1430,2],[1474,2],
    #ptd 2007 - 2012
    [261,2],[301,2],[341,2],[381,2],[421,2],[461,2],
    # cs_mets
    [1953,2],[2253,2],[2553,2],[2853,2],[3153,2],[3453,2],
    [3753,2],[4053,2],[4353,2],[4653,2]
    ]

def slice(string, start, length):
    s = start - 1 #adjust for zero based list
    return string[s:s+length].strip()

def insert_table(cursor, table_name, line, table_cols, text_cols):
    col_data = []
        tbl_cols = []
        count = 0
    for tcol in text_cols:
        data=slice(line,tcol[0],tcol[1])
            if data != "":
                tbl_cols.append(table_cols[count])
                #print table_cols[count] +": " + data
        col_data.append(data)
            count +=1

        table_string = ("INSERT INTO " +
            table_name + " (" + ",".join(tbl_cols))
        table_string += ") VALUES (" + ("%s," * (len(tbl_cols)-1)) +
"%s)"
    cursor.execute((table_string), col_data)


def insert_pdesaf(cursor, line, year):
    cols = []
```

```
    cols.append(slice(line,0,10))
    cols.append(slice(line,0,2))
    cols.append(slice(line,101,2))
    cols.append(slice(line,146,1))
    cols.append(slice(line,97,1))
    cols.append(slice(line,41,1))
    cols.append(slice(line,37,4))
    cols.append(slice(line,1881,3))
    cols.append(slice(line,1888,4))
    cols.append(slice(line,1886,2))
    #rsncd1
    cols.append(slice(line,43,1))
    cols.append(slice(line,1884,2))
    #ptbcnt 2007-2013
    cols.append(slice(line,1252,2))
    cols.append(slice(line,1296,2))
    cols.append(slice(line,1340,2))
    cols.append(slice(line,1384,2))
    cols.append(slice(line,1428,2))
    cols.append(slice(line,1472,2))
    #hmocnt 2007-2013
    cols.append(slice(line,1254,2))
    cols.append(slice(line,1298,2))
    cols.append(slice(line,1342,2))
    cols.append(slice(line,1386,2))
    cols.append(slice(line,1430,2))
    cols.append(slice(line,1474,2))
    #ptd 2007 - 2013
    cols.append(slice(line,261,2))
    cols.append(slice(line,301,2))
    cols.append(slice(line,341,2))
    cols.append(slice(line,381,2))
    cols.append(slice(line,421,2))
    cols.append(slice(line,461,2))
    # cs_mets
    cols.append(slice(line,1953,2))
    cols.append(slice(line,2253,2))
    cols.append(slice(line,2553,2))
    cols.append(slice(line,2853,2))
    cols.append(slice(line,3153,2))
    cols.append(slice(line,3453,2))
    cols.append(slice(line,3753,2))
    cols.append(slice(line,4053,2))
    cols.append(slice(line,4353,2))
    cols.append(slice(line,4653,2))
    cursor.execute((pedsf_string), cols)


try:
    cnx = mysql.connector.connect(
        user='root', password='',
        host='localhost',database='seer_db')
```

```
except mysql.connector.Error as err:
    if err.errno == errorcode.ER_ACCESS_DENIED_ERROR:
        print("Something is wrong with your user name or password")
    elif err.errno == errorcode.ER_BAD_DB_ERROR:
        print("Database does not exist")
    else:
        print(err)
    sys.exit(1)

cursor = cnx.cursor()
cursor.execute("truncate PEDSF;")

count = 0
with open(PEDSF_1_FILE) as infile:
    for line in infile:
        insert_table(cursor,"PEDSF",line,pedsf_tbl_cols,pedsf_txt_cols)
    count +=1
        if count % 500 == 0:
             print count
with open(PEDSF_2_FILE) as infile:
    for line in infile:
        insert_table(cursor,"PEDSF",line,pedsf_tbl_cols,pedsf_txt_cols)
    count +=1
        if count % 500 == 0:
             print count
cnx.commit()
cursor.close()
cnx.close()
```

seer_load_other.py

```
# -*- coding: utf-8 -*-
# seer_load_other.py
# MySQL Workbench Python script
# Written in MySQL Workbench 6.3.2
# Author: Andrew Schroeder  May 2nd 2015

import mysql.connector
from mysql.connector import errorcode
import sys
import re
import os
##########################################
# constants and data
##########################################
DATA_PATH = "/samba/seer/data"

DME_FILES = []
for i in range(7,13):
    DME_FILES.append('dme{:02d}.txt'.format(i))
```

```
dme_tbl_cols = ['patient_id', 'hcpcs', 'ndc_cd', 'claim_from_date']
dme_txt_cols = [[1,10],[93,5],[409,11],[32,8]]

PDE_FILES = []
for i in range(7,13):
    PDE_FILES.append('pdesaf{:02d}.txt'.format(i))
pde_tbl_cols = ['patient_id', 'brand', 'prod_srvc_id', 'service_date']
pde_txt_cols = [[1,10],[90,30],[35,11],[27,8]]

nch_file_rages = [0,1,2,3,4,5,7,8,9,9,9,9,7]
NCH_FILES = []
for i in range(7,13):
    for j in range(1,nch_file_rages[i]):
        NCH_FILES.append('nch{:02d}.file{:03d}.txt'.format(i, j))
nch_tbl_cols = ['patient_id', 'hcpcs', 'claim_from_date']
nch_txt_cols = [[1,10],[93,5],[32,8]]

op_file_rages = [0,1,2,3,4,5,6,6,7,7,8,8,7]
OP_FILES = []
for i in range(7,13):
    for j in range(1,op_file_rages[i]):
        OP_FILES.append('outsaf{:02d}.file{:03d}.txt'.format(i, j))
op_tbl_cols = ['patient_id', 'hcpcs', 'claim_from_date']
op_txt_cols = [[1,10],[241,5],[32,8]]

progress_text = "Row Count: {:}\t\tTable: {:%}\t\tTotal: {:%}\r"

##########################################
#   functions
##########################################
def calc_num_records(FILES):
    record_size = 0
    records = 0
    for file in FILES:
        path = os.path.join(DATA_PATH,file)
        if record_size == 0:
            with open(path) as infile:
                record_size = len(infile.readline())

        if not os.path.isfile(path):
            print "Error {0} does not exist!".format(path)
            sys.exit(1)

        if record_size != 0:
            records += os.path.getsize(path) / record_size
    return records

def slice(string, start, length):
    s = start - 1 #adjust for zero based list
    return string[s:s+length].strip()

def insert_table(cursor, table_name, line, table_cols, text_cols):
```

```
    col_data = []
    tbl_cols = []
    count = 0
    for tcol in text_cols:
        data=slice(line,tcol[0],tcol[1])
        #convert from MMDDYYYY to YYYYMMDD
        if table_cols[count] == 'claim_from_date' or
            table_cols[count] == 'service_date':
            date = data[4:] + "-" + data[:2]  + "-" + data[2:4]
            data = date
        if data != "":
            tbl_cols.append(table_cols[count])
            col_data.append(data)
        count +=1

    table_string = "INSERT INTO " + table_name + " (" +
",".join(tbl_cols)
    table_string += ") VALUES (" + ("%s," * (len(tbl_cols)-1)) + "%s)"
    cursor.execute((table_string), col_data)

##########################################
# main script body
##########################################

# Connect to database
try:
    cnx = mysql.connector.connect(user='root', password='',
        host='localhost',database='seer_db')
except mysql.connector.Error as err:
    if err.errno == errorcode.ER_ACCESS_DENIED_ERROR:
        print("Something is wrong with your user name or password")
    elif err.errno == errorcode.ER_BAD_DB_ERROR:
        print("Database does not exist")
    else:
        print(err)
    sys.exit(1)

# open data files and count the number of records
#   (used for progress reporting)
dme_records = calc_num_records(DME_FILES)
print "DME Records: {0}".format(dme_records)
pde_records = calc_num_records(PDE_FILES)
print "PDE Records: {0}".format(pde_records)
nch_records = calc_num_records(NCH_FILES)
print "NCH Records: {0}".format(nch_records)
op_records = calc_num_records(OP_FILES)
print "OP Records: {0}".format(op_records)

#import the DME data to the DME table
cursor = cnx.cursor()
cursor.execute("truncate DME;")
total_progress = 0
```

```
total_records = dme_records + pde_records + nch_records + op_records
count = 0
for file in DME_FILES:
    with open(os.path.join(DATA_PATH, file)) as infile:
        for line in infile:
            insert_table(cursor,"DME",line,dme_tbl_cols,dme_txt_cols)
            count +=1
            if count % 5000 == 0:
                sys.stdout.write(
                   progress_text.format(
                        count, float(count)/float(dme_records),
                        float(total_progress + count)
                        / float(total_records) ))
                sys.stdout.flush()
cnx.commit()
print "DME table committed."

#import the PDE data to the PDE table
cursor.execute("truncate PDE;")
total_progress += count
count = 0
for file in PDE_FILES:
    with open(os.path.join(DATA_PATH, file)) as infile:
        for line in infile:
            insert_table(cursor,"PDE",line, pde_tbl_cols,pde_txt_cols)
            count +=1
            if count % 5000 == 0:
                sys.stdout.write(
                    progress_text.format(
                        count, float(count)/float(dme_records),
                        float(total_progress + count)
                        / float(total_records) ))
                sys.stdout.flush()
cnx.commit()
print "PDE table committed."

#import the NCH data to the NCH table
cursor.execute("truncate NCH;")
total_progress += count
count = 0
for file in NCH_FILES:
    with open(os.path.join(DATA_PATH, file)) as infile:
        for line in infile:
            insert_table(cursor,"NCH",line,nch_tbl_cols,nch_txt_cols)
            count +=1
            if count % 5000 == 0:
                 sys.stdout.write(
                    progress_text.format(
                        count, float(count)/float(dme_records),
                        float(total_progress + count)
                        / float(total_records) ))
                 sys.stdout.flush()
```

```
cnx.commit()
print "NCH table committed."

#import the outpatient data into the OUTPATIENT database
cursor.execute("truncate OUTPATIENT;")
total_progress += count
count = 0
for file in OP_FILES:
    with open(os.path.join(DATA_PATH, file)) as infile:
        for line in infile:

insert_table(cursor,"OUTPATIENT",line,op_tbl_cols,op_txt_cols)
            count +=1
            if count % 5000 == 0:
                 sys.stdout.write(
                    progress_text.format(
                        count, float(count)/float(dme_records),
                        float(total_progress + count)
                        / float(total_records) ))
                 sys.stdout.flush()
cnx.commit()
print "OUTPATIENT table committed."

cursor.close()
cnx.close()
```

update_master_table.py

```
# -*- coding: utf-8 -*-
# update_master_table.py
# author: Andrew Schroeder May, 2015

import sys
import re
from datetime import datetime

class COLS:
    patient_id = 0
    reporting_id = 1
    rac_recb = 2
    census_pov_ind = 3
    urbrur = 4
    m_sex = 5
    date_yr = 6
    date_mo = 7
    birthyr = 8
    age_dx = 9
    rsncd1 = 10
    cs_met_1 = 11
    diagnosis_date = 12
    ccDME_hcpcs = 13
```

```
    ccDME_claim_date = 14
    ccDME_ndc_cd = 15
    mmDME_hcpcs = 16
    mmDME_claim_date = 17
    mmDME_ndc_cd = 18
    ccNCH_hcpcs = 19
    ccNCH_claim_date = 20
    mmNCH_hcpcs = 21
    mmNCH_claim_date = 22
    ccOP_hcpcs = 23
    ccOP_claim_date = 24
    mmOP_hcpcs = 25
    mmOP_claim_date = 26
    mmPDE_brand = 27
    mmPDE_srvc_date = 28
    mmPDE_prod_srvc_id = 29


MASTER_FILE_IN = "master_table.txt"
MASTER_FILE_OUT = "master_table_updated.txt"

future_date = datetime.strptime('2025-01-01', "%Y-%m-%d")

count = 0
with open(MASTER_FILE_OUT, 'w') as outfile:
    with open(MASTER_FILE_IN) as infile:
        outfile.write(
            next(infile).rstrip() + "\t".join(
            ["","days","first_chemo","chemo_type", "first_ma",
"ma_type"])
            + "\n")
        for line in infile:

            cols = line.split()
            print("{0}, {1}, {2}".format(
                cols[COLS.ccDME_claim_date],
                cols[COLS.ccNCH_claim_date],
                cols[COLS.ccOP_claim_date]))
            # chemo data
            if cols[COLS.ccDME_claim_date] == 'NULL':
                date1 = future_date
                data = "NULL"
            else:
                date1 = datetime.strptime(
                    cols[COLS.ccDME_claim_date], "%Y-%m-%d")

            if cols[COLS.ccNCH_claim_date] == 'NULL':
                date2 = future_date
            else:
                date2 = datetime.strptime(
                    cols[COLS.ccNCH_claim_date], "%Y-%m-%d")
```

```
if cols[COLS.ccOP_claim_date] == 'NULL':
    date3 = future_date
else:
    date3 = datetime.strptime(
        cols[COLS.ccOP_claim_date], "%Y-%m-%d")


if  not (date1 > date2 or date1 > date3):
    date_out = date1
    if cols[COLS.ccDME_ndc_cd] != "NULL":
        data_out = cols[COLS.ccDME_ndc_cd]
    else:
        data_out = cols[COLS.ccDME_hcpcs]
elif not (date2 > date1 or date2 > date3):
    date_out = date2
    data_out = cols[COLS.ccNCH_hcpcs]

elif not (date3 > date2 or date3 > date1):
    date_out = date3
    data_out = cols[COLS.ccOP_hcpcs]

# ma data
if cols[COLS.mmDME_claim_date] == 'NULL':
    ma_date1 = future_date
    ma_data = "NULL"
else:
    ma_date1 = datetime.strptime(
        cols[COLS.mmDME_claim_date], "%Y-%m-%d")

if cols[COLS.mmNCH_claim_date] == 'NULL':
    ma_date2 = future_date
else:
    ma_date2 = datetime.strptime(
        cols[COLS.mmNCH_claim_date], "%Y-%m-%d")

if cols[COLS.mmOP_claim_date] == 'NULL':
    ma_date3 = future_date
else:
    ma_date3 = datetime.strptime(
        cols[COLS.mmOP_claim_date], "%Y-%m-%d")

if cols[COLS.mmPDE_srvc_date] == 'NULL':
    ma_date4 = future_date
else:
    ma_date4 = datetime.strptime(
        cols[COLS.mmPDE_srvc_date], "%Y-%m-%d")



if  not (ma_date1 > ma_date2 or
        ma_date1 > ma_date3 or
        ma_date1 > ma_date4):
```

```
        ma_date_out = ma_date1;
        if cols[COLS.mmDME_ndc_cd] != "NULL":
            ma_data_out = cols[COLS.mmDME_ndc_cd]
        else:
            ma_data_out = cols[COLS.mmDME_hcpcs]
    elif not (ma_date2 > ma_date1 or
            ma_date2 > ma_date3 or
            ma_date2 > ma_date4):
        ma_date_out = ma_date2;
        ma_data_out = cols[COLS.mmNCH_hcpcs]
    elif not (ma_date3 > ma_date1 or
            ma_date3 > ma_date2 or
            ma_date3 > ma_date4):
        ma_date_out = ma_date3;
        ma_data_out = cols[COLS.mmOP_hcpcs]
    elif not (ma_date4 > ma_date1 or
            ma_date4 > ma_date2 or
            ma_date4 > ma_date3):
        ma_date_out = ma_date4;
        if cols[COLS.mmPDE_brand] != "NULL":
            ma_data_out = cols[COLS.mmPDE_brand]
        else:
            ma_data_out = cols[COLS.mmPDE_prod_srvc_id]
    else:
        print ('This should not happen')
        sys.exit(1)


    if date_out == future_date or
        cols[COLS.diagnosis_date] == "NULL":
        days_out = "NULL"
        date_out = "NULL"
    else:
        diagnosis_date = datetime.strptime(
            cols[COLS.diagnosis_date], "%Y-%m-%d")
        elapsed = date_out - diagnosis_date
        days_out = elapsed.days
        date_out = date_out.strftime("%Y-%m-%d")


    if ma_date_out == future_date or
            cols[COLS.diagnosis_date] == "NULL":
        ma_days_out = "NULL"
        ma_date_out = "NULL"
    else:
        diagnosis_date = datetime.strptime(
            cols[COLS.diagnosis_date], "%Y-%m-%d")
        ma_elapsed = ma_date_out - diagnosis_date
        ma_days_out = ma_elapsed.days
        ma_date_out = ma_date_out.strftime("%Y-%m-%d")

    newline = line.rstrip() + "\t".join(
```

```
                                                    ['',str(days_out),
                                                    date_out,data_out,
                                                    ma_date_out,
                                                    ma_data_out])
            print(newline)

            outfile.write(newline + "\n")
            count +=1
```

seer_create.sql

```
-- MySQL Script generated by MySQL Workbench
-- seer_create.sql
-- 05/01/15 17:33:43
-- MySQL Workbench Forward Engineering

SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0;
SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS, FOREIGN_KEY_CHECKS=0;
SET @OLD_SQL_MODE=@@SQL_MODE,
SQL_MODE='TRADITIONAL,ALLOW_INVALID_DATES';

-- -----------------------------------------------------
-- Schema seer_db
-- -----------------------------------------------------
DROP SCHEMA IF EXISTS `seer_db` ;

-- -----------------------------------------------------
-- Schema seer_db
-- -----------------------------------------------------
CREATE SCHEMA IF NOT EXISTS `seer_db`
DEFAULT CHARACTER SET utf8 COLLATE utf8_general_ci ;
USE `seer_db` ;

-- -----------------------------------------------------
-- Table `seer_db`.`PEDSF`
-- -----------------------------------------------------
DROP TABLE IF EXISTS `seer_db`.`PEDSF` ;

CREATE TABLE IF NOT EXISTS `seer_db`.`PEDSF` (
  `idPEDSF` INT NOT NULL AUTO_INCREMENT,
  `patient_id` VARCHAR(10) NULL,
  `reporting_id` VARCHAR(2) NULL,
  `rac_recb` TINYINT(2) UNSIGNED NULL,
  `census_pov_ind` TINYINT(1) UNSIGNED NULL,
  `urbrur` TINYINT(1) UNSIGNED NULL,
  `m_sex` TINYINT(1) UNSIGNED NULL,
  `birthyr` YEAR NULL,
  `age_dx` TINYINT(3) UNSIGNED NULL,
  `date_yr` YEAR NULL,
  `date_mo` TINYINT(2) UNSIGNED NULL,
  `rsncd1` TINYINT(1) UNSIGNED NULL,
```

```
  `seq_num` TINYINT(2) UNSIGNED NULL,
  `year` YEAR NULL,
  `ptbcnt_2007` TINYINT(2) NULL,
  `ptbcnt_2008` TINYINT(2) NULL,
  `ptbcnt_2009` TINYINT(2) NULL,
  `ptbcnt_2010` TINYINT(2) NULL,
  `ptbcnt_2011` TINYINT(2) NULL,
  `ptbcnt_2012` TINYINT(2) NULL,
  `hmocnt_2007` TINYINT(2) NULL,
  `hmocnt_2008` TINYINT(2) NULL,
  `hmocnt_2009` TINYINT(2) NULL,
  `hmocnt_2010` TINYINT(2) NULL,
  `hmocnt_2011` TINYINT(2) NULL,
  `hmocnt_2012` TINYINT(2) NULL,
  `ptd_2007` TINYINT(2) NULL,
  `ptd_2008` TINYINT(2) NULL,
  `ptd_2009` TINYINT(2) NULL,
  `ptd_2010` TINYINT(2) NULL,
  `ptd_2011` TINYINT(2) NULL,
  `ptd_2012` TINYINT(2) NULL,
  `cs_met_1` TINYINT(2) NULL,
  `cs_met_2` TINYINT(2) NULL,
  `cs_met_3` TINYINT(2) NULL,
  `cs_met_4` TINYINT(2) NULL,
  `cs_met_5` TINYINT(2) NULL,
  `cs_met_6` TINYINT(2) NULL,
  `cs_met_7` TINYINT(2) NULL,
  `cs_met_8` TINYINT(2) NULL,
  `cs_met_9` TINYINT(2) NULL,
  `cs_met_10` TINYINT(2) NULL,
  PRIMARY KEY (`idPEDSF`))
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `seer_db`.`NCH`
-- -----------------------------------------------------
DROP TABLE IF EXISTS `seer_db`.`NCH` ;

CREATE TABLE IF NOT EXISTS `seer_db`.`NCH` (
  `idNCH` INT NOT NULL AUTO_INCREMENT,
  `patient_id` VARCHAR(10) NOT NULL,
  `hcpcs` VARCHAR(5) NULL,
  `claim_from_date` DATE NOT NULL,
  PRIMARY KEY (`idNCH`))
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `seer_db`.`OUTPATIENT`
-- -----------------------------------------------------
DROP TABLE IF EXISTS `seer_db`.`OUTPATIENT` ;
```

```sql
CREATE TABLE IF NOT EXISTS `seer_db`.`OUTPATIENT` (
  `idOUTPATIENT` INT NOT NULL AUTO_INCREMENT,
  `patient_id` VARCHAR(10) NOT NULL,
  `hcpcs` VARCHAR(5) NULL,
  `claim_from_date` DATE NOT NULL,
  PRIMARY KEY (`idOUTPATIENT`))
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `seer_db`.`DME`
-- -----------------------------------------------------
DROP TABLE IF EXISTS `seer_db`.`DME` ;

CREATE TABLE IF NOT EXISTS `seer_db`.`DME` (
  `idDME` INT NOT NULL AUTO_INCREMENT,
  `patient_id` VARCHAR(10) NOT NULL,
  `hcpcs` VARCHAR(5) NULL,
  `ndc_cd` VARCHAR(11) NULL,
  `claim_from_date` DATE NOT NULL,
  PRIMARY KEY (`idDME`))
ENGINE = InnoDB;


-- -----------------------------------------------------
-- Table `seer_db`.`PDE`
-- -----------------------------------------------------
DROP TABLE IF EXISTS `seer_db`.`PDE` ;

CREATE TABLE IF NOT EXISTS `seer_db`.`PDE` (
  `idPDE` INT NOT NULL AUTO_INCREMENT,
  `patient_id` VARCHAR(10) NOT NULL,
  `brand` VARCHAR(30) NULL,
  `prod_srvc_id` VARCHAR(11) NULL,
  `service_date` DATE NULL,
  PRIMARY KEY (`idPDE`))
ENGINE = InnoDB;

SET SQL_MODE = '';
GRANT USAGE ON *.* TO krista;
 DROP USER krista;
SET SQL_MODE='TRADITIONAL,ALLOW_INVALID_DATES';
CREATE USER 'krista' IDENTIFIED BY 'M#4806d#5656';

GRANT ALL ON `seer_db`.* TO 'krista';
GRANT SELECT ON TABLE `seer_db`.* TO 'krista';
GRANT SELECT, INSERT, TRIGGER ON TABLE `seer_db`.* TO 'krista';
GRANT SELECT, INSERT, TRIGGER, UPDATE,
DELETE ON TABLE `seer_db`.* TO 'krista';

SET SQL_MODE=@OLD_SQL_MODE;
```

```
SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS;
SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS;
```

seer_queries.sql

```
-- seer_queries.sql
-- author: Andrew Schroeder
-- May 5th, 2015

-- NCH, OUTPATIENT, PDE, DME tables are filtered for specific criteria
-- m, c, mm, cc prefixes are used to distinguish filtered results from
-- original table. New tables are generated for performance
enhancements
-- on final "master_view" queries

-- c prefix indicates subset with matching chemotherapy code
-- cc prefix indicates subset of c with only first claim record
--    after diagnosis_date
-- m prefix indicated subset with matching monoclonal antibody code
-- mm prefix indicates subset of m with only first claim record
--    after diagnosis_date

-- create "c" and "m" tables
DROP TABLE IF EXISTS `seer_db`.`cNCH`;
CREATE TABLE cNCH AS
     SELECT NCH.idNCH, NCH.patient_id, NCH.hcpcs, NCH.claim_from_date
     FROM NCH
     WHERE NCH.hcpcs IN ('J9190','J8520','J8521',
                         'J0640','J9200','J9263','J9206');
DROP TABLE IF EXISTS `seer_db`.`mNCH`;
CREATE TABLE mNCH AS
     SELECT NCH.idNCH, NCH.patient_id, NCH.hcpcs, NCH.claim_from_date
     FROM NCH
     WHERE NCH.hcpcs IN ('J9035','J9055','J9303');

DROP TABLE IF EXISTS `seer_db`.`cOP`;
CREATE TABLE cOP AS
     SELECT OUTPATIENT.idOUTPATIENT, OUTPATIENT.patient_id,
        OUTPATIENT.hcpcs, OUTPATIENT.claim_from_date
     FROM OUTPATIENT
     WHERE OUTPATIENT.hcpcs IN ('J9190','J8520','J8521','J0640',
                                'J9200','J9263','J9206');
DROP TABLE IF EXISTS `seer_db`.`mOP`;
CREATE TABLE mOP AS
     SELECT OUTPATIENT.idOUTPATIENT, OUTPATIENT.patient_id,
            OUTPATIENT.hcpcs, OUTPATIENT.claim_from_date
     FROM OUTPATIENT
     WHERE OUTPATIENT.hcpcs IN ('J9035','J9055','J9303');

DROP TABLE IF EXISTS `seer_db`.`mPDE`;
```

```
CREATE TABLE mPDE AS
    SELECT PDE.idPDE, PDE.patient_id, PDE.brand, PDE.prod_srvc_id,
           PDE.service_date
    FROM PDE
    WHERE brand IN ('AVASTIN','ERBITUX','VECTIBIX') OR
          prod_srvc_id IN('5024206001','50242006101','66733094823',
                          '66733095823','55513095401','55513095601');

DROP TABLE IF EXISTS `seer_db`.`cDME`;
CREATE TABLE cDME AS
    SELECT DME.idDME, DME.patient_id, DME.hcpcs, DME.ndc_cd,
           DME.claim_from_date
    FROM DME
    WHERE DME.ndc_cd IN('00703301513', '00703301812', '00703301912',
          '10139006301', '10139006310', '10139006311', '10139006312',
          '10139006320', '10139006350', '63323011710', '63323011720',
          '63323011751', '63323011761', '66758004401', '66758004403',
          '00004110020', '54868414300', '54868414301', '54868414302',
          '00004110150', '00004110175', '54868526000', '54868526001',
          '54868526002', '54868526003', '54868526004', '54868526005',
          '54868526006', '54868526007', '54868526008', '54868526009',
          '00703279301', '00703279701', '00703514001', '00703514501',
          '00703514591', '00904231560', '25021081310', '25021081430',
          '25021081530', '25021081567', '25021081630', '25021081667',
          '55390000901', '55390005110', '55390005210', '55390005301',
          '55390005401', '55390081810', '55390082401', '55390082501',
          '55390082601', '62701090030', '62701090099', '62701090125',
          '63323071050', '63323071100', '55390013501', '63323014507',
          '00024059010', '00024059120', '00024059240', '00069006701',
          '00069007001', '00069007401', '00069101001', '00703398501',
          '00703398601', '25021021120', '25021021250', '41616017640',
          '41616017840', '47335017640', '47335017840', '61703036318',
          '61703036322', '63323017530', '63323017650', '63323065010',
          '63323065017', '63323065020', '63323065027', '00009111101',
          '00009111102', '00009752901', '00009752902', '00009752903',
          '00009752904', '00009752905', '00143970101', '00143970201',
          '00703443211', '00703443411', '10019093401', '10019093402',
          '10019093417', '10019093479', '10518010310', '18111000202',
          '18111000203', '23155017931', '23155017932', '25021020002',
          '25021020005', '25021021402', '25021021405', '55390029501',
          '55390029601', '59762752902', '61703034909', '61703034916',
          '61703034936', '63323019302', '63323019305', '66758004801',
          '66758004802')
          OR DME.hcpcs IN ('J9190','J8520','J8521','J0640',
                           'J9200','J9263','J9206');
DROP TABLE IF EXISTS `seer_db`.`mDME`;
CREATE TABLE mDME AS
    SELECT DME.idDME, DME.patient_id, DME.hcpcs, DME.ndc_cd,
           DME.claim_from_date
    FROM DME
    WHERE DME.ndc_cd IN( '50242006001', '50242006101', '66733094823',
                         '66733095823', '55513095401', '55513095601')
```

```
          OR DME.hcpcs IN ('J9035','J9055','J9303');


-- create "cc" and "mm" tables

-- create ccDME, uses cDME_temp for simpler SQL
DROP TABLE IF EXISTS `seer_db`.`cDME_temp`;
CREATE TABLE cDME_temp AS
SELECT cDME.*
FROM cDME
INNER JOIN (
    SELECT cDME.patient_id,MIN(cDME.claim_from_date) as min_date
    FROM cDME
    GROUP BY cDME.patient_id) tmp
    ON cDME.patient_id = tmp.patient_id
        AND cDME.claim_from_date = tmp.min_date;
DROP TABLE IF EXISTS `seer_db`.`ccDME`;
CREATE TABLE ccDME AS
SELECT c.*
FROM cDME_temp c
INNER JOIN (
    SELECT c.patient_id, MIN(c.idDME) as min_id
    FROM cDME_temp c
    GROUP BY c.patient_id) tmp
    ON c.patient_id = tmp.patient_id AND c.idDME = tmp.min_id;

-- create mmDME, uses mDME_temp for simpler SQL
DROP TABLE IF EXISTS `seer_db`.`mDME_temp`;
CREATE TABLE mDME_temp AS
SELECT mDME.*
FROM mDME
INNER JOIN (
    SELECT mDME.patient_id,MIN(mDME.claim_from_date) as min_date
    FROM mDME
    GROUP BY mDME.patient_id) tmp
    ON mDME.patient_id = tmp.patient_id
       AND mDME.claim_from_date = tmp.min_date;
DROP TABLE IF EXISTS `seer_db`.`mmDME`;
CREATE TABLE mmDME AS
SELECT m.*
FROM mDME_temp m
INNER JOIN (
    SELECT m.patient_id, MIN(m.idDME) as min_id
    FROM mDME_temp m
    GROUP BY m.patient_id) tmp
    ON m.patient_id = tmp.patient_id AND m.idDME = tmp.min_id;


-- create ccNCH, uses cNCH_temp for simpler SQL
DROP TABLE IF EXISTS `seer_db`.`cNCH_temp`;
CREATE TABLE cNCH_temp AS
SELECT cNCH.*
```

```
FROM cNCH
INNER JOIN (
    SELECT cNCH.patient_id,MIN(cNCH.claim_from_date) as min_date
    FROM cNCH
    GROUP BY cNCH.patient_id) tmp
    ON cNCH.patient_id = tmp.patient_id
       AND cNCH.claim_from_date = tmp.min_date;
DROP TABLE IF EXISTS `seer_db`.`ccNCH`;
CREATE TABLE ccNCH AS
SELECT c.*
FROM cNCH_temp c
INNER JOIN (
    SELECT c.patient_id, MIN(c.idNCH) as min_id
    FROM cNCH_temp c
    GROUP BY c.patient_id) tmp
    ON c.patient_id = tmp.patient_id AND c.idNCH = tmp.min_id;

-- create mmNCH, uses mNCH_temp for simpler SQL
DROP TABLE IF EXISTS `seer_db`.`mNCH_temp`;
CREATE TABLE mNCH_temp AS
SELECT mNCH.*
FROM mNCH
INNER JOIN (
    SELECT mNCH.patient_id,MIN(mNCH.claim_from_date) as min_date
    FROM mNCH
    GROUP BY mNCH.patient_id) tmp
    ON mNCH.patient_id = tmp.patient_id
       AND mNCH.claim_from_date = tmp.min_date;
DROP TABLE IF EXISTS `seer_db`.`mmNCH`;
CREATE TABLE mmNCH AS
SELECT m.*
FROM mNCH_temp m
INNER JOIN (
    SELECT m.patient_id, MIN(m.idNCH) as min_id
    FROM mNCH_temp m
    GROUP BY m.patient_id) tmp
    ON m.patient_id = tmp.patient_id AND m.idNCH = tmp.min_id;

-- create ccOP, uses cOP_temp for simpler SQL
DROP TABLE IF EXISTS `seer_db`.`cOP_temp`;
CREATE TABLE cOP_temp AS
SELECT cOP.*
FROM cOP
INNER JOIN (
    SELECT cOP.patient_id,MIN(cOP.claim_from_date) as min_date
    FROM cOP
    GROUP BY cOP.patient_id) tmp
    ON cOP.patient_id = tmp.patient_id
       AND cOP.claim_from_date = tmp.min_date;
DROP TABLE IF EXISTS `seer_db`.`ccOP`;
CREATE TABLE ccOP AS
SELECT c.*
```

```sql
FROM cOP_temp c
INNER JOIN (
    SELECT c.patient_id, MIN(c.idOUTPATIENT) as min_id
    FROM cOP_temp c
    GROUP BY c.patient_id) tmp
    ON c.patient_id = tmp.patient_id AND c.idOUTPATIENT = tmp.min_id;

-- create mmOP, uses mOP_temp for simpler SQL
DROP TABLE IF EXISTS `seer_db`.`mOP_temp`;
CREATE TABLE mOP_temp AS
SELECT mOP.*
FROM mOP
INNER JOIN (
    SELECT mOP.patient_id,MIN(mOP.claim_from_date) as min_date
    FROM mOP
    GROUP BY mOP.patient_id) tmp
    ON mOP.patient_id = tmp.patient_id
       AND mOP.claim_from_date = tmp.min_date;
DROP TABLE IF EXISTS `seer_db`.`mmOP`;
CREATE TABLE mmOP AS
SELECT m.*
FROM mOP_temp m
INNER JOIN (
    SELECT m.patient_id, MIN(m.idOUTPATIENT) as min_id
    FROM mOP_temp m
    GROUP BY m.patient_id) tmp
    ON m.patient_id = tmp.patient_id AND m.idOUTPATIENT = tmp.min_id;

-- create mmPDE, uses mPDE_temp for simpler SQL
DROP TABLE IF EXISTS `seer_db`.`mPDE_temp`;
CREATE TABLE mPDE_temp AS
SELECT mPDE.*
FROM mPDE
INNER JOIN (
    SELECT mPDE.patient_id,MIN(mPDE.service_date) as min_date
    FROM mPDE
    GROUP BY mPDE.patient_id) tmp
    ON mPDE.patient_id = tmp.patient_id
       AND mPDE.service_date = tmp.min_date;
DROP TABLE IF EXISTS `seer_db`.`mmPDE`;
CREATE TABLE mmPDE AS
SELECT m.*
FROM mPDE_temp m
INNER JOIN (
    SELECT m.patient_id, MIN(m.idPDE) as min_id
    FROM mPDE_temp m
    GROUP BY m.patient_id) tmp
    ON m.patient_id = tmp.patient_id AND m.idPDE = tmp.min_id;

-- create the master_view

CREATE OR replace view master_view as
```

```
select
    smPEDSF.patient_id, smPEDSF.reporting_id, smPEDSF.rac_recb,
    smPEDSF.census_pov_ind, smPEDSF.urbrur, smPEDSF.m_sex,
    smPEDSF.date_yr, smPEDSF.date_mo, smPEDSF.birthyr, smPEDSF.age_dx,
    smPEDSF.rsncd1, smPEDSF.cs_met_1,
    LAST_DAY(CONCAT_WS('-',smPEDSF.date_yr,smPEDSF.date_mo,'1'))
        as `diagnosis_date`,
    ccDME.hcpcs as ccDME_hcpcs,
    ccDME.claim_from_date as ccDME_claim_date,
    ccDME.ndc_cd as ccDME_ndc_cd,
    mmDME.hcpcs as mmDME_hcpcs,
    mmDME.claim_from_date as mmDME_claim_date,
    mmDME.ndc_cd as mmDME_ndc_cd,
    ccNCH.hcpcs as ccNCH_hcpcs,
    ccNCH.claim_from_date as ccNCH_claim_date,
    mmNCH.hcpcs as mmNCH_hcpcs,
    mmNCH.claim_from_date as mmNCH_claim_date,
    ccOP.hcpcs as ccOP_hcpcs,
    ccOP.claim_from_date as ccOP_claim_date,
    mmOP.hcpcs as mmOP_hcpcs,
    mmOP.claim_from_date as mmOP_claim_date,
    mmPDE.brand as mmPDE_brand,
    mmPDE.service_date as mmPDE_srvc_date,
    mmPDE.prod_srvc_id as mmPDE_prod_srvc_id
FROM smPEDSF
LEFT OUTER JOIN ccDME
    ON smPEDSF.patient_id = ccDME.patient_id
LEFT OUTER JOIN mmDME
    ON smPEDSF.patient_id = mmDME.patient_id
LEFT OUTER JOIN ccNCH
    ON smPEDSF.patient_id = ccNCH.patient_id
LEFT OUTER JOIN mmNCH
    ON smPEDSF.patient_id = mmNCH.patient_id
LEFT OUTER JOIN ccOP
    ON smPEDSF.patient_id = ccOP.patient_id
LEFT OUTER JOIN mmOP
    ON smPEDSF.patient_id = mmOP.patient_id
LEFT OUTER JOIN mmPDE
    ON smPEDSF.patient_id = mmPDE.patient_id;
```

export_master_view.sh

```
#!/usr/bin/bash
mysql -A seer_db -e "SELECT * FROM master_view ORDER BY patient_id;" >
master_table.txt
```

Appendix E: Variables Transferred to SPSS for Analysis

- Ten digit patient ID (column #1 in the PEDSF document; patient_id)
- Registry reporting the data (The first 2 digits in the patient ID; PEDSF column #1)
- Race (column #101 of the PEDSF document; rac_recb)
- Census Tract Poverty Indicator, measure of SES (column #146 of the PEDSF document; census_pov_ind)
- PEDSF re-code of the 2003 Rural/Urban Continuum Codes (column #97 of the PEDSF document; urbrur)
- Gender (PEDSF column #41; m_sex)
- Year of birth (PEDSF file column #37)
- Age at diagnosis (PEDSF column #1881)
- Year of diagnosis (PEDSF column #1888)
- Month of diagnosis (PEDSF column #1886)
- Reason for original Medicare entitlement (PEDSF column #43)
- Cancer sequence number (PEDSF column #1884)
- Medicare Part B coverage months 2007-2011 (PEDSF columns #548 and below)
- Medicare Part D coverage months 2007-2011 (PEDSF columns #221 and below)
- Medicare managed care plan (Medicare Part C) coverage months 2007-2011 (PEDSF columns #550 and below; HMO Months).
- Metastasis at Diagnosis (PEDSF column #1953; CS Mets at Dx)

- First Chemotherapy Claim and Date: These are two computed variables (first_chemo and chemo_type) as in 6b and 6c in Appendix D above. The variable first_chemo is the date of the first post diagnosis chemotherapy claim. The variable chemo_type is the NDC or HCPCS code for the first post diagnosis appearance of any of the following codes in any of the Part B files (Carrier Claims, DME or Outpatient Claims): HCPCS codes: J9190, J8520, J8521, J0640, J9200, J9263, J9206; NDC codes: 00703301513, 00703301812, 00703301912, 10139006301, 10139006310, 10139006311, 10139006312, 10139006320, 10139006350, 63323011710, 63323011720, 63323011751, 63323011761, 66758004401, 66758004403, 00004110020, 54868414300, 54868414301, 54868414302, 00004110150, 00004110175, 54868526000, 54868526001, 54868526002, 54868526003, 54868526004, 54868526005, 54868526006, 54868526007, 54868526008, 54868526009, 00703279301, 00703279701, 00703514001, 00703514501, 00703514591, 00904231560, 25021081310, 25021081430, 25021081530, 25021081567, 25021081630, 25021081667, 55390000901, 55390005110, 55390005210, 55390005301, 55390005401, 55390081810, 55390082401, 55390082501, 55390082601, 62701090030, 62701090099, 62701090125, 63323071050, 63323071100, 55390013501, 63323014507, 00024059010, 00024059120, 00024059240, 00069006701, 00069007001, 00069007401, 00069101001, 00703398501, 00703398601, 25021021120, 25021021250, 41616017640, 41616017840, 47335017640, 47335017840, 61703036318,

61703036322, 63323017530, 63323017650, 63323065010, 63323065017, 63323065020, 63323065027, 00009111101, 00009111102, 00009752901, 00009752902, 00009752903, 00009752904, 00009752905, 00143970101, 00143970201, 00703443211, 00703443411, 10019093401, 10019093402, 10019093417, 10019093479, 10518010310, 18111000202, 18111000203, 23155017931, 23155017932, 25021020002, 25021020005, 25021021402, 25021021405, 55390029501, 55390029601, 59762752902, 61703034909, 61703034916, 61703034936, 63323019302, 63323019305, 66758004801, 66758004802. The locations for these codes and dates in each of the files are outlined below.

- Carrier Claims:
  - HCPCS codes (column #93)
  - Date of treatment for the previous code (column #32)
- Outpatient Claims:
  - HCPCS code (column #241)
  - Date of treatment for the previous code (column #32)
- DME:
  - HCPCS code (column #93)
  - NDC code (column #409)
  - Date of treatment for the previous code (column #32)

- First Monoclonal Antibody Claim and Date: These are two computed variables (first_ma and ma_type) as in 6d and 6e in Appendix D above. The variable first_ma is the date of the first post diagnosis monoclonal antibody claim. The variable ma_type is the NDC, HCPCS code, or Brand Name code for the first post diagnosis appearance of any of the following codes in any of the Part B files (Carrier Claims, DME or Outpatient Claims): (Carrier Claims, Outpatient Claims, or PDE): HCPCS codes: J9035, J9055, J9303; NDC codes: 50242006001, 50242006101, 66733094823, 66733095823, 55513095401, 55513095601; or Brand Name codes: Avastin, Erbitux, Vectibix. The locations for these codes and dates in each of the files are outlined below.
    - Carrier Claims:
        - HCPCS code (column #93)
        - Date of treatment for the previous code (column #32)
    - Outpatient Claims:
        - HCPCS code (column #241)
        - Date of treatment for the previous code (column #32)
    - PDE:
        - Brand name code (column #90)
        - NDC code (column #35)
        - PDE drug dispense date (column #27)

Appendix F: SEER-Medicare Data Approve to Use

6/14/2015 Walden University Mail - SEER-Medicare request

Gmail by Google

Krista Schroeder [redacted]

SEER-Medicare request

Yanisko, Elaine (IMS [redacted] Mon, Mar 9, 2015 at 2:21 PM
To: Raymond Panas <[redacted]
Cc: Krista Schroeder <[redacted]>

Dr. Panas

I have attached the reviewers' comments on your application. You do not need to respond to these comments unless you decide you want the additional files.

I have not received the hard copy of the signed DUA. You should send that to me at the address below.

Please complete the attached checklist and return it to me by email. When I have the hard copy of the DUA, I can send the completed checklist to be invoiced. Invoices are sent by regular mail, but I can provide a PDF if you prefer. Once we receive payment the checklist is placed in the queue to be completed. Our turnaround time is 3-6 weeks, but is dependent on the number of requests in the queue. Completed datasets are shipped UPS ground. Please be sure to provide a valid UPS delivery address in section 4 of the checklist

Reminders:

1. Release of SEER-Medicare data is project specific. You may only access these data to work on the project as it was described in the approved application. Any other analysis must be submitted as a new application. No work can begin until all approvals have been secured.

2. All manuscripts must be submitted for review prior to submitting for publication.

I recommend any staff with access to these files sign the DUA for your records as documentation they have agreed to abide by the rules / requirements for using these data. As the PI, you will be held responsible for any violations of the DUA.

Please note this checklist includes the 2013 claims files (except for the Part D Event file). If you want this additional year of data, please be sure that your DUA includes this year before you mail it to me. There is a delay in the release of the 2013 files. We hope this will be resolved quickly, but we will not invoice for these data until this has been resolved.

Please let me know if you have any questions.

Thanks

Elaine

Elaine Yanisko

IMS, Inc.

Information in this e-mail may be confidential. It is intended only for the addressee(s) identified above. If you are not the addressee(s), or an employee or agent of the addressee(s), please note that any dissemination, distribution, or copying of this communication is strictly prohibited. If you have received this e-mail in error, please notify the sender of the error.

2 attachments

Panas colorectal 02_26_2015 review.docx
17K

2014-Checklist-022815.doc
49K

Appendix G: Logistic Regression Syntax; Monoclonal Antibody Disparities by Race

Monoclonal Antibody Treatment and Race Logistic Regression Syntax

```
DATASET ACTIVATE DataSet1.
USE ALL.
COMPUTE filter_$=(NOT rac_recb = 3 and NOT rac_recb = 9 and NOT rac_recb = 12).
VARIABLE LABELS filter_$ 'NOT rac_recb = 3 and NOT rac_recb = 9 and NOT
rac_recb = 12 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

LOGISTIC REGRESSION VARIABLES MA_treatment
  /METHOD=ENTER Race_new
  /CONTRAST (Race_new)=Indicator(1)
  /PRINT=CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Appendix H: Logistic Regression Syntax; Monoclonal Antibody Disparities by SES

Monoclonal Antibody Treatment and SES Logistic Regression Syntax

```
DATASET ACTIVATE DataSet1.
USE ALL.
COMPUTE filter_$=(NOT census_pov_ind = 9 ).
VARIABLE LABELS filter_$ 'NOT census_pov_ind = 9  (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

LOGISTIC REGRESSION VARIABLES MA_treatment
  /METHOD=ENTER census_pov_ind
  /CONTRAST (census_pov_ind)=Indicator(1)
  /PRINT=CI(95)
 /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Appendix I: Logistic Regression Syntax; Monoclonal Antibody Disparities by Neighborhood Characteristics

Monoclonal Antibody Treatment and Neighborhood Characteristics Logistic Regression Syntax

```
DATASET ACTIVATE DataSet1.
USE ALL.
COMPUTE filter_$=(NOT urbrur = 9 ).
VARIABLE LABELS filter_$ 'NOT urbrur = 9  (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

LOGISTIC REGRESSION VARIABLES MA_treatment
  /METHOD=ENTER urbrur
  /CONTRAST (urbrur)=Indicator(1)
  /PRINT=CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Appendix J. Logistic Regression Syntax; Model Including All Significant Independent Variables and Covariates

Monoclonal Antibody Treatment and All Significant Variables Logistic Regression

```
DATASET ACTIVATE DataSet1.
FILTER OFF.
USE ALL.
EXECUTE.

LOGISTIC REGRESSION VARIABLES MA_treatment
  /METHOD=ENTER reporting_id age_dx
    /CONTRAST (reporting_id)=Indicator(1)
  /PRINT=CI(95)
 /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```